Guest Lecture
Machine Learning in Healthcare

Narges Razavian
Assistant Professor
Departments of Radiology & Population Health NYUMC
narges.razavian@nyumc.org

Machine Learning
November 1st, 2018

This Lecture

Overview of healthcare & landscape of healthcare data

Some snapshots of research on machine learning in healthcare
 Early Disease Prediction using EHR time series
 Medical Imaging:
  Radiology (X-Rays, Mammograms, MRI, Ultrasound)
  Pathology (Histopathology)
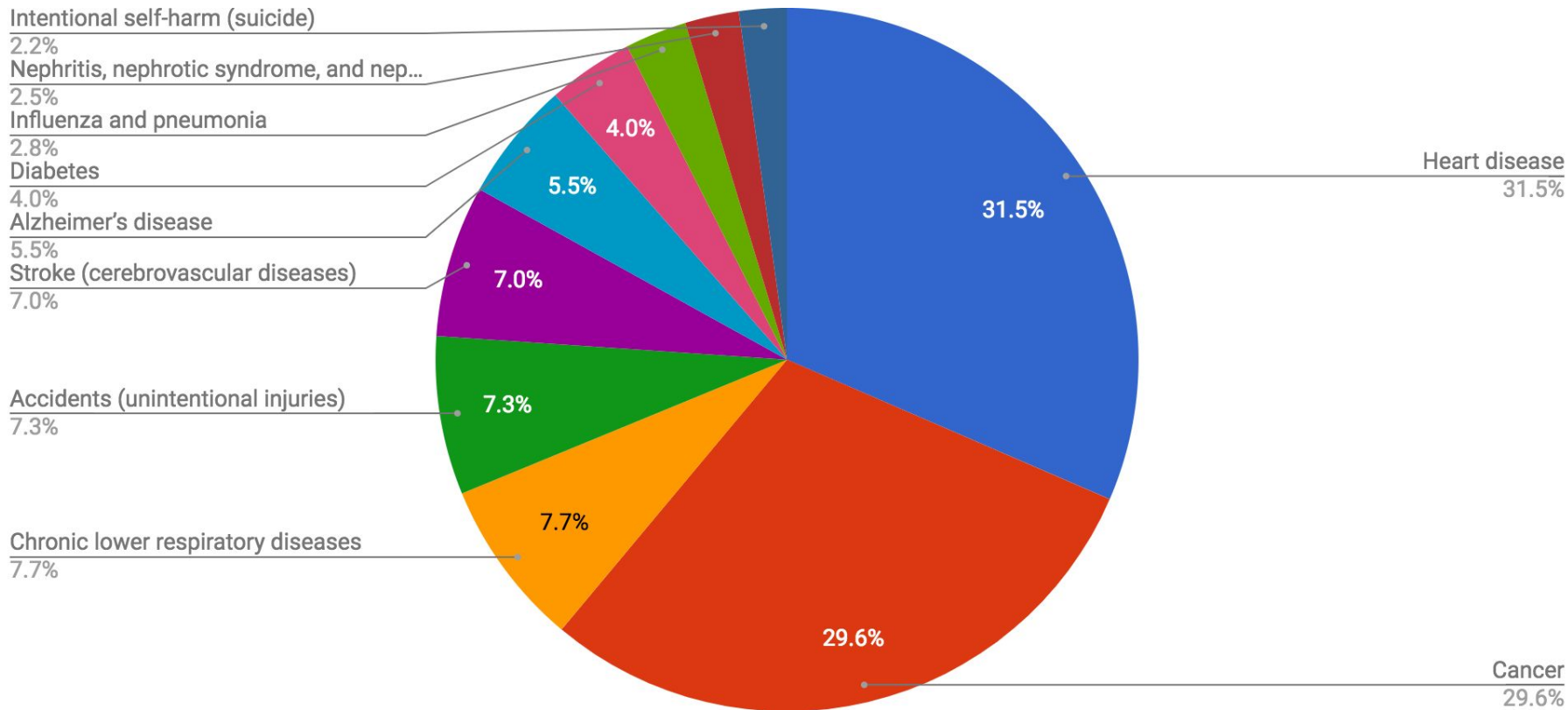 *Microscopy*
 Genomics and sequences and text

Thoughts on research trends in short and long term in this field.
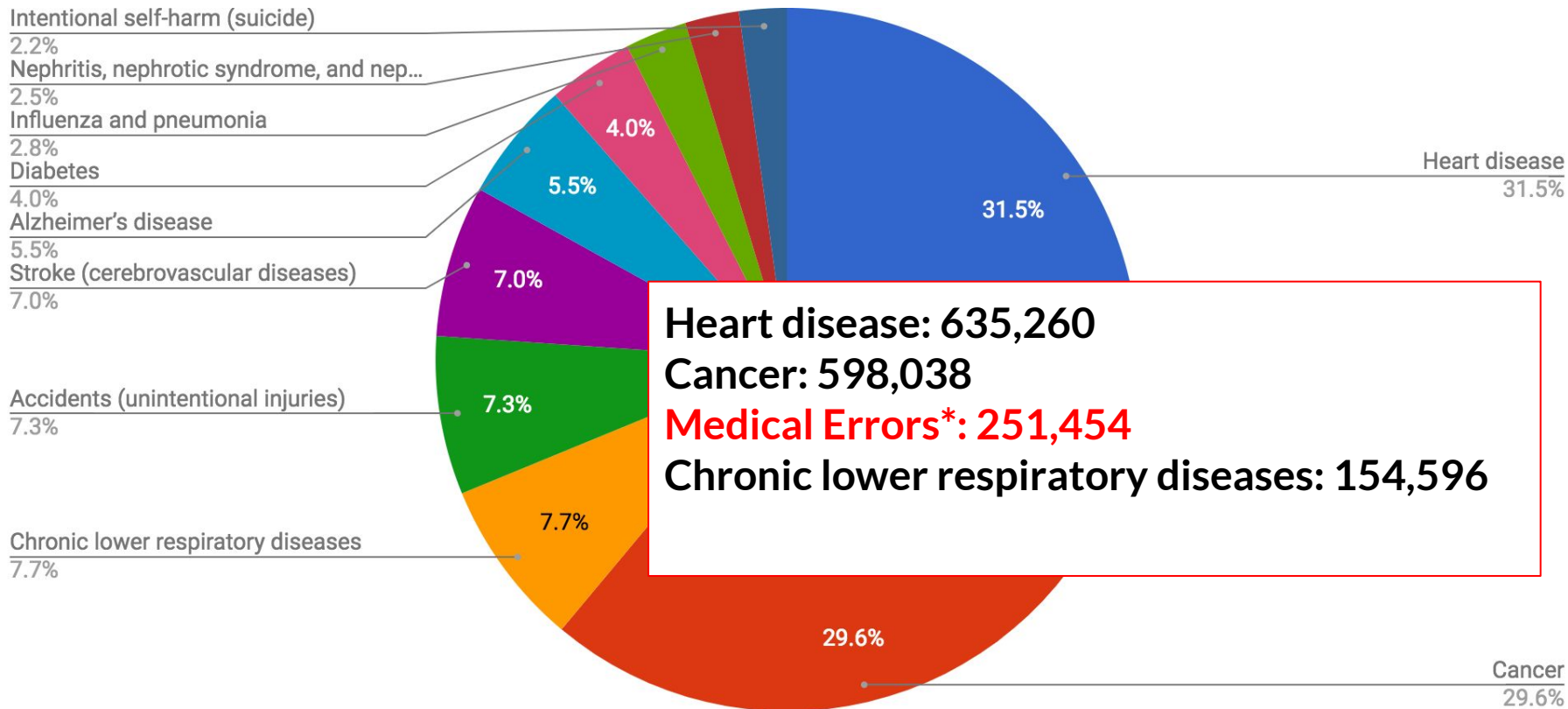
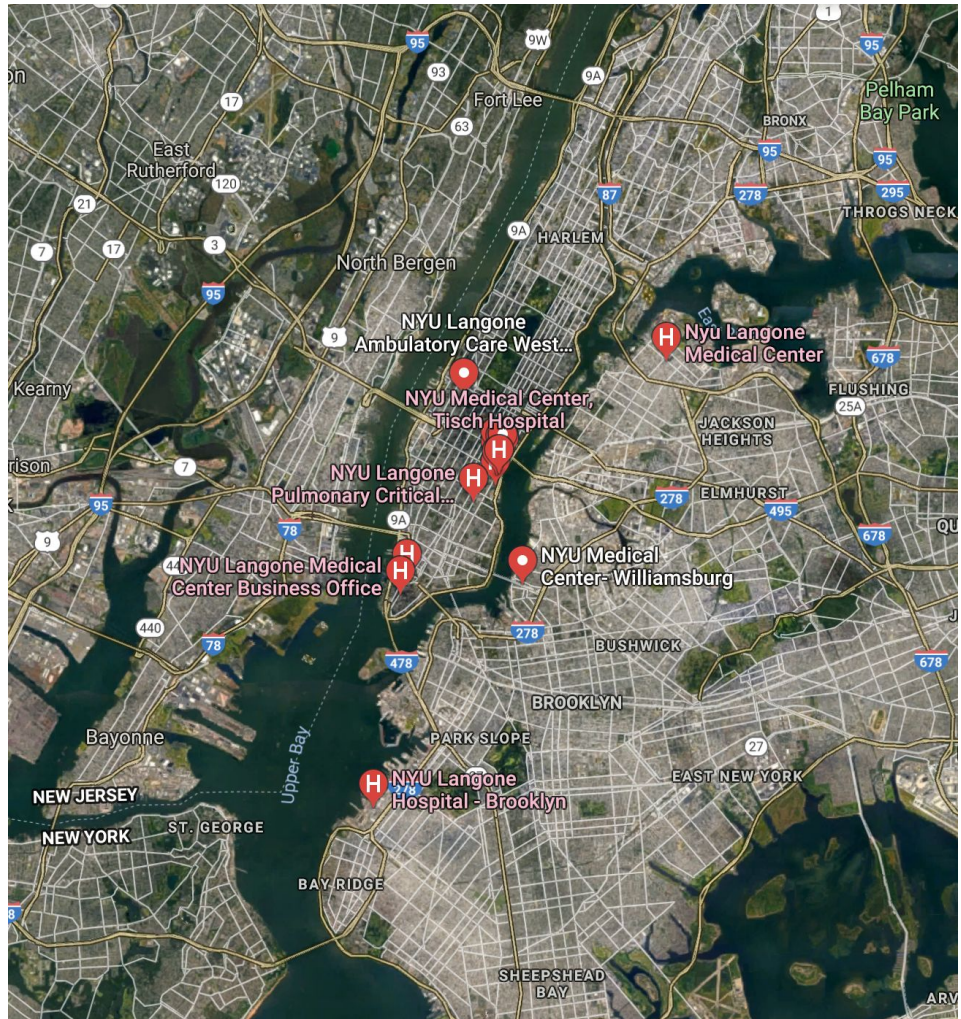# Healthcare in Numbers

What are the top killer diseases?
What are the diseases people go to doctors for?
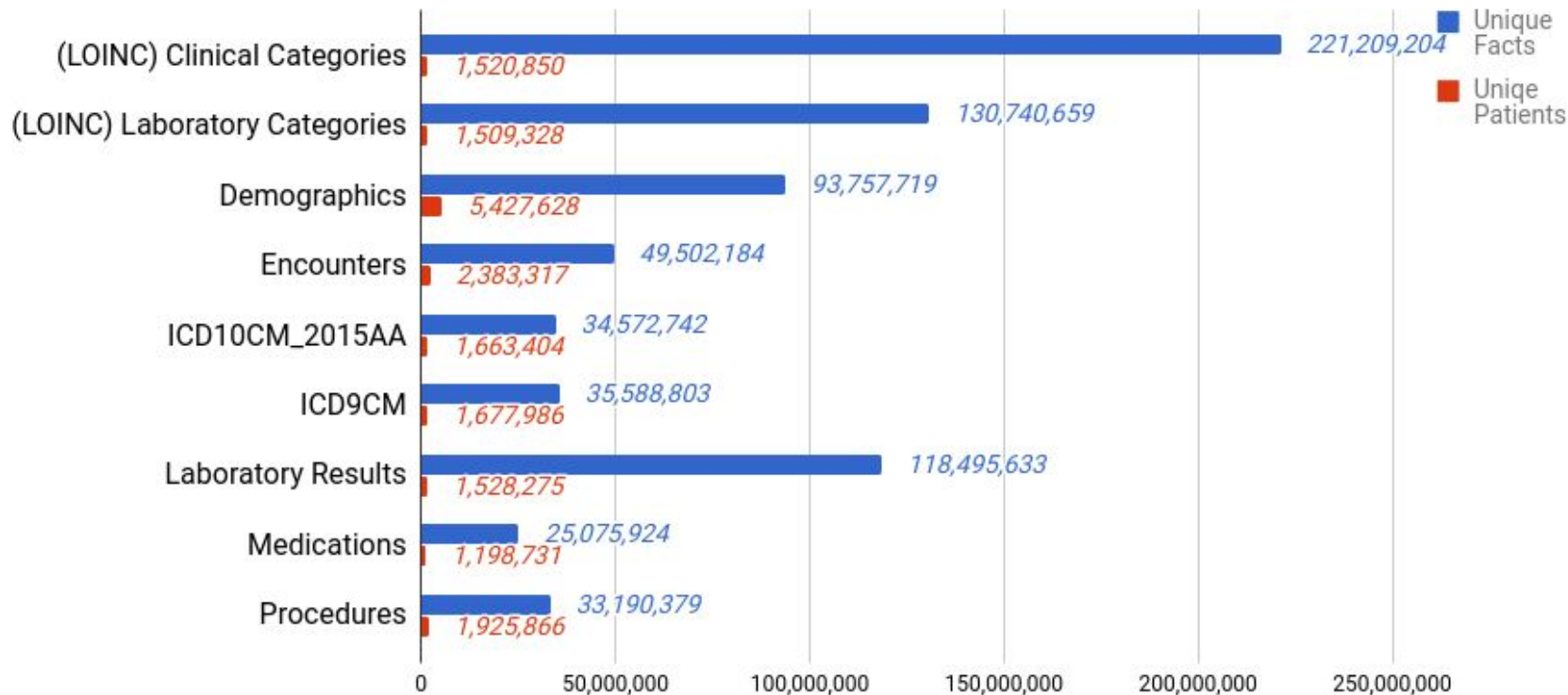
"Immature" Causes of Death in 2016, USA

Intentional self-harm (suicide)
2.2%
Nephritis, nephrotic syndrome, and nep...
2.5%
Influenza and pneumonia
2.8%
Diabetes
4.0%
Alzheimer's disease
5.5%
Stroke (cerebrovascular diseases)
7.0%

Accidents (unintentional injuries)
7.3%

Chronic lower respiratory diseases
7.7%

Heart disease
31.5%

31.5%

4.0%

5.5%

7.0%

7.3%

7.7%

29.6%

Cancer
29.6%

"Immature" Causes of Death in 2016, USA

Intentional self-harm (suicide)
2.2%

Nephritis, nephrotic syndrome, and nep...
2.5%

Influenza and pneumonia
2.8%

Diabetes
4.0%

Alzheimer's disease
5.5%

Stroke (cerebrovascular diseases)
7.0%

Accidents (unintentional injuries)
7.3%

Chronic lower respiratory diseases
7.7%

Heart disease
31.5%

Cancer
29.6%

4.0%

5.5%

7.0%

7.3%

7.7%

31.5%

29.6%

Heart disease: 635,260
Cancer: 598,038
Medical Errors*: 251,454
Chronic lower respiratory diseases: 154,596

NYU Medical School - de-identified database i2b2 (2 years ago)

Diseases

Disease	Blue	Red
Certain conditions originating in the perinatal period	43,986	10,818
Certain infectious and parasitic diseases	451,418	133,280
Congenital malformations, deformations and chromosomal abnormalities	262,603	48,712
Diseases of the blood/blood-forming organs and disorders of the immune mechanism	617,466	109,009
Diseases of the circulatory system	4,089,695	349,945
Diseases of the digestive system	1,259,833	245,571
Diseases of the ear and mastoid process	354,945	86,429
Diseases of the eye and adnexa	246,308	71,768
Diseases of the genitourinary system	1,473,933	280,470
Diseases of the musculoskeletal system and connective tissue	4,963,905	555,033
Diseases of the nervous system	1,455,389	239,126
Diseases of the respiratory system	1,215,401	260,885
Diseases of the skin and subcutaneous tissue	739,027	150,515
Endocrine, nutritional and metabolic diseases	3,153,053	386,539
External causes of morbidity	47,041	26,912
Factors influencing health status and contact with health services	4,469,121	712,131
Injury, poisoning and certain other consequences of external causes	1,081,864	245,880
Mental, behavioral and neurodevelopmental disorders	1,072,284	171,461
Neoplasms	2,385,994	189,302
Pregnancy, childbirth and the puerperium	322,517	35,391

Procedures

Category	Unique Facts	Unique Patients
Anesthesia	335,130	194,458
Category II Codes	143,120	14,781
Category III Codes	32,091	11,833
Evaluation and Management Servic...	11,666,119	1,517,413
Medicine Services and Procedures	6,828,091	902,899
Pathology and Laboratory Procedur...	3,100,177	464,409
Radiology Procedures	13,399,297	1,304,062
Surgery	3,960,338	813,604

Legend:
- Unique Facts
- Unique Patients

X-axis: 0, 5,000,000, 10,000,000, 15,000,000

Healthcare in Action

What happens Where and When?
What's the constraints of each location?

Overview of Healthcare in Action

Emergency Dept: Triage & Stabilization
➔ Bleeding/pain/etc
➔ internal/external problems
➔ Patient awake or unconscious
➔ Quick diagnosis needed
➔ Localization of main cause
➔ Quick action to give patient time
➔ Can be: Fast, Noisy, Loud, Mechanical

Overview of Healthcare in Action
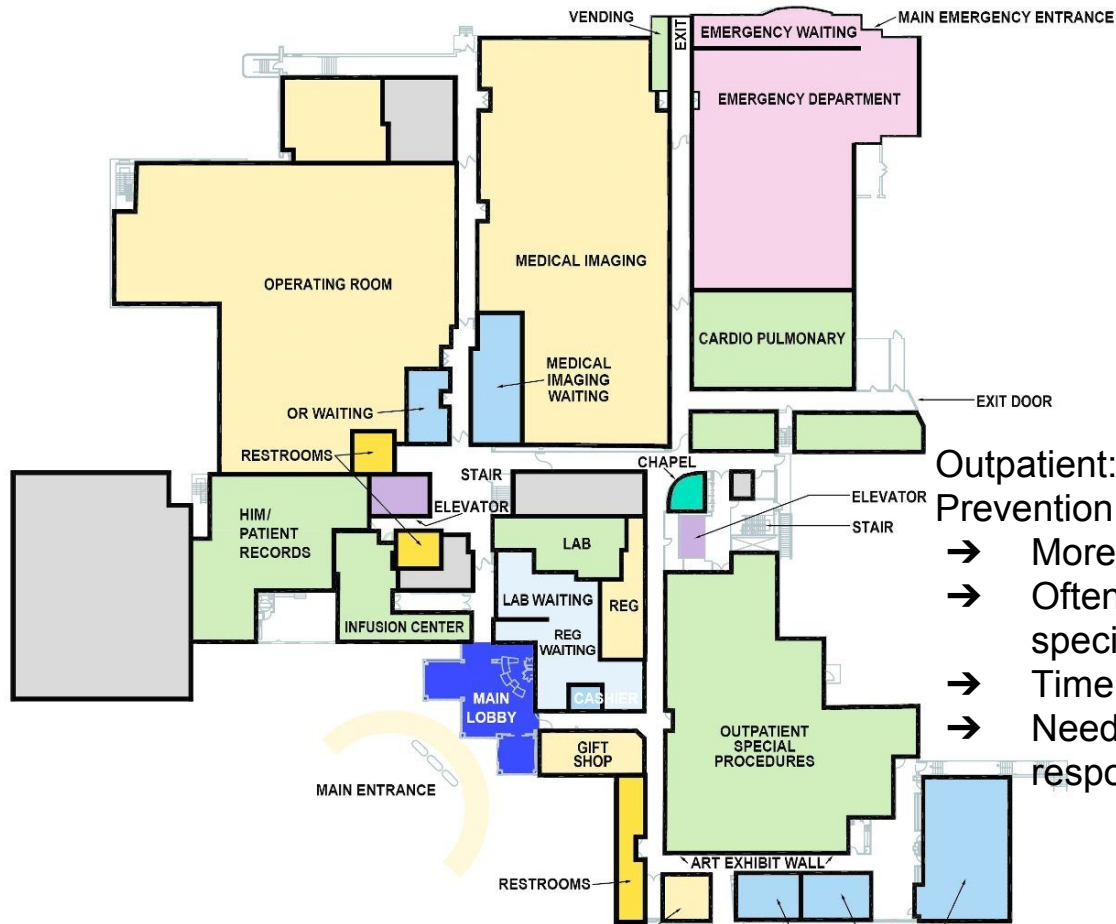
Outpatient: Diagnosis, Curing and Prevention
- ➔ More time to diagnose
- ➔ Often symptoms aren't specific/strong enough
- ➔ Time to do (diagnostic) tests
- ➔ Need to track medication response or Prevent s.th.

Overview of Healthcare in Action

Surgery: Either Emergency or Elected

➔ Invasive and need to be complete in one session
➔ For biopsy(diagnosis) or treatment
➔ *Robotic Surgery:* less invasive.
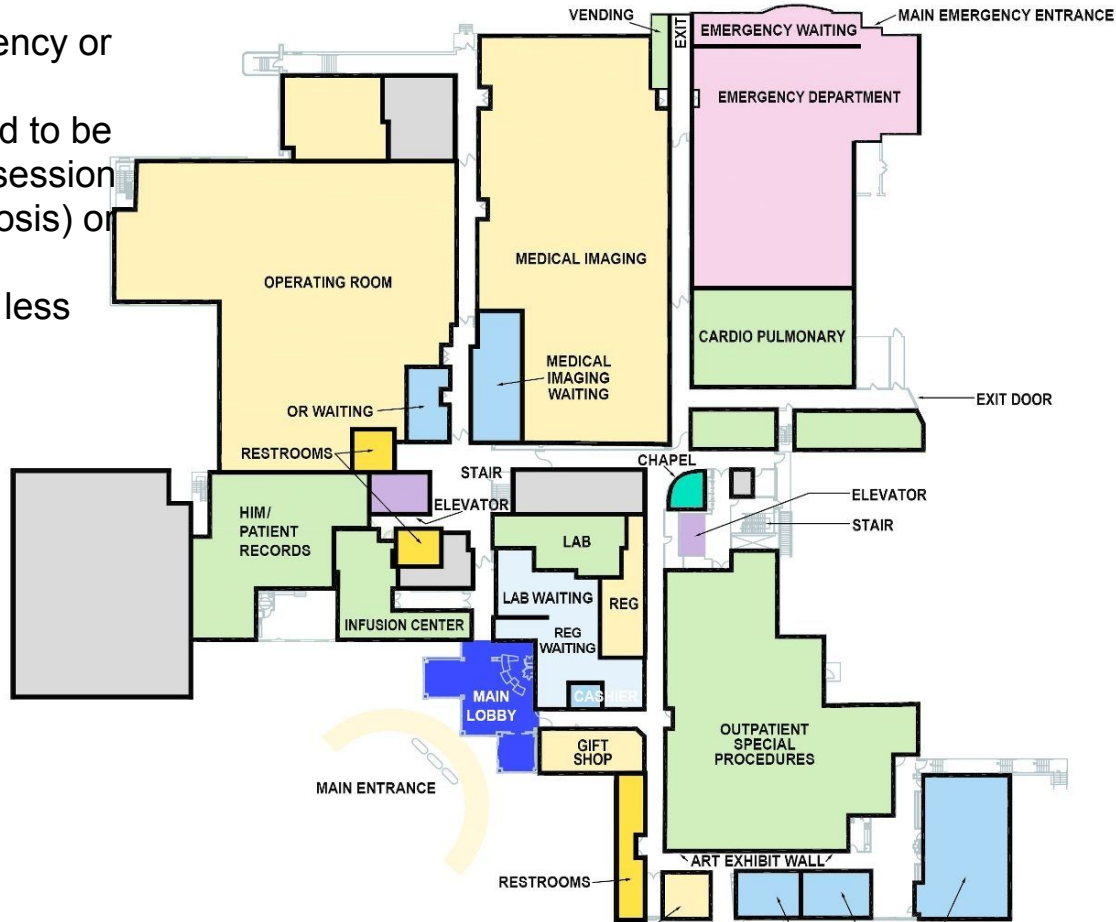
Overview of Healthcare in Action
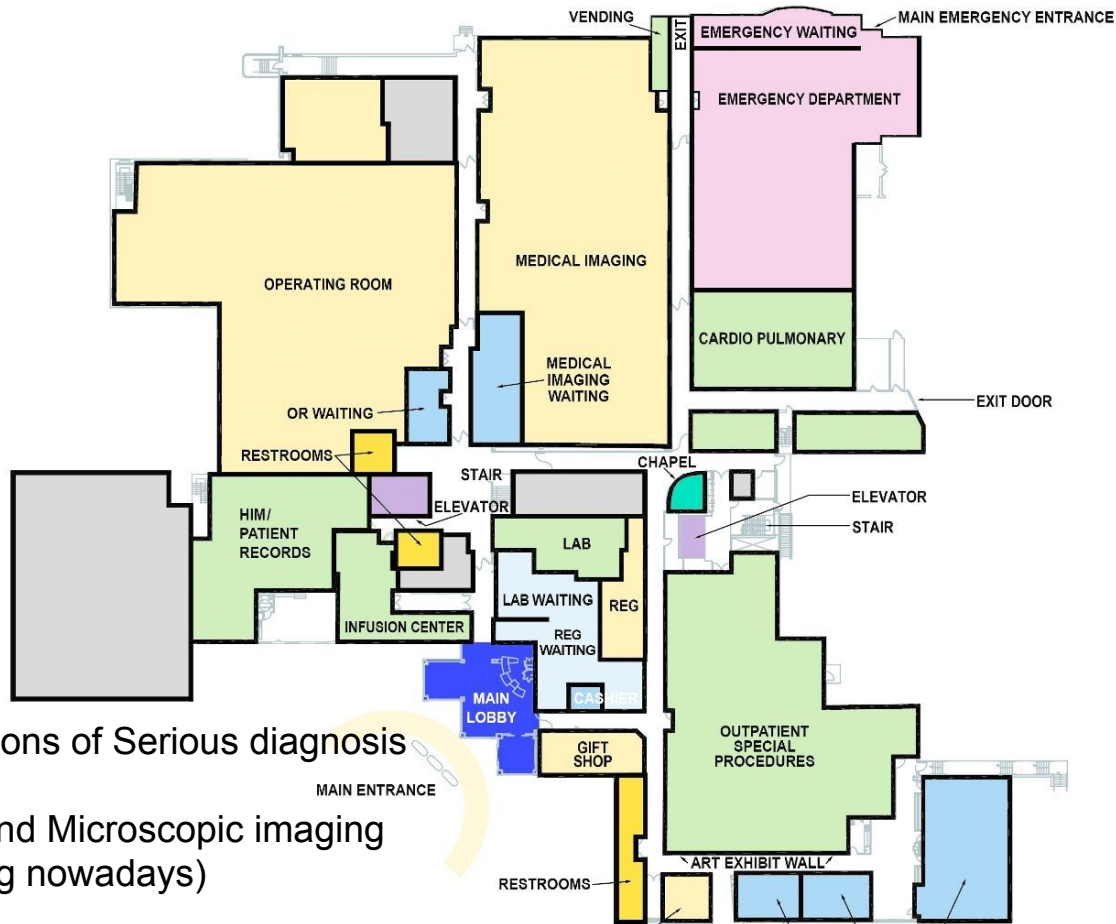
Pathology: Confirmations of Serious diagnosis
- ➔ Most cancers,
- ➔ Tissues, cells and Microscopic imaging
- ➔ (Genetic reading nowadays)

Diverse Data Modalities

Diverse Modalities: Text and Structured data Time Series (NYU Data)
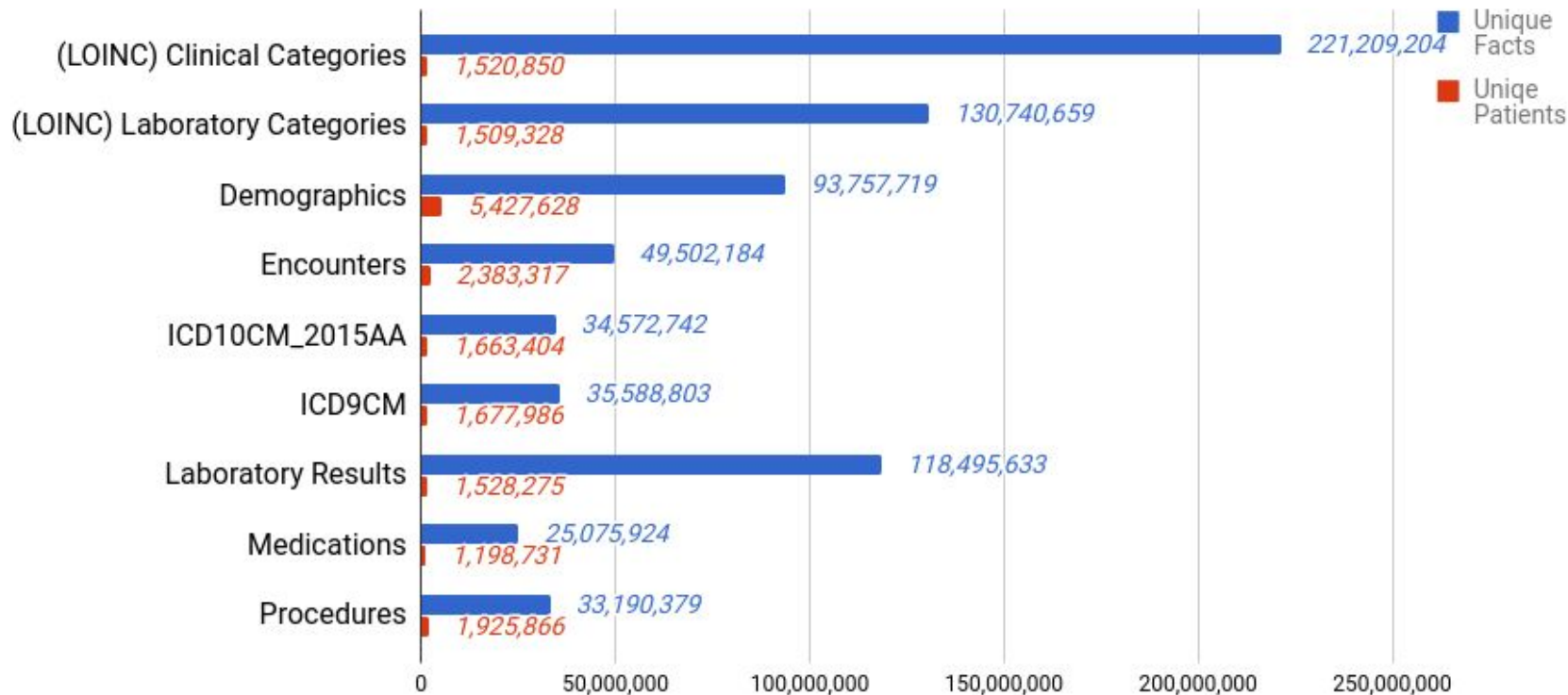
Diverse Modalities: Images (NYU data)

Computed tomography, maxillofacial area
0.8%

Magnetic resonance angiography, head
0.9%

Magnetic resonance (eg, proton) imaging, abdomen
2.0%

Radiologic examination, abdomen
2.2%

Computed tomography, head or brain
3.3%

Dual-energy X-ray absorptiometry (DXA), bone density study, 1 or more sites
3.9%

Diagnostic Nuclear Medicine Procedures
4.6%

Bone/Joint Studies
4.8%

Magnetic resonance (eg, proton) imaging, brain (including brain stem)
5.2%

Computed tomography, thorax
5.7%

Computed tomography, abdomen and pelvis
5.8%

Diagnostic Ultrasound Procedures
39.8%

Breast, Mammography
15.4%

2,886,893

1,117,349

423,278

413,525

378,170

351,016

330,748

Diverse Modalities: Genomics (Public GDC data)
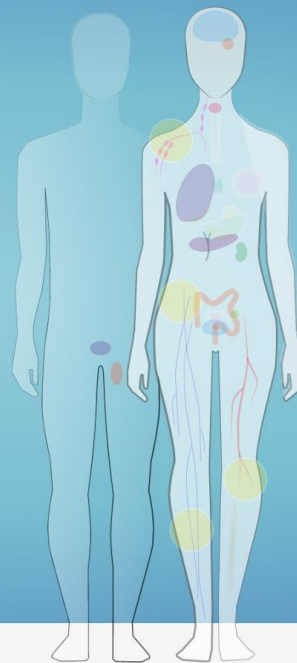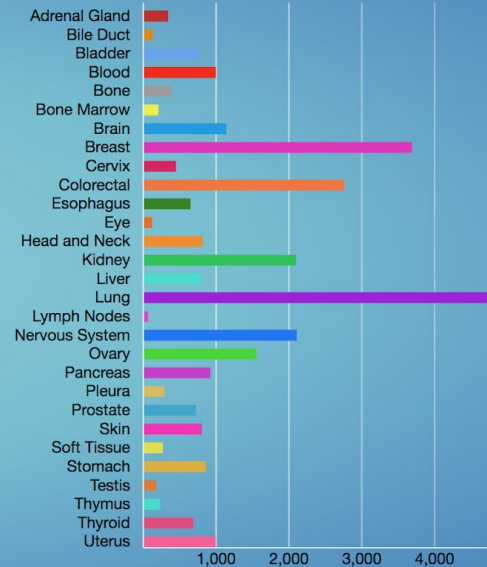
What else?

Questions that Could Use More ML in Healthcare

Early detection, Detection, and Prevention

  Automated/Augmented Diagnosis/screening & Lowering medical errors

  Finding new bio-makers, less invasive, more specific & sensitive, scalable

  Better clinical trial recruitment - faster drug design

Tracking Treatment Response and Disease Progression

  Finding, measuring, and visualizing biomarker & changes over time

Low resource settings & where time is limited i.e. ED department

  Prioritization of patients

  Lowering missed diagnosis - augmented diagnosis, automations, etc

What else?

Some snapshots of research on machine learning in healthcare

Early Disease Prediction using EHR time series

Electronic Health Records

Demographic and lifestyle ——————————————————————————————————→ Time

Medications:
-NDC code (drug name)
-Quantity
-Date of fill

Encounters
-Free Text Notes
-Diagnosis code (ICD10s)
-Procedure (CPTs)
-Specialty
-Location of service
-Service Provider ID
-Inpatient/outpatient
-Cost

Radiology Imaging:
- MRI, CT, PET, etc.
- Free Text (Radiology notes)
- Assessment codes

Pathology:
- Microscopic images (histopathology)
- Genetic test
- Free text assessments

Lab Tests:
-LOINC code (urine or blood test name)
-Results (actual values/Flags)
-Date

Electronic Health Records

Demographic and lifestyle

Time

Medications:
-NDC code (drug name)
-Quantity
-Date of fill

Encounters
-Free Text Notes
-Diagnosis code (ICD10s)
-Procedure (CPTs)
-Specialty
-Location of service
-Service Provider ID
-Inpatient/outpatient
-Cost

Radiology Imaging:
- MRI, CT, PET, etc.
- Free Text (Radiology notes)
- Assessment codes

Pathology:
- Microscopic images (histopathology)
- Genetic test
- Free text assessments

Lab Tests:
-LOINC code (urine or blood test name)
-Results (actual values/Flags)
-Date

Electronic Health Records

Demographic and lifestyle ———————————————————————→ Time

Medications:
-NDC code (drug name)
-Quantity
-Date of fill

Encounters
- **Free Text Notes**
-Diagnosis code (ICD10s)
-Procedure (CPTs)
-Specialty
-Location of service
-Service Provider ID
-Inpatient/outpatient
-Cost

Radiology Imaging:
- MRI, CT, PET, etc.
- **Free Text (Radiology notes)**
- Assessment codes

Lab Tests:
-LOINC code (urine or blood test name)
-Results (actual values/Flags)
-Date

Pathology:
- Microscopic images (histopathology)
- Genetic test
- **Free text assessments**

Electronic Health Records

Time →

Demographic and lifestyle

Medications:
-NDC code (drug name)
-Quantity
-Date of fill

Encounters
-Free Text Notes
-Diagnosis code (ICD10s)
-Procedure (CPTs)
-Specialty
-Location of service
-Service Provider ID
-Inpatient/outpatient
-Cost

Radiology Imaging:
- **MRI, CT, PET, etc.**
- Free Text (Radiology notes)
- Assessment codes

Pathology:
- **Microscopic images (histopathology)**
- Genetic test
- Free text assessments

Lab Tests:
-LOINC code (urine or blood test name)
-Results (actual values/Flags)
-Date

Disease Prediction/Forecasting

Space of machine learning methods

Feature interactions

	Specified by human experts	+Learned
Specified by human experts	- Standard Regression - Rule Based Expert Systems - Bayesian networks *Parameters: Few* *Data Needed: Small*	- Decision Trees - Bayesian networks with structure learning - Random Forests *Parameters: Medium* *Data Needed: Medium/large*
+Learned	- Bayesian networks with hidden variables - Dimensionality reduction - PCA/ICA *Parameters: Medium* *Data Needed: Medium*	- Deep learning *Parameters: Larges* *Data Needed: Large/X-Large*

Complex features

Disease Prediction/Forecasting

Electronic Health Records

Time

Demographic and lifestyle

Medications:
-NDC code (drug name)
-Quantity
-Date of fill

Encounters
-Free Text Notes
-Diagnosis code (ICD10s)
-Procedure (CPTs)
-Specialty
-Location of service
-Service Provider ID
-Inpatient/outpatient
-Cost

Radiolog
- MRI, CT, PET, etc.
- Free Text (Radiology notes)
- Assessment codes

The Model

Lab Tests:
-LOINC code (urine or blood test name)
-Results (actual values/Flags)
-Date

Pathology:
- Microscopic images (histopathology)
- Genetic test
- Free text assessments

Feature Engineering: ~42,000 features

22	39	990	16,632	233	224	7x1000	228	32

990: indicator for using Medication groups

16,632: indicator for each icd9 diagnosis

39: coverage

233: indicator for each ICD-9 procedures group

224: indicator for each CPT group

22: Diabetes known risk factors

7x1000: Laboratory indicators for:
Test request
Test value high
Test value low
Test value normal
Test value increasing
Test value decreasing
Test value fluctuating

228: Indicator for each specialty

32: Indicator for each service place

- All variables except ICD-9 diagnosis evaluated in 6 months, 2 years and entire history prior to T2D onset.

Learning features and Deep Learning/Multitask learning

Input labs

Time

A
B
C
D
E

Lab Combination Subnetwork: Vertical convolution to combine labs

Vertical Convolution (+Relu+batchnorm)
(Kernel sizes: |Labs| x 1)

Vertical Convolution (+Relu+batchnorm)
(Kernel sizes: |previous layer filters | x 1)

Temporal Max pool

Temporal Subnetwork: Temporal pooling and temporal convolution

Temporal Convolution (+ Relu +BatchNorm)

2 Layers of Dropout + Fully connected +ReLU

batchnorm +Log Softmax

$P(Y_1=1|input)$

$P(Y_3=1|input)$

$P(Y_M=1|input)$

Long Short Term
Memory Recurrent
Units

batchnorm
+Log Softmax

$P(Y_1=1|input)$

$P(Y_3=1|input)$

$P(Y_M=1|input)$

Input
labs

A
B
C
D
E

Time

2 Layers of
Dropout + Fully connected
+ReLU
Connected to the **last**
LSTM memory unit

Prediction Quality on the test set of size 98,000 individuals

ICD9 Code and disease description	LR	LSTM	CNN1	CNN2	Ens	Pos
585.6 End stage renal disease	0.886	0.917	0.910	0.916	0.920	837
285.21 Anemia in chr kidney dis	0.849	0.866	0.868	0.880	0.879	1598
585.3 Chr kidney dis stage III	0.846	0.851	0.857	0.858	0.864	2685
584.9 Acute kidney failure NOS	0.805	0.820	0.828	0.831	0.835	3039
250.01 DMI wo cmp nt st uncntrl	0.822	0.813	0.819	0.825	0.829	1522
250.02 DMII wo cmp uncntrld	0.814	0.819	0.814	0.821	0.828	3519
593.9 Renal and ureteral dis NOS	0.757	0.794	0.784	0.792	0.798	2111
428.0 CHF NOS	0.739	0.784	0.786	0.783	0.792	3479
V053 Need prphyl vc vrl hepat	0.731	0.762	0.752	0.780	0.777	862
790.93 Elvtd prstate spcf antgn	0.666	0.758	0.761	0.768	0.772	1477
185 Malign neopl prostate	0.627	0.757	0.751	0.761	0.768	761
274.9 Gout NOS	0.746	0.761	0.764	0.757	0.767	1529
362.52 Exudative macular degen	0.687	0.752	0.750	0.757	0.765	538
607.84 Impotence, organic orign	0.663	0.739	0.736	0.748	0.752	1372
511.9 Pleural effusion NOS	0.708	0.736	0.742	0.746	0.749	2701
616.10 Vaginitis NOS	0.692	0.736	0.736	0.746	0.747	440
600.01 BPH w urinary obs/LUTS	0.648	0.737	0.737	0.738	0.747	1681
285.29 Anemia-other chronic dis	0.672	0.713	0.725	0.746	0.739	1075
346.90 Migrne unsp wo ntrc mgrn	0.633	0.736	0.710	0.724	0.732	471
427.31 Atrial fibrillation	0.687	0.725	0.728	0.733	0.736	3766
250.00 DMII wo cmp nt st uncntr	0.708	0.718	0.708	0.719	0.728	3125
425.4 Prim cardiomyopathy NEC	0.683	0.718	0.719	0.722	0.726	1414
728.87 Muscle weakness-general	0.683	0.704	0.718	0.722	0.723	4706
620.2 Ovarian cyst NEC/NOS	0.660	0.720	0.700	0.711	0.719	498
286.9 Coagulat defect NEC/NOS	0.690	0.694	0.709	0.715	0.718	958

Overview of some results so far on *general* NYUMC patient cohort

Disease	AUC Retrospective	AUC prospective
10ccs186:DM in preg	95.95%	94.50%
10ccs29:Prostate can	94.97%	94.01%
10ccs174:Fem infertil	90.88%	91.56%
10ccs171:Menstrual dx	90.13%	90.51%
10ccs108:chf;nonhp	87.67%	89.78%
10ccs654 lang and dev disorders	90.61%	89.58%
10ccs158:Chr ren fail	89.14%	88.16%
10ccs158.5:Chr ren fail stage 5+	91.38%	88.11%
10ccs50:DiabMel w/cm	90.37%	87.20%
10ccs157:Ac renl fail	84.04%	86.26%
10ccs655 anxiety disorder	87.47%	86.21%
10ccs26:Cervix cancr	88.15%	85.59%
10ccs173:Menopausl dx	85.55%	85.44%
10ccs652 hyperactivity	78.94%	85.09%
10ccs25:Uterus cancr	80.96%	84.49%
10ccs653: Alzheimer's and Dementia	85.61%	83.12%
10ccs24:Breast cancr	85.06%	82.93%
10ccs199:Ulcer skin	86.42%	82.38%
10ccs101:Coron athero	84.45%	82.29%
10ccs79:Parkinson-s	87.62%	82.28%
10ccs99:Htn complicn	86.65%	82.17%
10ccs172:Ovarian cyst	82.11%	81.14%
10ccs107:Cardia arrst	75.35%	80.94%
10ccs40:Mult myeloma	81.19%	80.38%

Applicable to many more outcomes and tasks

- Early prediction of childhood obesity
- Predicting diabetes complications
- Predicting risk of re-hospitalization
- Detecting undocumented but existing diseases
- Using lab values only to predict future diseases
- Predicting medication adherence
- Predicting no-shows
- Etc. etc. etc.…

- Many industries interested: Hospitals, Insurance companies, Government Medicare/Medicaid, Center for Disease Control, etc.

Medical Imaging:
Radiology (X-rays, Mammograms, MRI, Ultrasound)
Pathology
Microscopy

Plain X-Rays or Radiographs

Most common & oldest type of radiology image.

Great to show *Carbon vs. Calcium*

Good for: Bones, Teeth, Chest X-Rays, Mammography, Abdominal X-ray.

Result: 2D image

Risks: Radiation exposure

Opportunities in research:
- Augmented/automatic Diagnosis
- Lowering X-ray dosage

Projectional radiography

X-ray generator

Object

X-ray detector
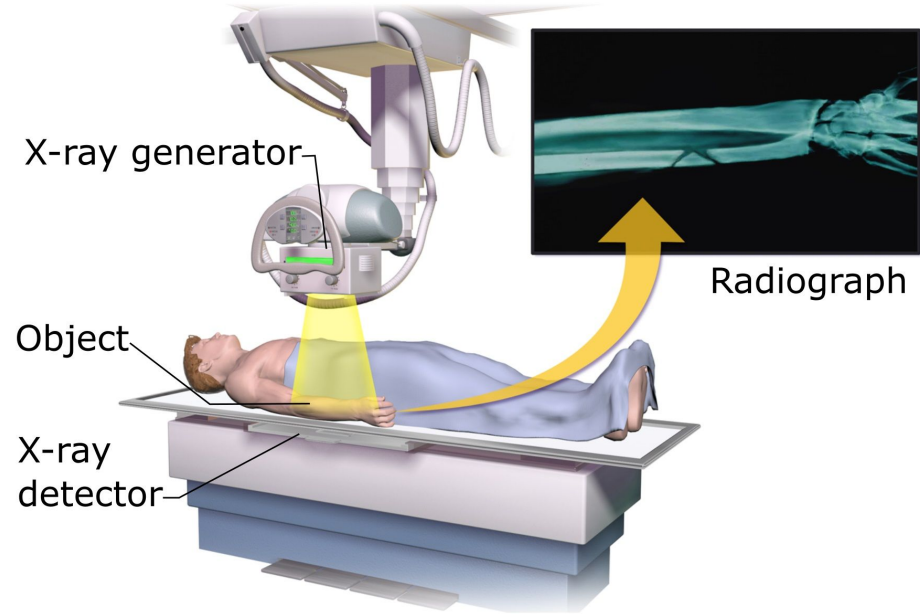
Radiograph

Related Papers on Bone X-Ray Radiographs

MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs

MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs

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

[*, †]Equal Contribution

[1] Department of Computer Science
Stanford University
{pranavsr, jirvin16}@cs.stanford.edu

[2]Department of Medicine
Stanford University
rball@stanford.edu

[3]Department of Radiology
Stanford University
mlungren@stanford.edu

MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs

Task: determining whether an X-ray study is normal or abnormal.

Motivation:
- Musculoskeletal conditions affect more than 1.7 billion people worldwide,
- 30 million emergency department visits annually

Data (Public):
- 14,863 studies from 12,173 patients, with a total of 40,561 multi-view radiographic images.
- Includes: elbow, finger, forearm, hand, humerus, shoulder, and wrist
- Labels from Stanford Hospital (from 2001 to 2012)

Baseline:
- **DenseNet-169** with **Multi-task Cross Entropy Loss**

Evaluation:
- Cohen's kappa statistic
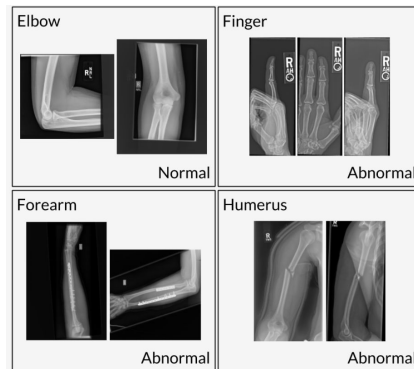
MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs

Leaderboard

Will your model perform as well as radiologists in detecting abnormalities in musculoskeletal X-rays?

Rank	Date	Model	Kappa
		Best Radiologist Performance *Stanford University* Rajpurkar & Irvin et al., 17	0.778
1	Aug 22, 2018	base-comb3(ensemble) *jtz Availink*	**0.805**
2	Sep 14, 2018	double_res(ensemble model) *SCU_MILAB*	0.804
3	Jul 24, 2018	he_j	0.775
4	Aug 19, 2018	ianpan (ensemble) *RIH 3D Lab*	0.774
5	Jul 24, 2018	he_j	0.774
6	Jun 17, 2018	gcm (ensemble) *Peking University*	0.773
6	Sep 10, 2018	ty101 *single model*	0.773
7	Aug 31, 2018	he_j	0.764
7	Aug 31, 2018	AIAPlus (ensemble) *Taiwan AI Academy* http://aiacademy.tw	0.764
8	Sep 04, 2018	SER_Net_Baseline (single model) *SJTU*	0.764
9	Jul 14, 2018	Trs (single model) *SCU_MILAB*	0.763

Related paper on Chest X-rays

"ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases"

ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases

Xiaosong Wang[1], Yifan Peng [2], Le Lu [1], Zhiyong Lu [2], Mohammadhadi Bagheri [1], Ronald M. Summers [1]
[1]Department of Radiology and Imaging Sciences, Clinical Center,
[2] National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, Bethesda, MD 20892
{xiaosong.wang,yifan.peng,le.lu,luzh,mohammad.bagheri,rms}@nih.gov

Abstract

The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). On the other side, it is still an open question how this type of hospital-size knowledge database containing invaluable imaging informatics (i.e., loosely labeled) can be used to facilitate the data-hungry deep learning paradigms in building truly large-scale high precision computer-aided diagnosis (CAD) systems.
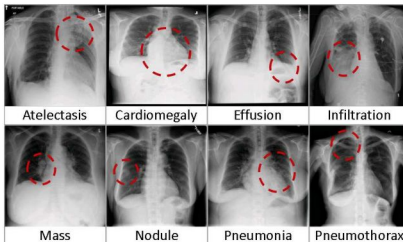
Figure 1. Eight common thoracic diseases observed in chest X-rays that validate a challenging task of fully-automated diagnosis.

"ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases"

Task: Identification & Localization of Thorax Diseases.

Motivation: Reducing medical errors and improving "incidental finding" success.

The data:
- 108,948 frontal view X-ray images of 32,717 unique patients
- Labels from radiology reports. (8 disease labels)

Evaluation: AUC

Baseline: Standard imaging models up to 2017

"ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases"

Item #	OpenI	Ov.	ChestX-ray8	Ov.
Report	2,435	-	108,948	-
Annotations	2,435	-	-	-
Atelectasis	315	122	5,789	3,286
Cardiomegaly	345	100	1,010	475
Effusion	153	94	6,331	4,017
Infiltration	60	45	10,317	4,698
Mass	15	4	6,046	3,432
Nodule	106	18	1,971	1,041
Pneumonia	40	15	1,062	703
Pneumothorax	22	11	2,793	1,403
Normal	1,379	0	84,312	0

Setting	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
Initialization with different pre-trained models								
AlexNet	0.6458	0.6925	0.6642	0.6041	**0.5644**	0.6487	0.5493	0.7425
GoogLeNet	0.6307	0.7056	0.6876	0.6088	0.5363	0.5579	0.5990	0.7824
VGGNet-16	0.6281	0.7084	0.6502	0.5896	0.5103	0.6556	0.5100	0.7516
ResNet-50	**0.7069**	**0.8141**	**0.7362**	**0.6128**	0.5609	**0.7164**	**0.6333**	**0.7891**
Different multi-label loss functions								
CEL	0.7064	0.7262	0.7351	0.6084	0.5530	0.6545	0.5164	0.7665
W-CEL	0.7069	0.8141	0.7362	0.6128	0.5609	0.7164	0.6333	0.7891

Table 3. *AUCs of ROC curves for multi-label classification in different DCNN model setting.*

T(IoBB)	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
T(IoBB) = 0.1								
Acc.	0.7277	0.9931	0.7124	0.7886	0.4352	0.1645	0.7500	0.4591
AFP	0.0823	0.0487	0.0589	0.0426	0.0691	0.0630	0.0691	0.0264
T(IoBB) = 0.25 (Two times larger on both x and y axis than ground truth B-Boxes)								
Acc.	0.5500	0.9794	0.5424	0.5772	0.2823	0.0506	0.5583	0.3469
AFP	0.1666	0.1534	0.1189	0.0914	0.0975	0.0741	0.1250	0.0487
T(IoBB) = 0.5								
Acc.	0.2833	0.8767	0.3333	0.4227	0.1411	0.0126	0.3833	0.1836
AFP	0.2703	0.2611	0.1859	0.1422	0.1209	0.0772	0.1768	0.0772
T(IoBB) = 0.75								
Acc.	0.1666	0.7260	0.2418	0.3252	0.1176	0.0126	0.2583	0.1020
AFP	0.3048	0.3506	0.2113	0.1737	0.1310	0.0772	0.2184	0.0873
T(IoBB) = 0.9								
Acc.	0.1333	0.6849	0.2091	0.2520	0.0588	0.0126	0.2416	0.0816
AFP	0.3160	0.3983	0.2235	0.1910	0.1402	0.0772	0.2317	0.0904

Table 4. *Pathology localization accuracy and average false positive number for 8 disease classes.*

Follow-up: CheXNet (Also a DenseNet model)

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.

Input
Chest X-Ray Image

CheXNet
121-layer CNN

Output
Pneumonia Positive (85%)
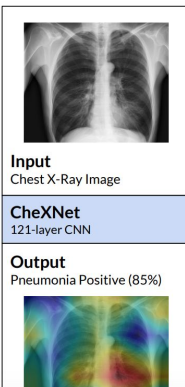
1. Introduction

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**

	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

Criticism of the Dataset (Applies to most datasets)

Labels aren't accurate

In all the images, red = clearly wrong label. Orange = I doubt it, I wouldn't report it, but I can't really exclude it. Correlate clinically :p

Atelectasis
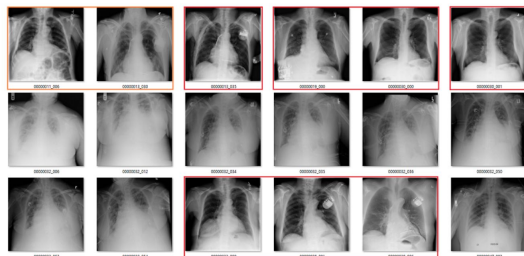
Fibrosis

Cardiomegaly

Read:https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/
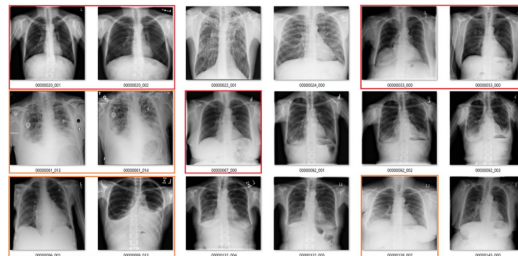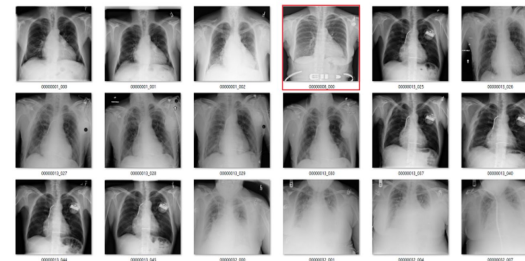
Mammograms: Low-dose X-Rays

Screening Mammograms: 4 images
Diagnostic Mammograms: More than
4 images

Currently recommended once every 2
years for every 50-74 yo women.

Does not work for dense breasts.
(Many young patients or asian
ethnicities)
- Ultrasound

Related paper on automatic Mammography *Screening*

High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks

Krzysztof J. Geras[1,3], Stacey Wolfson[3], Yiqiu Shen[1], Nan Wu[1], S. Gene Kim[3,4], Eric Kim[3], Laura Heacock[3], Ujas Parikh[3], Linda Moy[3,4], Kyunghyun Cho[1,2,5]

Abstract—Advances in deep learning for natural images have prompted a surge of interest in applying similar techniques to medical images. The majority of the initial attempts focused on replacing the input of a deep convolutional neural network with a medical image, which does not take into consideration the fundamental differences between these two types of images. Specifically, fine details are necessary for detection in medical images, unlike in natural images where coarse structures matter most. This difference makes it inadequate to use the existing network architectures developed for natural images, because they work on heavily downscaled images to reduce the memory requirements. This hides details necessary to make accurate predictions. Additionally, a single exam in medical imaging often comes with a set of views which must be fused in order to reach a correct conclusion. In our work, we propose to use a multi-view deep convolutional neural network that handles a set of high-resolution medical images. We evaluate it on large-scale mammography-based breast cancer screening (BI-RADS prediction) using 886,000 images. We focus on investigating the impact of the training set size and image size on the prediction accuracy. Our results highlight that performance increases with the size of training set, and that the best performance can only be achieved using the original resolution. In the reader study, performed on a random subset of the test set, we confirmed the efficacy of our model, which achieved performance comparable to a committee of radiologists when presented with the same data.

Index Terms—breast cancer screening, deep convolutional neural networks, deep learning, machine learning, mammography

screening interval for mammograms has been the subject of public debate with different professional societies offering varying guidelines for mammographic screening [2], [3], [4], [5]. In particular, there has been public discussion regarding the potential harms of screening. These harms include false positive recalls and false positive biopsies as well as anxiety caused by recall for diagnostic testing after a screening exam. Overall, the recall rate following a screening mammogram is between 10-15%. This equates to about 3.3 to 4.5 million callback exams for additional testing [6].

The vast majority of the women asked to return following an inconclusive mammogram undergo another mammogram and/or ultrasound for clarification. Most of these false positive findings are found to represent normal breast tissue. Only 10% to 20% of women who have an abnormal screening mammogram are recommended to undergo a biopsy. Only 20-40% of these biopsies yield a diagnosis of cancer [7]. In 2014, over 39 million screening and diagnostic mammography exams were performed in the US. Therefore, in addition to the anxiety from undergoing a false positive mammogram, there are significant costs associated with unnecessary follow ups and biopsies. Clearly, there is an unmet need to shift the balance of routine breast cancer screening towards more benefit and less harm.

High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks

Data: 886,000 images, 129,208 unique patients

	BI-RADS 0	BI-RADS 1	BI-RADS 2
Training	21946 / 95471	74832 / 327035	67446 / 298680
Validation	2634 / 11471	11542 / 50627	10376 / 46178
Test	1341 / 5871	5986 / 26213	5595 / 24891

Labels: BI-RADs scores

Baseline: Custom CNN

Evaluation: AUC & Reader Study

	radiologists	MV-DCN	radiologists + MV-DCN
0 vs. others	0.650	0.547	0.653
1 vs. others	0.765	0.757	0.792
2 vs. others	0.699	0.759	0.759
macAUC	0.704	0.688	0.735

layer	kernel size	stride	#maps	repetition
global average pooling			256	
convolution	3×3	1×1	256	×3
max pooling	2×2	2×2	128	
convolution	3×3	1×1	128	× 3
max pooling	2×2	2×2	128	
convolution	3×3	1×1	128	× 3
max pooling	2×2	2×2	64	
convolution	3×3	1×1	64	× 2
convolution	3×3	2×2	64	
max pooling	3×3	3×3	32	
convolution	3×3	2×2	32	
input			1	

Fig. 2. Description of one deep convolutional network column for a single view. It transforms the input view (a gray-scale image) into a 256-dimensional vector.

Classifier $p(y|x)$			
Fully connected layer (1024 hidden units)			
Concatenation (256×4 dim)			
DCN	DCN	DCN	DCN
L-CC	R-CC	L-MLO	R-MLO
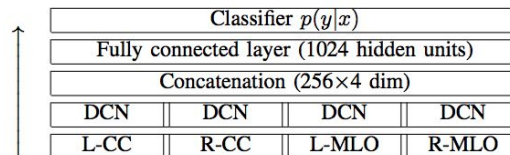
Fig. 3. An overview of the proposed multi-view deep convolutional network. DCN refers to the convolutional network network column from Figure 2. The arrow indicates the direction of information flow.

Magnetic Resonance Imaging (MRI)

Watch (25 mins): **https://www.youtube.com/watch?v=djAxjtN_7VE**

- Protons (Hydrogen nuclei) rotate randomly.
- A rotating positive charge creates magnetic field.
- If put under a bigger magnetic field, the proton spins somewhat lines-up.
- If exposed to radio-frequency proportional to the magnetic field, they flip.
- As the radio-frequency is removed, they emit a measurable signal (Phase & Frequency & Magnitude) as they go back.
  - **Fat** has different reaction to this removal vs **Water**
  - **Pulse Sequence:** Order of applying and removing radio-frequency.
  - Can localize each measured signal by creating asymmetric large magnetic waves.
  - MRI signal is originally captured in Fourier Space
  - Currently 1.5 T, 3 T, 7 Tesla clinically available.
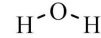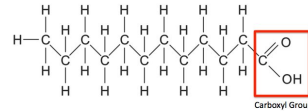
Pulse Sequences: T1 vs T2 vs FLAIR vs DTI vs ...
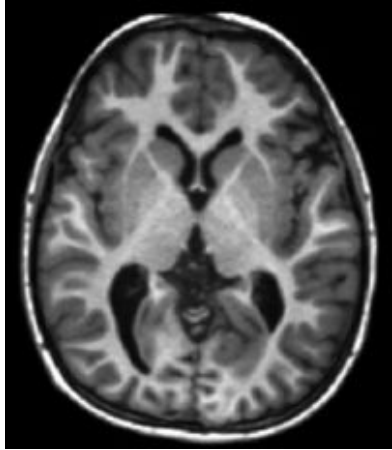
T1:
Brighter: Fat and Contrast agents
Darker: Higher water content: (edema, tumor, infarction, inflammation, infection, hemorrhage)
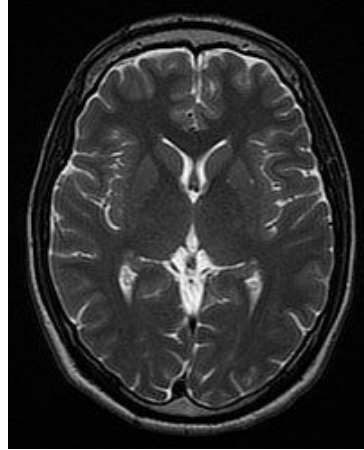
T2:
Brighter: Water
Darker: Fat tissue

FLAIR: High signal in stroke, multiple sclerosis (MS) plaques, subarachnoid haemorrhage and meningitis.
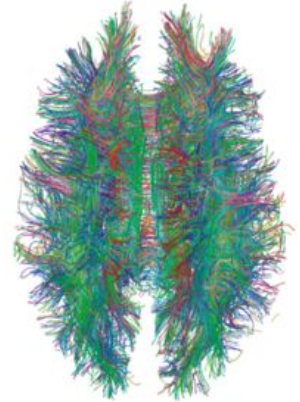
DTI:
Measures of Brownian motion of water molecules
Can image direction of nerve fibers
Useful for tumor deformation studies

MRI is originally in **Fourier Space** - called K-Space

(a)

(b)
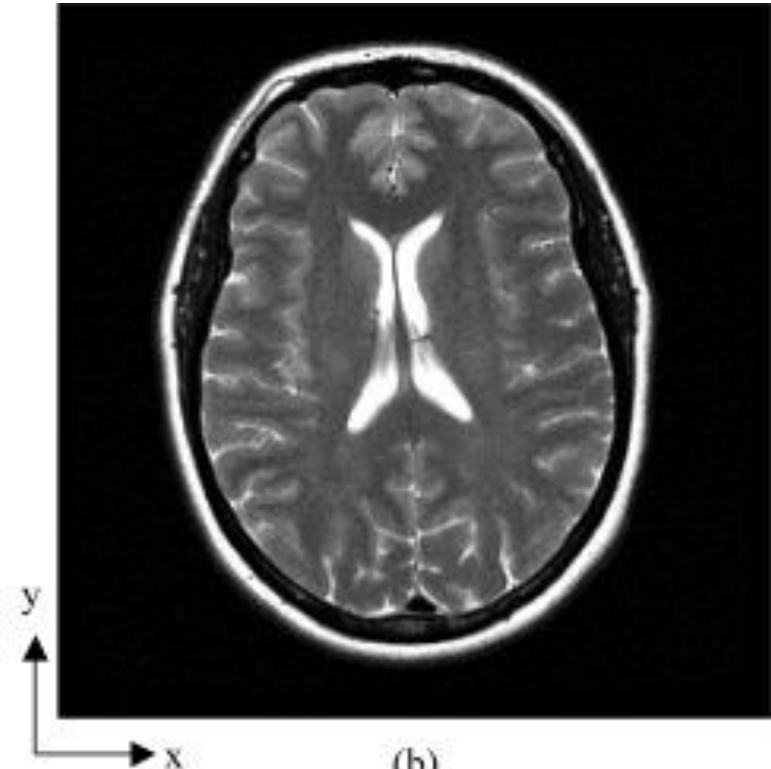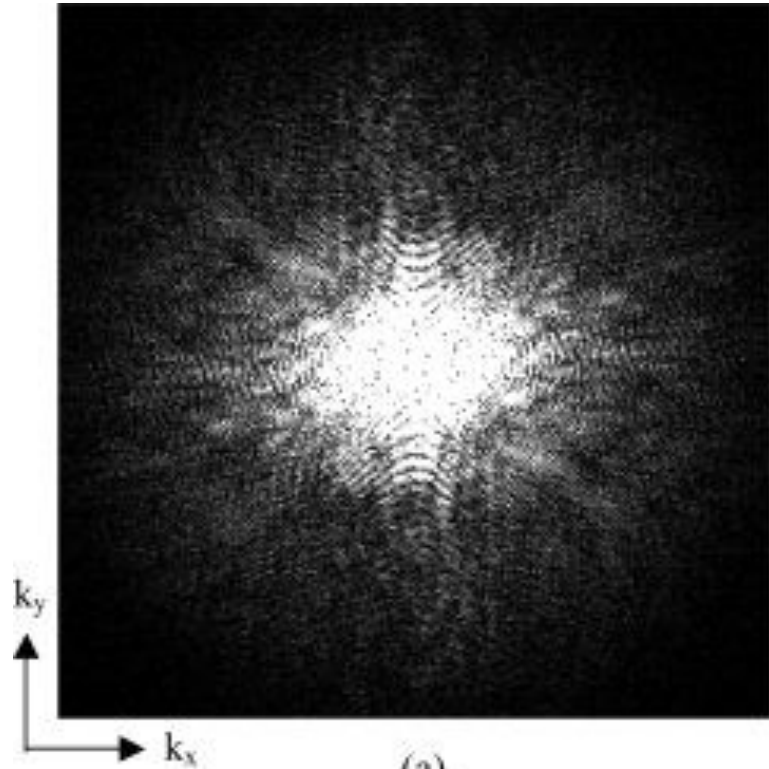
Missing data in K-space leads to pixel space artifacts
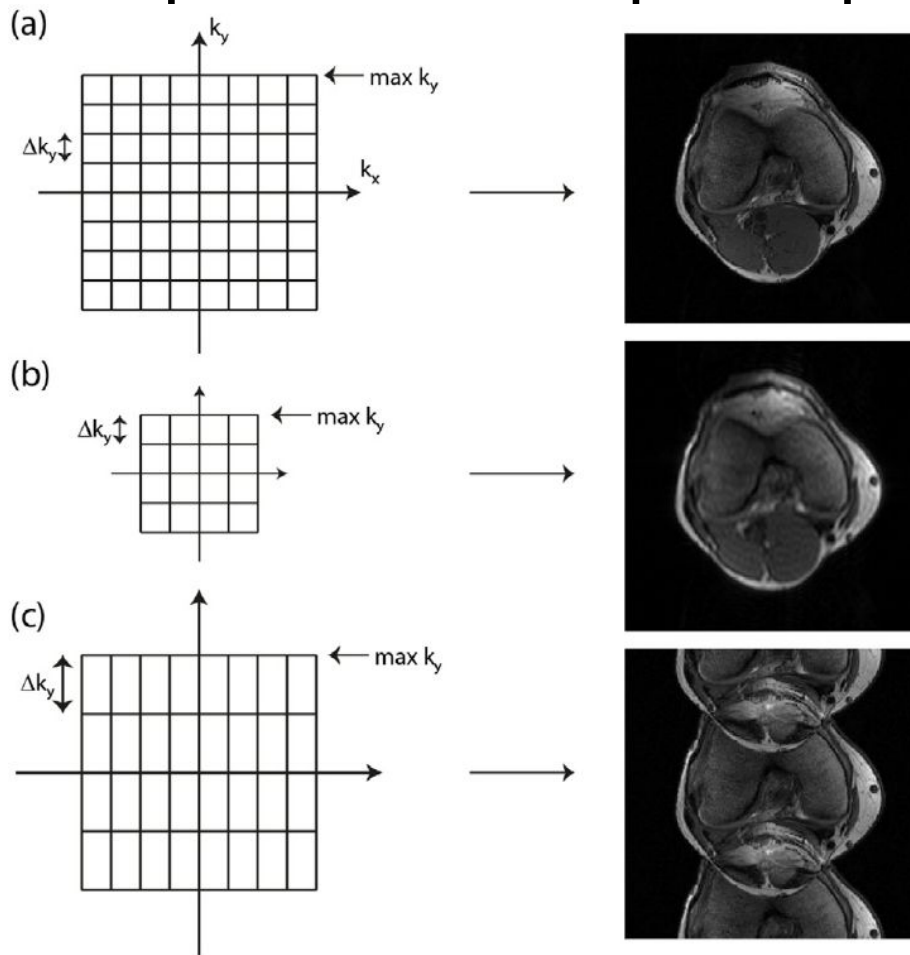
Issues and Potentials for Research

Improving Acquisition time & Image reconstruction

   15/20 minutes stuck inside a tube: too long!

Diagnosis and automation:

   2D and 3D classifiers, localization, segmentation

   Time series alignment, classification, visualization

Advanced Imaging Invention

   MRI fingerprinting and diagnosis

Segmentation of MRIs: Brain

"QuickNAT: Segmenting MRI Neuroanatomy in 20 seconds"

QuickNAT: Segmenting MRI Neuroanatomy in 20 seconds

Abhijit Guha Roy, Sailesh Conjeti, Nassir Navab and Christian Wachinger

Affiliations:

A. Guha Roy, C. Wachinger
Laboratory for Artificial Intelligence in Medical Imaging (AI-Med),
Department of Child and Adolescent Psychiatry, Psychosomatic and Psychotherapy,
Ludwig-Maximilian-University,
Waltherstr. 23, 80337 Munich, Germany

A. Guha Roy, S. Conjeti, N. Navab
Chair for Computer Aided Medical Procedures, Department of Informatics
Technical University of Munich,
Boltzmannstr. 3, 85748 Garching, Munich, Germany.

S. Conjeti
German Center for Neurodegenerative Diseases (DZNE)

"QuickNAT: Segmenting MRI Neuroanatomy in 20 seconds"

Motivation:
- Accurate brain structural segmentation is central to nearly all neuroimaging analyses.
- Freesurfer takes 2-4 hours to segment a volume.

Task: Segmentation of 40+ regions per volume

Data: ADNI Auxiliary data & MICCAI brain segmentation challenge (30 manual segmented volumes)

Baseline: Variant of U-net

Loss function: Weighted cross entropy & Weighted Dice loss $Dice = \frac{2 \cdot |mask \cap prediction|}{|mask| + |prediction|}$

(i) Ground Truth

(a) QuickNAT Architecture

Coronal

Axial

Sagittal

3D MRI T1 volume

View Aggregation Stage

3D Whole Brain Segmentation
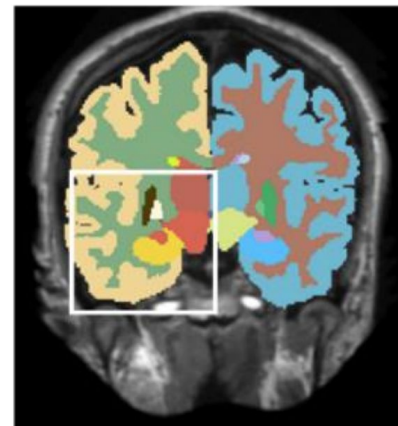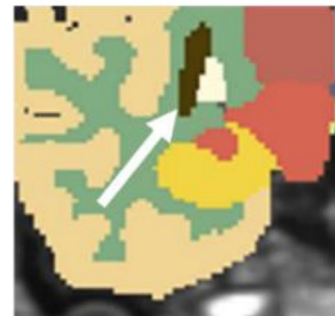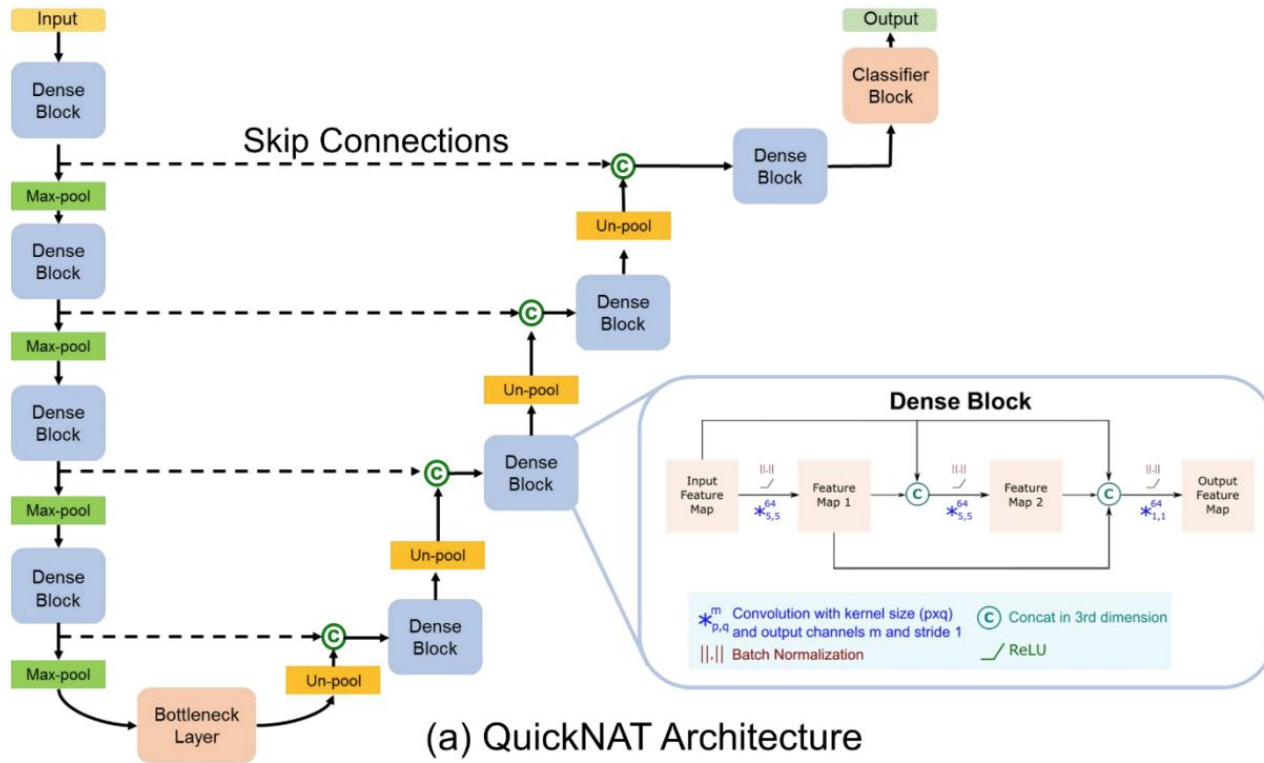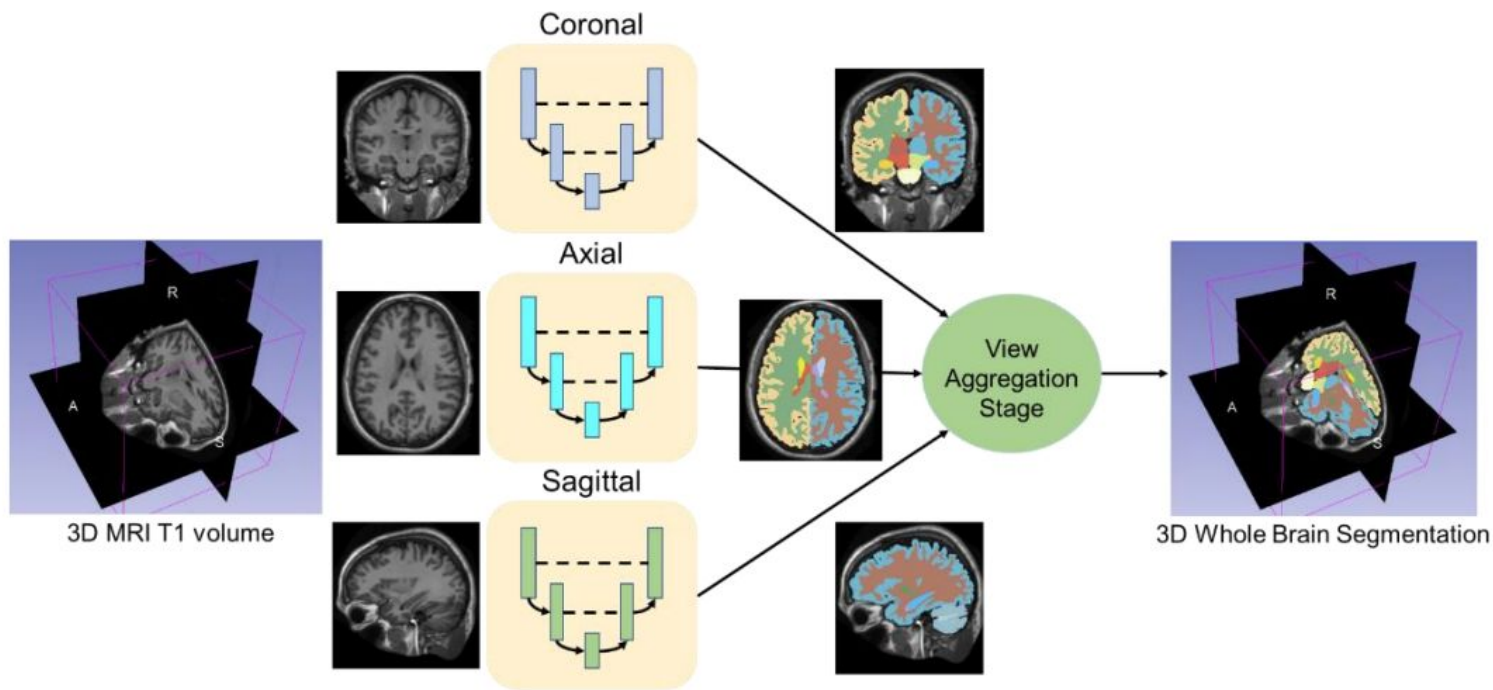
(b) Multi-View Aggregation for final prediction
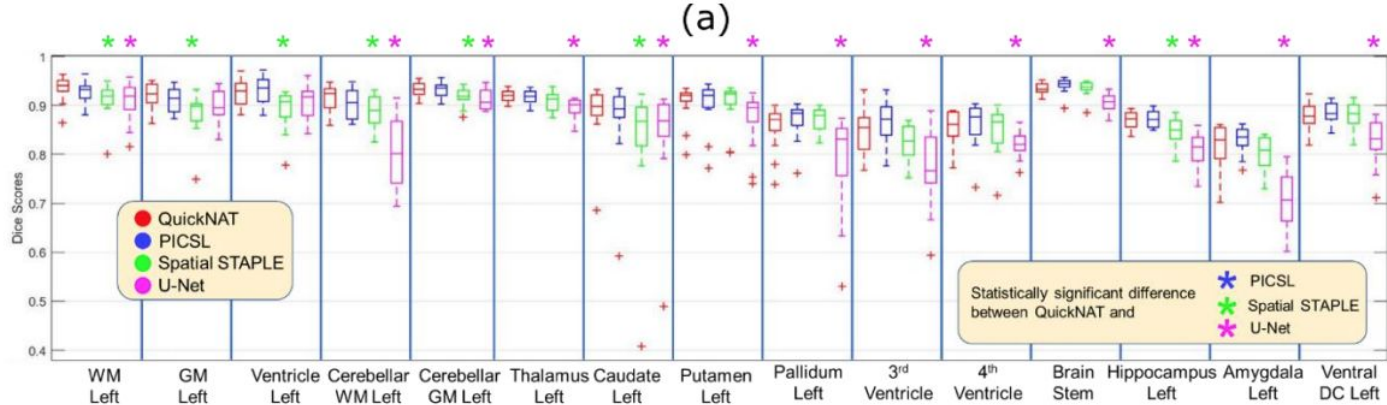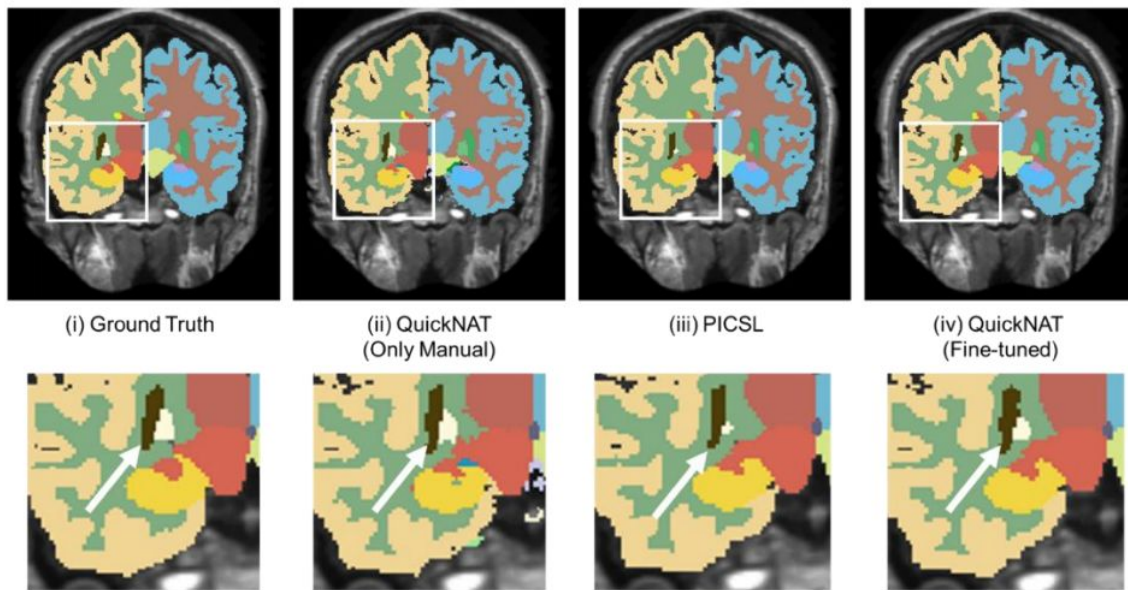
(a)

(b)

(i) Ground Truth

(ii) QuickNAT (Only Manual)

(iii) PICSL

(iv) QuickNAT (Fine-tuned)

"End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification"

https://arxiv.org/pdf/1810.00523.pdf

End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification

Soheil Esmaeilzadeh[1], Dimitrios Ioannis Belivanis[1],
Kilian M. Pohl[2], and Ehsan Adeli[1]

[1]Stanford University   [2]SRI International
{soes, dbelivan, eadeli}@stanford.edu   kilian.pohl@sri.com

Abstract. As shown in computer vision, the power of deep learning lies in automatically learning relevant and powerful features for any perdition task, which is made possible through end-to-end architectures. However, deep learning approaches applied for classifying medical images do not adhere to this architecture as they rely on several pre- and post-processing steps. This shortcoming can be explained by the relatively small number of available labeled subjects, the high dimensionality of neuroimaging data, and difficulties in interpreting the results of deep learning methods. In this paper, we propose a simple 3D Convolutional Neural Networks and exploit its model parameters to tailor the end-to-end architecture for the diagnosis of Alzheimer's disease (AD). Our model can diagnose AD with an accuracy of 94.1% on the popular ADNI dataset using only MRI data, which outperforms the previous state-of-the-art. Based on the learned model, we identify the disease biomarkers, the results of which were in accordance with the literature. We further transfer the learned model to diagnose mild cognitive impairment (MCI), the prodromal stage of AD, which yield better results compared to other methods.

3v1 [cs.CV] 1 Oct 2018

"End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification"

Task: Differentiate between AD, MCI, Normal

Dataset: ADNI (publicly available) - small-ish

Table 1: ADNI-1 subjects demographic information.

Class	Sex	Count	Age						
			mean±std	min	25%	50%	75%	max
AD	M	97	75.0±7.9	55.2	70.8	75.3	80.4	91.0
	F	103	76.1±7.4	56.5	71.1	77.0	82.3	87.9
MCI	M	265	75.4±7.3	54.6	71.0	75.4	80.7	89.8
	F	146	73.6±7.5	55.2	69.1	74.3	79.7	86.2
NC	M	112	76.1±4.7	62.2	72.5	75.8	78.5	89.7
	F	118	75.8±5.2	60.0	72.1	75.6	79.1	87.7

Architecture: 3D CNN - vanilla 3D
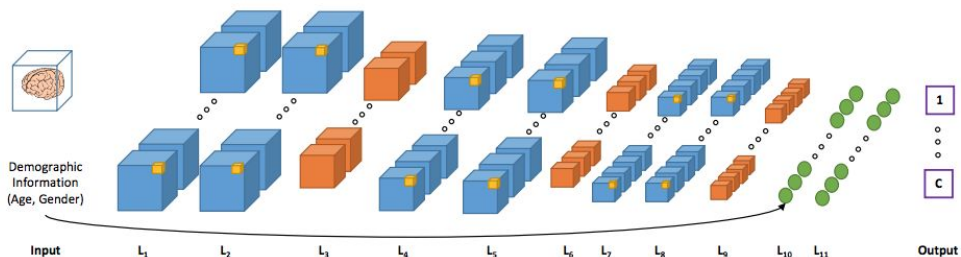
Fig. 2: 3D-CNN architecture used in this paper. The blue cubes (L_1, L_2, L_4, L_5, L_7, and L_8) are convolutional layers; Orange cubes (L_3, L_6, and L_9) are max-pooling layers; and the last two layers are fully connected (FC) layers.

Results & Visualizations

Table 2: Ablation tests: testing performance comparison of different models (last row is our model). The comparison includes the Accuracy (Acc), F_2 score, Precision (Pre), and Recall (Rec) of all methods (Reg: Regularization, D/O: Drop-Out, Aug: Augmentation).

Model	Simple				Complex			
	Acc%	F_2	Pre	Rec	Acc%	F_2	Pre	Rec
3D-CNN	68.7	0.71	0.68	0.72	66.5	0.69	0.67	0.70
3D-CNN+Reg	77.6	0.77	0.74	0.78	77.4	0.75	0.72	0.76
3D-CNN+Reg+D/O	83.1	0.811	0.78	0.82	79.7	0.82	0.79	0.84
3D-CNN+Reg+D/O+Aug (Ours)	**94.1**	**0.93**	**0.92**	**0.94**	88.3	0.89	0.88	0.91

Table 3: Comparisons with prior works for AD diagnosis.

Method	Modalities	Acc%	Sen	Spe
[12]	MRI+PET	85.7	0.99	0.54
[3]	MRI	90.8	N/A	N/A
[10]	MRI	91.1	0.88	0.93
[5]	MRI	93.9	0.94	0.93
Ours	MRI	**94.1**	0.94	0.91
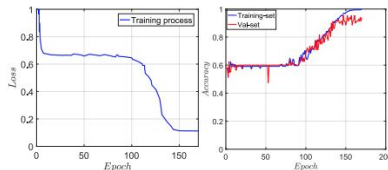
Fig. 3: (Left) training loss and (Right) training-validation accuracies with respect to the number of epochs for our 3D-CNN.
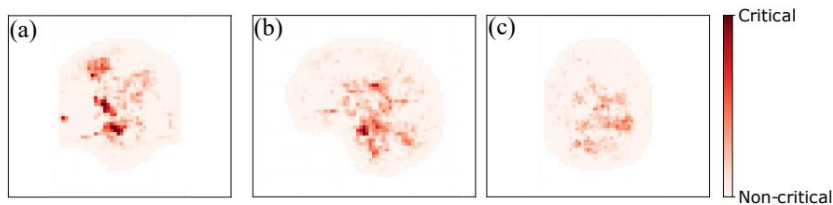
Fig. 4: Relative importance of different voxels associated with AD diagnosis.

Table 4: Testing performance for three-class Alzheimer classification.

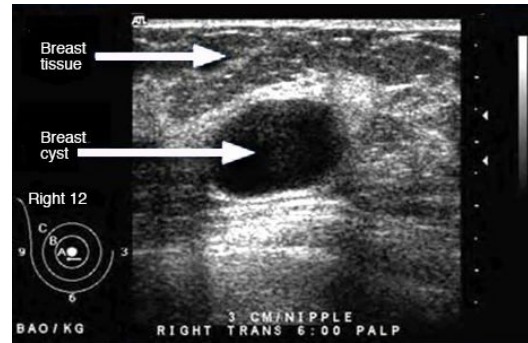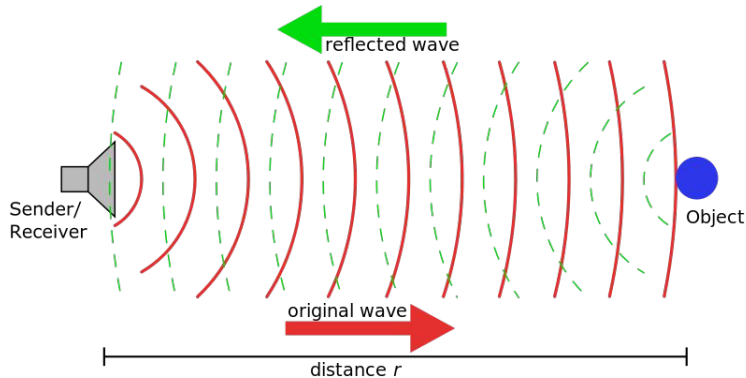| Method	Simple				Complex			
	Acc%	F_2	Pre	Rec	Acc%	F_2	Pre	Rec
3D-CNN+D/O+Reg+with learning transfer	**61.1**	**0.62**	**0.59**	**0.63**	57.2	0.59	0.55	0.61
3D-CNN+D/O+Reg+w/o learning transfer	0.54	53.4	0.49	0.55	48.3	0.50	0.45	0.52

Ultrasound Imaging or Sonography

Sound waves with frequencies - higher than those audible to humans (>20,000 Hz)

provides images in real-time

No radiation and portable

Limits on its field of view: Difficult to 'see' behind Bones and Air (for now)

Can be used to see: Elasticity of tissue, 3D shape, Tissue maps

reflected wave

Sender/
Receiver

Object

original wave

distance r

Breast tissue

Breast cyst

Right 12

9   C   3
    B
    A
        6

3 CM/NIPPLE
BAO/KG   RIGHT TRANS 6:00 PALP

© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

Related work on Segmenting Tumors in Ultrasound

"Automated and real-time segmentation of suspicious breast masses using convolutional neural network"

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5955504/

Automated and real-time segmentation of suspicious breast masses using convolutional neural network

Viksit Kumar, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing,[#1] Jeremy M. Webb, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing,[#2] Adriana Gregory, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing,[2] Max Denis, Conceptualization, Investigation, Methodology, Writing – original draft,[2,¤] Duane D. Meixner, Data curation, Visualization, Writing – original draft,[2] Mahdi Bayat, Data curation, Visualization,[1] Dana H. Whaley, Data curation, Project administration, Supervision, Visualization, Writing – review & editing,[2] Mostafa Fatemi, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing,[1] and Azra Alizad, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing[2,*]

Yong Fan, Editor

[1] Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, United States of America
[2] Department of Radiology, Mayo Clinic College of Medicine and Science, Rochester, Minnesota, United States of America
University of Pennsylvania Perelman School of Medicine, UNITED STATES
[#] Contributed equally.
Competing Interests: The authors have declared that no competing interests exist.
[¤] Current address: Physical and Life Sciences Solutions LLC, Randolph, Massachusetts, United States of America
* E-mail: alizad.azra@mayo.edu

Author information ▼ Article notes ► Copyright and License information ► Disclaimer

"Automated and real-time segmentation of suspicious breast masses using convolutional neural network"

Motivation: Detection and Localization of tumors

Model: Standard U-Net

Data:

BI-RADS distribution of patients in training/validation and test sets.

BI-RADS	No. of Patients in Training and Validation Set	No. of Patients in Testing Set
2	3	1
3	2	10
4	155	35
5	41	15
6	6	0

Evaluation: Dice Loss

Metrics		All cases (n = 61)	Benign (n = 39)	Malignant (n = 22)	IDC (n = 14)	Fibroadenoma (n = 23)
Dice coefficient	MU	0.82±0.10	0.81±0.11	0.83±0.09	0.81±0.10	0.84±0.09
	DRLS	0.84±0.09	0.82±0.10	0.87±0.07	0.87±0.06	0.84±0.06
	OU	0.52±0.27	0.48±0.28	0.57±0.24	0.55±0.28	0.48±0.27
TPF[a]	MU	0.84±0.15	0.80±0.16	0.89±0.11	0.90±0.13	0.80±0.14
	DRLS	0.79±0.12	0.76±0.12	0.83±0.12	0.83±0.12	0.77±0.10
	OU	0.61±0.06	0.55±0.06	0.70±0.05	0.68±0.07	0.57±0.04
FPF[b]	MU	0.01±0.02	0.01±0.02	0.02±0.02	0.02±0.02	0.01±0.01
	DRLS	0.01±0.02)	0.01±0.02	0.01±0.02	0.01±0.02	0.01±0.01
	OU	0.31±0.06	0.31±0.05	0.27±0.07	0.29±0.09	0.32±0.04
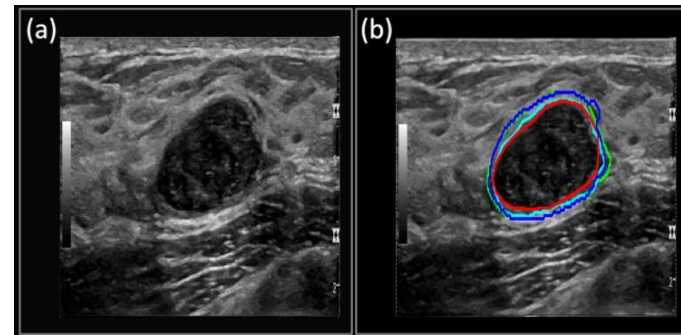
Pathology

Typical Cancer Diagnosis Process

Initial: Radiological Images

- X-Ray, CT scans, MRIs, PET

Confirmation & staging/subtyping: Pathology

- No Surgery: Needle biopsy - fine needle aspiration (FNA) or core biopsy
- Surgery and General Anesthesia: FFPE or Frozen - 1cm^3 cube or more tissue
  - FFPE: Formalin; Paraffin; Slicing; Staining with H&E
  - Frozen: Faster and takes few minutes - during surgery
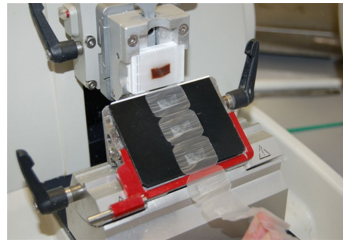
The Data: Public TCGA (The Cancer Genome Atlas)
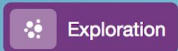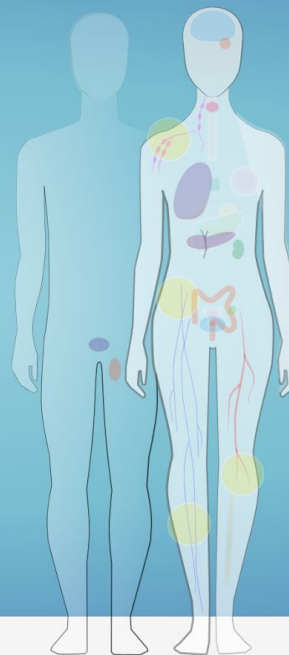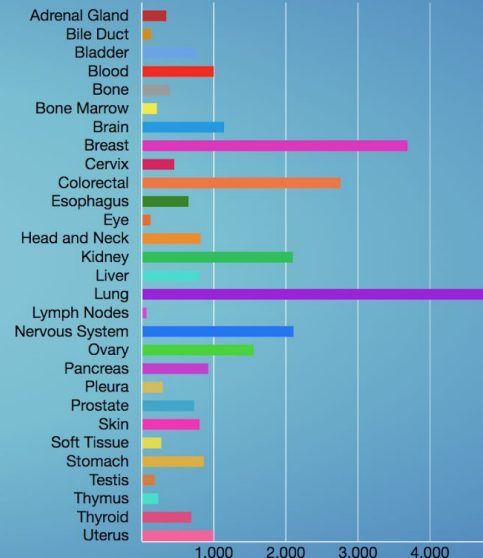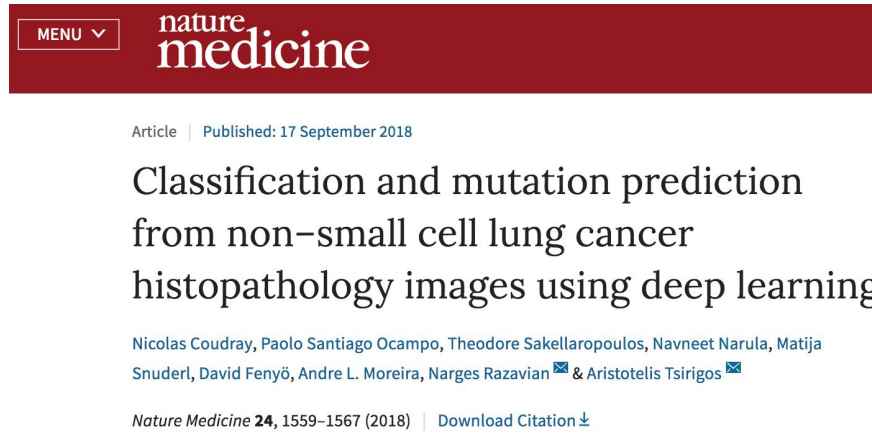
Related work: Classification of Histopathology Images

"Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning"

https://www.nature.com/articles/s41591-018-0177-5

nature medicine

Article | Published: 17 September 2018

Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning

Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L. Moreira, Narges Razavian ✉ & Aristotelis Tsirigos ✉

Nature Medicine **24**, 1559–1567 (2018) | Download Citation ⬇

Abstract

Visual inspection of histopathology slides is one of the main methods used by pathologists to assess the stage, type and subtype of lung

Lung Cancer:
Second most common cancer, and leading cause of cancer death

234,000 new cases in 2018

154,000 deaths[1]

80% are *Non-Small Cell Lung Cancer*[2]

EGFR mutations

20% in USA/Europe

60% in East Asia[3-4]

Approved Molecularly Targeted Therapies for EGFR-mutant lung cancers[5-6]

[1] USA 2018 Stats, The American Cancer Society, https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html
[2] The American Cancer Society, https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/what-is-non-small-cell-lung-cancer.html
[3] Rosell, Rafael, et al. *New England Journal of Medicine* 361.10 (2009): 958-967.
[4] https://www.mycancergenome.org/content/disease/lung-cancer/egfr/
[5] Shi, Yuankai, et al. *Journal of thoracic oncology* 9.2 (2014): 154-162.
[6] https://www.curetoday.com/articles/treatment-for-egfr-mutant-lung-cancer-is-rapidly-expanding

The Data

1,634 whole-slide images (1,176 tumor tissues and 459 normal tissues)
- For Adenocarcinoma, there are also mutations available
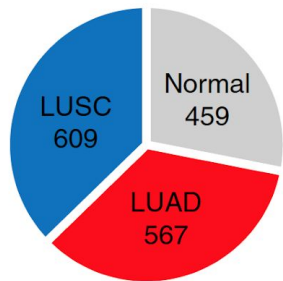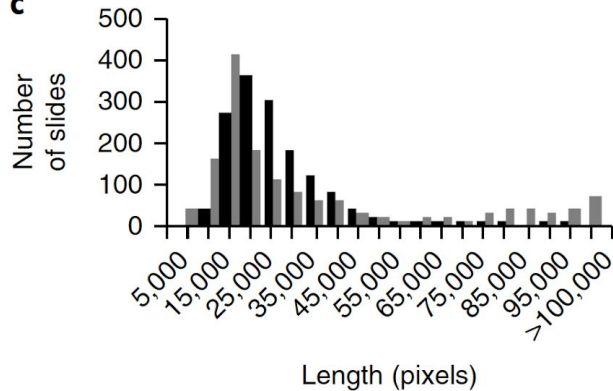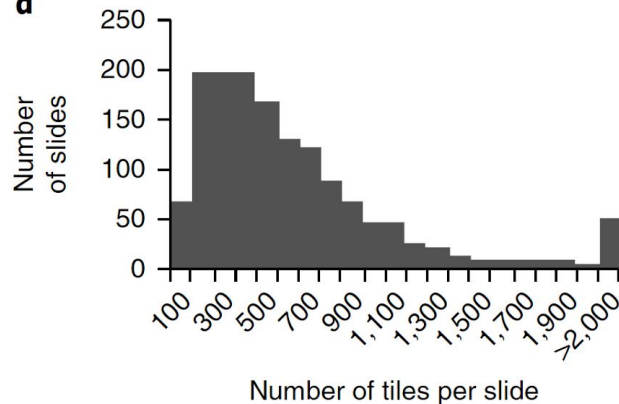
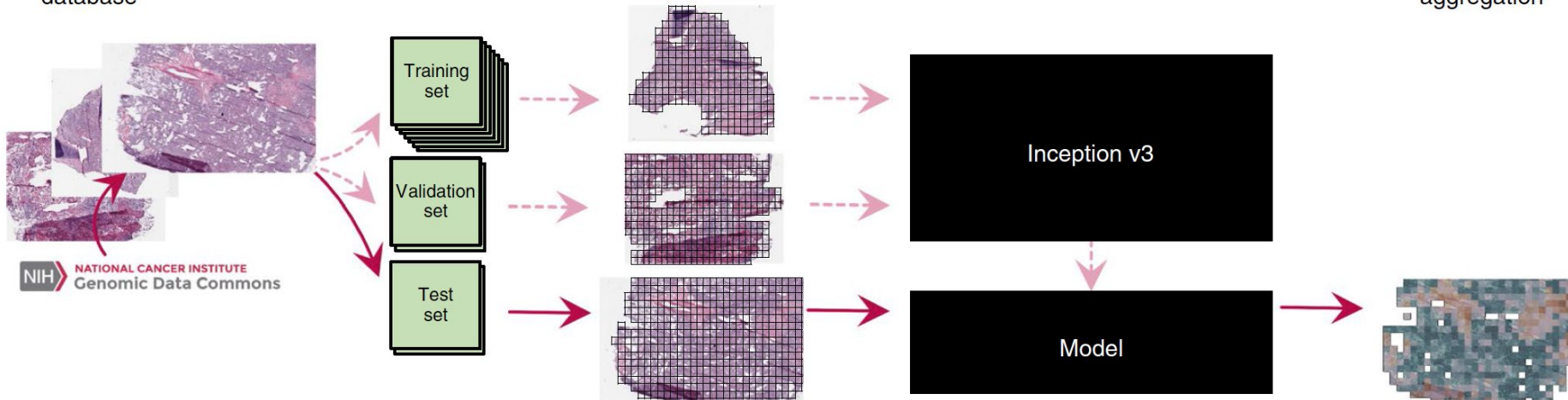# Training, Validation, Test, Aggregation

b

(i) Download from GDC database

(ii) Separate in 3 datasets

(iii) Tile and filter out background tiles

(iv) Per-tile training

(v) Testing and per-slide tile aggregation

NIH | NATIONAL CANCER INSTITUTE Genomic Data Commons

Training set

Validation set

Test set

Inception v3

Model

Results

Classification	Information	AUC after aggregation by… ... average predicted probability	AUC after aggregation by… ... percentage of positively classified tiles
Normal vs Tumor (20x tiles)	a) Inception v3, fully-trained	0.993 [0.974-1.000]	0.990 [0.969-1.000]
LUAD vs LUSC (20x tiles)	b) Inception v3, transfer learning	0.847 [0.782-0.906]	0.844 [0.777-0.904]
LUAD vs LUSC (20x tiles)	c) Inception v3, fully-trained	0.950 [0.913-0.980]	0.947 [0.911-0.978]
LUAD vs LUSC (20x tiles)	d) Same as (c) but aggregation done solely on tiles classified as "tumor" by A	0.952 [0.915-0.981]	0.949 [0.912-0.980]
LUAD vs LUSC (5x tiles)	Inception v3, fully-trained	**0.942 [0.907-0.971]**	0.906 [0.851-0.951]
3 classes. Normal vs LUAD vs LUSC at 20x	Normal	0.984 [0.947-1.000]	0.985 [0.953-1.000]
3 classes. Normal vs LUAD vs LUSC at 20x	LUAD	**0.969 [0.933-0.994]**	0.970 [0.937-0.993]
3 classes. Normal vs LUAD vs LUSC at 20x	LUSC	0.966 [0.935-0.990]	0.964 [0.932-0.989]
3 classes. Normal vs LUAD vs LUSC at 20x	Micro-average	0.970 [0.950-0.986]	0.969 [0.949-0.985]
3 classes. Normal vs LUAD vs LUSC at 20x	Macro-average	0.976 [0.949-0.993]	0.976 [0.950-0.993]
3 classes. Normal vs LUAD vs LUSC at 5x	Normal	**0.997 [0.993-0.998]**	0.988 [0.962-1.000]
3 classes. Normal vs LUAD vs LUSC at 5x	LUAD	0.965 [0.942-0.983]	0.938 [0.896-0.971]
3 classes. Normal vs LUAD vs LUSC at 5x	LUSC	**0.977 [0.960-0.991]**	0.964 [0.937-0.986]
3 classes. Normal vs LUAD vs LUSC at 5x	Micro-average	**0.980 [0.972-0.987]**	0.966 [0.948-0.980]
3 classes. Normal vs LUAD vs LUSC at 5x	Macro-average	**0.981 [0.968-0.991]**	0.964 [0.939-0.980]

n=244 slides for LUAD vs LUSC classifiers and n=170 slides for the others, all from 137 patients.

Predicting gene mutational status from whole-slide images
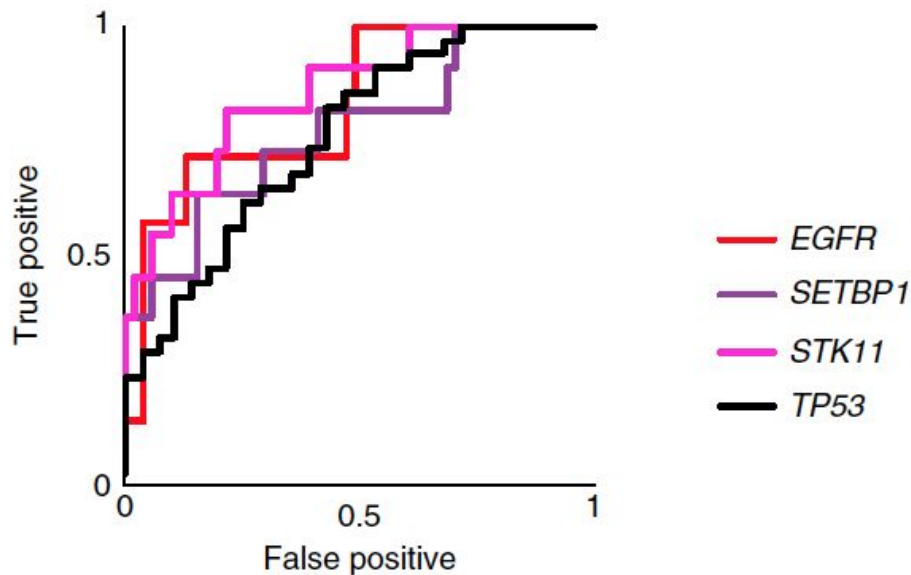
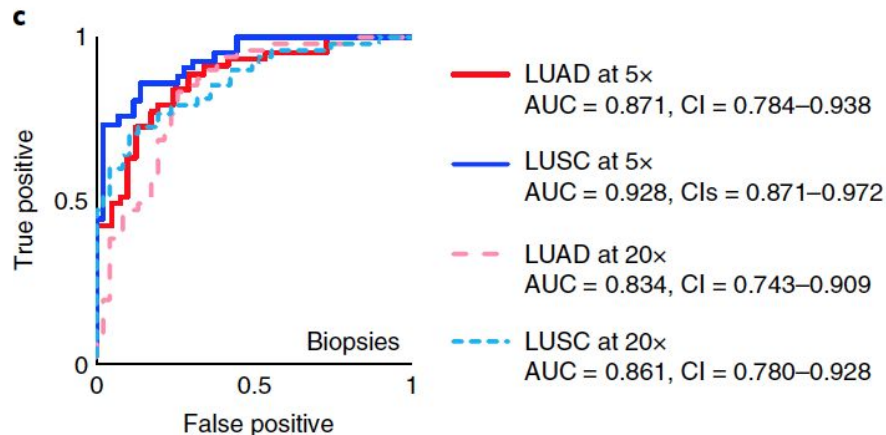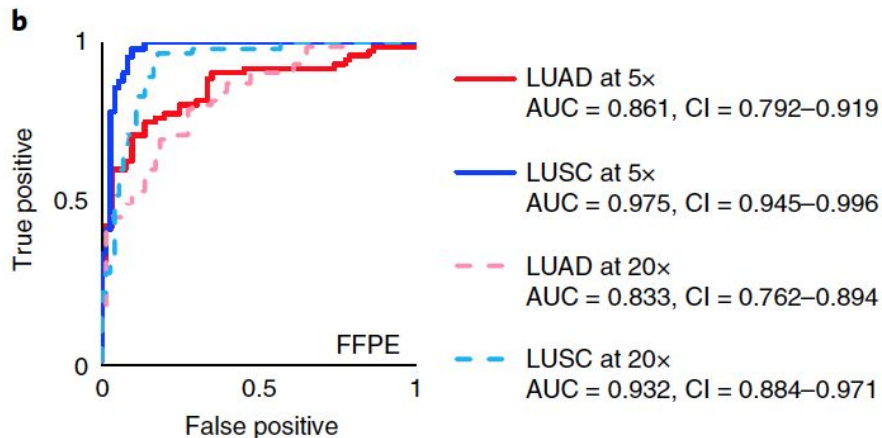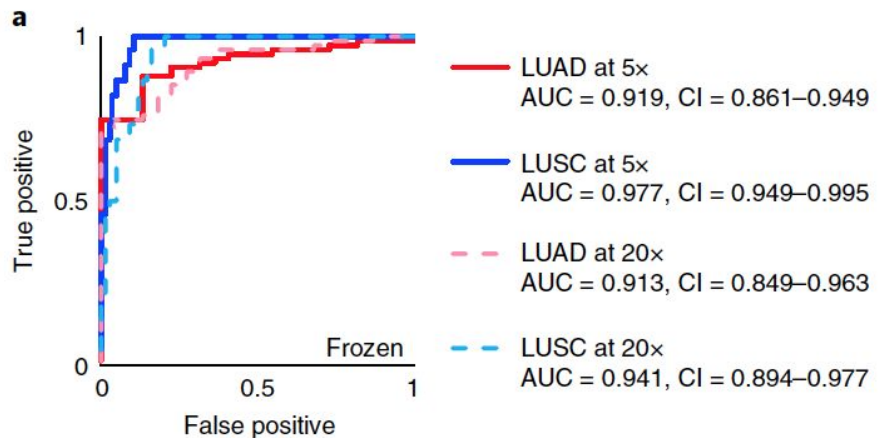**Table 1 | AUC achieved by the network trained on mutations (with 95% CIs)**

Mutations	Per-tile AUC	Per-slide AUC after aggregation by…	
		… average predicted probability	… percentage of positively classified tiles
STK11	0.845 (0.838–0.852)	0.856 (0.709–0.964)	0.842 (0.683-0.967)
EGFR	0.754 (0.746–0.761)	0.826 (0.628–0.979)	0.782 (0.516-0.979)
SETBP1	0.785 (0.776–0.794)	0.775 (0.595–0.931)	0.752 (0.550-0.927)
TP53	0.674 (0.666–0.681)	0.760 (0.626–0.872)	0.754 (0.627-0.870)
FAT1	0.739 (0.732–0.746)	0.750 (0.512–0.940)	0.750 (0.491-0.946)
KRAS	0.814 (0.807–0.829)	0.733 (0.580–0.857)	0.716 (0.552-0.854)
KEAP1	0.684 (0.670–0.694)	0.675 (0.466–0.865)	0.659 (0.440-0.856)
LRP1B	0.640 (0.633–0.647)	0.656 (0.513–0.797)	0.657 (0.512-0.799)
FAT4	0.768 (0.760–0.775)	0.642 (0.470–0.799)	0.640 (0.440-0.856)
NF1	0.714 (0.704–0.723)	0.640 (0.419–0.845)	0.632 (0.405-0.845)

$n = 62$ slides from 59 patients.

Generalization to Other Cohorts

NYULMC DATA

- Frozen sections (98 slides)
- FFPE sections (140 slides)
- Needle biopsies (102 slides)

a Frozen

- LUAD at 5×
  AUC = 0.919, CI = 0.861–0.949
- LUSC at 5×
  AUC = 0.977, CI = 0.949–0.995
- LUAD at 20×
  AUC = 0.913, CI = 0.849–0.963
- LUSC at 20×
  AUC = 0.941, CI = 0.894–0.977

b FFPE

- LUAD at 5×
  AUC = 0.861, CI = 0.792–0.919
- LUSC at 5×
  AUC = 0.975, CI = 0.945–0.996
- LUAD at 20×
  AUC = 0.833, CI = 0.762–0.894
- LUSC at 20×
  AUC = 0.932, CI = 0.884–0.971

c Biopsies

- LUAD at 5×
  AUC = 0.871, CI = 0.784–0.938
- LUSC at 5×
  AUC = 0.928, CIs = 0.871–0.972
- LUAD at 20×
  AUC = 0.834, CI = 0.743–0.909
- LUSC at 20×
  AUC = 0.861, CI = 0.780–0.928

Comparison to Pathologists

Supplementary Table 3. Inter-pathologists and binary deep-learning method variability estimated with the Cohen's Kappa statistic.

	Pathologist 1*	Pathologist 2**	Pathologist 3*	Consensus between pathologists	Deep-learning
TCGA	0.67 CIs=[0.56-8.78]	0.70 CIs=[0.60-0.81]	0.70 CIs=[0.59-0.81]	0.78 CIs=[0.69-0.88]	0.82 CIs=[0.74-0.91]
Pathologist 1		0.52 CIs=[0.39-0.65]	0.55 CIs=[0.42-0.67]	0.56 CIs=[0.44-0.69]	0.64 CIs=[0.52-0.75]
Pathologist 2			0.78 CIs=[0.69-0.88]	0.65 CIs=[0.54-0.77]	0.63 CIs=[0.52-0.75]
Pathologist 3				0.75 CIs=[65-0.86]	0.60 CIs=[0.48-0.72]
Consensus between 3 pathologists					0.77 CIs=[0.68-0.87]

n=170 slides from 137 patients
* thoracic pathologists; ** anatomic pathologist

Microscopy and Super-resolutions

Cellular Imaging - Latest Updates

Recent advances in *fluorescence microscopy*:
- Tagging 100s of RNAs (corresponding to genes), Proteins, etc. in live cells
- "Seeing" across time and space at much higher resolution
- Limits on amount of light that can be given to each batch
- Light is proportional to Resolution (Similar to X-Ray radiation dose)

Will change the way we understand drug response

Will change the way we understand cellular behaviour

Applications for All Cancers, Alzheimer's disease, Neurological conditions, etc.

Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy

https://www.biorxiv.org/content/early/2018/07/03/236463

Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy

Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Maurício Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, Eugene W. Myers
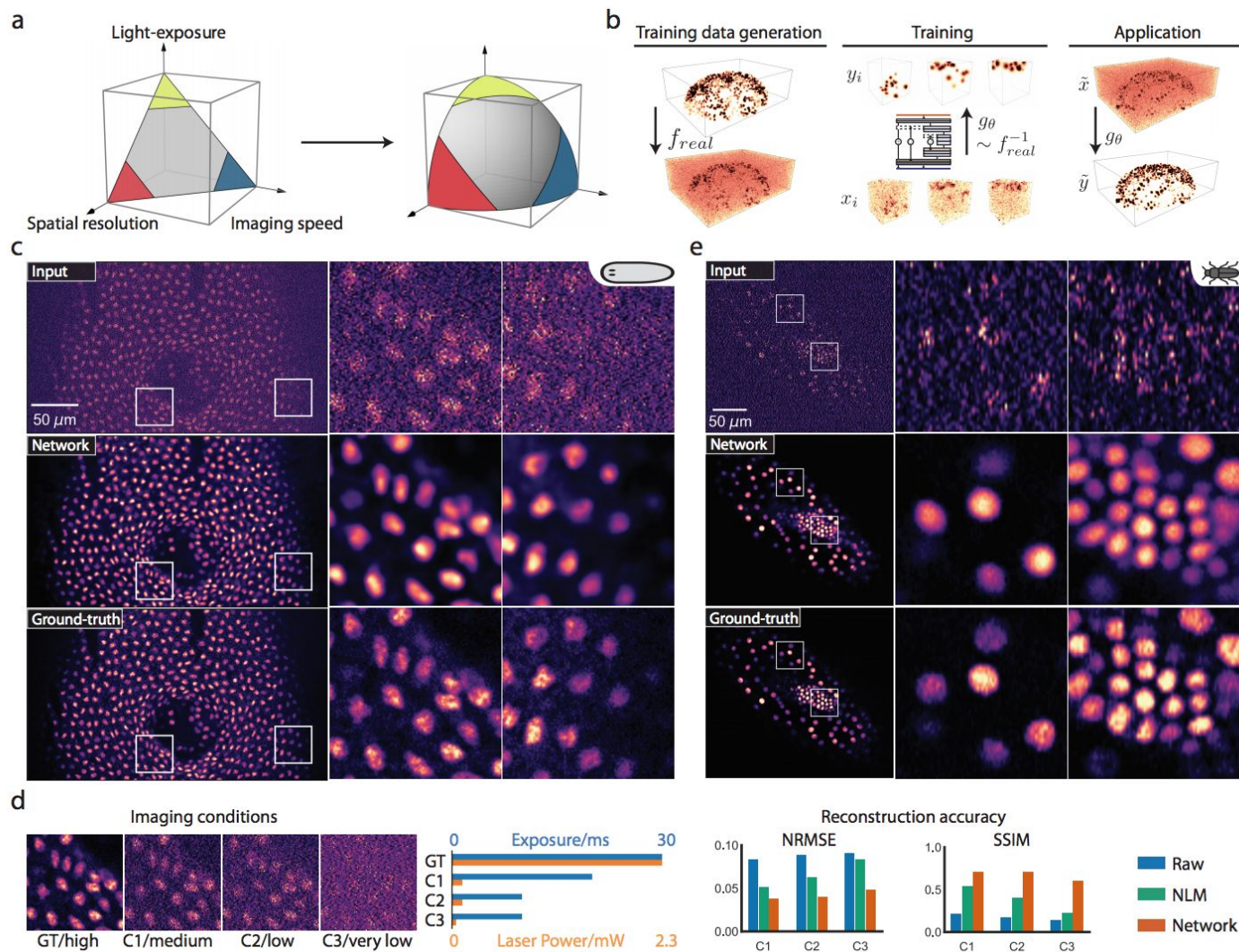
Abstract   Info/History   Metrics   Supplementary material   🗋 Preview PDF

Abstract

Fluorescence microscopy is a key driver of discoveries in the life-sciences, with observable phenomena being limited by the optics of the microscope, the chemistry of the fluorophores, and the maximum photon exposure tolerated by the sample. These limits necessitate trade-offs between imaging speed, spatial resolution, light exposure, and imaging depth. In this work we show how image restoration based on deep learning extends the range of biological phenomena observable by microscopy.

a Light-exposure, Spatial resolution, Imaging speed

b Training data generation, Training, Application

y_i

$\downarrow f_{real}$

$\uparrow g_\theta \sim f_{real}^{-1}$

x_i

$\tilde{x}$

$\downarrow g_\theta$

$\tilde{y}$

c Input, Network, Ground-truth

50 μm

e Input, Network, Ground-truth

50 μm

d Imaging conditions

GT/high, C1/medium, C2/low, C3/very low

Exposure/ms — 0 to 30

GT, C1, C2, C3

Laser Power/mW — 0 to 2.3

Reconstruction accuracy

NRMSE — 0.00 to 0.10 — C1, C2, C3

SSIM — 0.0 to 1.0 — C1, C2, C3

Raw, NLM, Network

Models for Sequences and Genomics

Biomarkers from Sequential Convolutional Nets

Babak Alipanahi, Andrew Delong, Matthew T Weirauch & Brendan J Frey,
"*Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning.*" Nature biotechnology (2015)

Collaboration: UToronto

Objective: Discover DNA/RNA motifs that bind to many binding proteins, and predict protein-binding in multiple tasks (in vitro and in vivo)

Data: 240,000 RNA sequences and 207 binding proteins; 40,000 DNA sequences and 86 binding proteins (transcription factors)

Convolution Model for discovering Motifs and Position Weight Matrices
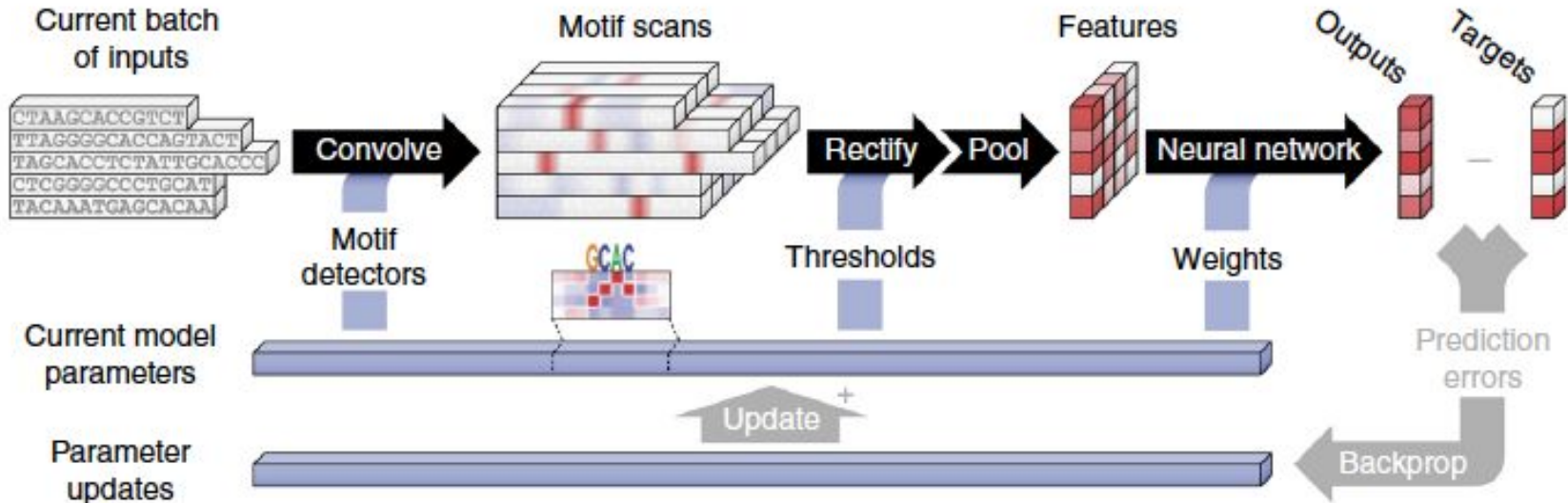
Results

In vitro:

- DNA Specificity prediction; Average AUC 0.726
- RNA Specificity prediction: Average AUC 0.84
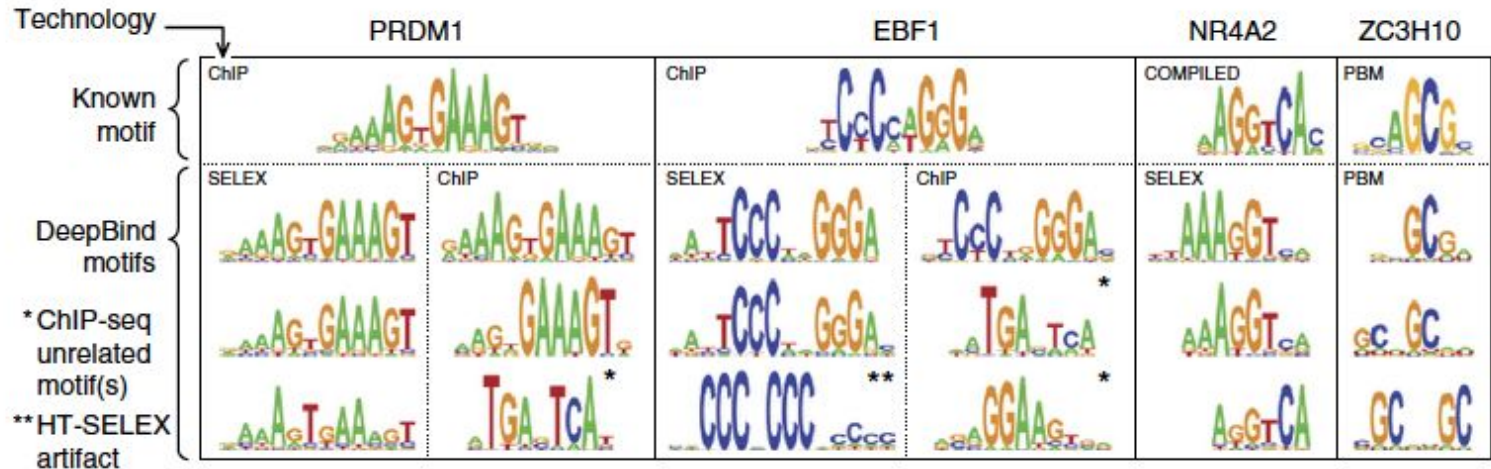
State of Research In ML for Healthcare
# Short term and Long term

Short term: many many standard supervised learning

It's natural & necessary to build several new baselines
- Healthcare has recently joined data-heavy fields.
- Most baselines in other fields haven't even been tried here.
- We do need to build many many baselines.
- New architectures/models aren't necessarily needed
- Need to understand what tasks are *actually* harder and need more ML innovations

Outcome of this stage:
- Models that can be deployed in practice: shift focus to integration & system changes & industry change
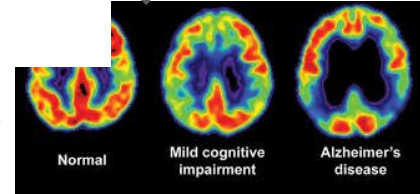- Identification of medical tasks that are actually difficult!
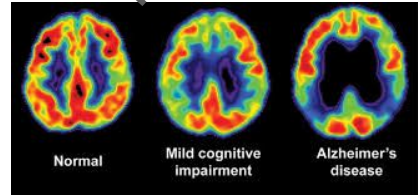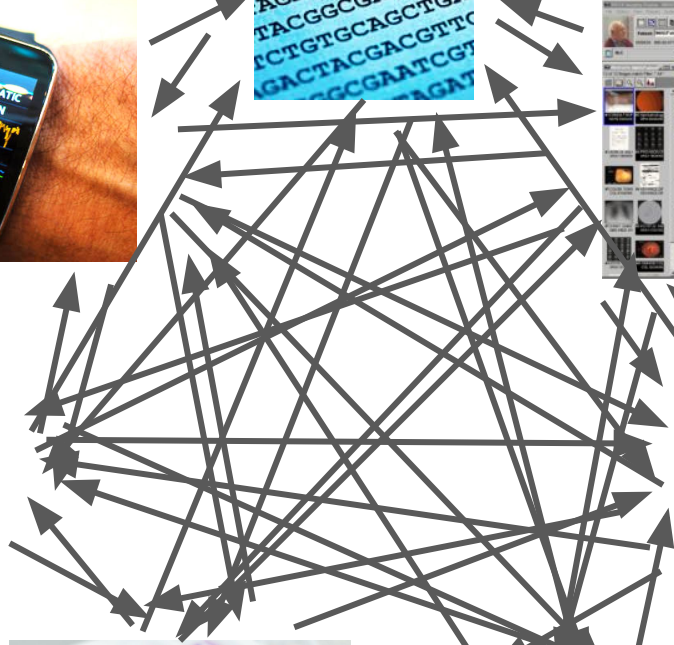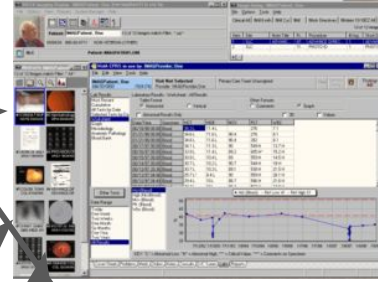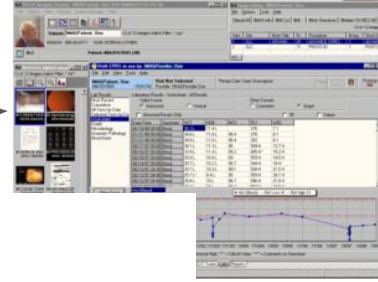
Each of these arrows learned
- Will save lives
- Will discover new hypothesis
- Will save money
- Will change industries

What is difficult today?

Tracking and representing and modeling *changes* over time
- Predicting it, predicting with it, disentangling factors, etc.
- Even ML tools aren't mature in this area.

Recommending treatments:
- Counterfactual inference & personalized medicine

Rare diseases..

Beyond current tools:
- New sensors & hardwares - Physics & Chemistry!
- Repurposing existing hardware (i.e. MRI pulse sequences, Ultrasounds, etc)
- Embedded sensors

That's it for now!

Email me with follow ups and questions:

Narges.Razavian@nyumc.org

Also, take the next semester's class:

Deep Learning for Medicine

BMSC-GA 4493 or BMIN-GA 3007