The Promise of AI for Personalized Medicine based on Medical Images

Les promesses de l'IA pour une médecine personnalisée basée sur les images médicales

Tal Arbel, PhD

Professor, McGill University

Canada CIFAR AI Chair, Mila

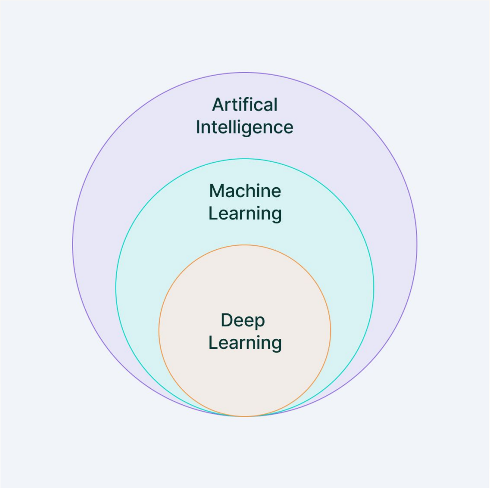
Department of Electrical and Computer Engineering

Director Probabilistic Vision Group, Medical Imaging Lab

Centre for Intelligent Machines

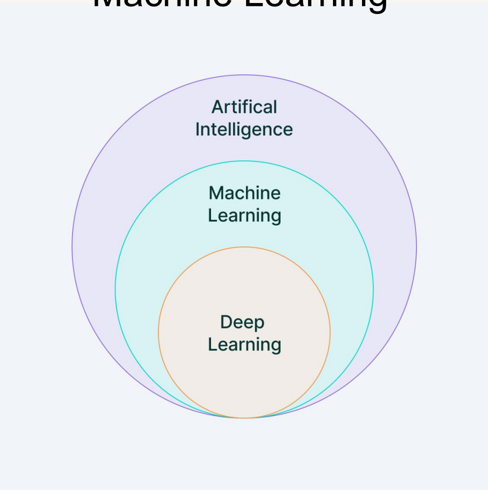
Prof. Tal Arbel

Machine Learning

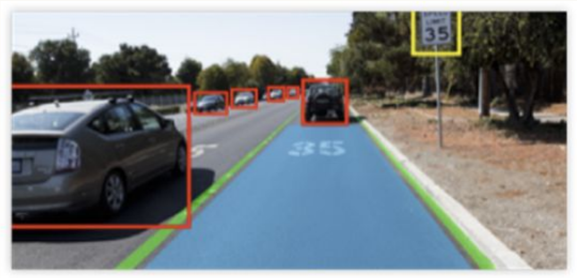


Prof. Tal Arbel

Machine Learning

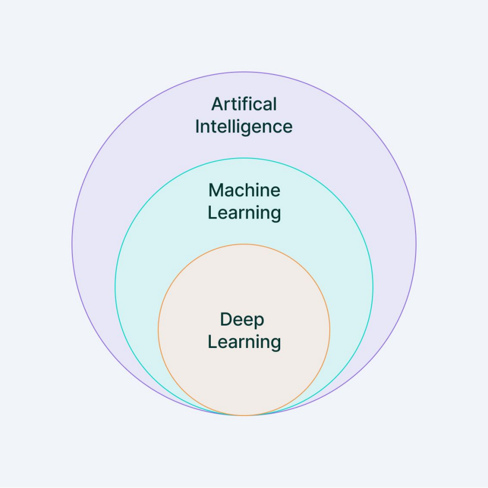


Computer Vision

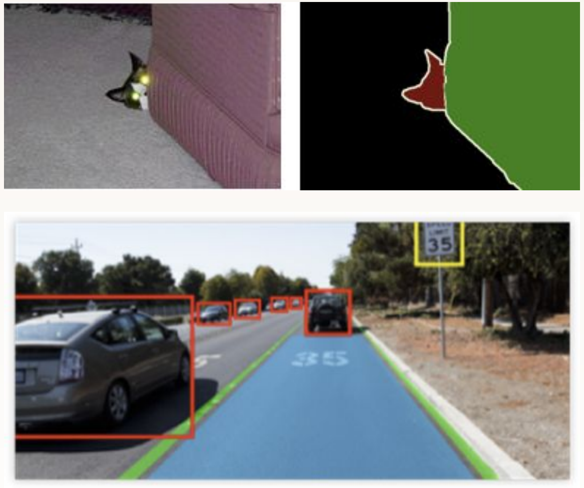


Prof. Tal Arbel

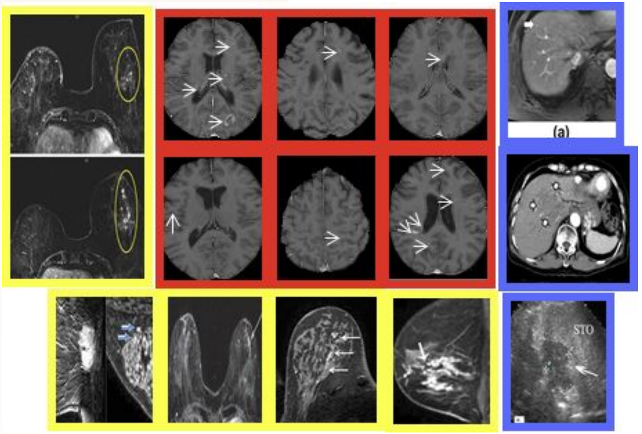
Machine Learning



Computer Vision

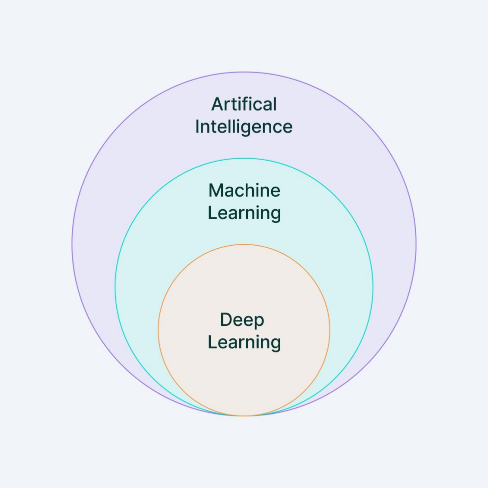


Medical Image Analysis

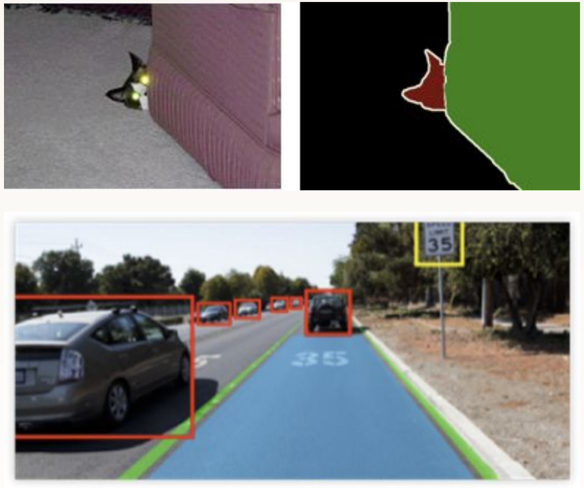


Prof. Tal Arbel

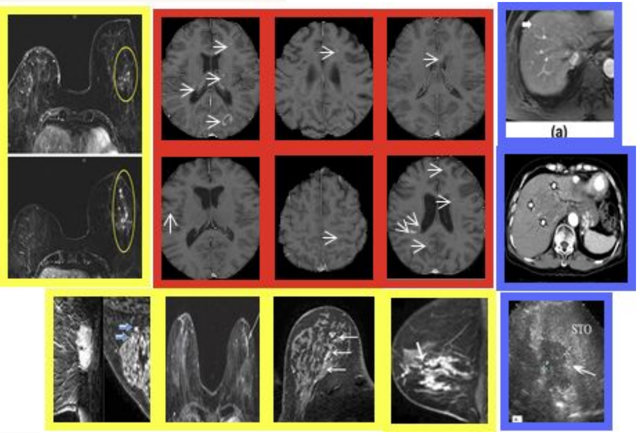
Machine Learning



Computer Vision



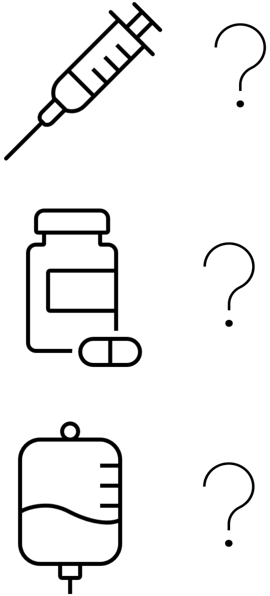
Medical Image Analysis



Real clinical applications: AI medical image analysis tools developed in my lab used in analysis of clinical trials for *almost all* new treatments for multiple sclerosis

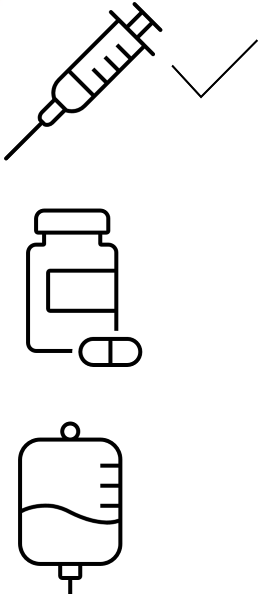
Clinical Scenario - Current Practice

Clinical Scenario - Current Practice



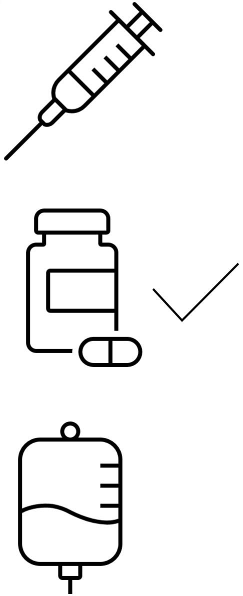
- Variety of treatments available for this patient’s illness

Clinical Scenario - Current Practice



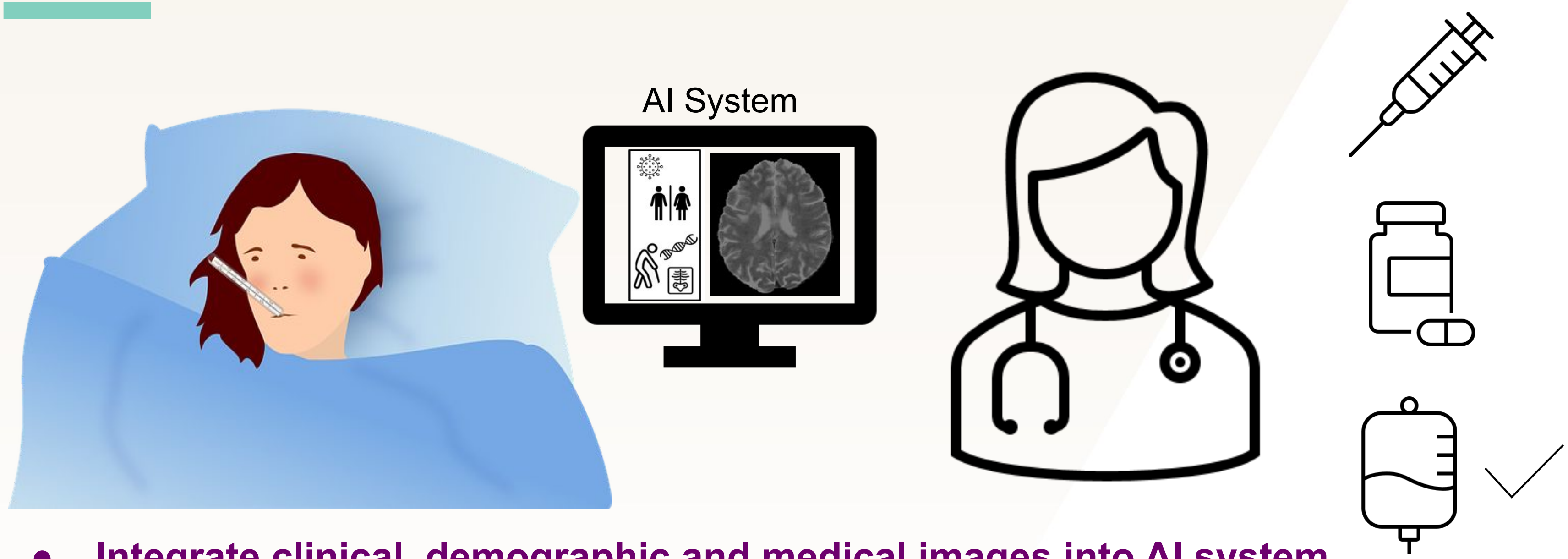
- Variety of treatments available for this patient’s illness
- **Treatment decision based on efficacy:**
  - Average treatment efficacy across population, choose highest

Clinical Scenario – Personalized Medicine



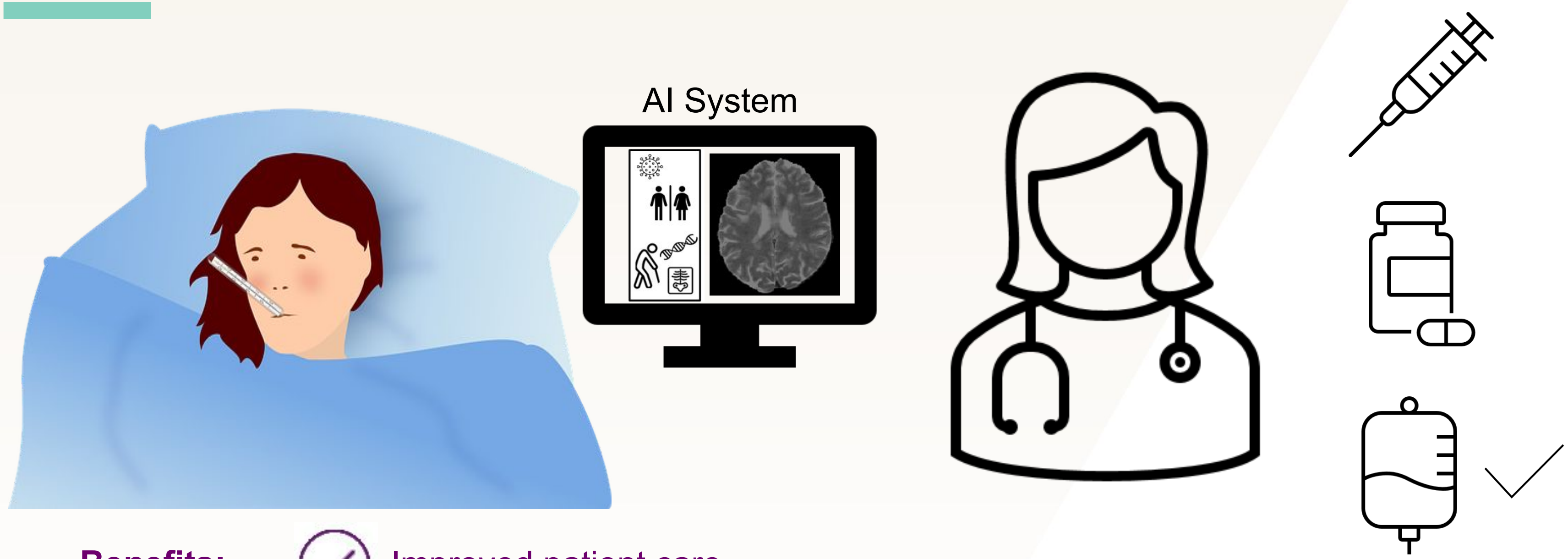
- Clinical and demographic information available
- **Treatment decision:** Average treatment efficacy conditioned on sub-group statistics

The Promise of AI for Image-Based Personalized Medicine



- **Integrate clinical, demographic and medical images into AI system**
- Provide clinicians with an AI tool which predicts future individual treatment response on several treatments using *discovered* image features

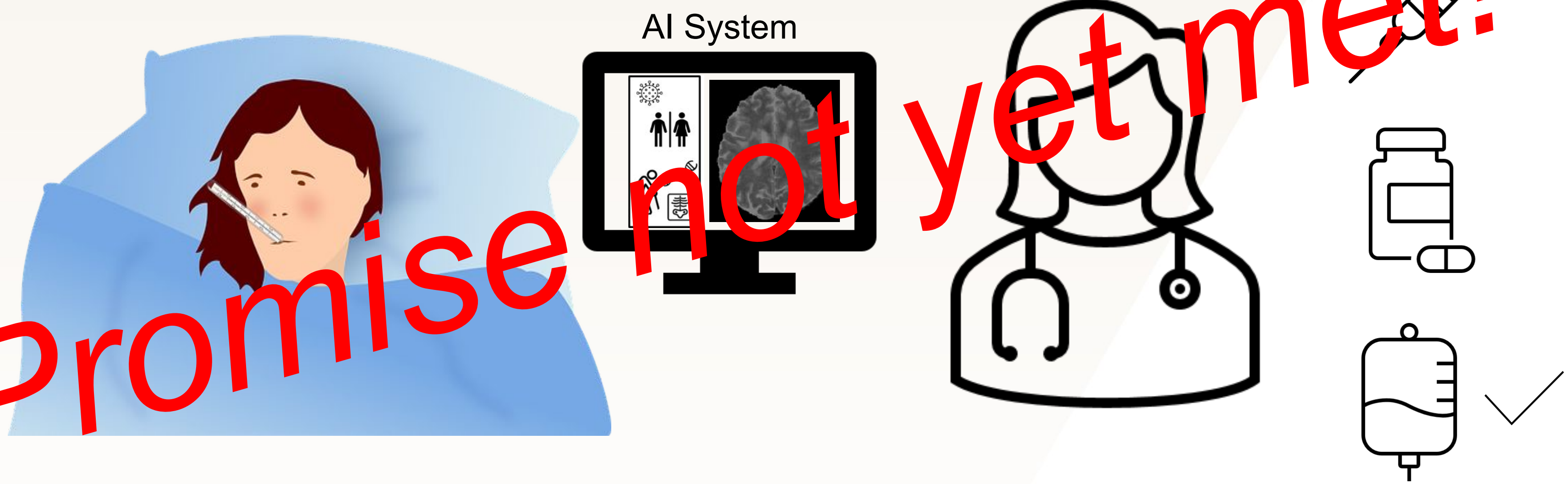
The Promise of AI for Image-Based Personalized Medicine



Benefits:

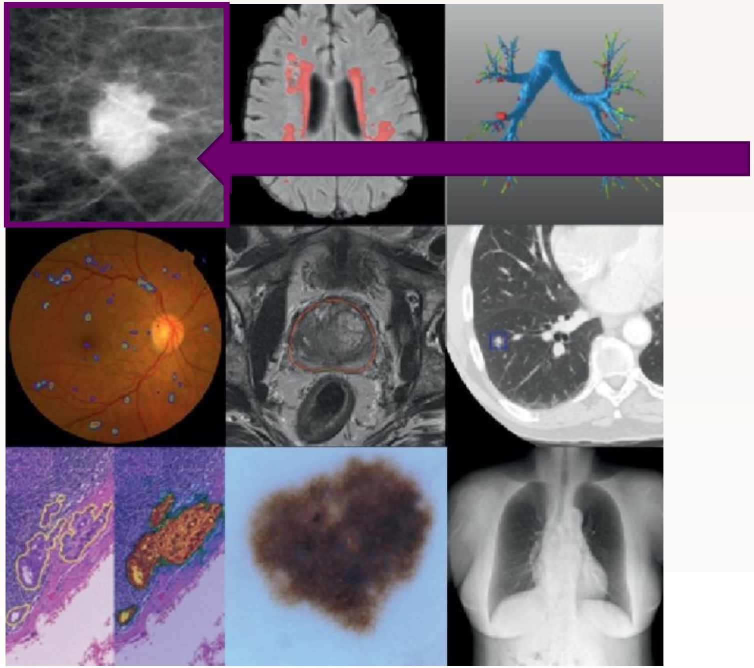
- Improved patient care
- Clinical trial enrichment

The Promise of AI for Image-Based Personalized Medicine



Promise not yet met!

Deep Learning for Medical Image Analysis



Medical imaging applications where deep learning models have achieved the SOTA

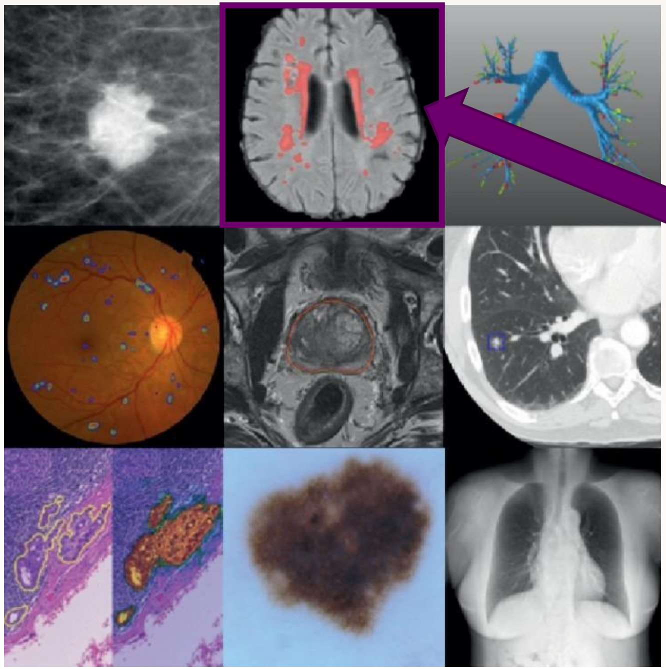
Mammographic mass classification

Segmentation of lesions in the brain

Breast cancer metastases detection

https://www.sciencedirect.com/science/article/pii/S1361841517301135

Deep Learning for Medical Image Analysis



Medical imaging applications where deep learning models have achieved the SOTA

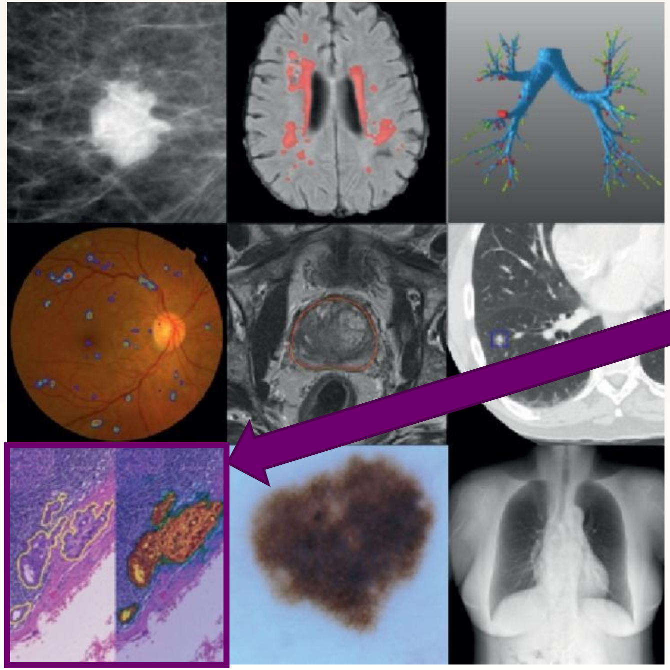
Mammographic mass classification

Segmentation of lesions in the brain

Breast cancer metastases detection

https://www.sciencedirect.com/science/article/pii/S1361841517301135

Deep Learning for Medical Image Analysis



Medical imaging applications where deep learning models have achieved the SOTA

Mammographic mass classification

Segmentation of lesions in the brain

Breast cancer metastases detection

Why haven't they been widely integrated into clinical workflow?

https://www.sciencedirect.com/science/article/pii/S1361841517301135

Unique Challenges: Deep Learning in the Clinic

Very large medical images but

... Lack of relevant training data

Unique Challenges: Deep Learning in the Clinic

Very large medical images but

... Lack of relevant training data

- Not robust/generalizable to real patient/data/label variability

ARTIFICIAL INTELLIGENCE

Hundreds of AI tools have been built to catch covid. None of them helped.

https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/

Unique Challenges: Deep Learning in the Clinic

Very large medical images but

... Lack of relevant training data

- Not robust/generalizable to real patient/data/label variability

- DL for medical imaging make (potentially deadly) mistakes.

Researchers Investigate When, How Healthcare AI Models Will Fail

https://healthitanalytics.com/news/researchers-investigate-when-how-healthcare-ai-models-will-fail

Unique Challenges: Deep Learning in the Clinic

Very large medical images but

... Lack of relevant training data

- Not robust/generalizable to real patient/data/label variability

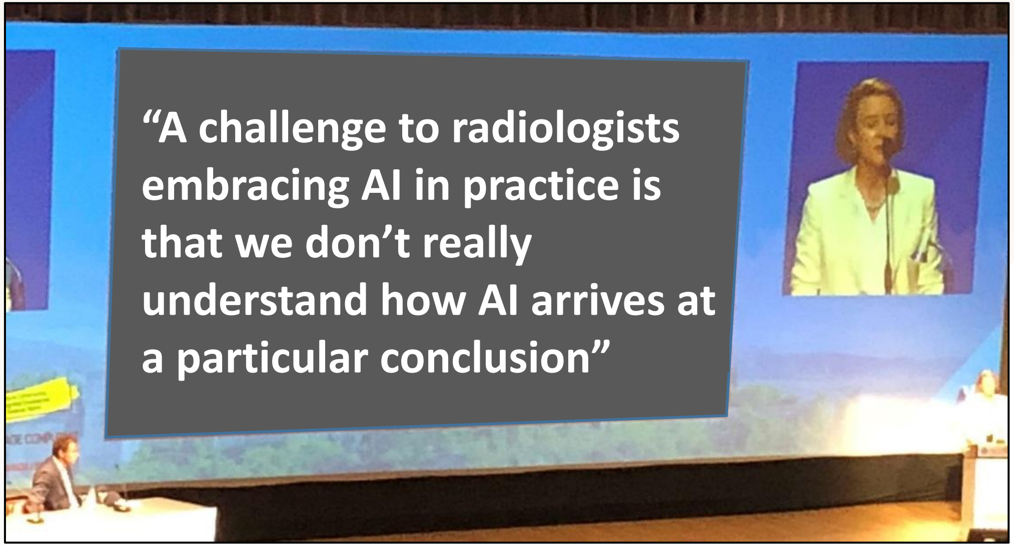
- DL for medical imaging make (potentially deadly) mistakes.

Researchers Investigate When, How Healthcare AI Models Will Fail

Need to convey uncertainty in AI predictions!

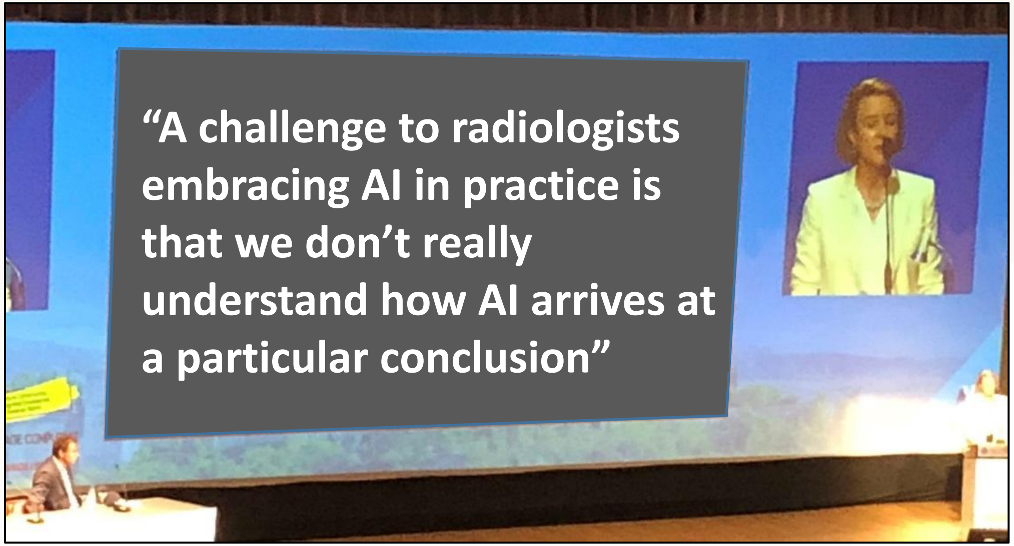
https://healthitanalytics.com/news/researchers-investigate-when-how-healthcare-ai-models-will-fail

Lack of Interpretability of Deep Learning Models



Dr. McGinty, Chair of the American College Radiology Board of Chancellors, *MICCAI 2018*

Lack of Interpretability of Deep Learning Models



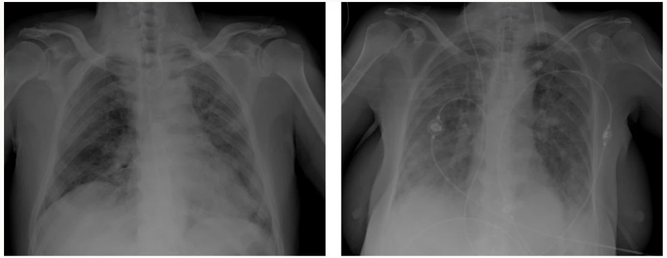
Dr. McGinty, Chair of the American College Radiology Board of Chancellors, *MICCAI 2018*

→ Need to open up the black box!

Deep Learning Models Can Be Biased

BRIEF REPORT | APPLIED MATHEMATICS |

Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis



(a) Male (b) Female

AI skin cancer diagnoses risk being less accurate for dark skin – study

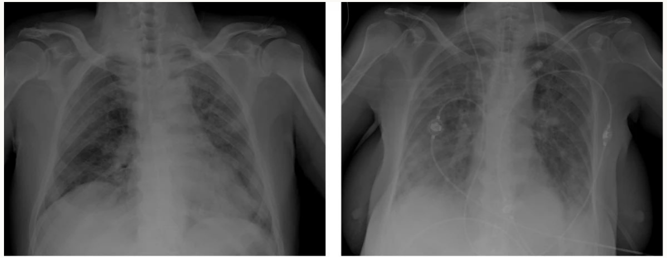
Research finds few image databases available to develop technology contain details on ethnicity or skin type

https://www.pnas.org/doi/10.1073/pnas.1919012117
https://www.theguardian.com/society/2021/nov/09/ai-skin-cancer-diagnoses-risk-being-less-accurate-for-dark-skin-study

Deep Learning Models Can Be Biased

BRIEF REPORT | APPLIED MATHEMATICS

Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis



(a) Male (b) Female

AI skin cancer diagnoses risk being less accurate for dark skin – study

Research finds few image databases available to develop technology contain details on ethnicity or skin type

→ Need to mitigate the biases

https://www.pnas.org/doi/10.1073/pnas.1919012117
https://www.theguardian.com/society/2021/nov/09/ai-skin-cancer-diagnoses-risk-being-less-accurate-for-dark-skin-study

Overview of the Talk

First deep learning model for personalized prediction from patients images

<u>Trustworthiness</u> and <u>reliability</u> of deep learning models needed in clinical applications:

Overview of the Talk

First deep learning model for personalized prediction from patients images

<u>Trustworthiness</u> and <u>reliability</u> of deep learning models needed in clinical applications:

- Uncertainty Estimation

Overview of the Talk

First deep learning model for personalized prediction from patients images

<u>Trustworthiness</u> and <u>reliability</u> of deep learning models needed in clinical applications:

- Uncertainty Estimation
- Explainability & Discovery of predictive image markers

Overview of the Talk

First deep learning model for personalized prediction from patients images

<u>Trustworthiness</u> and <u>reliability</u> of deep learning models needed in clinical applications:

- Uncertainty Estimation
- Explainability & Discovery of predictive image markers
- Improving fairness

Overview of the Talk

First deep learning model for personalized prediction from patients images

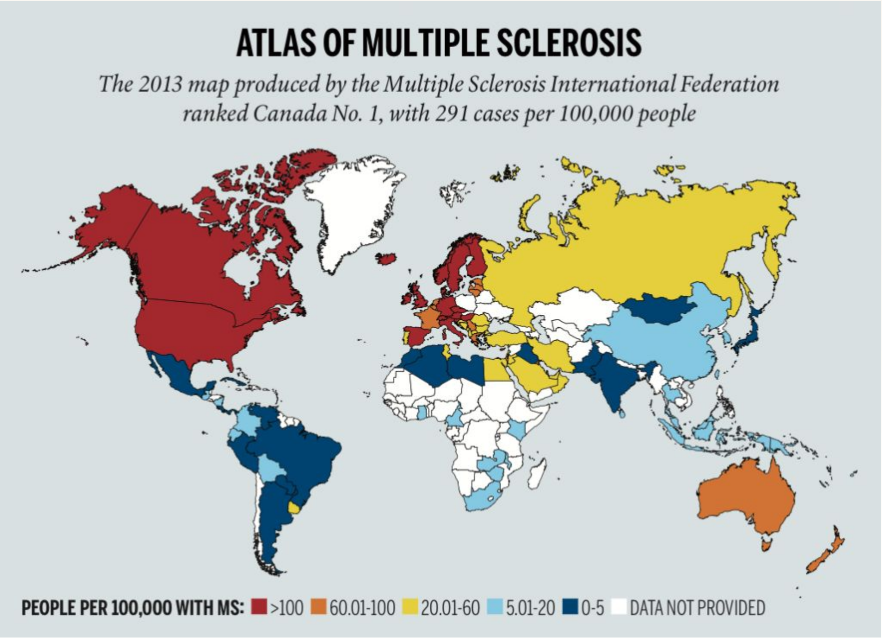
Trustworthiness and reliability of deep learning models needed in clinical applications:

- Uncertainty Estimation
- Explainability & Discovery of predictive image markers
- Improving fairness

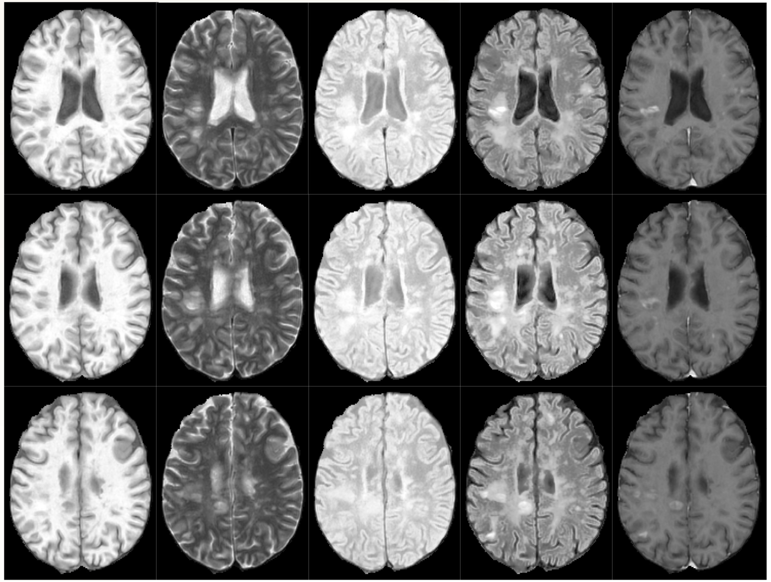
Case study: Multiple Sclerosis - Long term, complex neurological disease evolution

Case Study: Multiple Sclerosis

Most common neurological disease affecting young people; Canada has highest rate per capita.



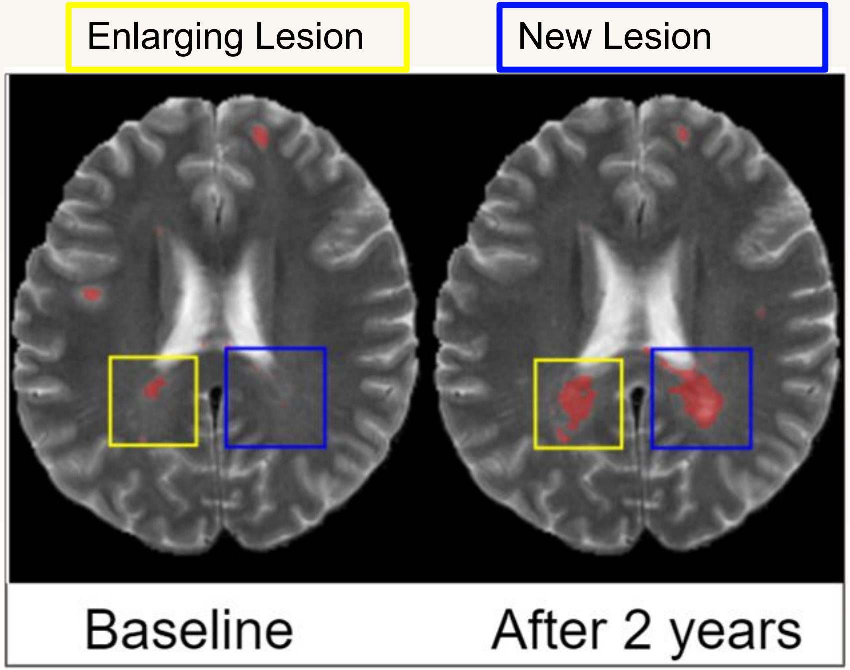
Multi-focal brain lesions visible on MRI



T1w T2w PDw FLR T1c

https://macleans.ca/society/health/could-canada-cause-multiple-sclerosis/

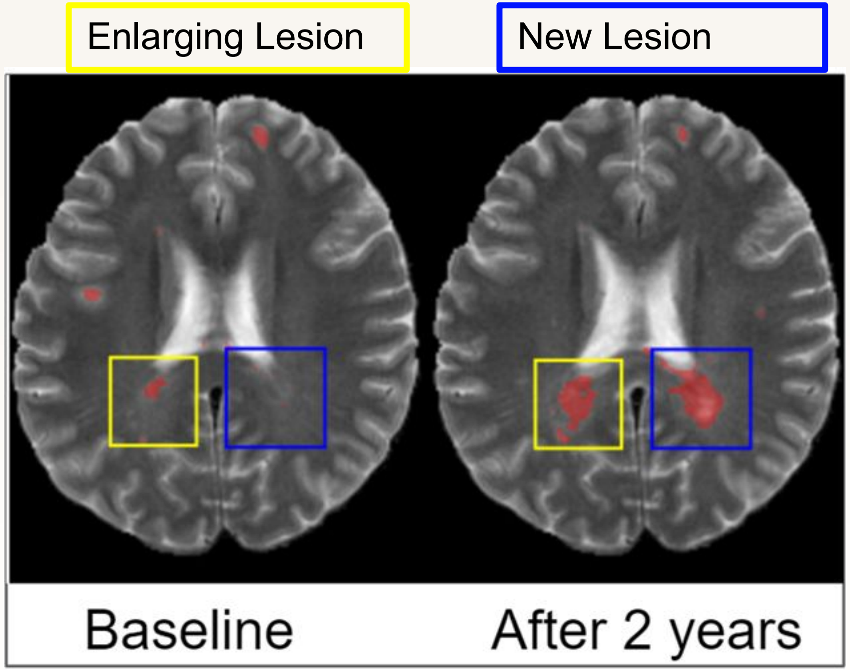
Case Study: Multiple Sclerosis



Appearance of new/enlarging (NE) lesions on successive MRI scans important:

- MRI markers of new disease activity since previous scan

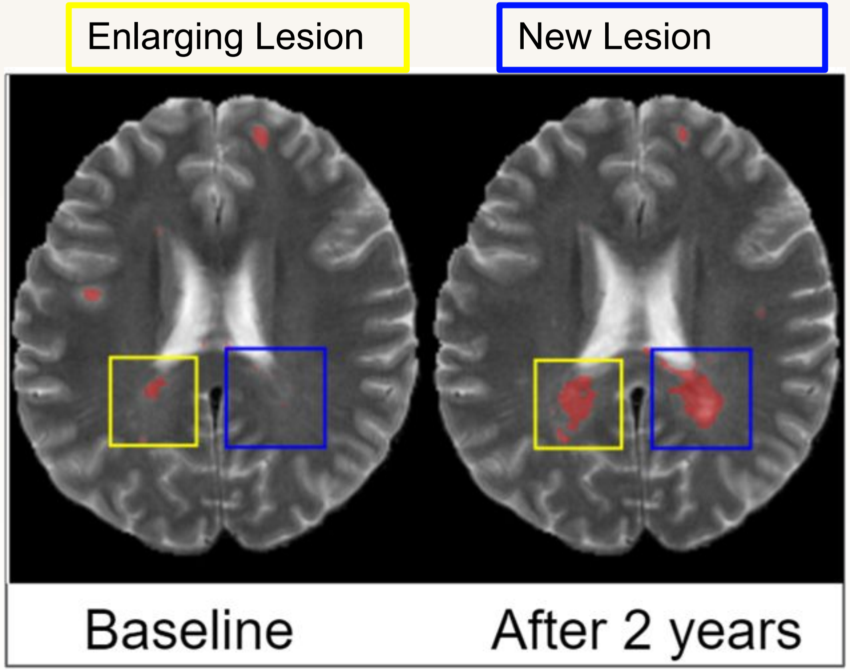
Case Study: Multiple Sclerosis



Appearance of new/enlarging (NE) lesions on successive MRI scans important:

- MRI markers of new disease activity since previous scan
- Treatments exist to help suppress new lesions, manage symptoms (not to stop progression)
  - Different efficacies
  - Risk profiles, etc.

Case Study: Multiple Sclerosis

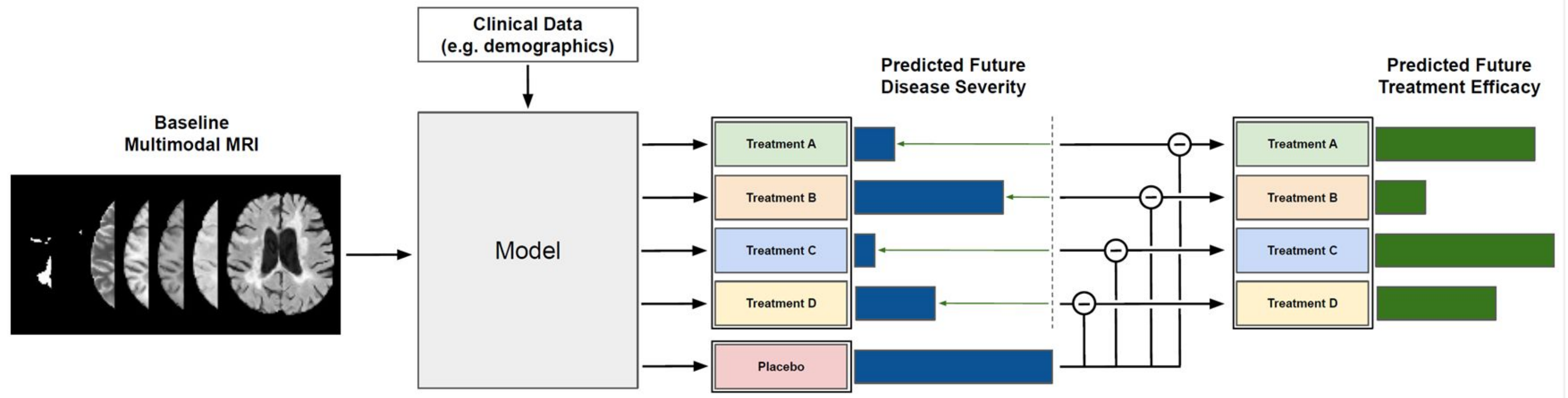


Appearance of new/enlarging (NE) lesions on successive MRI scans important:

- MRI markers of new disease activity since previous scan
- Treatments exist to help suppress new lesions, manage symptoms (not to stop progression)
  - Different efficacies
  - Risk profiles, etc.
- No Cure.

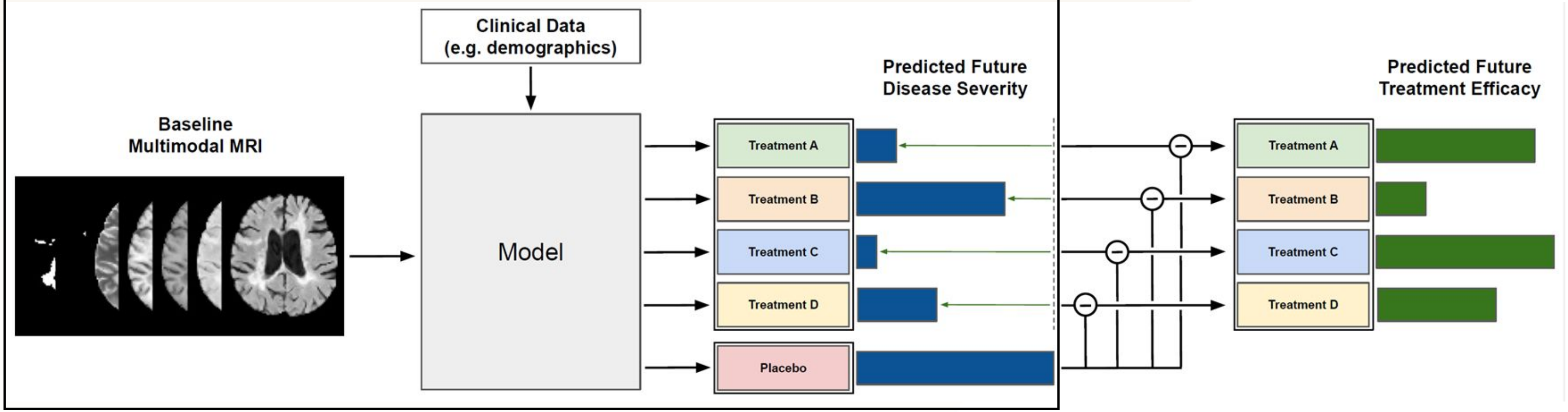
Deep Learning for Image-Based Precision Medicine

DL model that learns data driven imaging markers predictive of future disease progression for individual patients on and off treatment



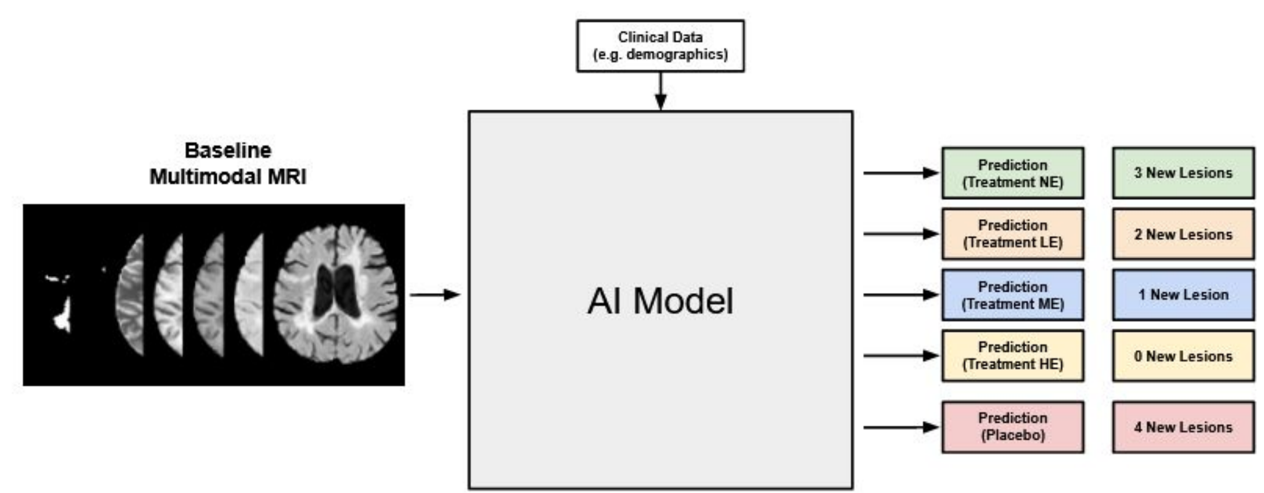
Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

Deep Learning for Personalized Prediction of Future Outcomes on and off Treatment from Images (Part-1)



Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

AI for Personalized Predictions of Future New Lesion Counts on and off Treatments from Images



System provides estimates of ***all*** treatment outcomes: **factual** and **counterfactual** (regardless of the true assignment)

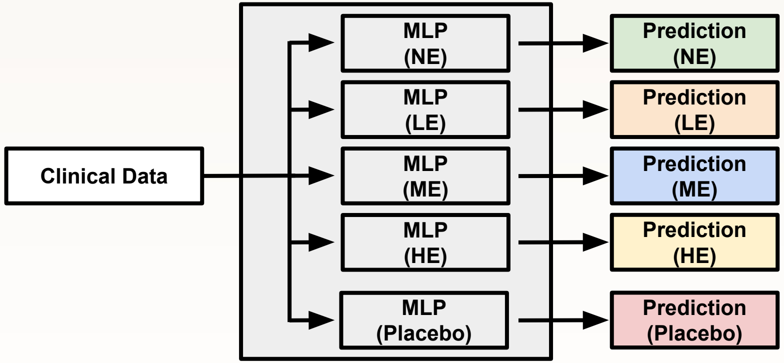
NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy

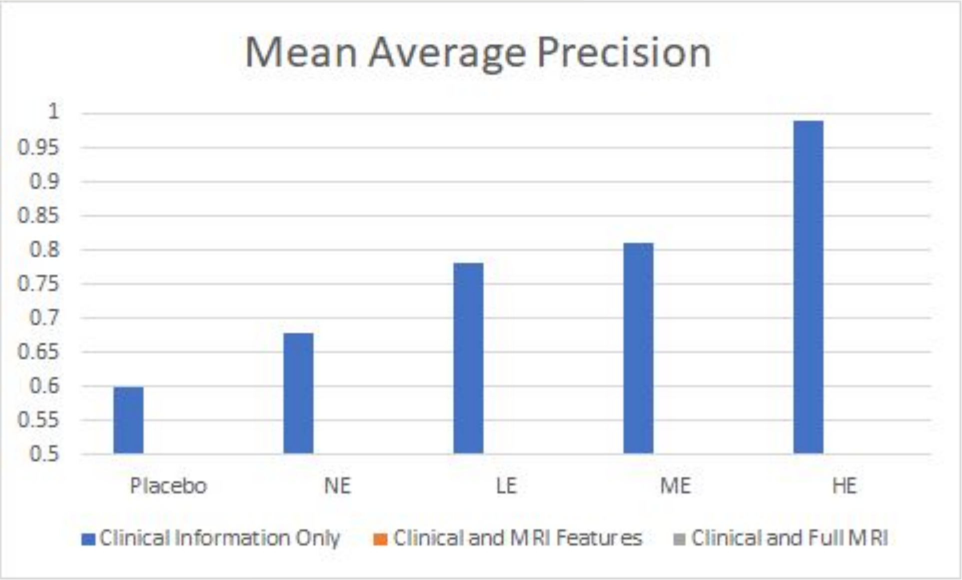
Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

Factual Model Results-Binarized Regression

Baseline 1:

Clinical features (Age, Sex, Baseline Disability)

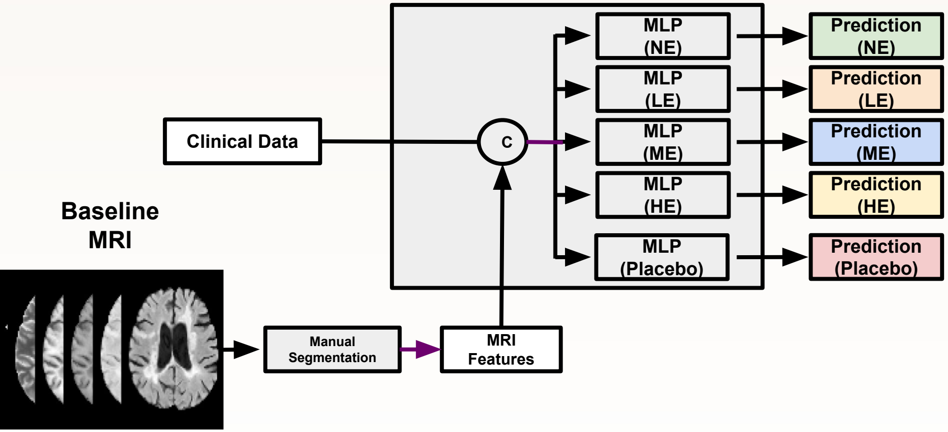


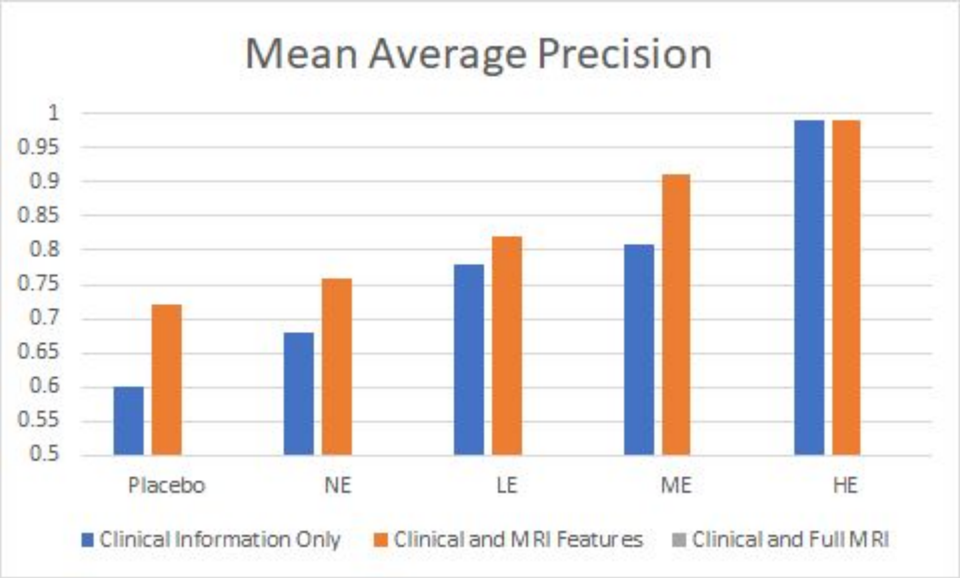


NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy

Factual Model Results-Binarized Regression

Baseline 2:
Clinical (Age, Sex, Baseline Disability)
+ MRI Derived Features (T2 Lesion Volume, Gad lesion count)





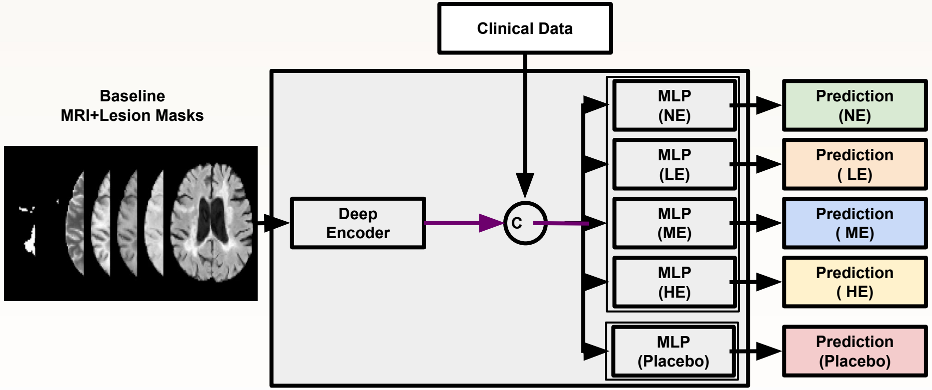
NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy

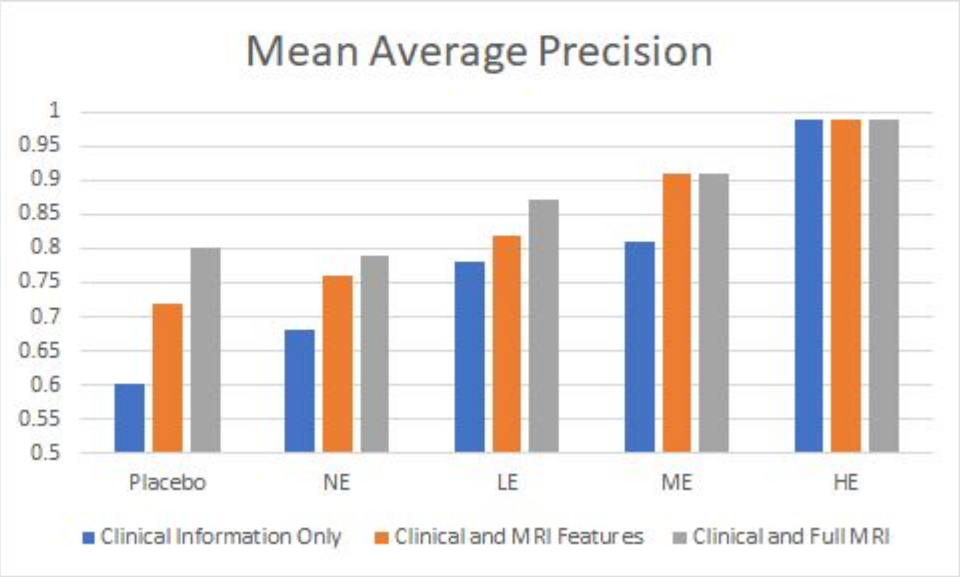
Factual Model Results-Binarized Regression

Our Model:

Clinical (Age, Sex, Baseline Disability)

+ Deep MRI Encoder

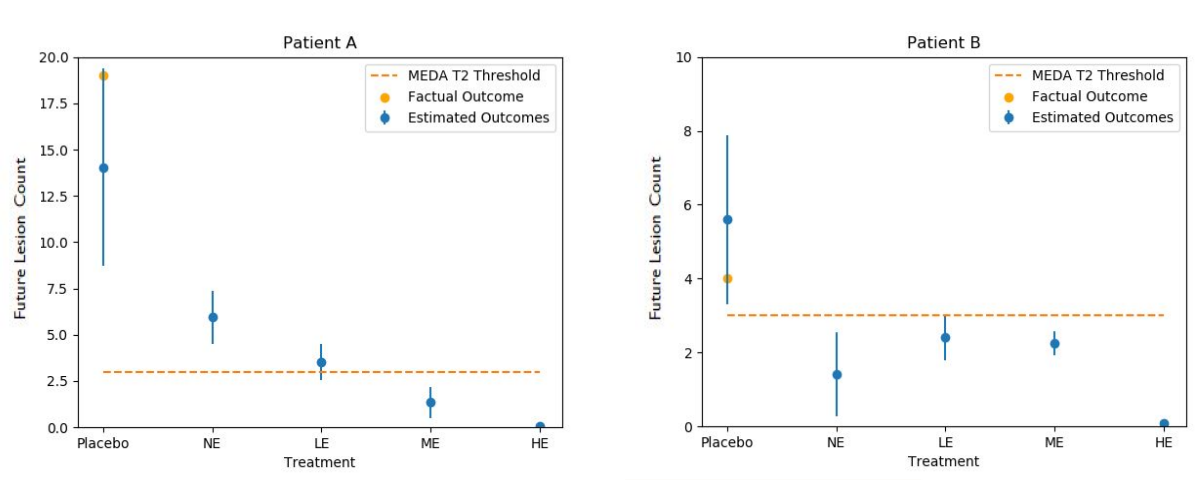




NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy

Deep Learning for Clinical Decision Support (Part 1)

Factual and Counterfactual Treatment Outcome Estimates

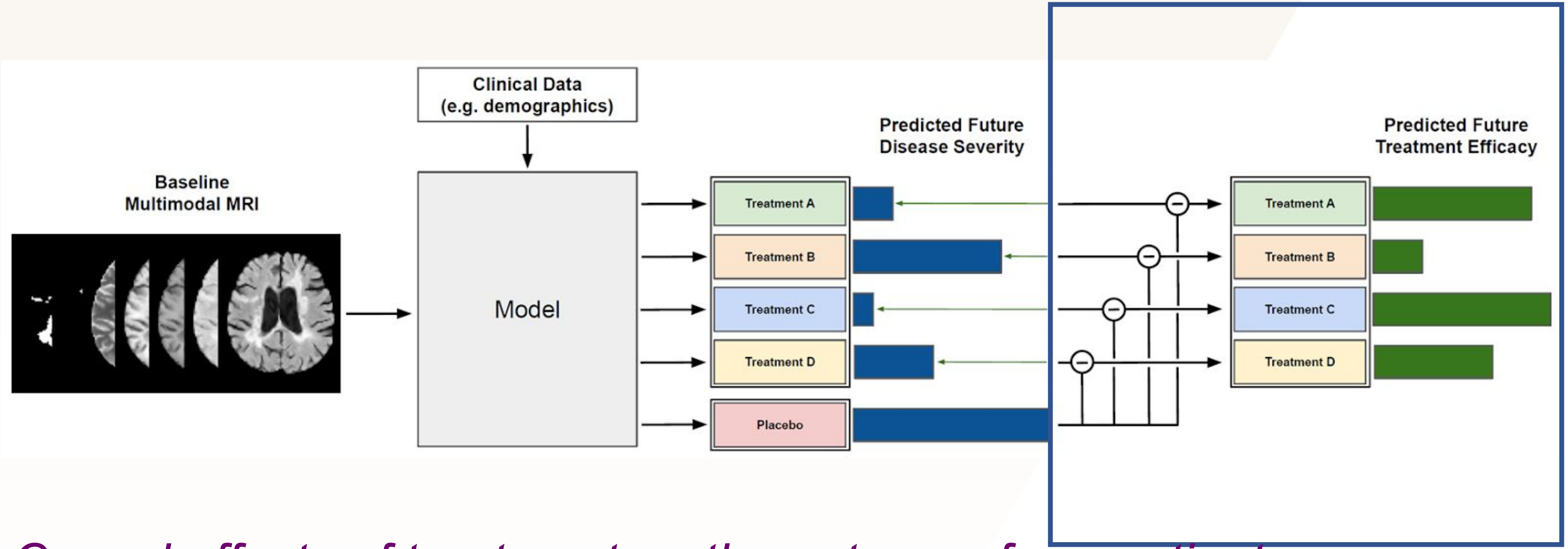


NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy
MEDA T2 threshold: >=3 NE lesions actionable number for DMT (therapy) escalation

Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

Deep Learning for Clinical Decision Support (Part 2)

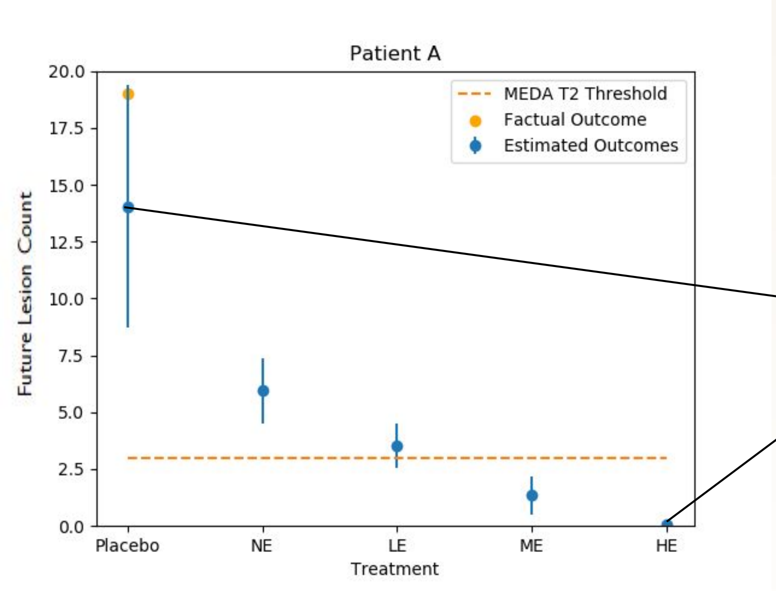
Predicting Future Treatment Effects



Causal effects of treatment on the outcome for a patient

Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

Estimating Future Personalized Treatment Response



Predicted Future Individual Treatment Response
= Reduction in future new lesion count relative to placebo (no treatment)

NE: No proven efficacy, **LE**: Lesser efficacy , **ME**: Moderate efficacy, **HE**: High efficacy

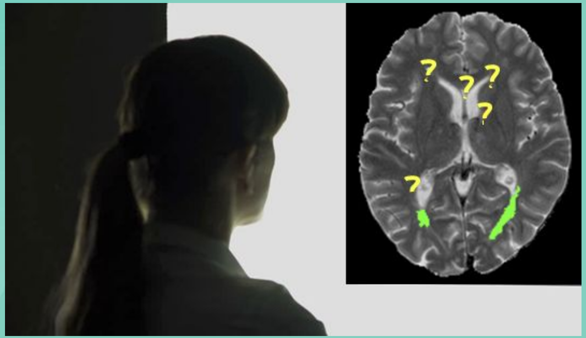
MEDA T2: 3 Future New T2 Lesions

Durso-Finley *et. al*, Conference on Medical Imaging with Deep Learning (MIDL) 2022

Great! Are we ready for clinical deployment?

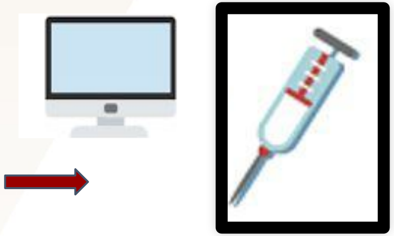
Trustworthy Image-Based Personalized Medicine

- Uncertainty Estimation
- Explainability
- Improving fairness



Trustworthy Image-Based Personalized Medicine

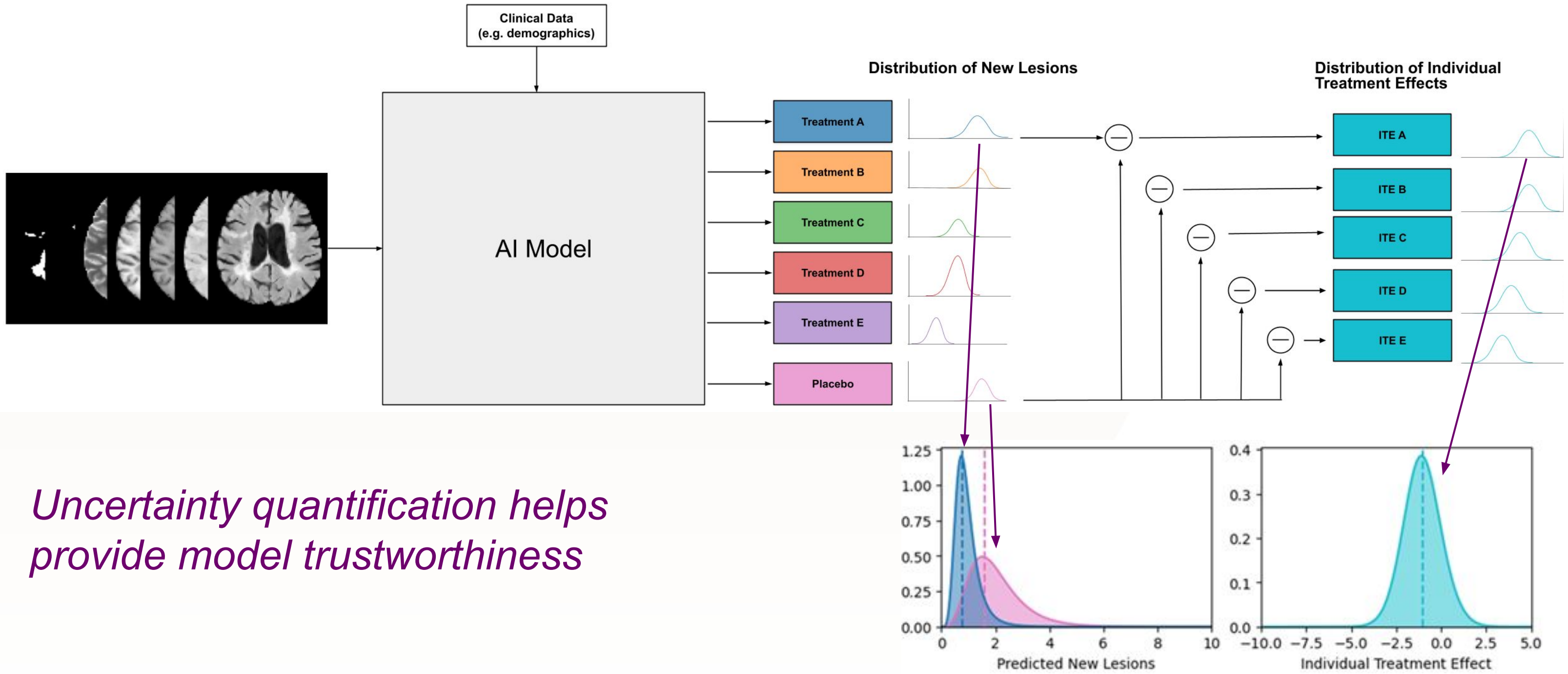
AI makes mistakes! High risk in handing over to clinician



Patient will be a future responder to Treatment A

What if we could quantify reliability of predictions made in form of uncertainty?

Trustworthy Treatment Effect Estimation



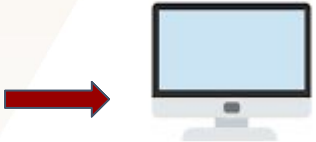
Uncertainty quantification helps provide model trustworthiness

Durso-Finley *et. al*, In submission, 2023

Trustworthy Image-Based Personalized Medicine

- Uncertainty Estimation
- Explainability
- Improving fairness

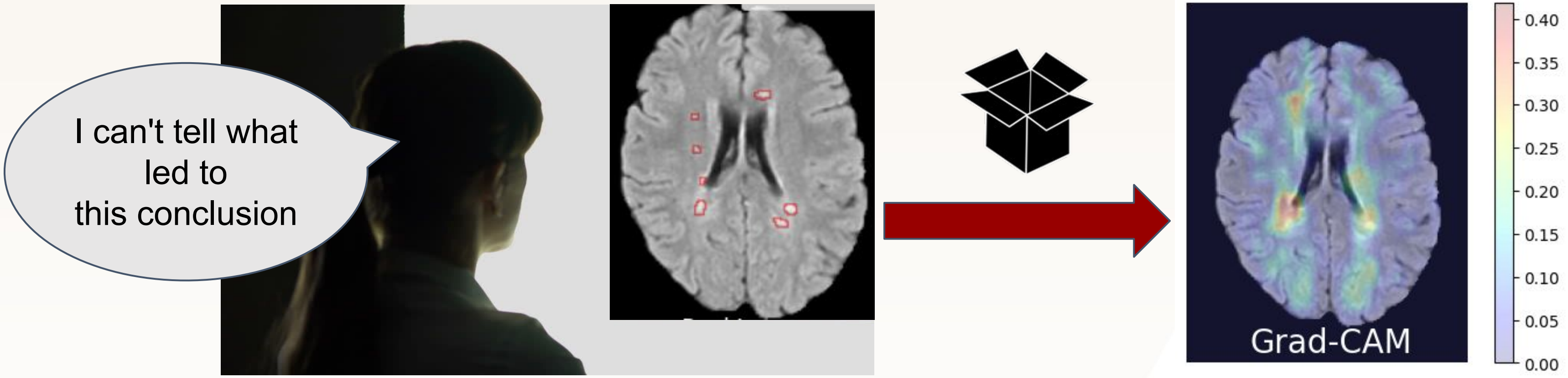
Trustworthy Image-Based Personalized Medicine



Patient will be a future responder to Treatment A

Explainable Deep Learning Models for Image Based Personalized Medicine – Opening up the black box

Future ***responder*** to treatment A



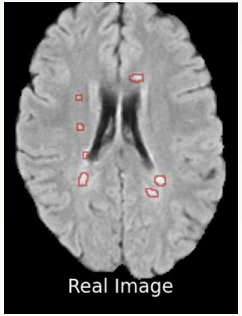
Where was the model looking at when it made its prediction?

***What** are the patient specific image markers that are predictive of future response?*

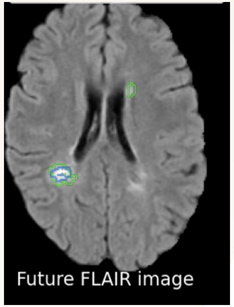
Kumar *et. al*, MIABID Workshop, MICCAI 2022

Explainability via Counterfactual Synthesis

Baseline



Year 1

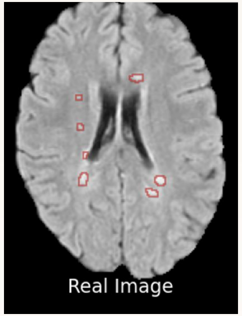


How would the patient's current (baseline) image change were it to have a different future disease outcome?

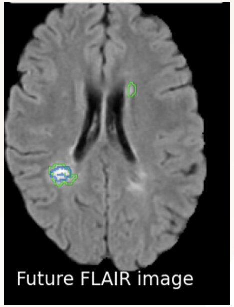
Kumar *et. al*, MIABID Workshop, MICCAI 2022

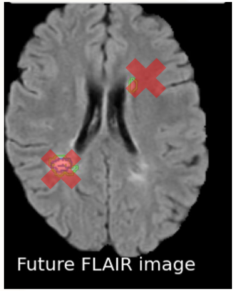
Explainability via Counterfactual Synthesis

Baseline



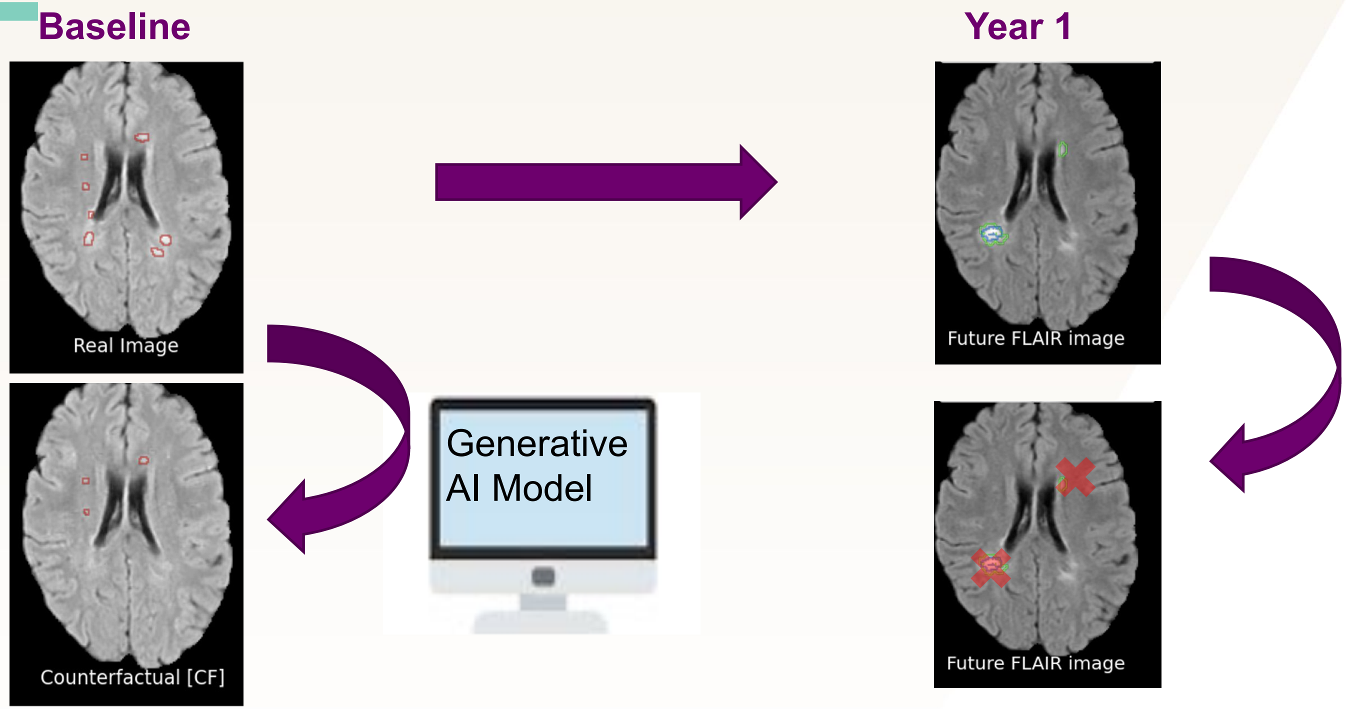
Year 1





Kumar *et. al*, MIABID Workshop, MICCAI 2022

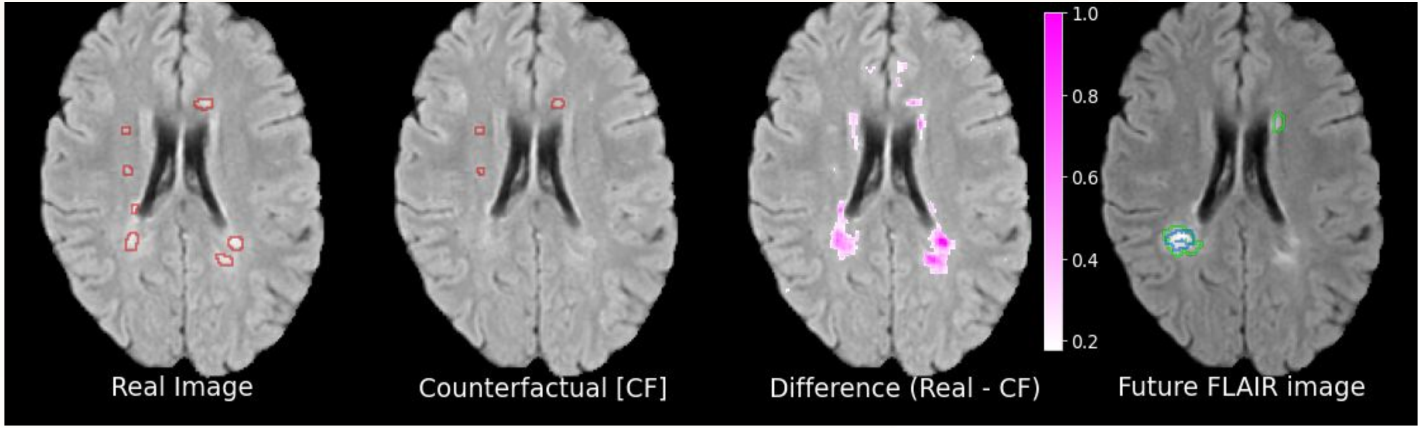
Explainability via Counterfactual Synthesis



Kumar *et. al*, MIABID Workshop, MICCAI 2022

Explainability via Counterfactual Synthesis

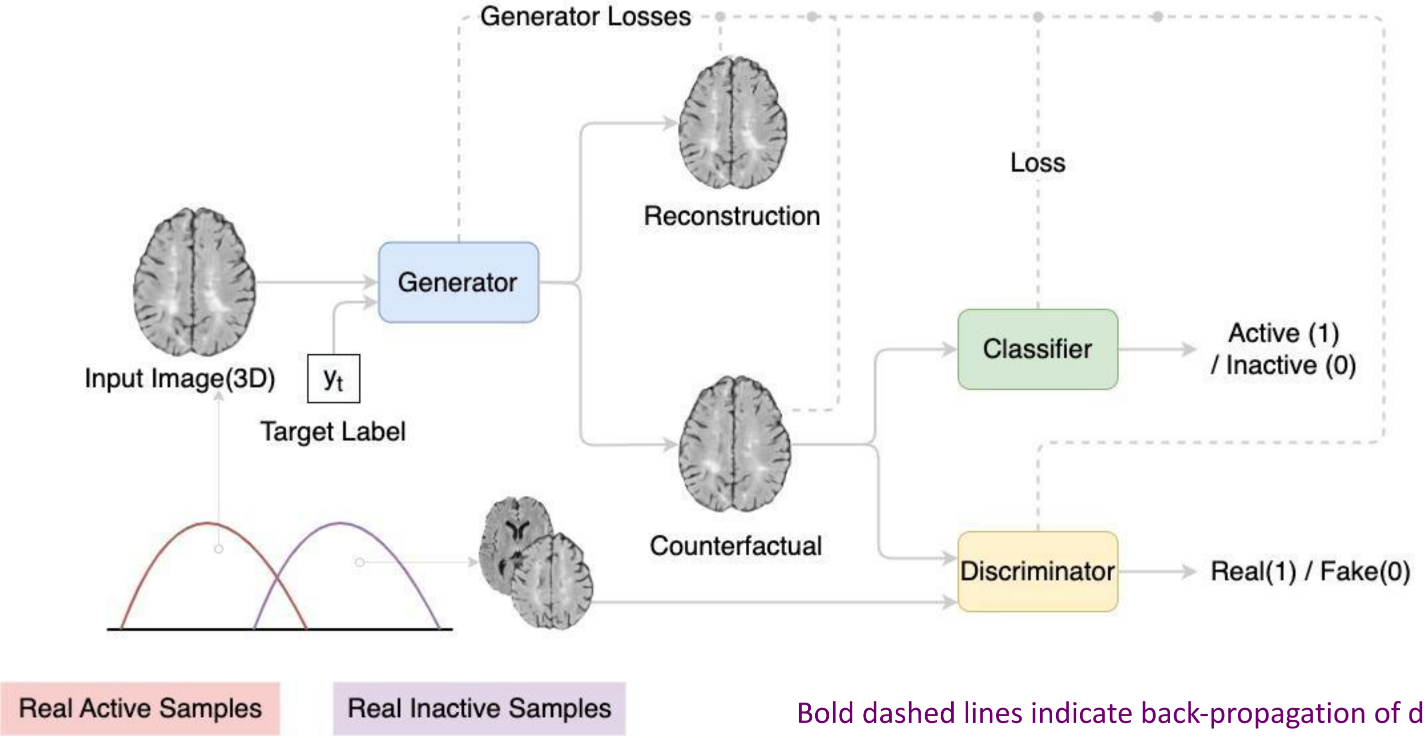
Identification of Personalized Image Markers Predictive of Future Patient Outcomes



Kumar *et. al*, MIABID Workshop, MICCAI 2022

Counterfactual Synthesis Model

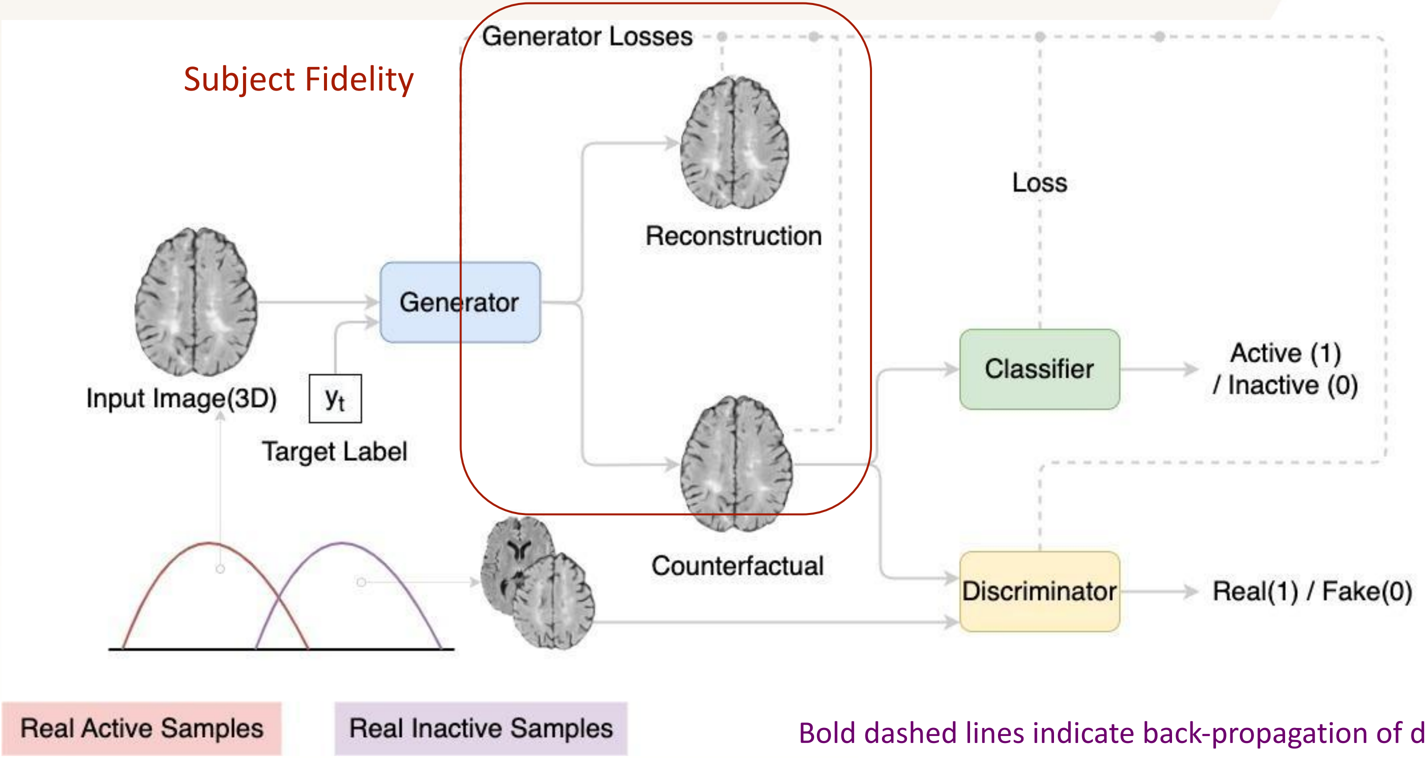
- Goal: CF image should maintain (i) subject fidelity, (ii) target class, (iii) realism



Bold dashed lines indicate back-propagation of different losses

Kumar *et. al*, MICCAI 2022 MIABID Workshop

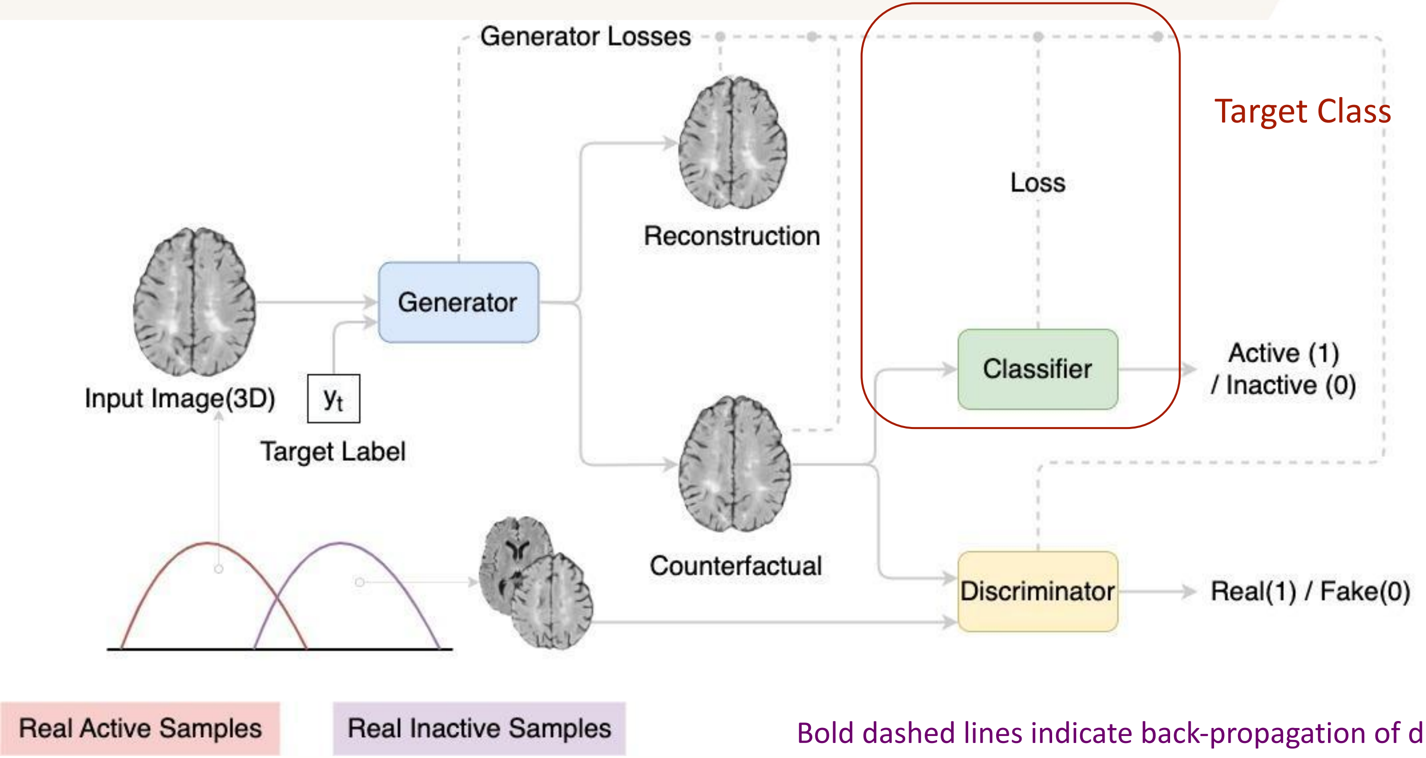
Counterfactual Synthesis Model



Bold dashed lines indicate back-propagation of different losses

Kumar *et. al*, MICCAI 2022 MIABID Workshop

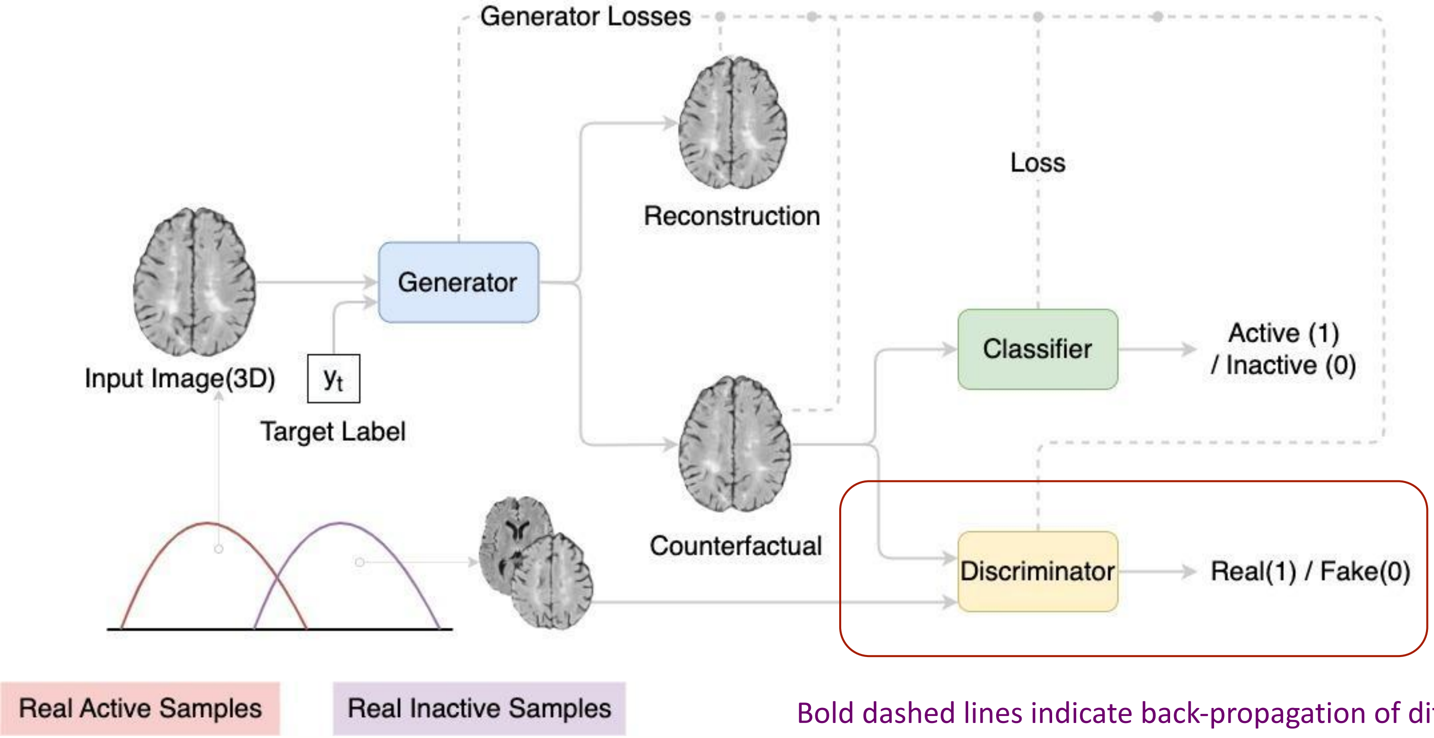
Counterfactual Synthesis Model



Bold dashed lines indicate back-propagation of different losses

Kumar *et. al*, MICCAI 2022 MIABID Workshop

Counterfactual Synthesis Model



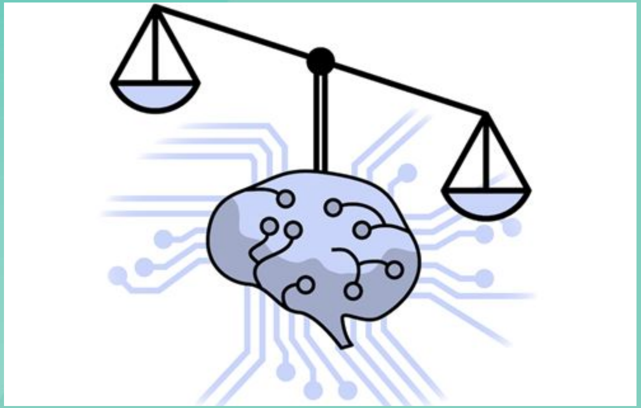
Realistic

Bold dashed lines indicate back-propagation of different losses

Kumar *et. al*, MICCAI 2022 MIABID Workshop

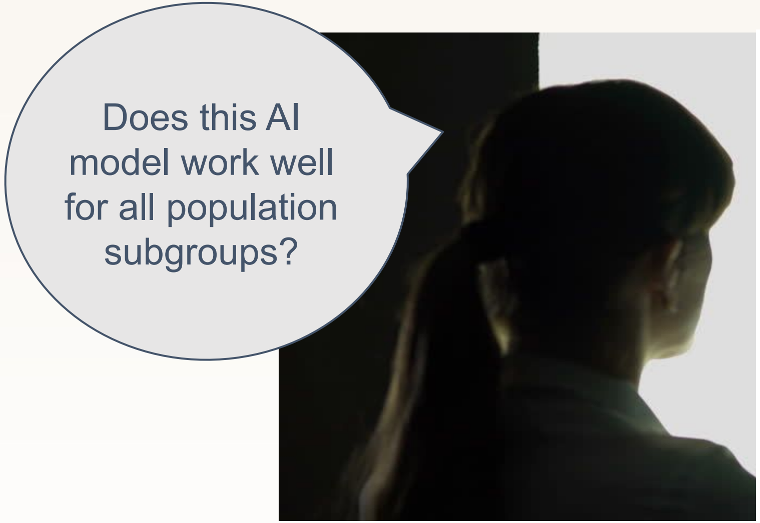
Trustworthy Image-Based Personalized Medicine

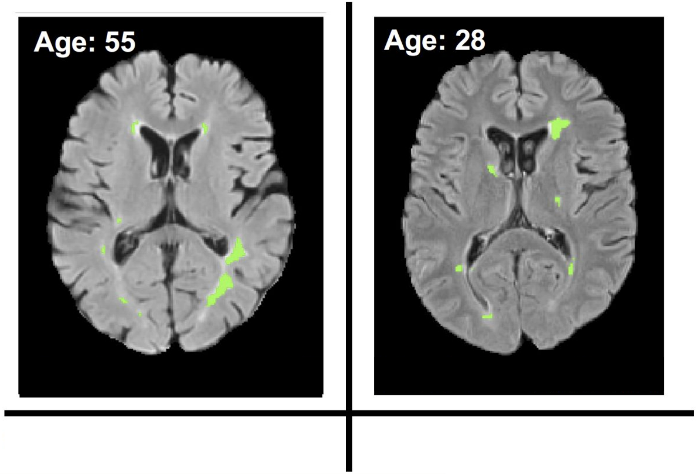
- Uncertainty Estimation
- Explainability
- Improving fairness



Trustworthy Image-Based Personalized Medicine

Biases in DL models across population subgroups

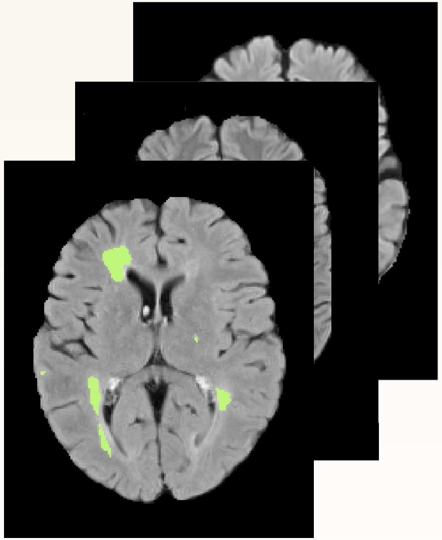


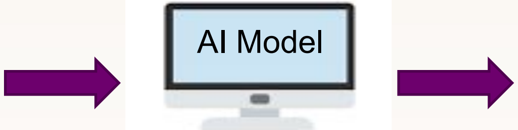


Patient will have new lesions in the future

Mehta et. al, MIDL 2023
Shui et. al, NeurIPS 2022
Shui *et. al*, In submission.

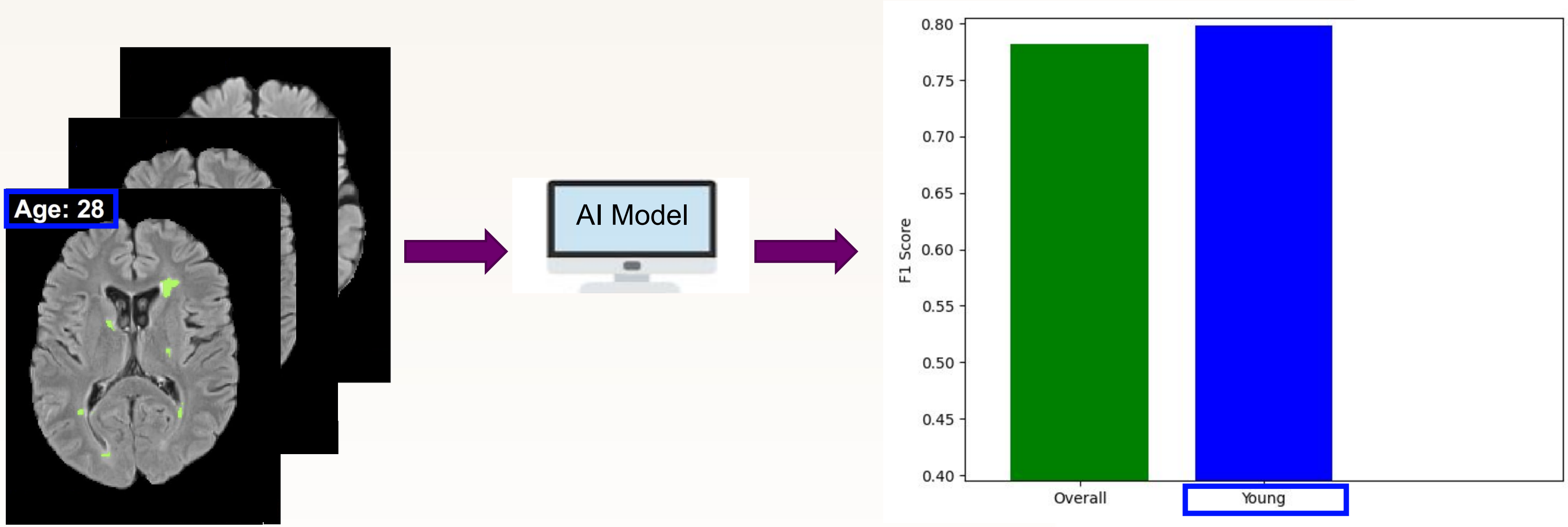
Mitigating Bias in AI Model Predictions





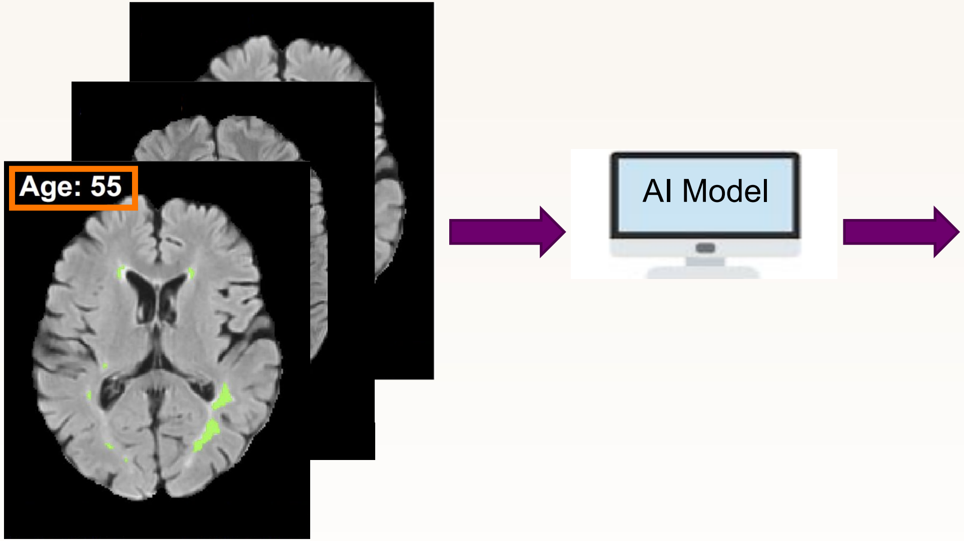
Mehta et. al, MIDL 2023

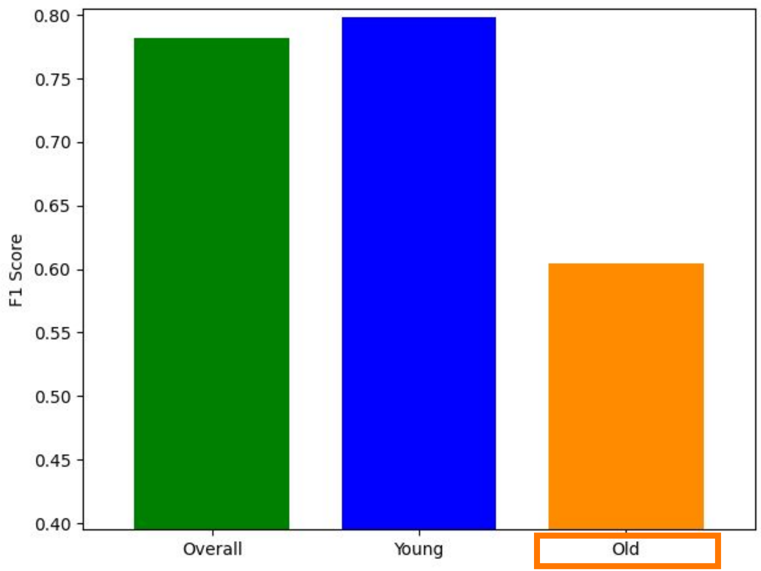
Mitigating Bias in AI Model Predictions



Mehta et. al, MIDL 2023

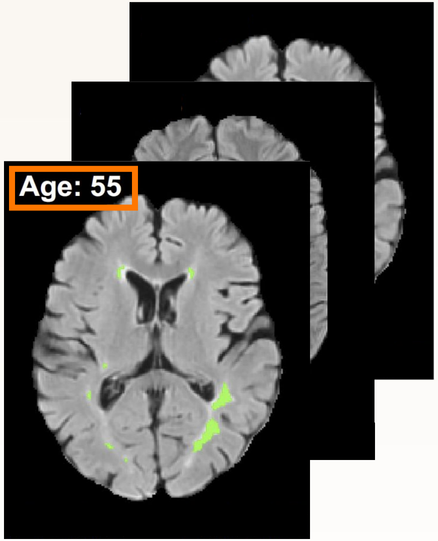
Mitigating Bias in AI Model Predictions



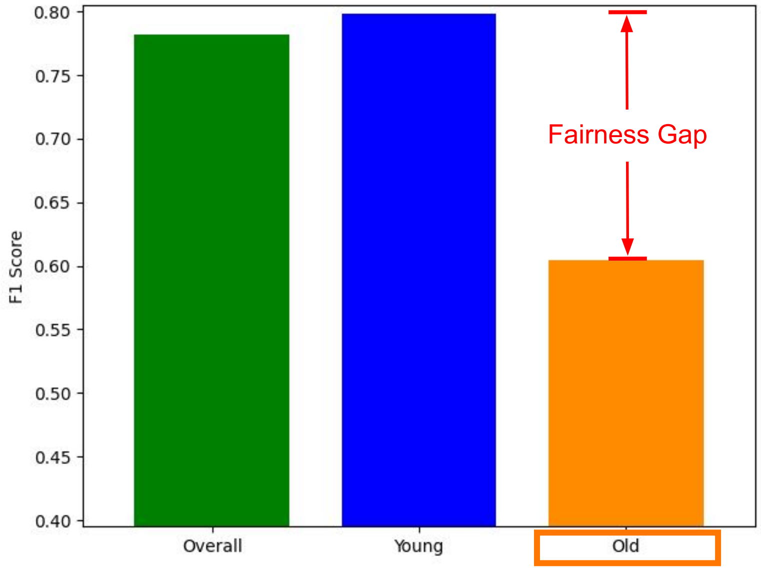


Mehta et. al, MIDL 2023

Mitigating Bias in AI Model Predictions

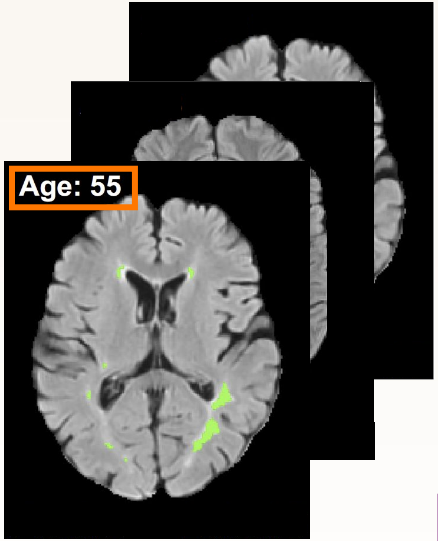


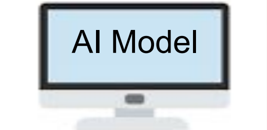
AI Model

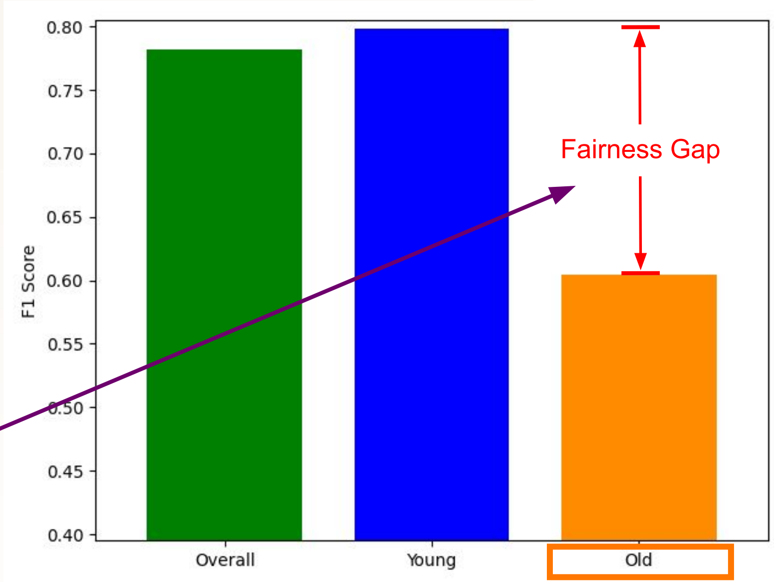


Mehta et. al, MIDL 2023

Mitigating Bias in AI Model Predictions





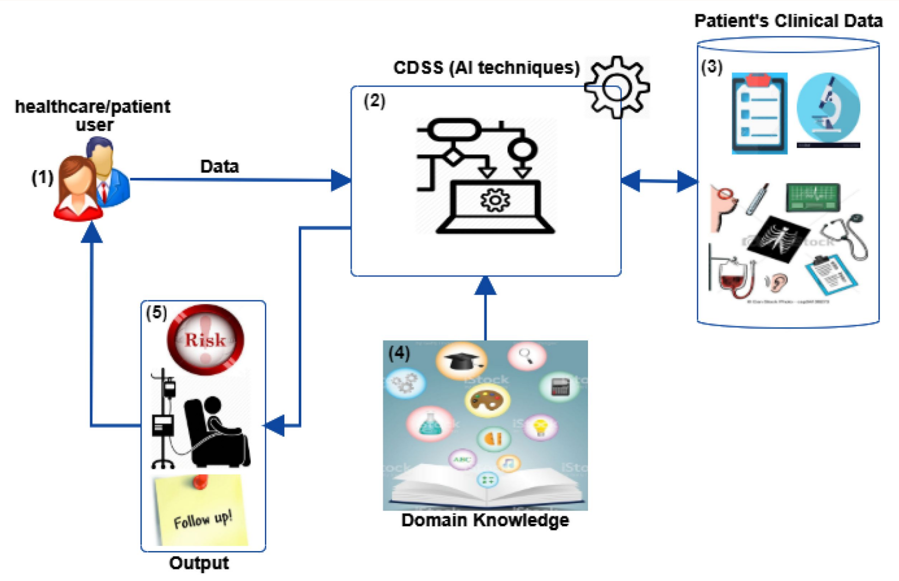


AI method reduces fairness gap by 28%

Mehta et. al, MIDL 2023

The Promise of AI for Clinical Decision Support

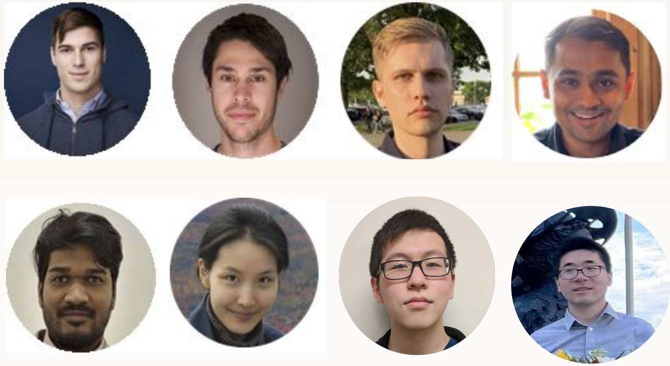
Provide clinicians with trustworthy AI tools to predict future individual treatment response on different treatments using *medical* images



https://www.mdpi.com/2072-6694/12/2/369

Thank you for your attention!

Probabilistic Vision Group



Sponsors

Collaborators

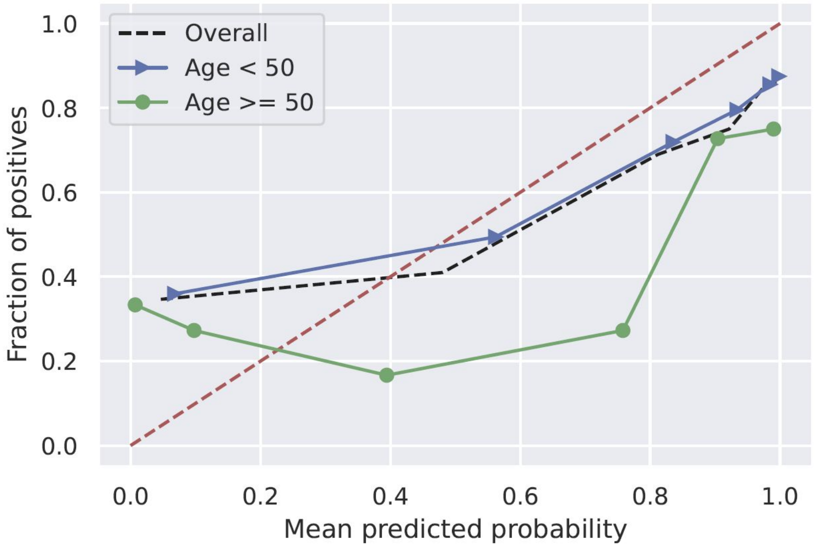
Douglas L. Arnold, Montreal Neurological Institute
Sotirios Tsaftaris, University of Edinburgh
Nick Powlowski, Microsoft Research
Yarin Gal, Oxford University

This work was made possible by Biogen, BioMS, MedDay, Novartis, Roche / Genentech, and Teva who generously provided the data

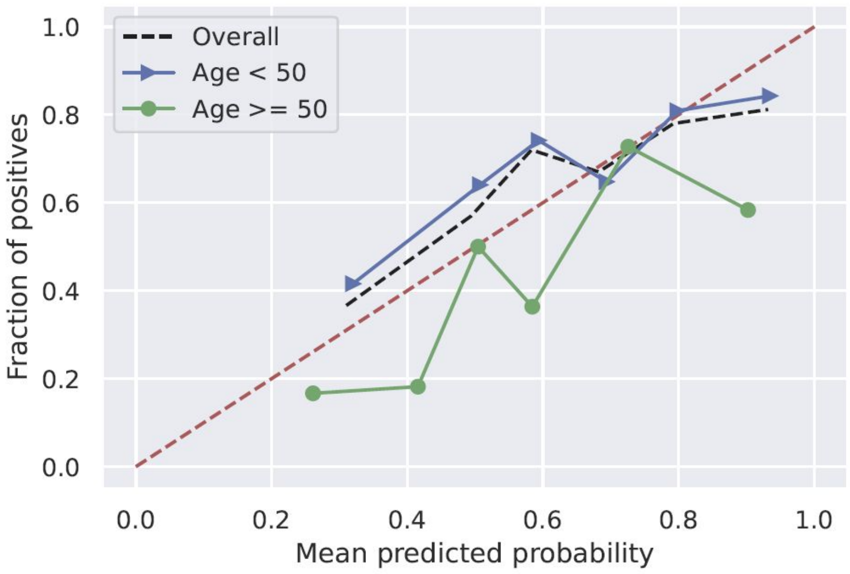
Extra Slides

Mitigating Bias in AI Model Predictions

Reliability Curve: With Bias



Reliability Curve: Mitigated Bias

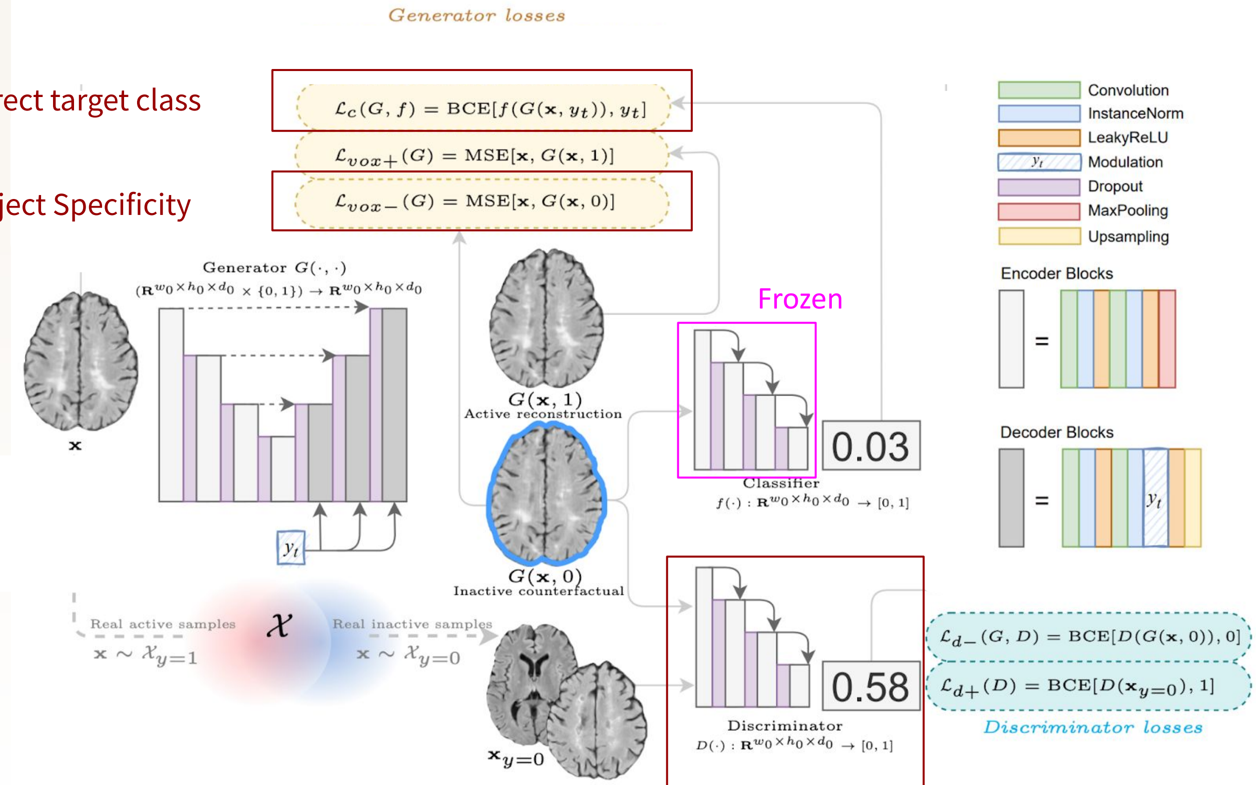


Shui *et. al*, In submission.

Counterfactual Synthesis

Correct target class

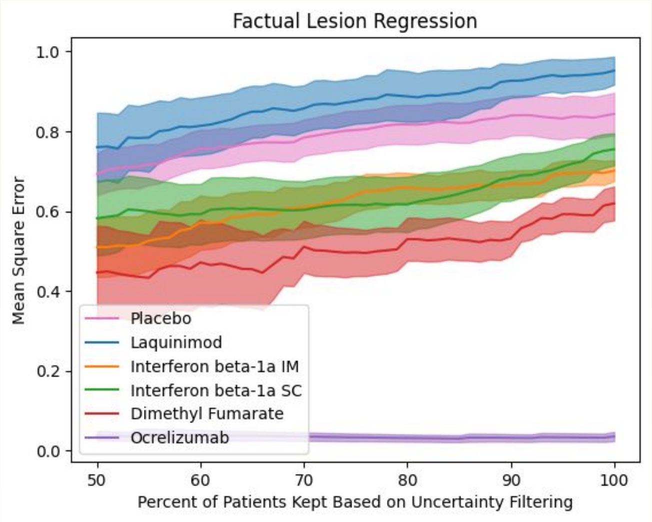
Subject Specificity

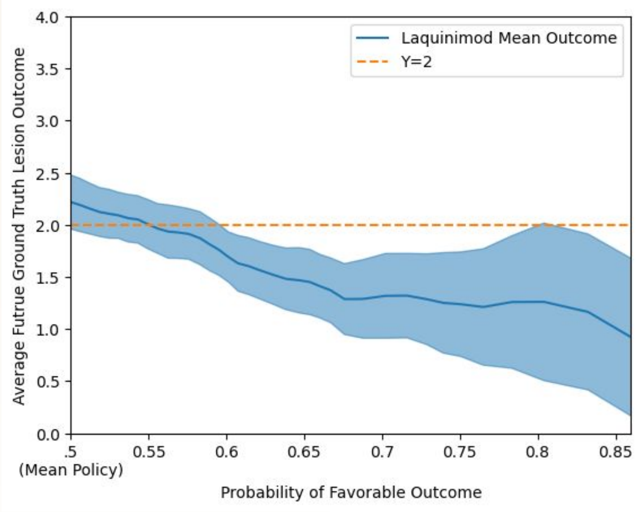


Sampled from real inactive distribution

f(.) : binary future lesion activity classifier
G(., .): Conditioned Generative module
D(., .): Discriminator

Trustworthy Treatment Effect Estimation





Uncertainty quantification helps provide model trustworthiness

Durso-Finley *et. al*, In submission, 2023