VIRGINIA TECH

Multimodal Explainable AI Models for Clinical Reasoning

Ismini Lourentzou

Assistant Professor

Computer Science | Virginia Tech

Perception and LANguage (PLAN) Lab

https://isminoula.github.io/

SANGHANI CENTER FOR ARTIFICIAL INTELLIGENCE & DATA ANALYTICS
VIRGINIA TECH

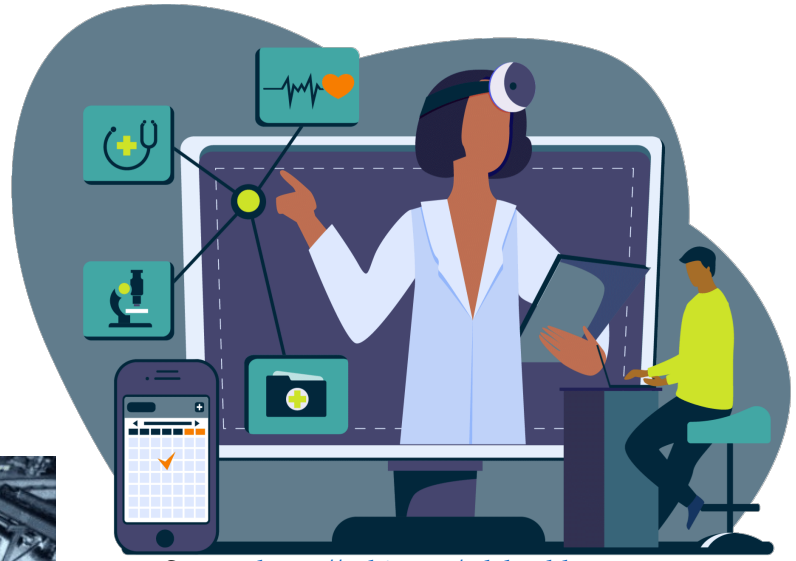
NATIONAL SECURITY INSTITUTE

COLLEGE OF ENGINEERING
COMPUTER SCIENCE
VIRGINIA TECH

Progress in AI/ML Adoption

Source: https://dribbble.com/shots/6781718-Smartwatch by **Laurens van Vliet**

Source: https://www.fastcompany.com/3045397/how-cars-are-made-in-20-gifs

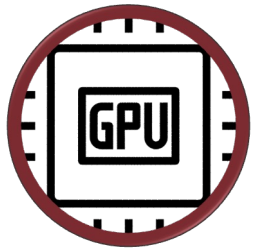


Source: https://ochin.org/telehealth

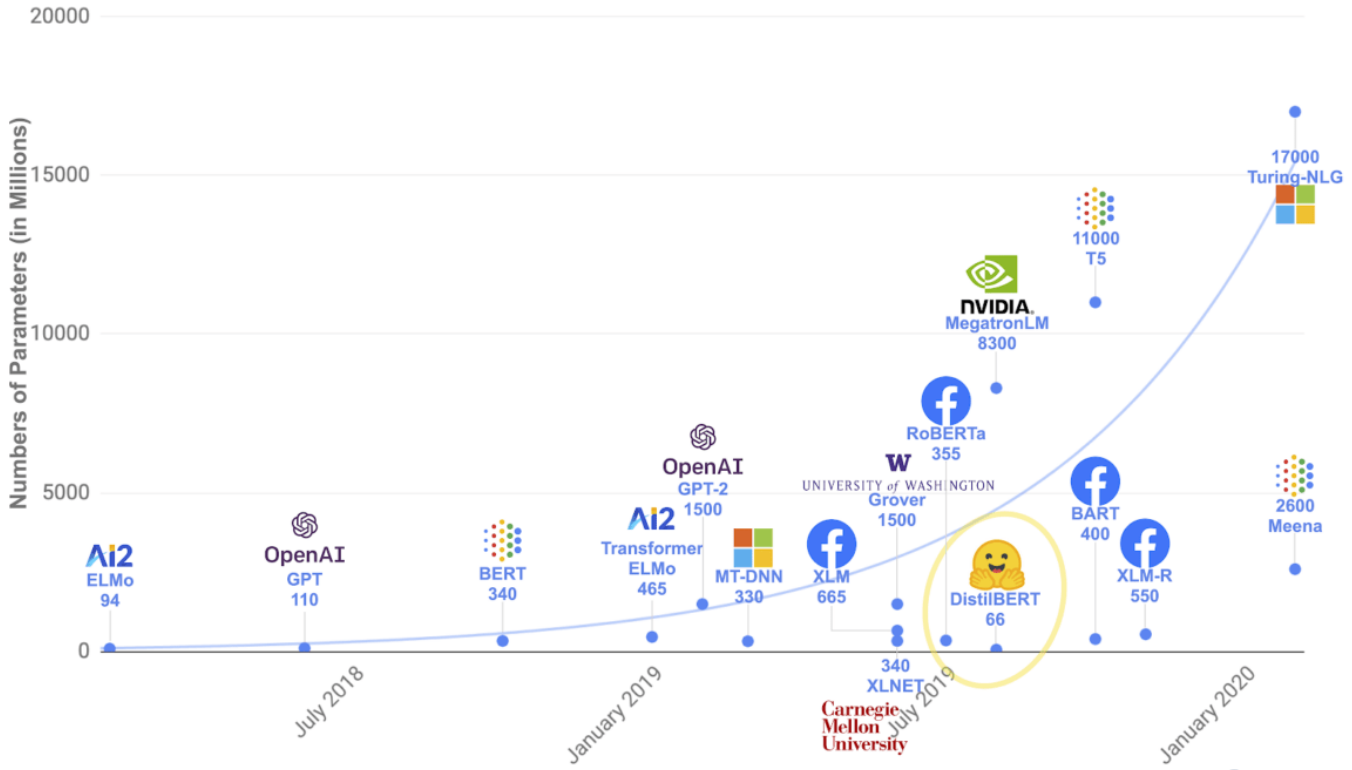
Key ML Ingredients

Data

+



Computing Power



Source: https://research.aimultiple.com/gpt/

What do NNs really learn?

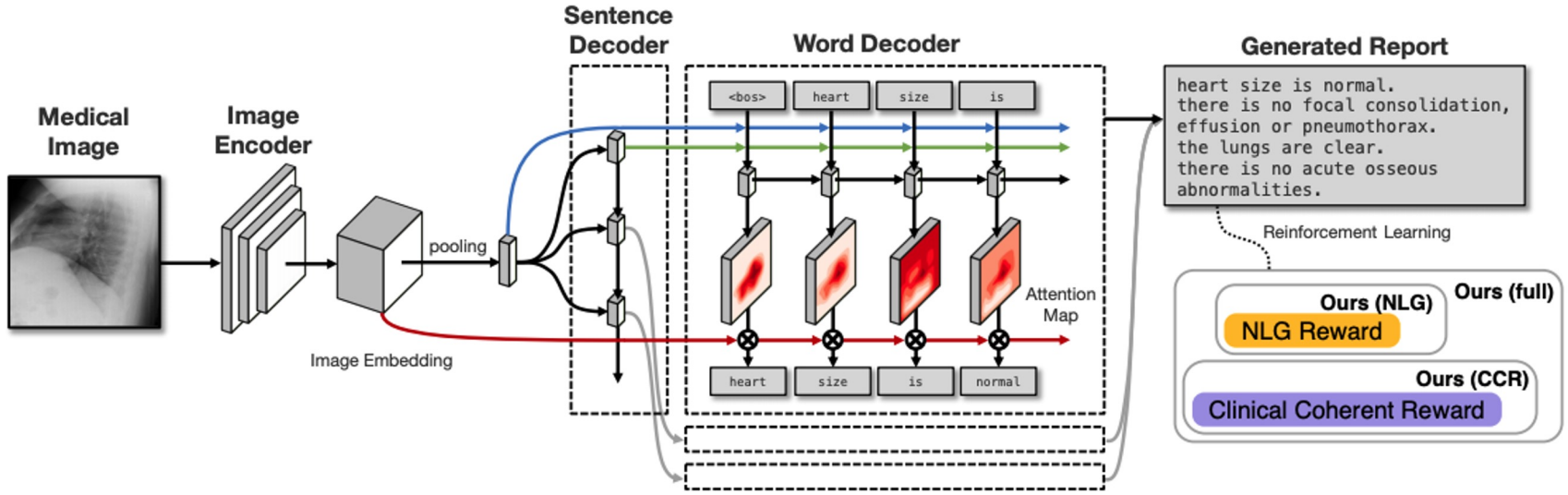
A herd of sheep grazing on a lush green h
Tags: grazing, sheep, mountain, cattle,

NeuralTalk2: A flock of birds flying in the air
Microsoft Azure: A group of giraffe standing next to a tree
Image: Fred Dunn, https://www.flickr.com/photos/gratapictures - CC-BY-NC

Source: https://www.aiweirdness.com/do-neural-nets-dream-of-electric-18-03-02/

What do NNs really learn?

as compared to the previous radiograph, there is no relevant change. tracheostomy tube is in place. there is a layering pleural effusions. NAME bilateral pleural effusion and compressive atelectasis at the right base. there is no pneumothorax.



CXR report generation [1]

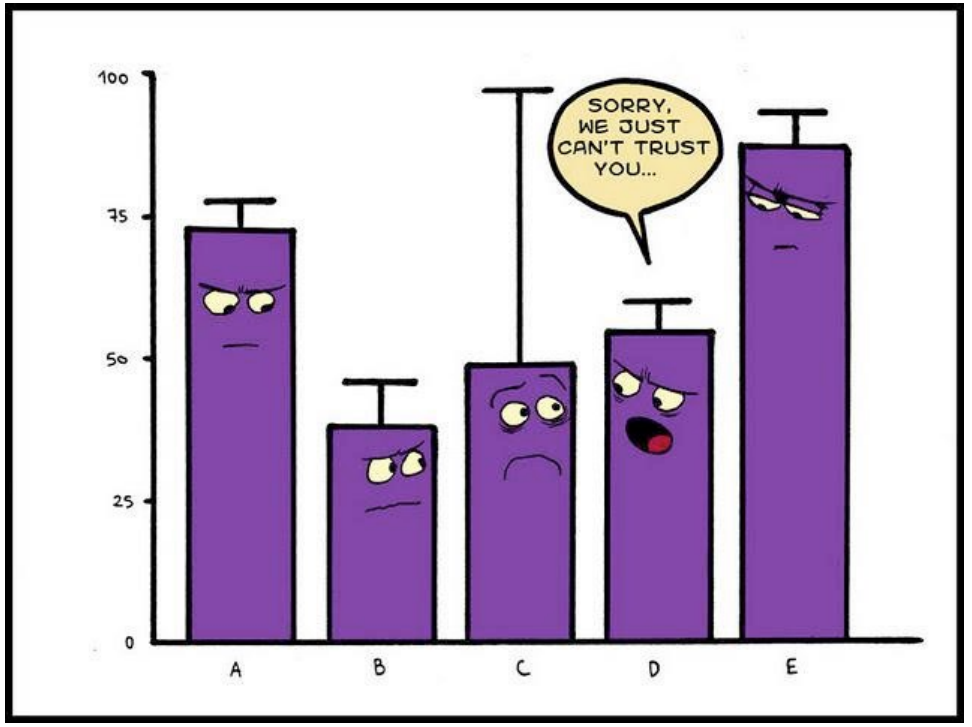
How can we gain confidence and trust in machine learning models deployed in healthcare?

[1] Liu, G., Hsu, T.M.H., McDermott, M., et. al., Clinically accurate chest x-ray report generation. ML4H 2019

Challenges with ML models

Limited Information
Models are not aware of anything else other than what is in their input data

Verifiable & Trustworthy Information
Non-trivial to sanitize (and trust) output of an ML model



Source:https://datascience.stackexchange.com/questions/42621/data-science-related-funny-quotes

Current Medical Imaging Datasets

Existing Chest X-Ray (CXR) datasets

Dataset	Annotation Level	Annotation Method	Num Labels	Anatomy Labeled	Graph	Dataset Size	Temporal Labels	Reports
SIIM-ACR Pneumothorax Segmentation [13]	Segmentation	Manual + augmented	1	No	No	12,047	No	No
RSNA Pneumonia Detection Challenge [38]	Bounding Boxes	Manual	1	No	No	30,000	No	No
Indiana University Chest X-ray collection [12]	Global	Automated	10	No	No	3,813	No	Yes
NIH CXR dataset [42]	Global	Automated	14	No	No	112,120	No	No
PLCO [40]	Global	Automated	24	Yes	No	236,000	Yes	No
Stanford CheXpert [20]	Global	Automated	14	No	No	224,316	No	No
MIMIC-CXR [22]	Global	Automated	14	No	No	377,110	No	Yes
Dutta [10]	Global	Manual	10	Yes	Yes	2,000	No	Yes
PadChest [5]	Global	Manual + automated	297	Yes	No	160,868	No	Yes
Montgomery County Chest X-ray [21]	Segmentation	Manual	1	Yes	No	138	No	No
Shenzen Hospital Chest X-ray [21]	Segmentation	Manual	1	Yes	No	662	No	No
Chest ImaGenome	Bounding Boxes	Automated	131	Yes	Yes	242,072	Yes	Yes

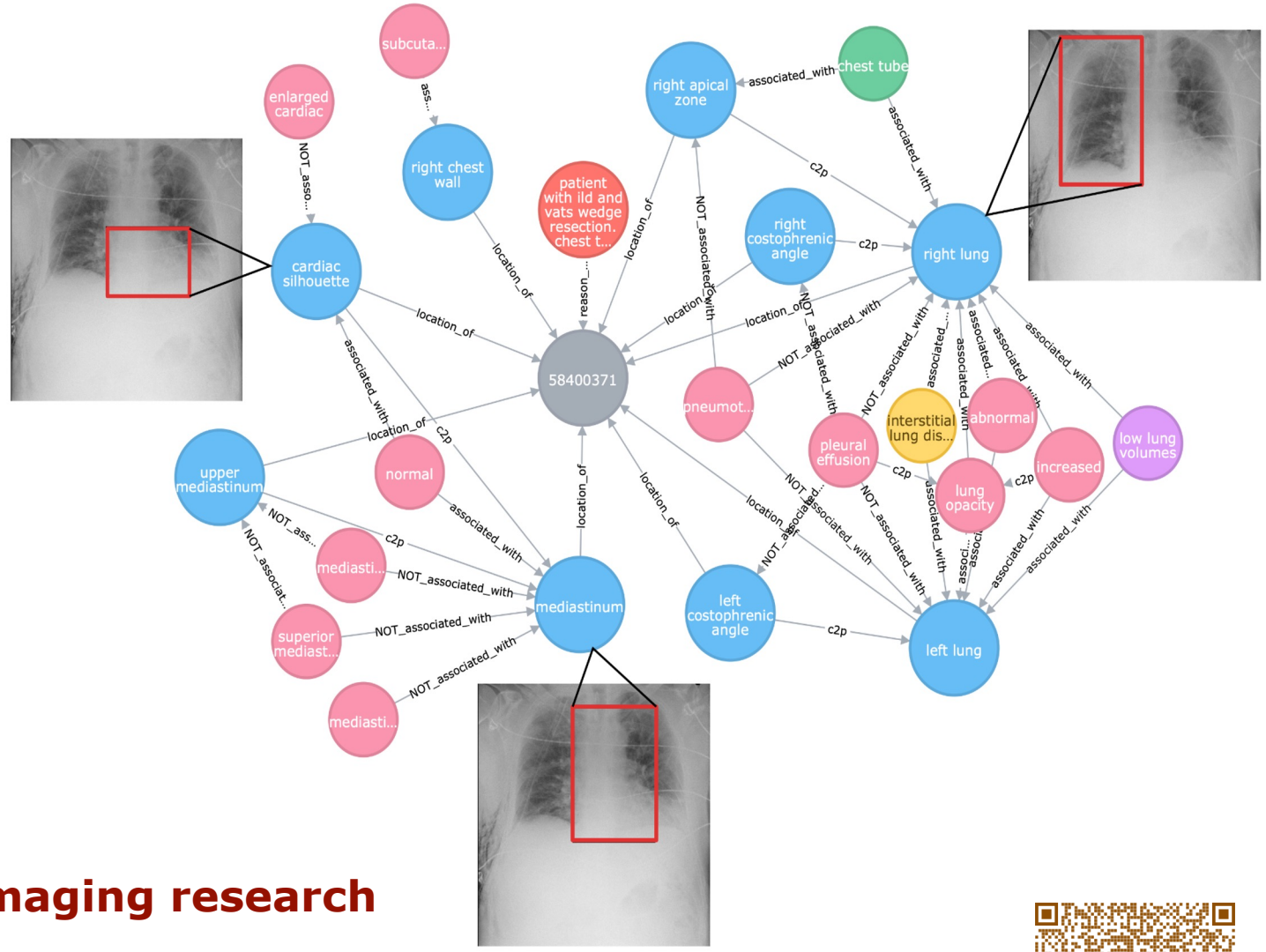
- **Existing datasets** focused on supporting classification, text generation, single abnormality detection or segmentation problems
- **Chest ImaGenome** Longitudinal Visual Genome-like dataset
  Supports classification, text generation, multiple abnormality detection, temporal comparison and anatomically aware clinical reasoning problems.
  Automatically generated with a dual-validated subset.

[2] Wu, J.T., Agu, N.N., Lourentzou, I., Sharma, A., Paguio, J.A., Yao, J.S., Dee, E.C., Mitchell, W.G., Kashyap, S., Giovannini, A. and Celi, L.A., Chest ImaGenome Dataset for Clinical Reasoning. NeurIPS Datasets and Benchmarks Track (2021)

Chest ImaGenome

Longitudinal Scene Graph Dataset that simplifies anatomically aware clinical reasoning:

- **242,072** graphs
- 29 object types
- 76 attribute types
- 4.9M relations



Support more explainable medical imaging research

[2] Wu, J.T., Agu, N.N., Lourentzou, I., Sharma, A., Paguio, J.A., Yao, J.S., Dee, E.C., Mitchell, W.G., Kashyap, S., Giovannini, A. and Celi, L.A., Chest ImaGenome Dataset for Clinical Reasoning. NeurIPS Datasets and Benchmarks Track (2021)

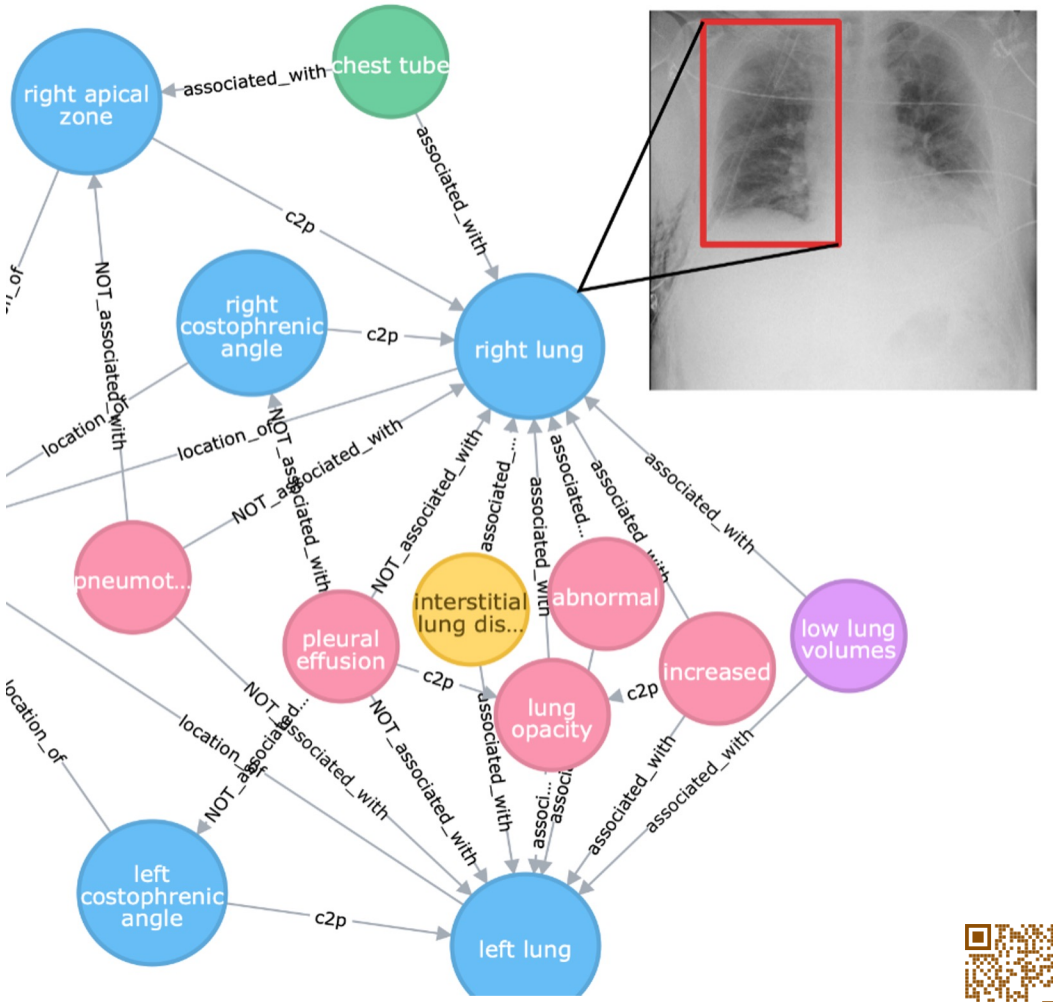
https://tinyurl.com/ChestImaGenome2021

SCAN ME

Object-attribute relations

{right lung} has {chest tube}

- 29 CXR anatomical objects
- 76 attribute types
- 6 semantic categories
- 1256 unique types of object-to-attribute relations
- 4.3M object-attribute relations
- Embedded parent-child & child-parent relations



[2] Wu, J.T., Agu, N.N., Lourentzou, I., Sharma, A., Paguio, J.A., Yao, J.S., Dee, E.C., Mitchell, W.G., Kashyap, S., Giovannini, A. and Celi, L.A., Chest ImaGenome Dataset for Clinical Reasoning. NeurIPS Datasets and Benchmarks Track (2021)

https://tinyurl.com/ChestImaGenome2021

SCAN ME

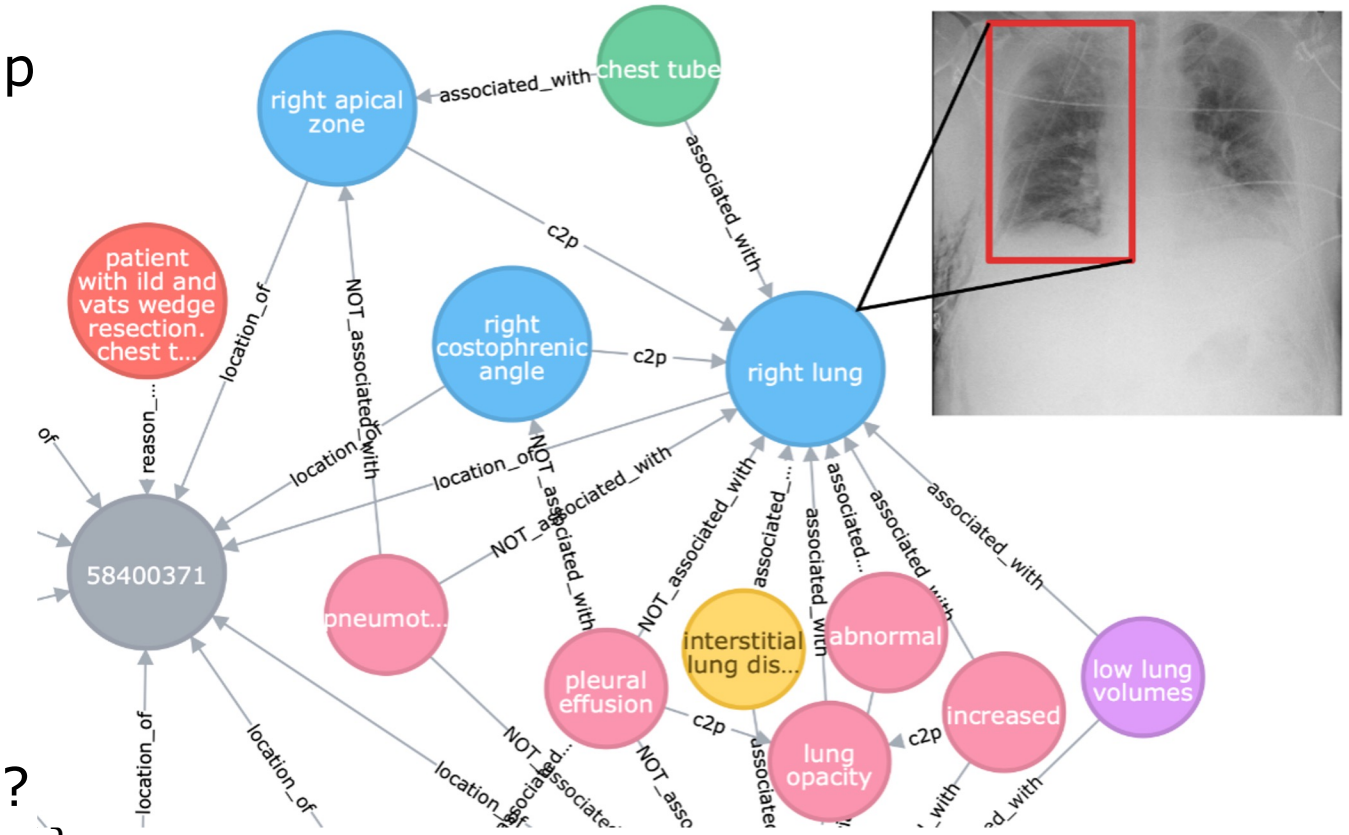
Object-attribute relations

One clinical question per scene grap

242,072 graphs

Other Clinical Q's:

- ↳ **Where is A?**
- ↳ **Does Obj have A?**
- ↳ **Is A1 associated with A2 at Obj?**
- ↳ Has (Obj,t) {C} since (Obj,t-1)?
  C = {**improved**, **worsened**, **no change**}

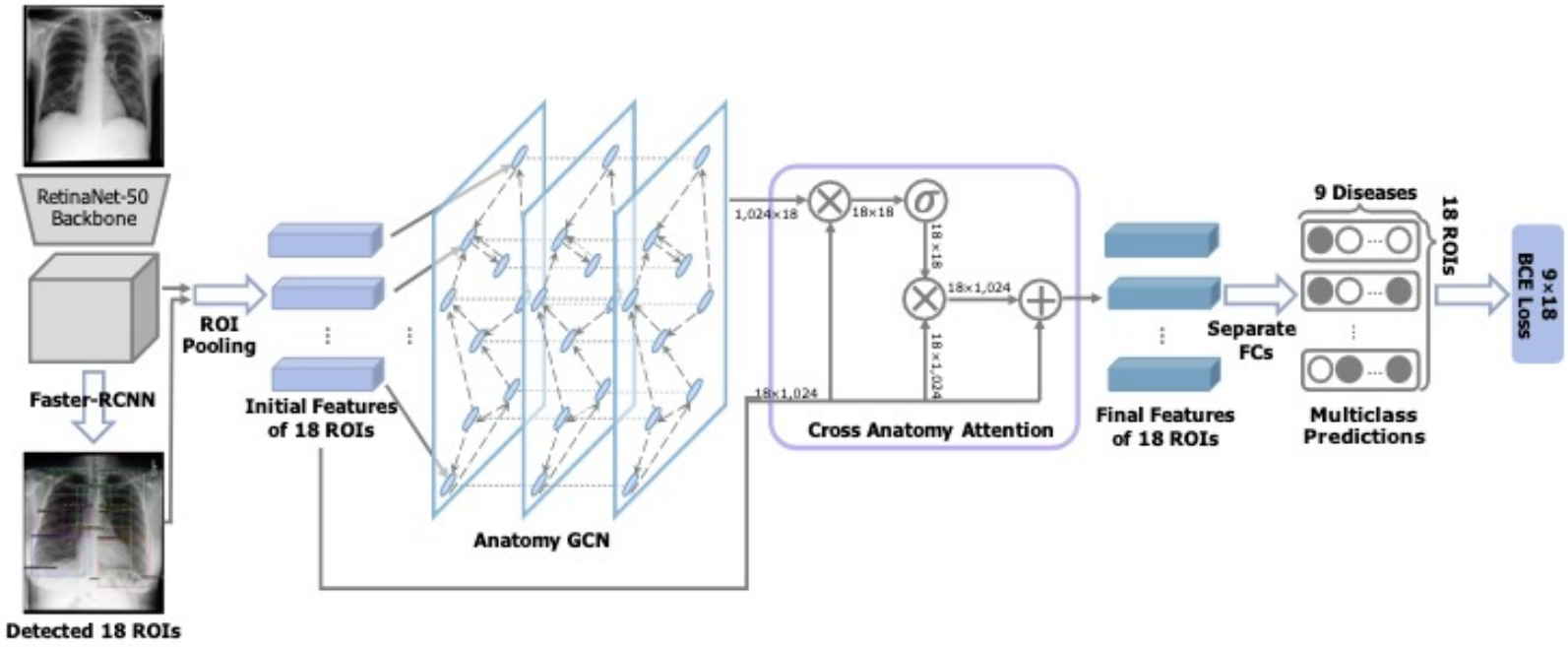


[2] Wu, J.T., Agu, N.N., Lourentzou, I., Sharma, A., Paguio, J.A., Yao, J.S., Dee, E.C., Mitchell, W.G., Kashyap, S., Giovannini, A. and Celi, L.A., Chest ImaGenome Dataset for Clinical Reasoning. NeurIPS Datasets and Benchmarks Track (2021)

https://tinyurl.com/ChestImaGenome2021

SCAN ME

Anatomy-aware X-ray Network (AnaXNet)



AnaXNet [3]: We extract anatomical regions of interest (ROIs) and their corresponding features, feed their vectors to a Graph Convolutional Network that learns their inter-dependencies, and combine the output with an attention mechanism, to perform the final classification with a dense layer.

[3] Agu, N.N., Wu, J.T., Chao, H., Lourentzou, I., Sharma, A., Moradi, M., Yan, P. and Hendler, J., Anaxnet: Anatomy aware multi-label finding classification in chest x-ray. MICCAI (2021)

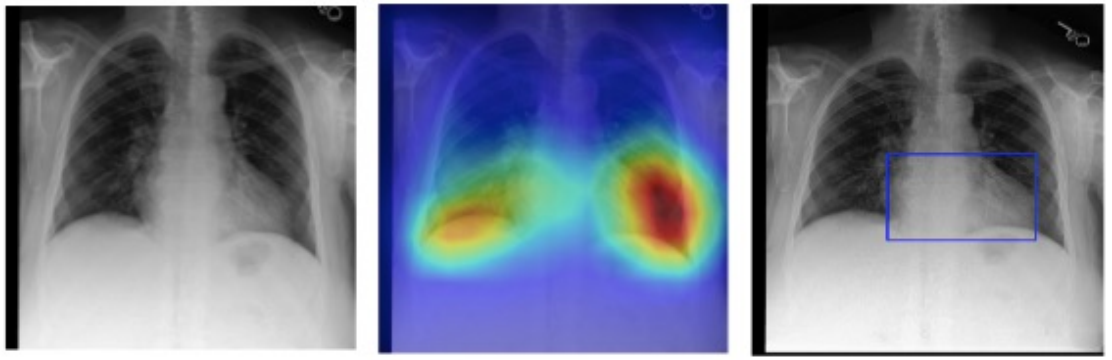
https://tinyurl.com/AnaXNet2021

SCAN ME

Anatomy-aware X-ray Network (AnaXNet)

Comparison against baselines (AUC scores)

Method	L1	L2	L3	L4	L5	L6	L7	L8	L9	**AVG**
Faster R-CNN	0.84	0.89	0.77	0.85	0.87	0.77	0.75	0.81	0.71	0.80
GlobalView	**0.91**	0.94	0.86	0.92	0.92	**0.93**	0.86	0.87	0.84	0.89
CheXGCN	0.86	0.90	0.91	0.94	**0.95**	0.75	**0.89**	**0.98**	0.88	0.90
AnaXNet (ours)	0.88	**0.96**	**0.92**	**0.99**	**0.95**	0.80	**0.89**	**0.98**	**0.97**	**0.93**



(a) Original Image (b) GlobalView (Grad-CAM) (c) AnaXNet

Enlarged cardiac silhouette example

[3] Agu, N.N., Wu, J.T., Chao, H., Lourentzou, I., Sharma, A., Moradi, M., Yan, P. and Hendler, J., Anaxnet: Anatomy aware multi-label finding classification in chest x-ray. MICCAI (2021)

https://tinyurl.com/AnaXNet2021

Object-Object Comparison Relations

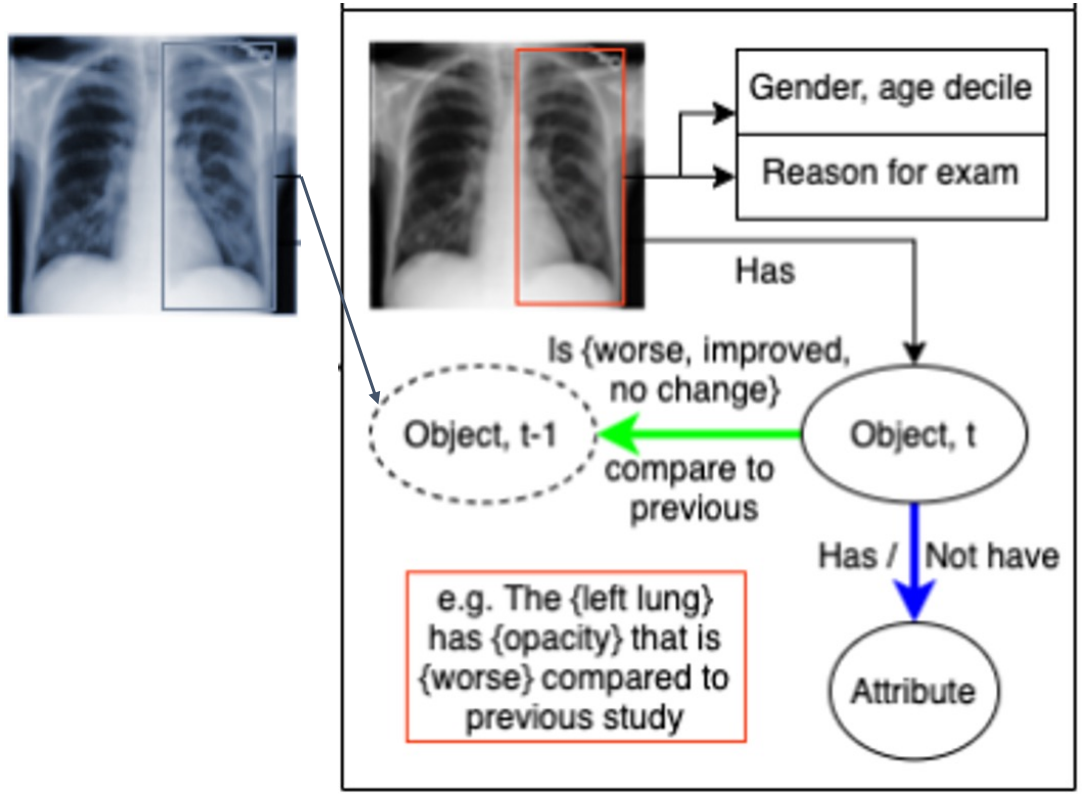
Classify **disease progression** over two sequential CXR exams

Over 670k localized comparison relations

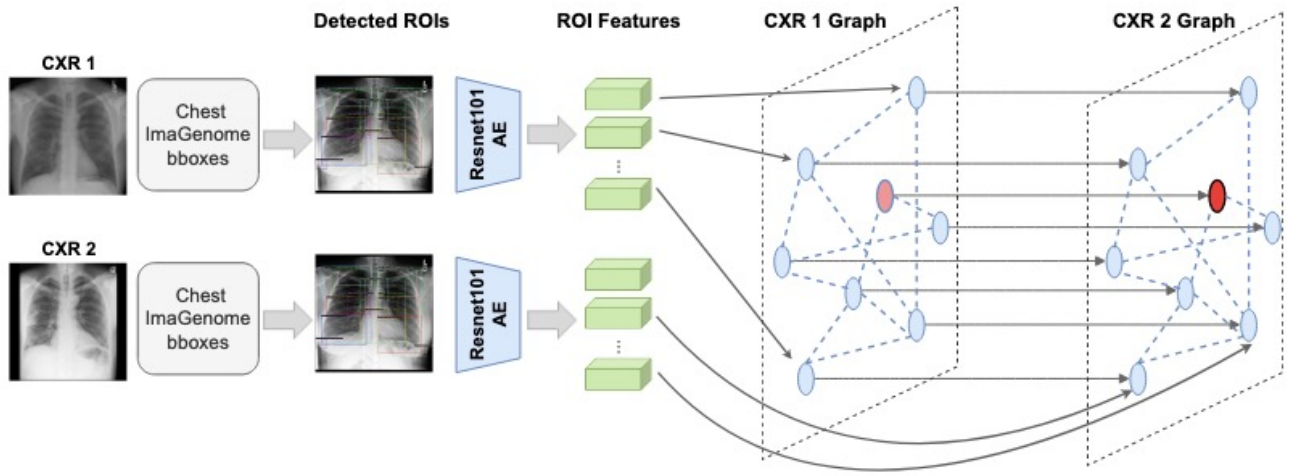
Clinical Question:
Has (Obj,t) {C} since (Obj,t-1)?

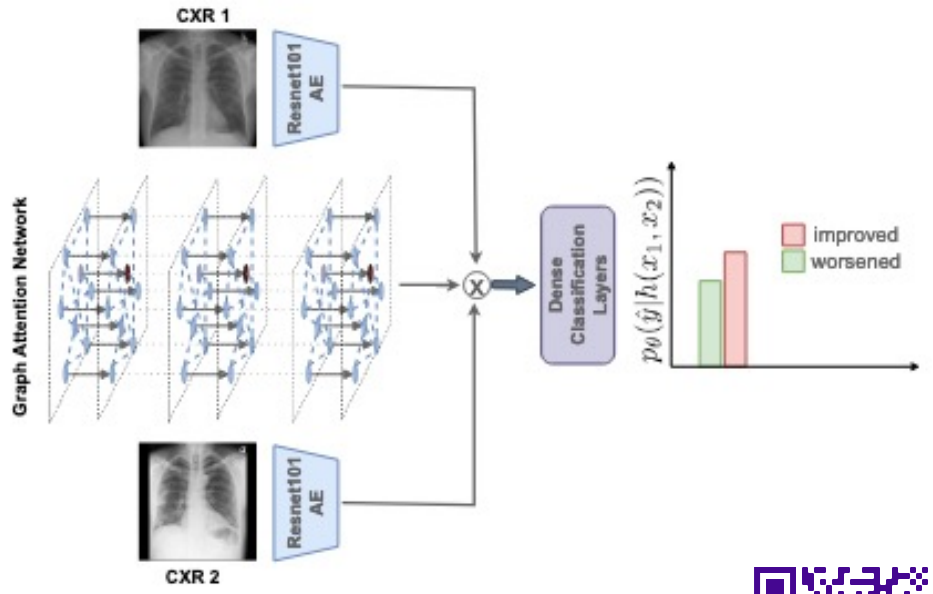
Comparison relations
C = {**improved**, **worsened**, **no change**}

- **Objects:** anatomical regions, such as. 'left lung' 'right lung', etc.
- **Attributes**: anatomical findings, such as 'pulmonary edema/hazy opacity', 'fluid overload/heart failure', etc.



CheXRelNet: Object-Object Comparison Relations





- Incorporate **local** and **global** visual features
- Capture **intra-image region-disease co-occurrences**
- Capture **inter-image region-based disease progression**

Learn **dependencies between anatomical region attributes** to accurately predict disease progression

[4] Karwande G.*, Mbakwe A.*, Wu J.T., Moradi M., Celi L. A., Lourentzou I., CheXRelNet: An Anatomy-Aware Model for Tracking Longitudinal Relationships between Chest X-Rays, MICCAI (2022)

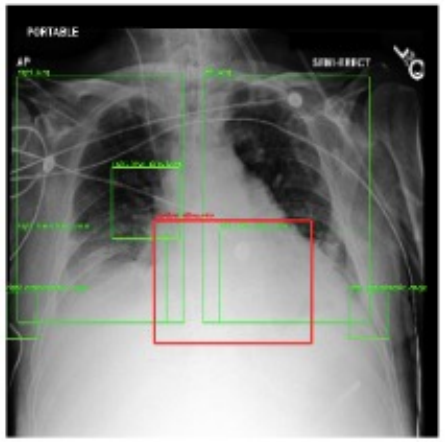
https://tinyurl.com/CheXRelNet

SCAN ME

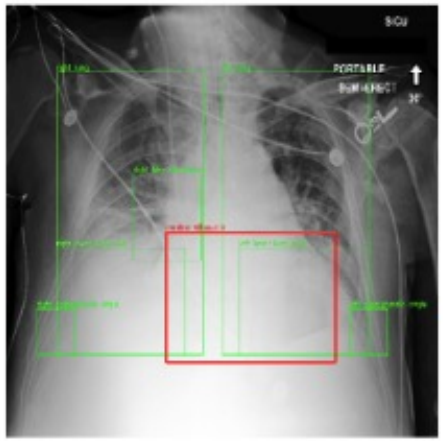
Results: Object-Object Comparison

Comparison against baselines (accuracy)

Method	D1	D2	D3	D4	D5	D6	D7	D8	D9	**All**
Local	0.59	0.53	0.60	0.47	0.56	0.46	0.61	0.47	0.63	0.60
Global	0.67	**0.69**	0.64	0.74	0.71	0.50	0.65	0.69	0.67	0.67
`CheXRelNet`	0.67	0.68	**0.66**	**0.75**	0.71	**0.52**	**0.67**	**0.73**	0.67	**0.68**



CXR 1



CXR 2

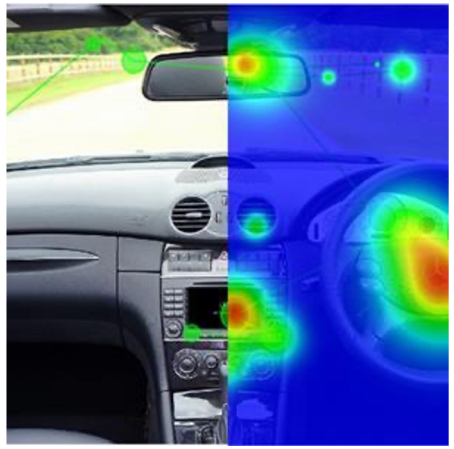
Pathology Fluid Overload/Heart Failure, class: Worsened

[4] Karwande G.*, Mbakwe A.*, Wu J.T., Moradi M., Celi L. A., Lourentzou I., CheXRelNet: An Anatomy-Aware Model for Tracking Longitudinal Relationships between Chest X-Rays, MICCAI (2022)

https://tinyurl.com/CheXRelNet

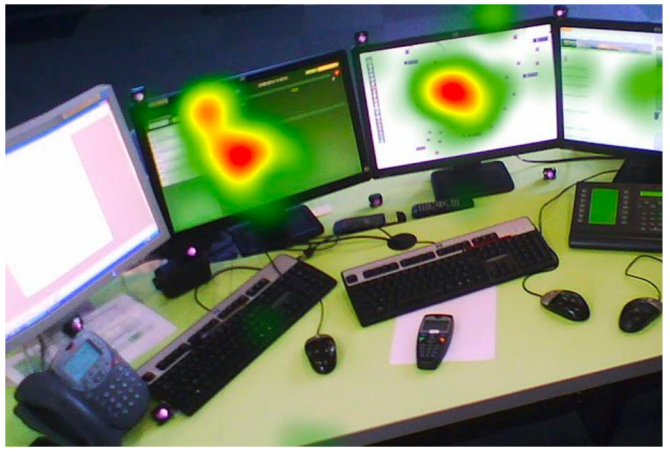
SCAN ME

Utilizing Eye Gaze

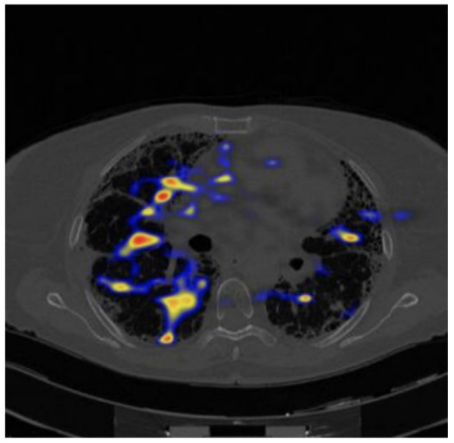


Eye tracking heatmaps for driver's attention

Consumer research [5]



Eye tracking traffic control room operators [6]



Eye gaze in CT [7]

Fairly unbiased view of what the humans are looking for in a scenario

Motivated by the high resemblance of activation maps used in deep learning for saliency with gaze fixation studies for other fields.

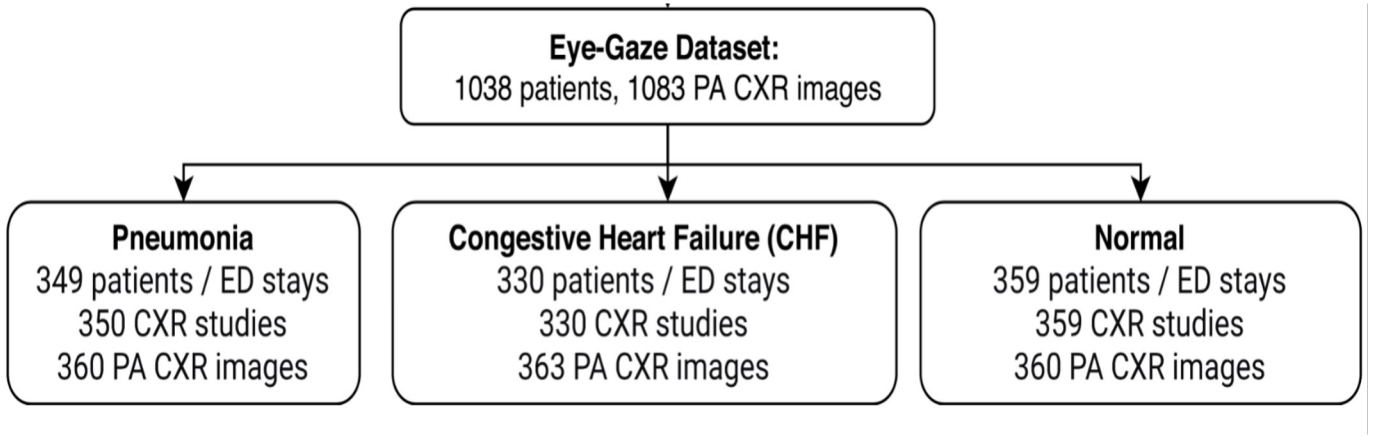
[5] https://phys.org/news/2012-07-consumer-product-giants-eye-trackers-size.html
[6] https://www.tobiipro.com/applications/scientific-research/case-studies/university-birmingham/
[7] Khosravan, N., Celik, H., Turkbey, B., Jones, E.C., Wood, B. and Bagci, U., A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning. Medical image analysis, 2019

Utilizing Eye Gaze in Radiology

Can we imitate the radiologist with attention supervision?

Utilize eye gaze information to

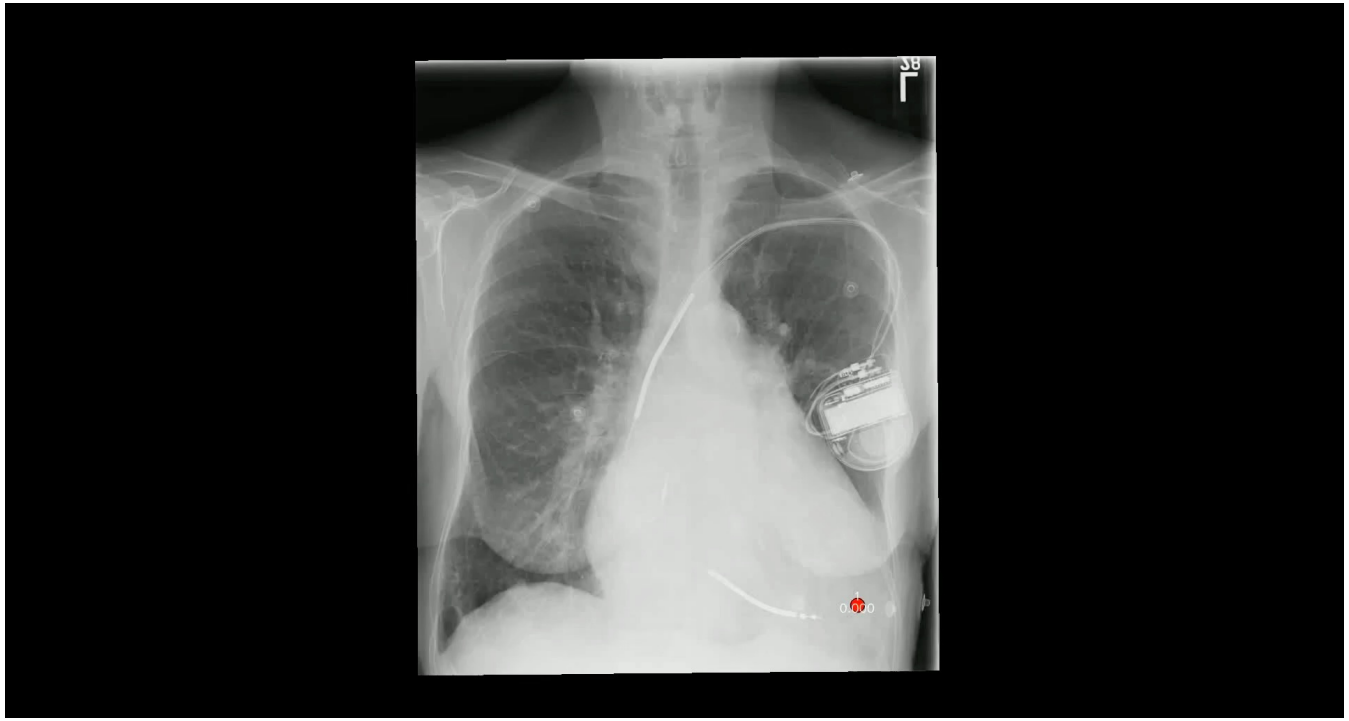
- Train **better radiology models**
- Improve **interpretability**
- Incorporate **domain knowledge** without explicit pixel-level labels



[8] Karargyris, A., Kashyap, S., Lourentzou, I., Wu, J.T., Sharma, A., Tong, M., Abedin, S., Beymer, D., Mukherjee, V., Krupinski, E.A. and Moradi, M., Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development. Nature Scientific data (2021)

https://tinyurl.com/EyeGaze2021

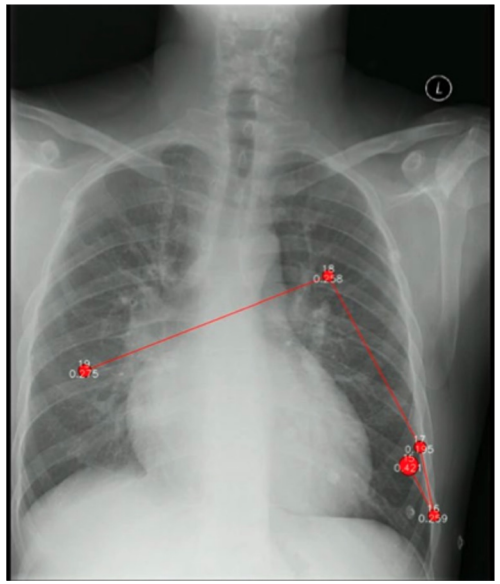
Data Sample

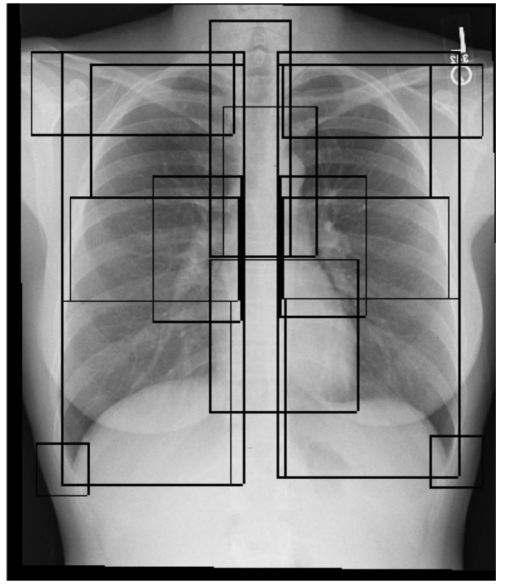


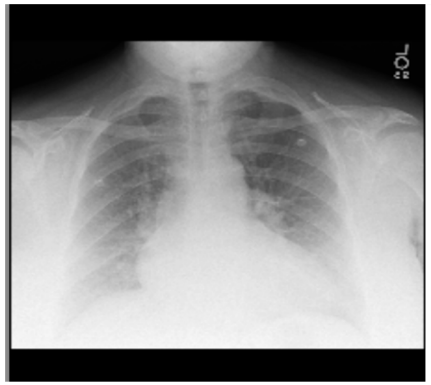
Radiology Report: Left-sided cardiac defibrillator with leads projecting over the right atrium and ventricle. Cardiomegaly. Aortic calcifications. Left-sided pleural effusion with atelectasis. prominent right breast shadow. The lungs appear hyper-inflated suggesting COPD. There is some scarring at the right base.

Data Sample

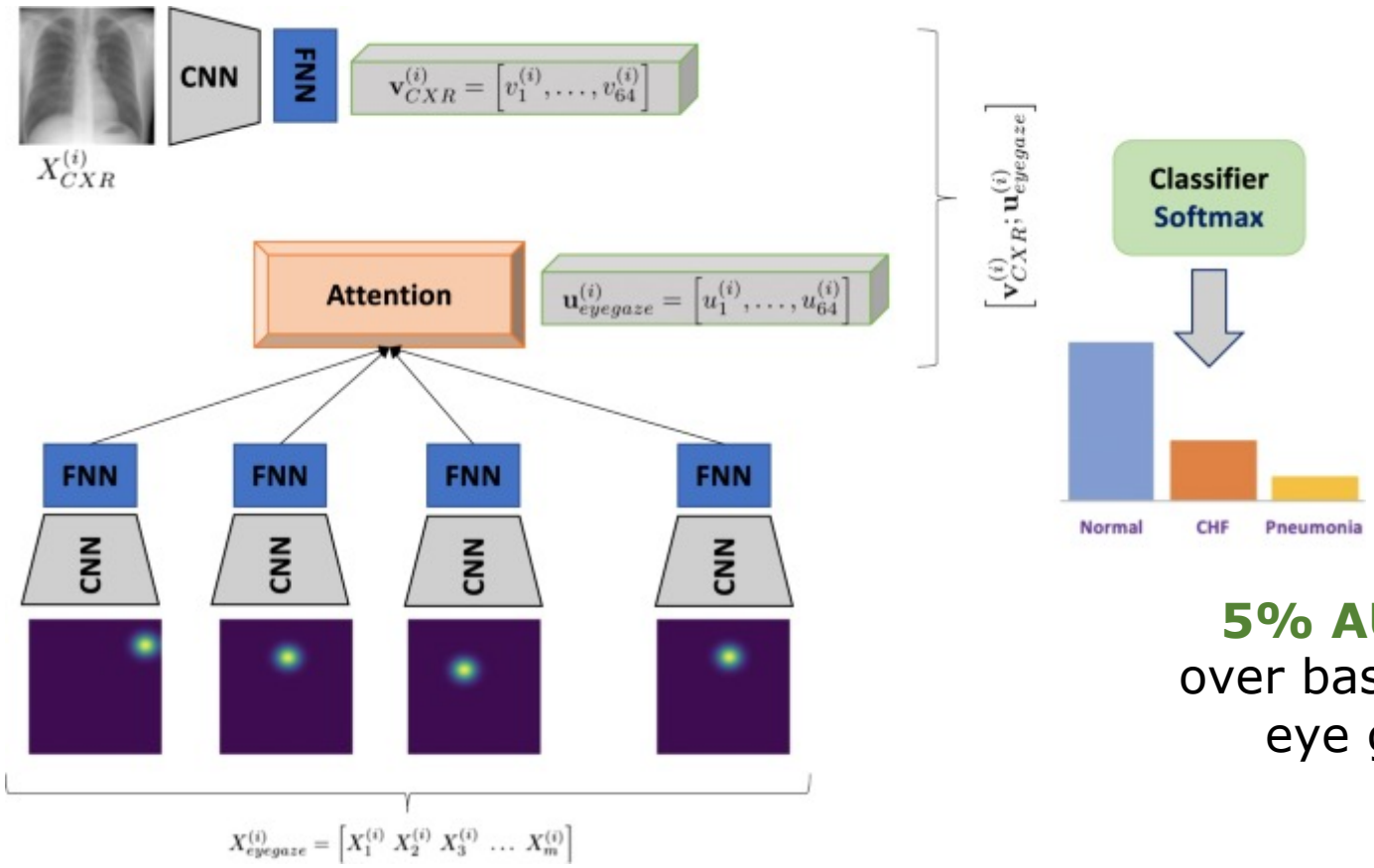
- **Eye gaze** data
- **Audio** (synced with eye gaze)
- **Transcript** (time-stamped)
- **Bounding boxes** of anatomies
  17 **anatomical regions**
- **Segmentation maps**
  of key anatomies







ML Experiment #1: Temporal Gaze as Input



5% AUC improvement
over baseline model without
eye gaze information

[8] Karargyris, A., Kashyap, S., Lourentzou, I., Wu, J.T., Sharma, A., Tong, M., Abedin, S., Beymer, D., Mukherjee, V., Krupinski, E.A. and Moradi, M., Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development. Nature Scientific data (2021)

https://tinyurl.com/EyeGaze2021

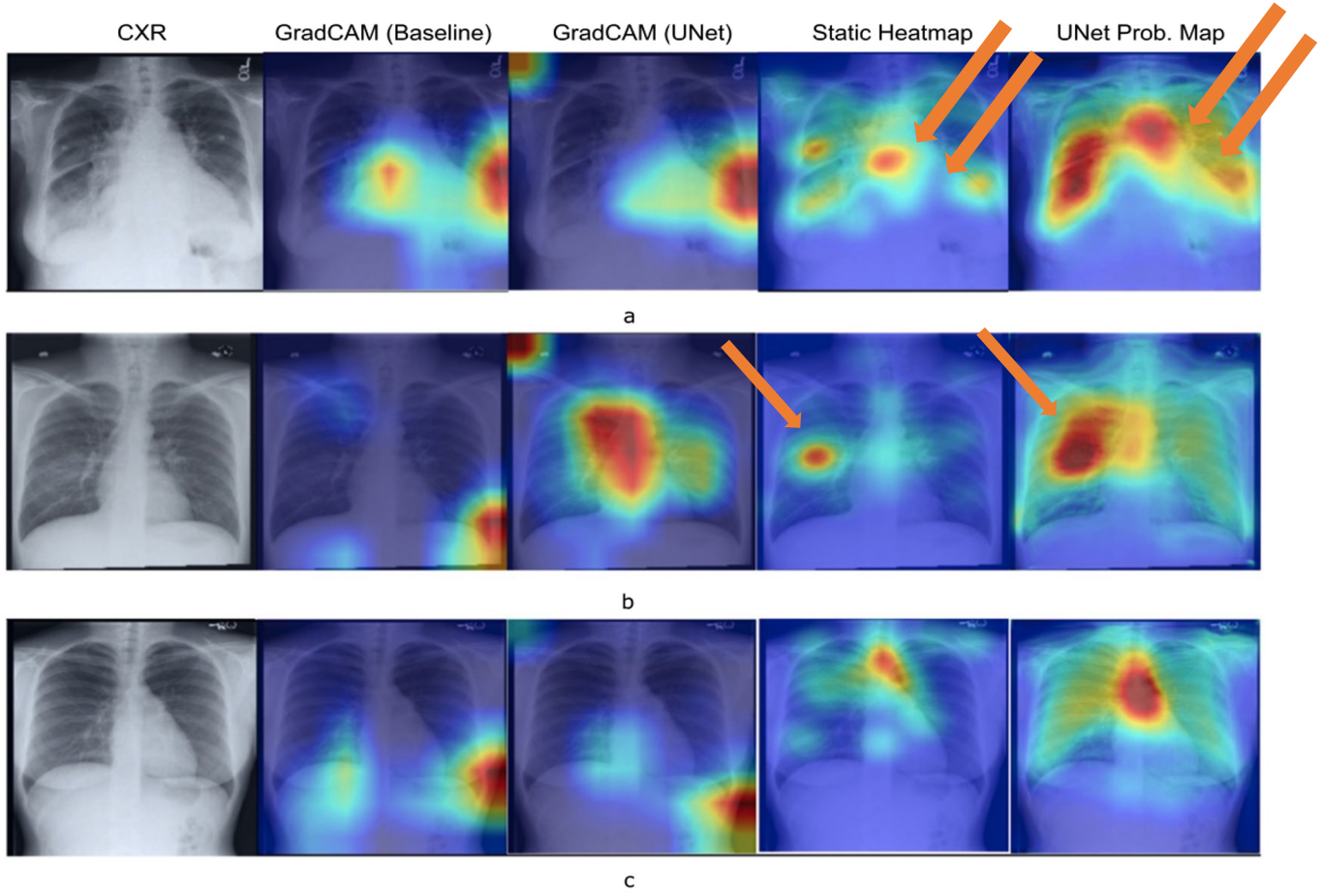
SCAN ME

ML Experiment #2: Static Gaze as Output

CHF: Gaze fixations increase around the hilar and heart region

Pneumonia: Gaze fixations focus on the lung opacity

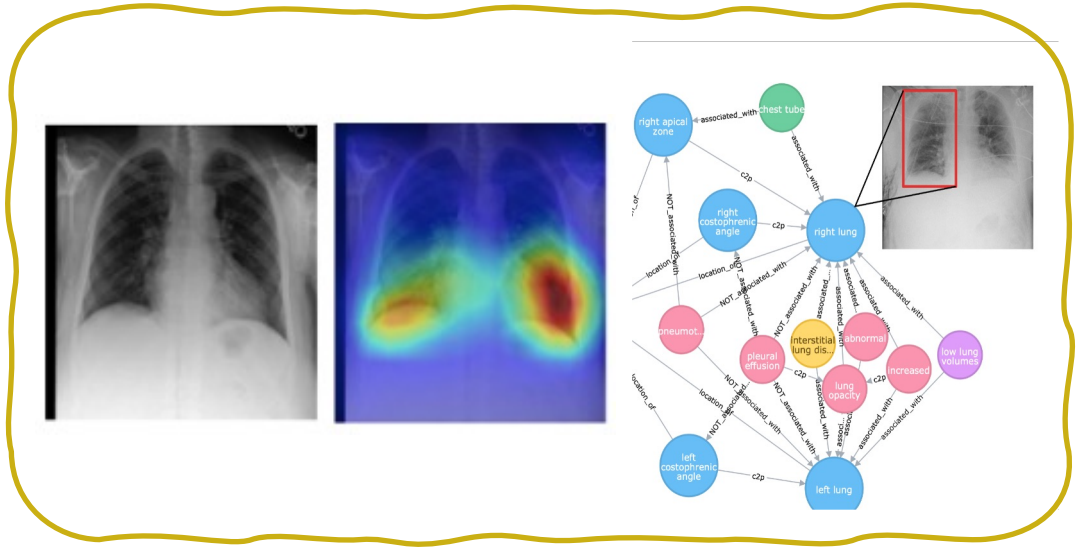
Normal: Gaze fixations skip around the image.

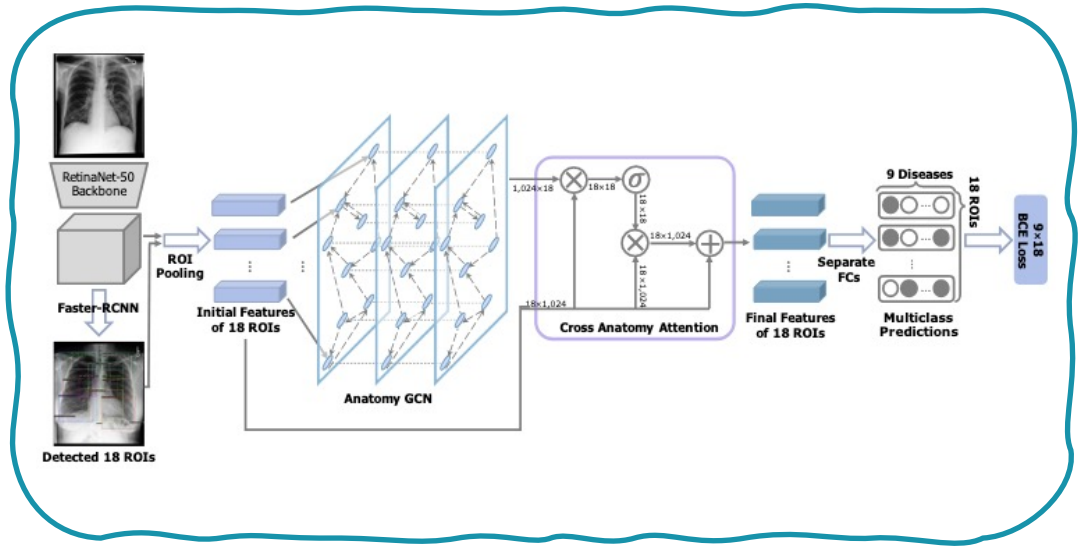


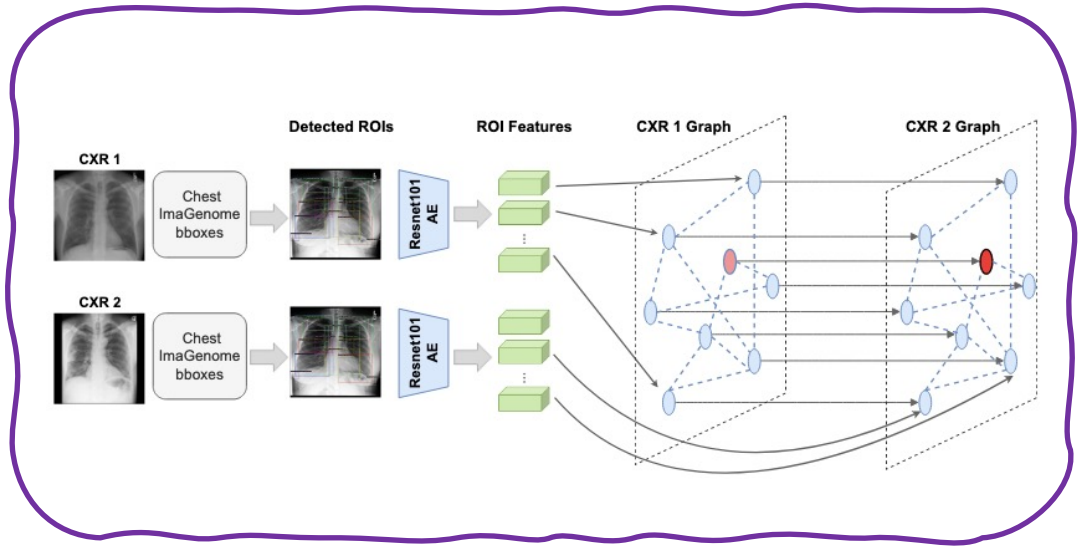
[8] Karargyris, A., Kashyap, S., Lourentzou, I., Wu, J.T., Sharma, A., Tong, M., Abedin, S., Beymer, D., Mukherjee, V., Krupinski, E.A. and Moradi, M., Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development. Nature Scientific data (2021)

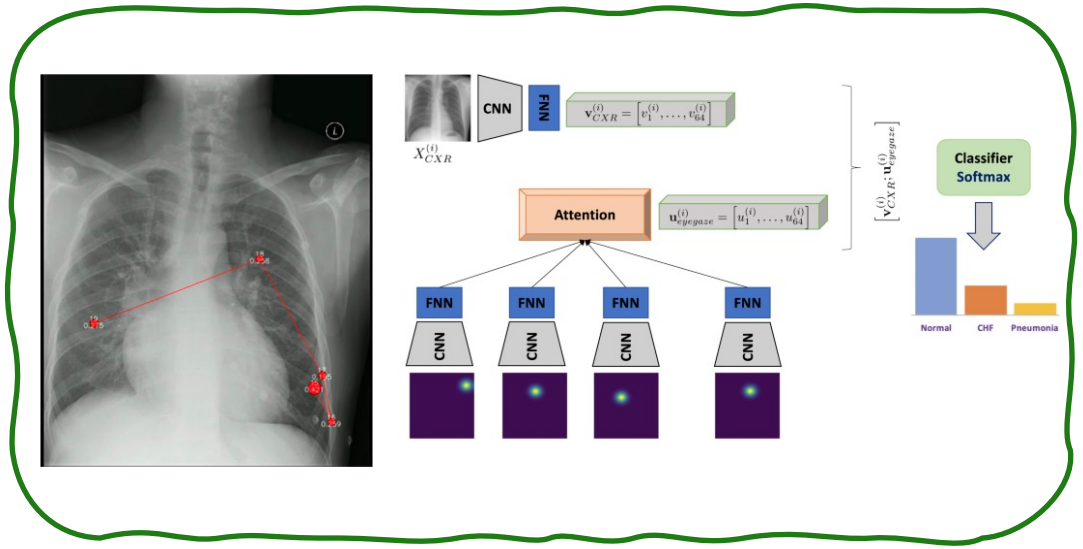
https://tinyurl.com/EyeGaze2021

To summarize ...









ML still fails to see the bigger picture

Source: https://blog.pepid.com/wp-content/uploads/2018/05/RobotsForNursesHero.png

Under-specification: Average case model performance on fixed test data

Data routinely collected might be influenced by social, institutional, and other types of biases

ML decisions cannot be same across all users

When designing new ML methods, we need to be able to …

Source: https://blog.pepid.com/wp-content/uploads/2018/05/RobotsForNursesHero.png

- ✓ Identify failure modes: **worst-case** performance
- ✓ Systematically assess **model** AND **data quality**
- ✓ Explicitly control for various biases
- ✓ Incorporate ways to define and assess model behavior when **deployed in the real-world**, when no ground-truth data are available

Thank you to...


Gaurang Karwande


Amarachi Mbakwe


Leo A. Celi


Joy T. Wu


Alexandros Karargyris

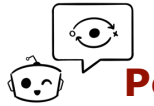

Nkechinyere N. Agu


Satyananda Kashyap


Mehdi Moradi


Pingkun Yan


James Hendler



Perception and LANguage (PLAN) Lab @VT

https://plan-lab.github.io/

VIRGINIA TECH

https://tinyurl.com/ChestImaGenome2021

Chest ImaGenome Dataset for Clinical Reasoning (NeurIPS 2021 Datasets)

https://tinyurl.com/AnaXNet2021

Anaxnet: Anatomy aware Multi-label Finding Classification in Chest X-ray

https://tinyurl.com/CheXRelNet

CheXRelNet: An Anatomy-Aware Model for Tracking Longitudinal Relationships between Chest X-Rays (MICCAI 2022)

https://tinyurl.com/EyeGaze2021

Creation and Validation of a Chest X-ray Dataset with Eye-Tracking and Report Dictation for AI Development.

IBM Research

Rensselaer

McMaster University

imes INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

MIT

Beth Israel Deaconess Medical Center

HARVARD MEDICAL SCHOOL TEACHING HOSPITAL

Stanford MEDICINE

National Science Foundation

amazon | science

DARPA

Commonwealth Cyber Initiative