Explanatory AI in Treatment Planning

Data-Driven Automation and Decision Making and Explanatory AI

Tom Purdie, PhD, MCCPM
Princess Margaret Cancer Centre

UNIVERSITY OF TORONTO

TECHNA

UHN
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab

Disclosures

Method and System for Automated Planning of Radiation Therapy technology patented in PCT/CA2011/001130

Automated Quality Assurance (QA) and Planning technology patented in WO2014197994 A1

Receive royalties from RaySearch Laboratories for license of technology for Automated Breast Treatment Planning and Machine Learning-based Automated Treatment Planning

Have an equity interest in bridge7, licensee of technology for Machine Learning-based for Automated Quality Assurance in Radiation Oncology

Acknowledgements

L. Conroy

C. McIntosh

NSERC CRSNG

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

TECHINA

RaySearch Laboratories

The Princess Margaret Cancer Foundation UHN

Outline

AI as a Black Box

How to Achieve Explainability

Similar Patients ➔ Source for Explainability

Atlas Selection Learning

Human | AI Collaboration

Recognizing Digits

Can think of this as a Black Box … But should we?

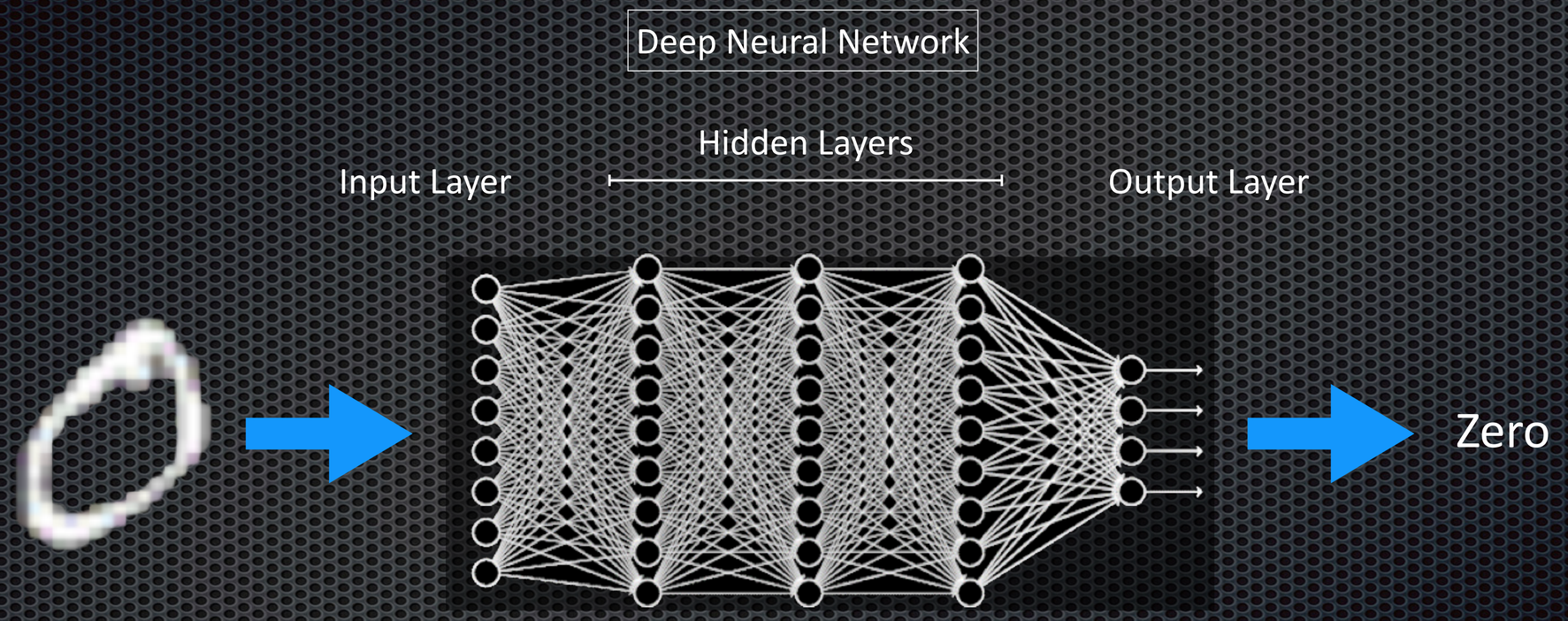
Black Box

Zero

Courtesy of Chris McIntosh

Recognizing Digits

Can think of this as a Black Box … But should we?



Courtesy of Chris McIntosh

Recognizing Digits

Can think of this as a Black Box … But should we?

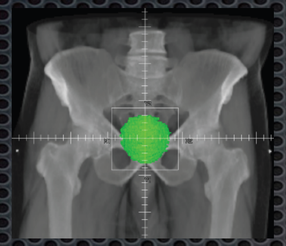
Zhong, Sheng-hua, *et al. Proceedings of the 19th ACM international conference on Multimedia*, 2011

Recognizing Patients

Typically patient features are concatenated into a vector

e.g. raw image or image features

Vector is used to quantify the patient



Correspondence between Patient Anatomy

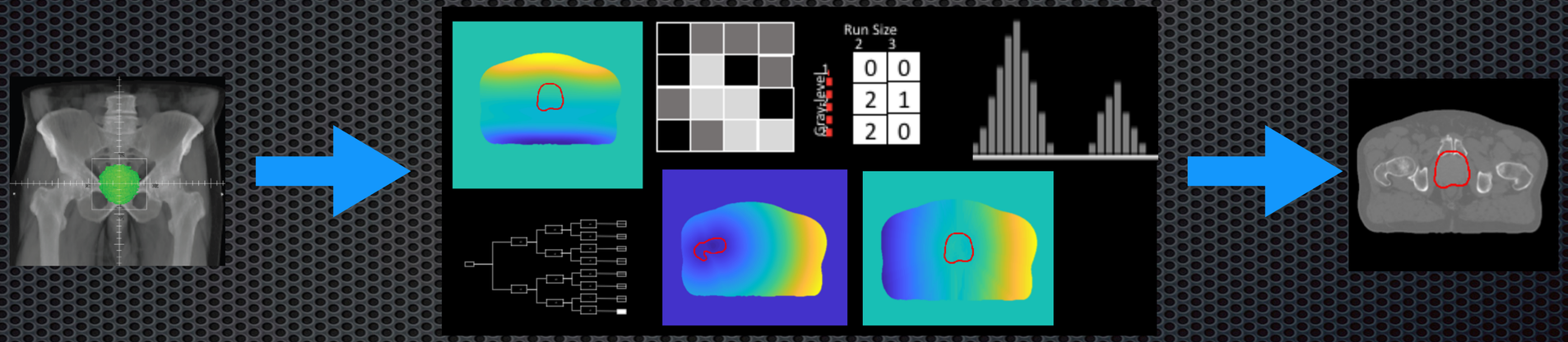
Courtesy of Chris McIntosh

Recognizing Patients

Typically patient features are concatenated into a vector

e.g. raw image or image features

Vector is used to quantify the patient



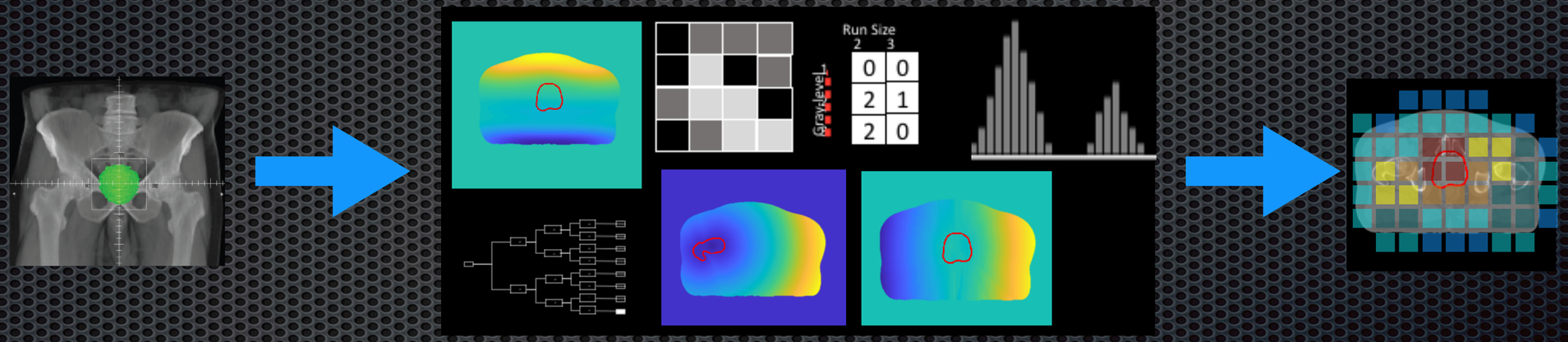
Courtesy of Chris McIntosh

Recognizing Patients

Typically patient features are concatenated into a vector

e.g. raw image or image features

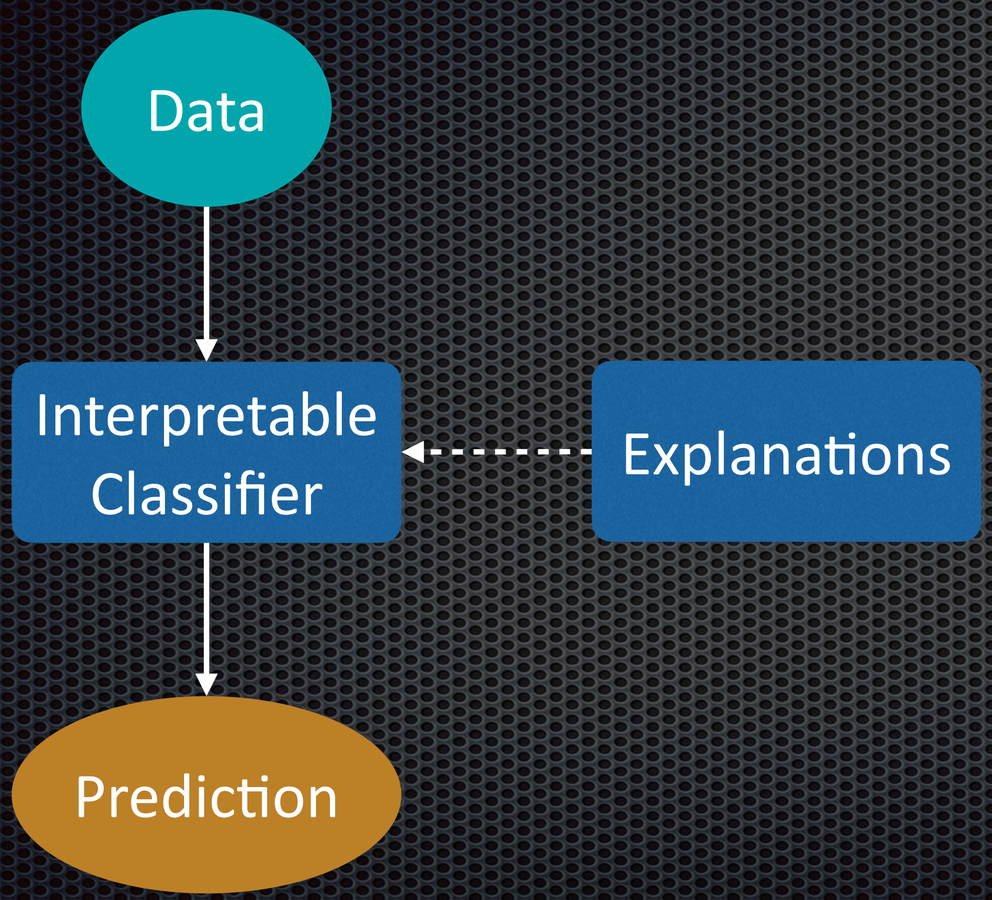
Vector is used to quantify the patient



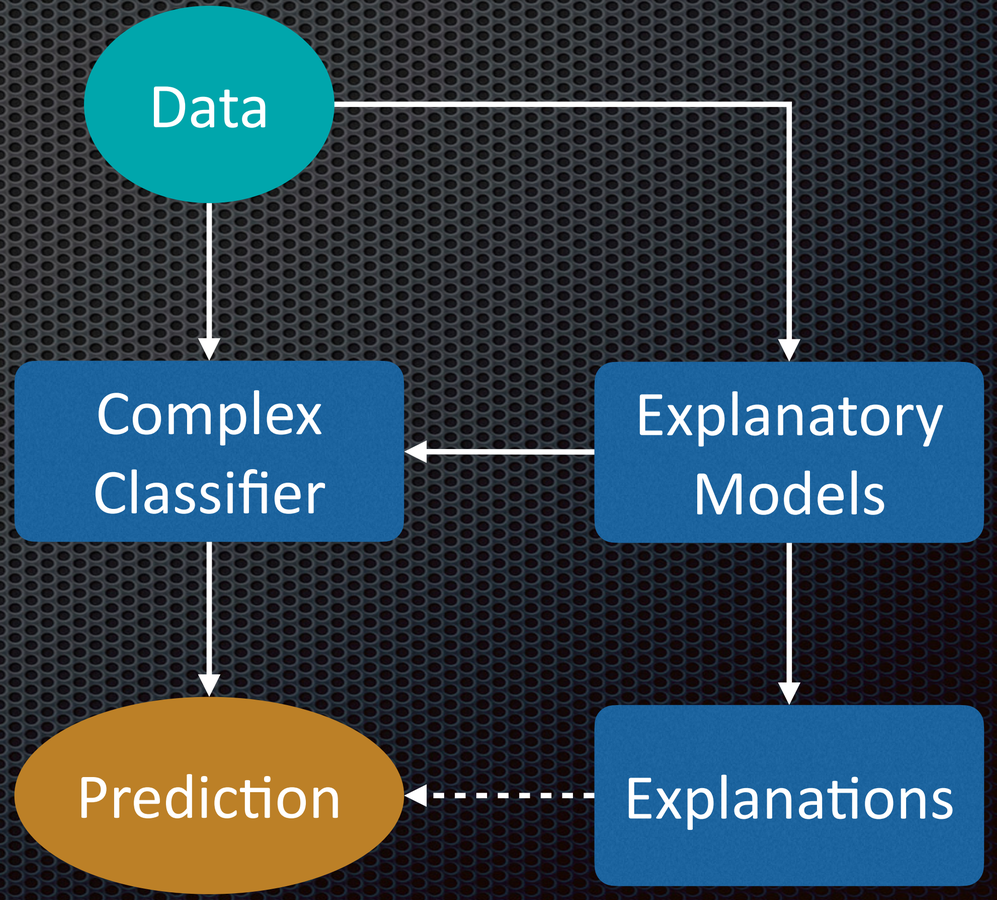
Courtesy of Chris McIntosh

How to Achieve Explainability

Intrinsic Interpretable Models
➔ Examines Classifier Structure



Post-hoc Interpretable Models
➔ Examines Feature Prediction Interactions

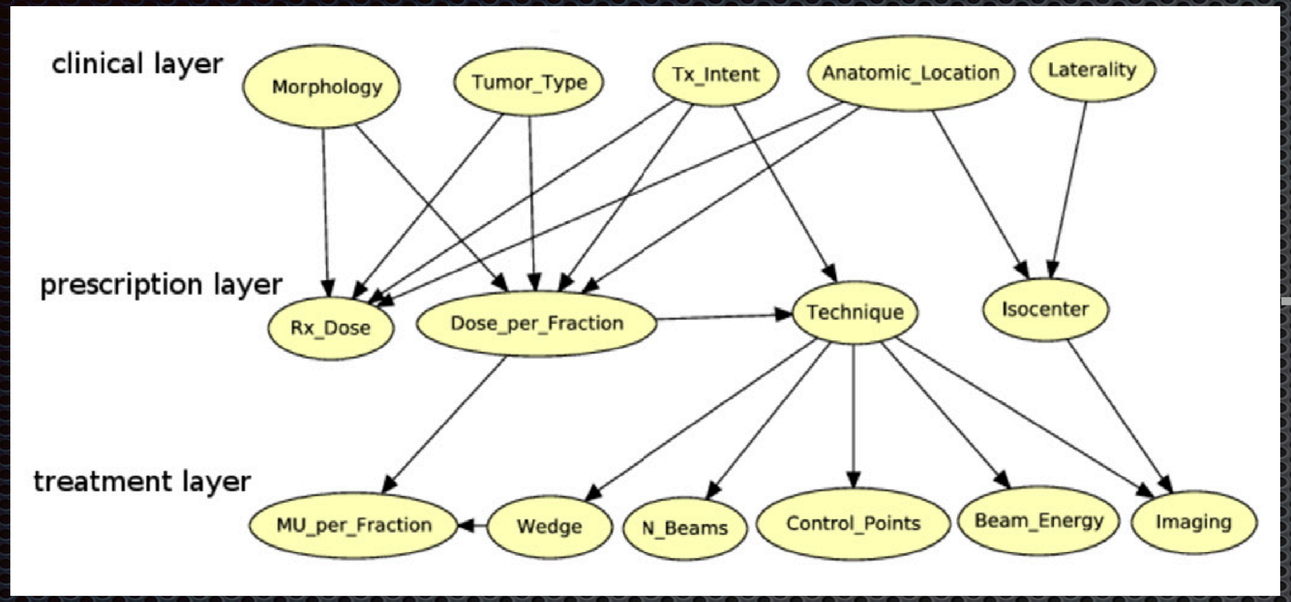


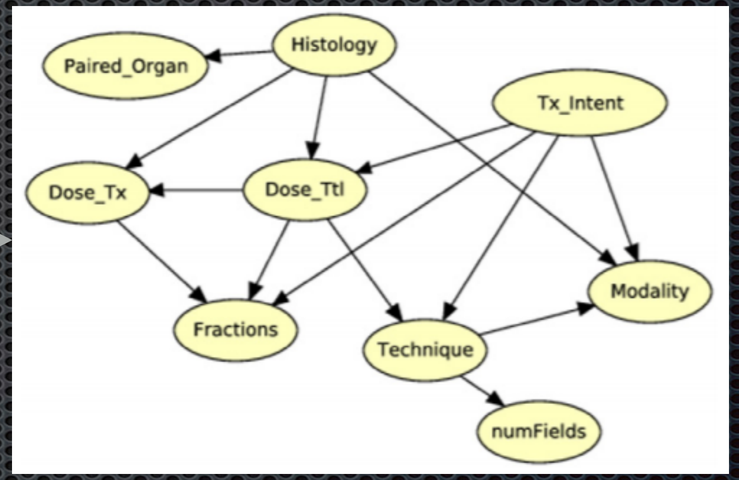
Courtesy of Yunsheng Chen

Intrinsic Interpretable Model

Bayesian Network

Joint probability distributions to define probability of one event given set of other known information

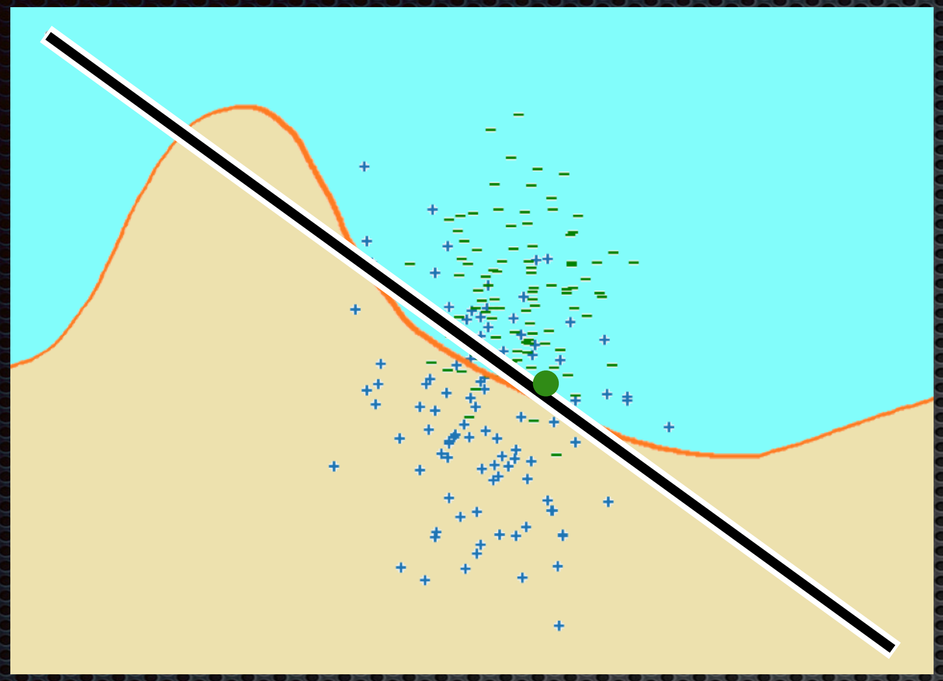




Kalet *et al*, PMB 2015

Post-hoc Interpretable Model

Local Interpretable Model-agnostic Explanations (LIME)

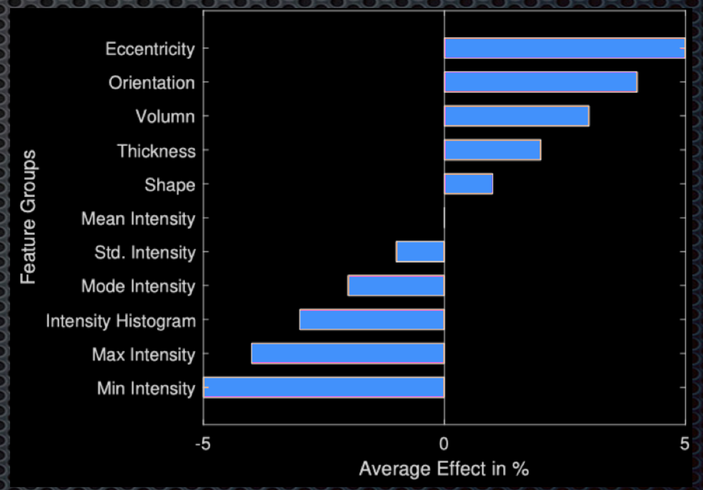


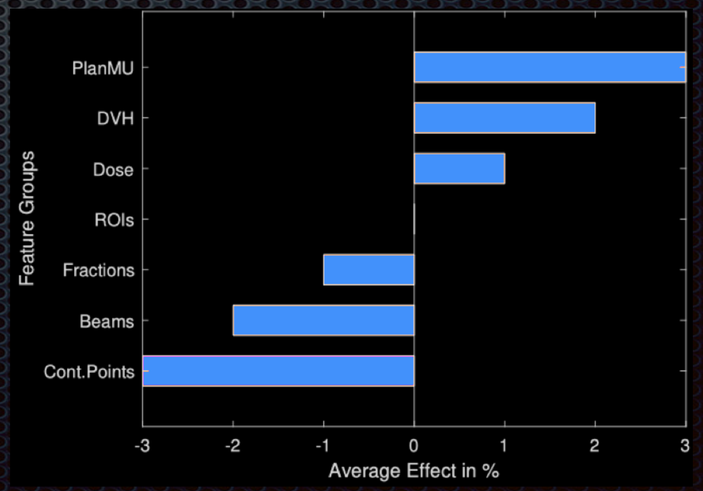
Learned Decision Boundary
Linear Approximation
Data point
+ / −
Perturbed samples
(data point +/− class)

Changes in feature values

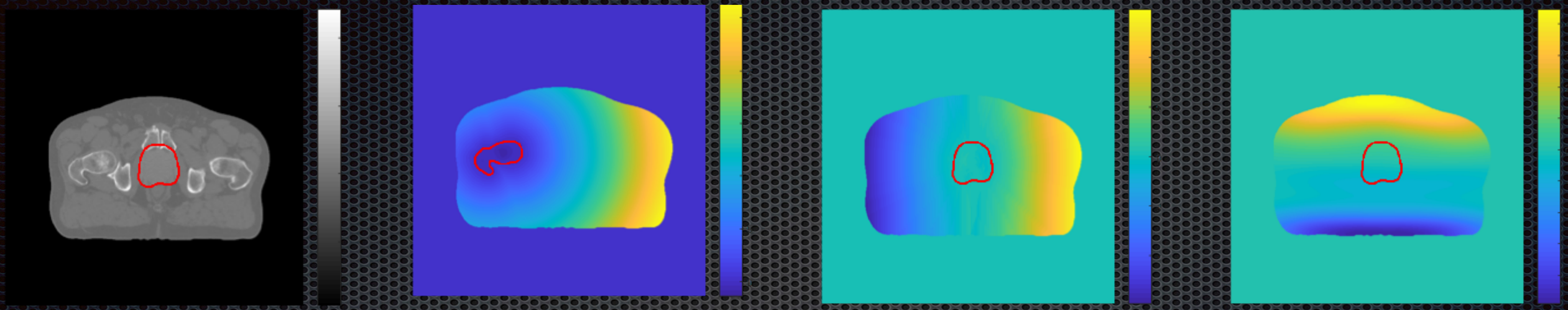
↓

Changes in classifier prediction





Derived Features Can Provide Insight



Input CT Image

x Distance to Prostate

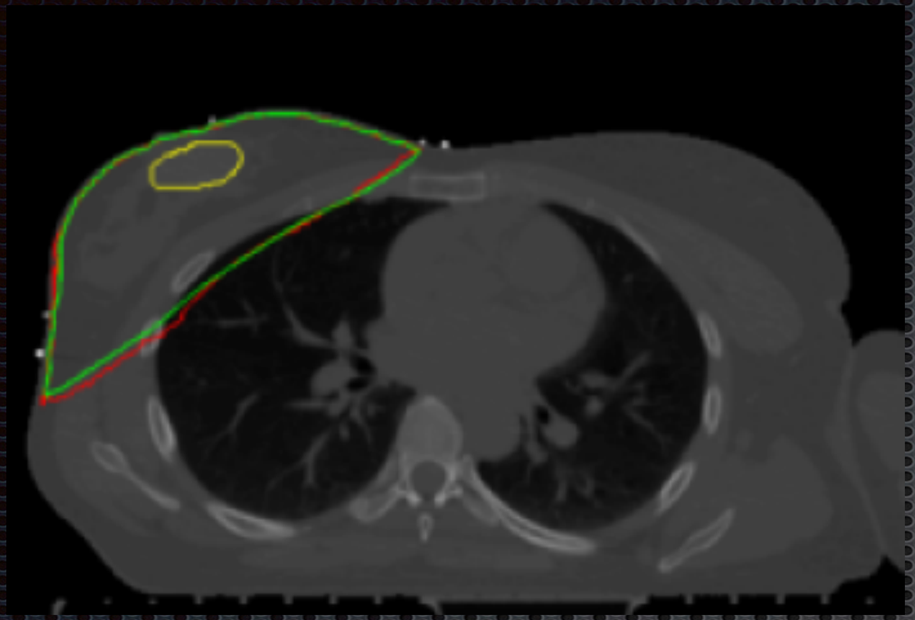
Distance to Femur

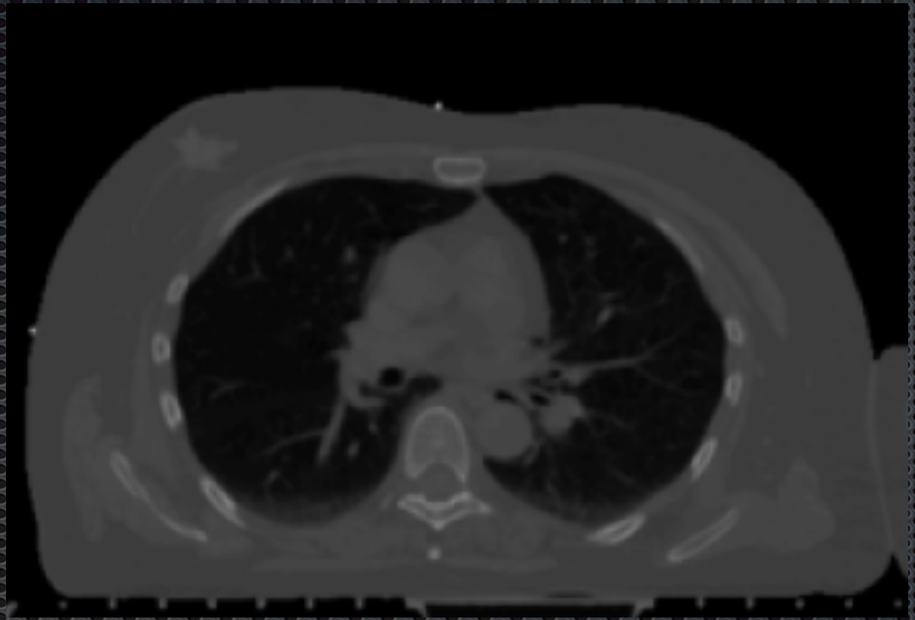
y Distance to External

Courtesy of Chris McIntosh

Patient Similarity

What does similar mean between patients?

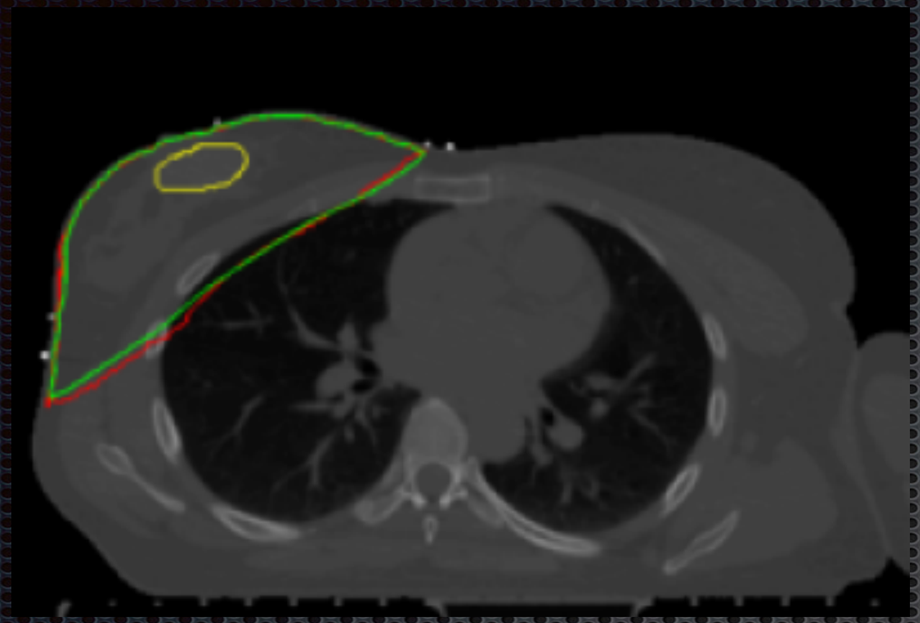


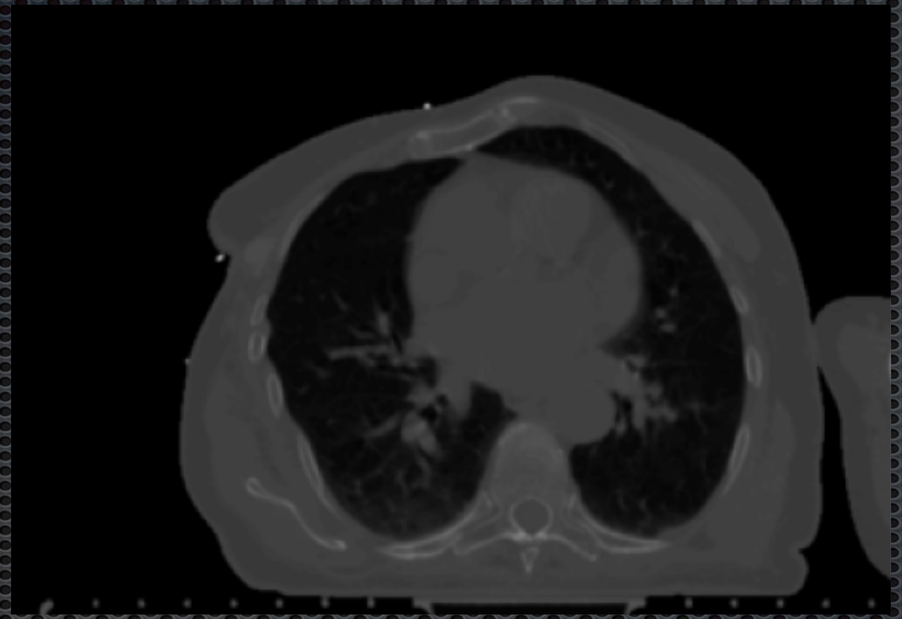


Courtesy of Chris McIntosh

Patient Similarity

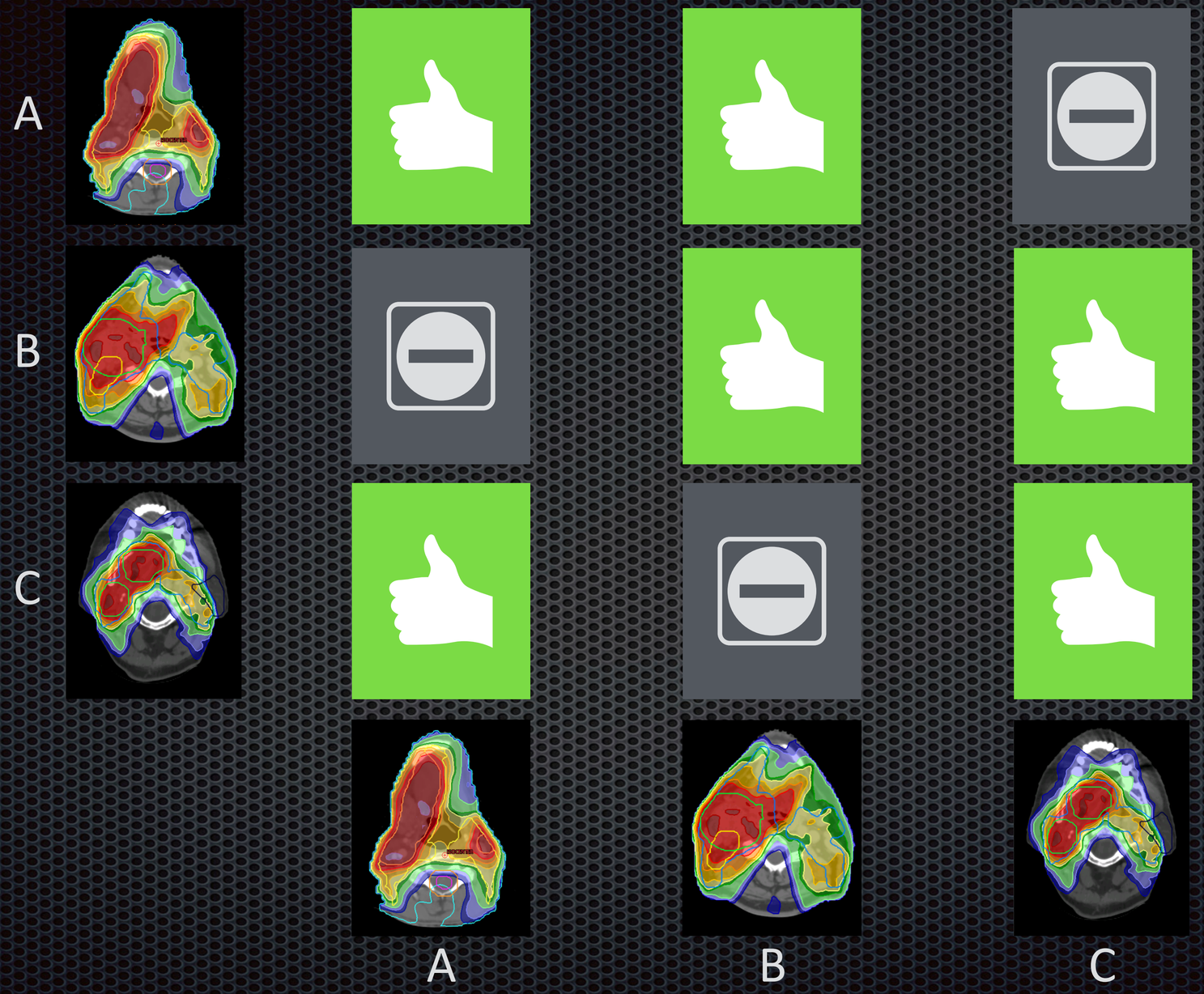
What does similar mean between patients?





Courtesy of Chris McIntosh

Learning Similarity



McIntosh, Purdie IEEE TMI 2016

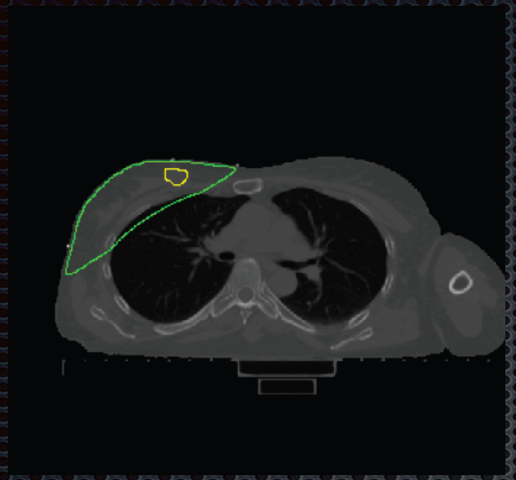
Features for Similarity using Atlas Regression Forests

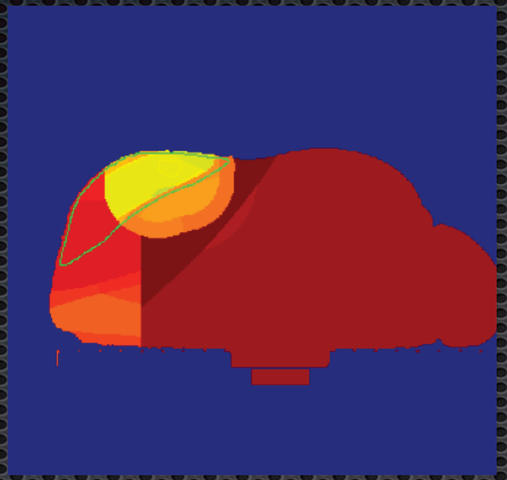
Image

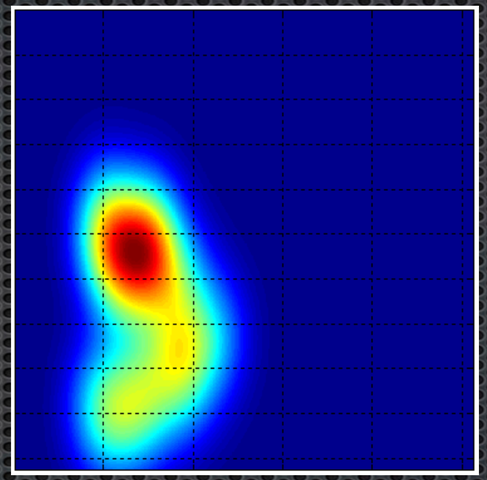
Leaf Segmentation
(Homogeneous Dose)

Encoding
(Density Estimation)

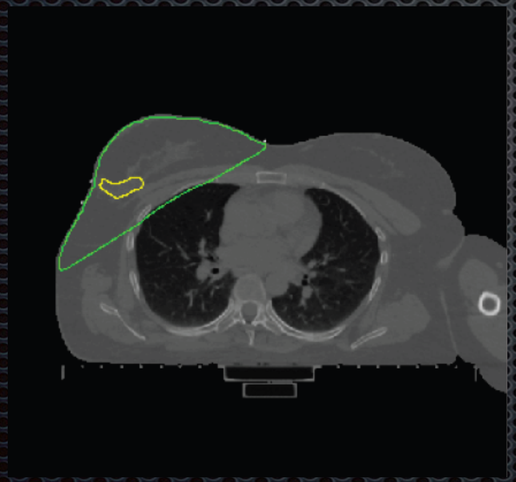
Learned ARF Patient A

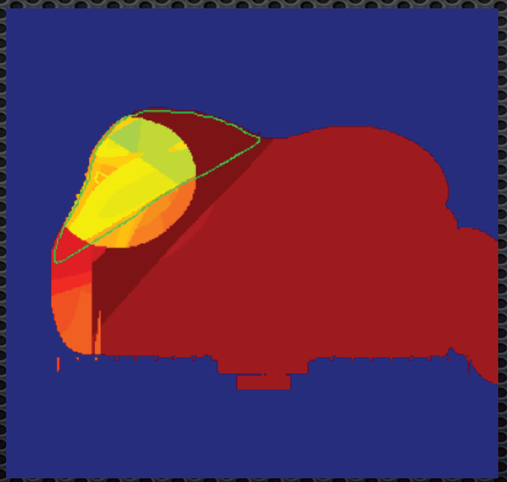


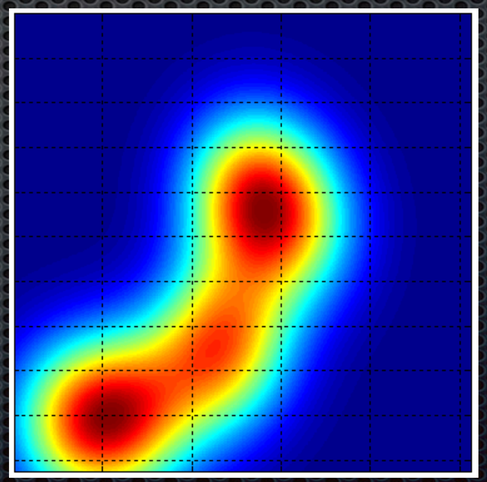




ARF-A on Novel

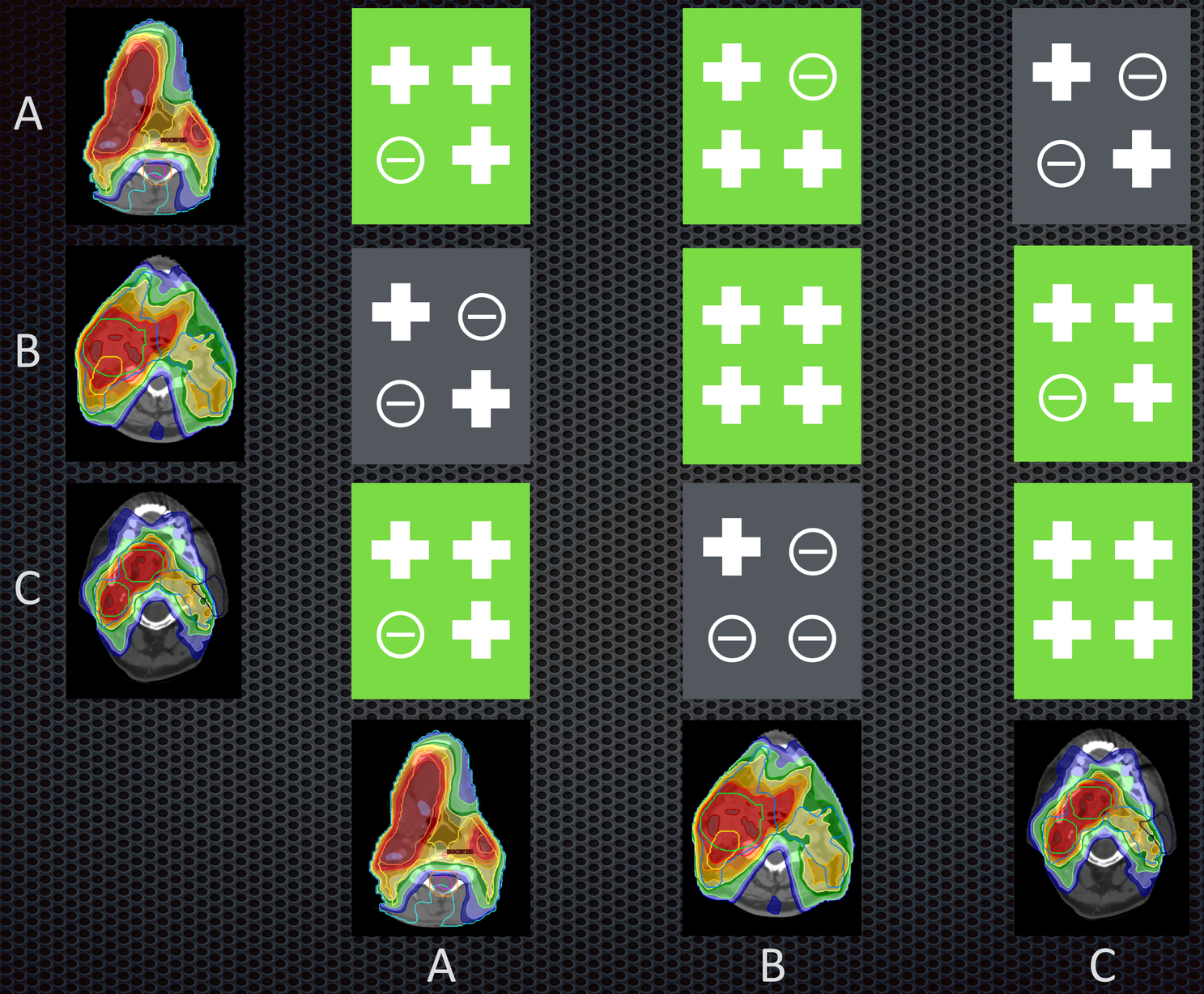






McIntosh, Purdie IEEE TMI 2016

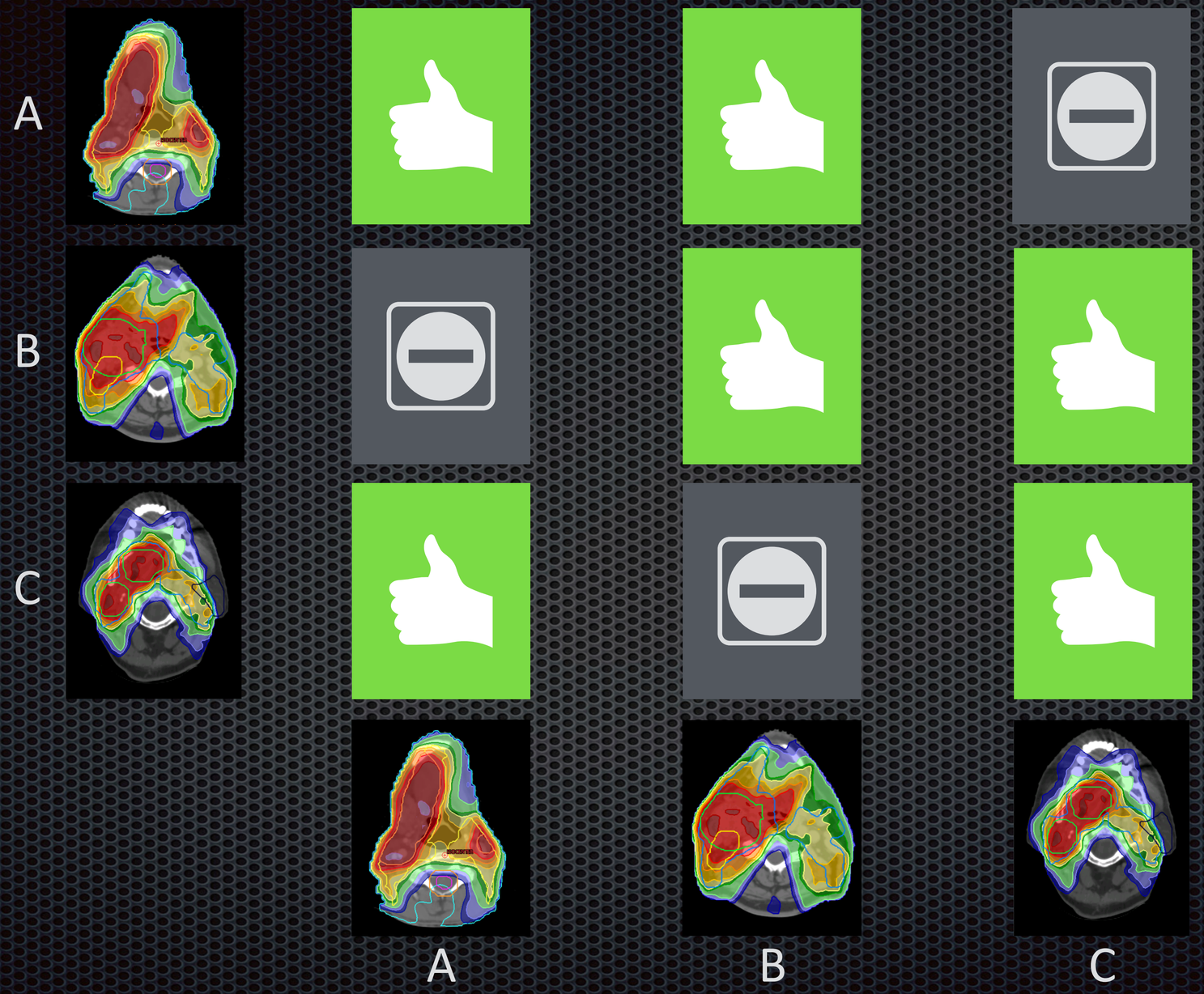
Learning Similarity



Bhattacharyya Distance between learned and observed density models at each leaf

McIntosh, Purdie IEEE TMI 2016

Learning Similarity

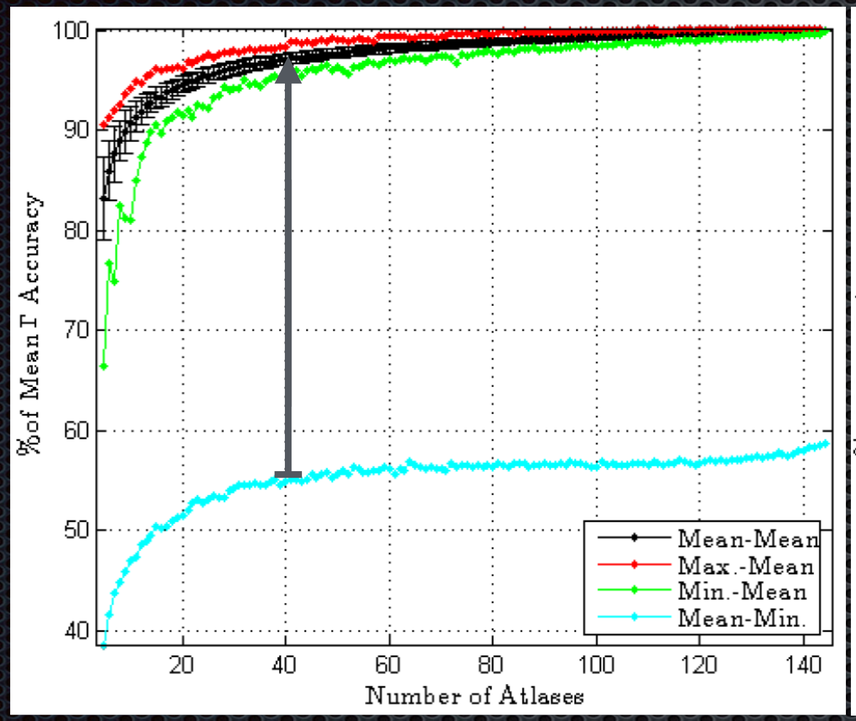


McIntosh, Purdie IEEE TMI 2016

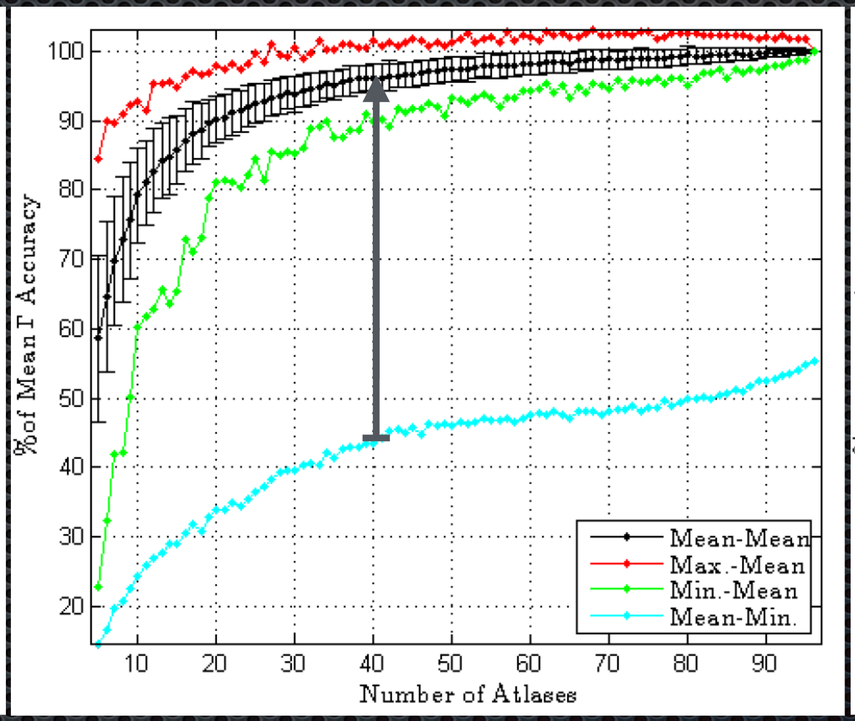
How many samples do I need in my training dataset?

Accuracy | Number of Training Atlases

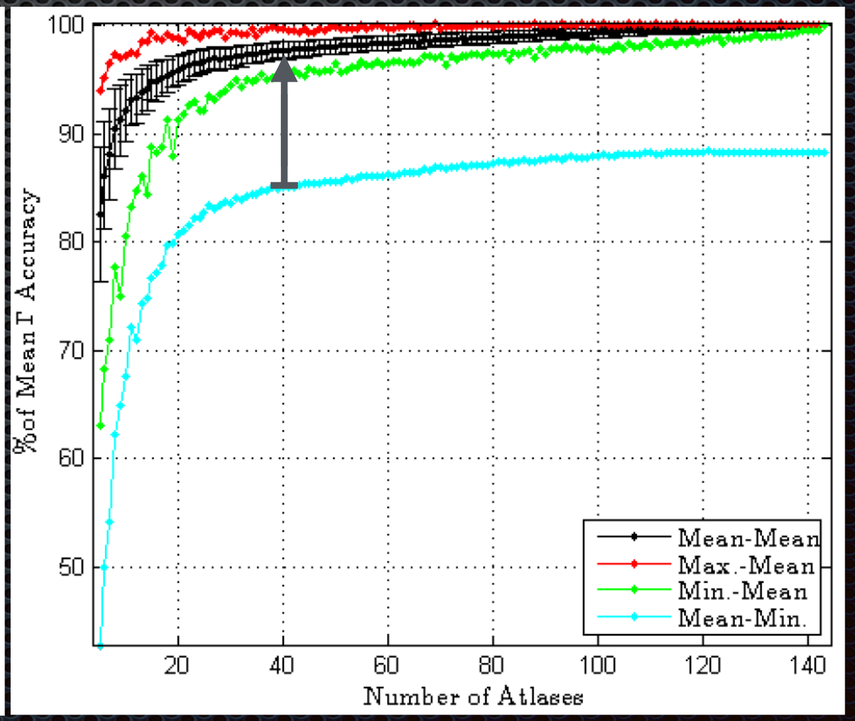
Accuracy is the % of voxels with similar radiation dose to ground truth



Whole Breast



Breast Cavity (Boost)

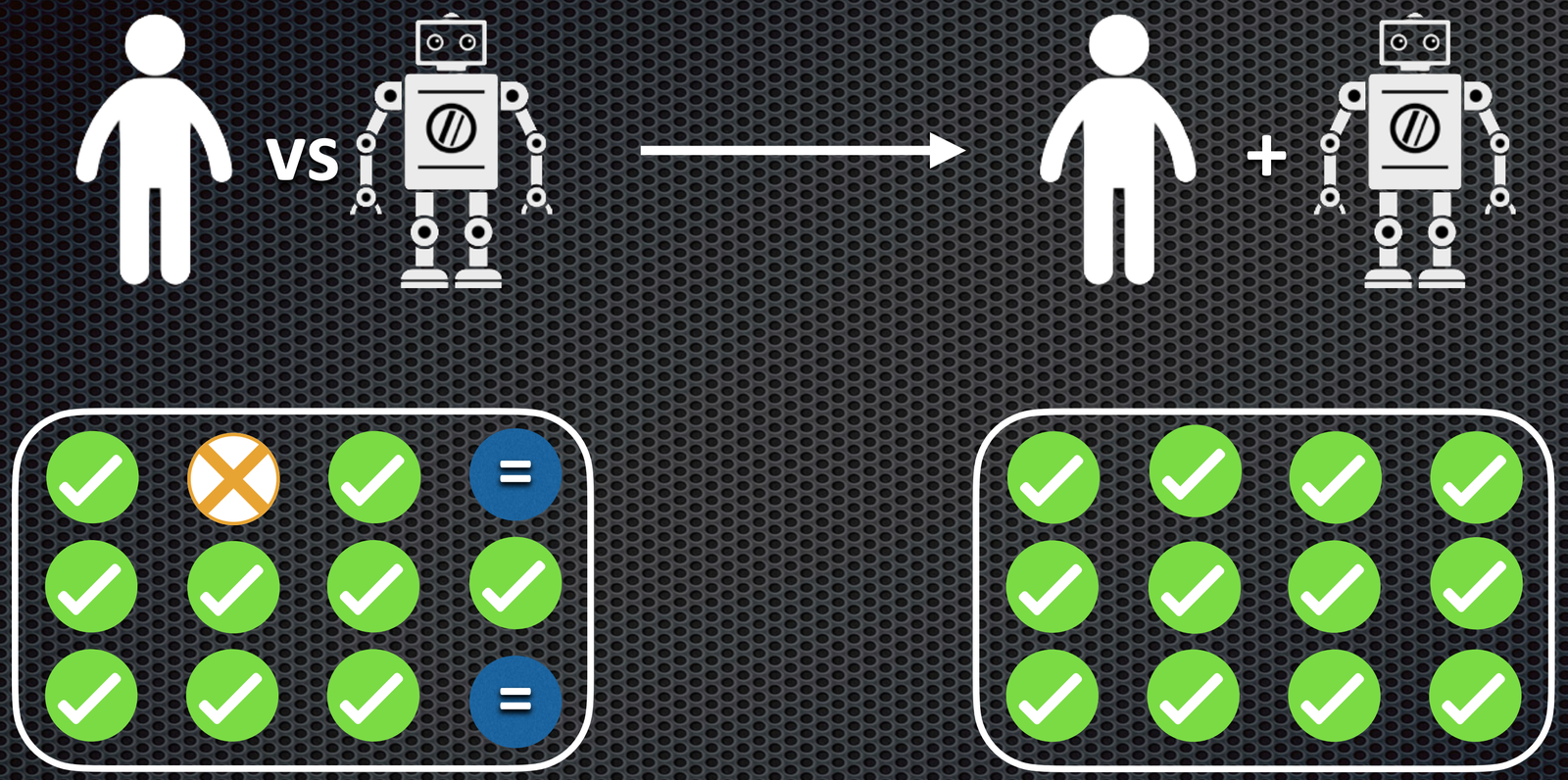


Prostate

Courtesy of Chris McIntosh

Explainable AI for Human | AI Collaboration

Improve Adoption of AI for Clinical Use



Courtesy of Leigh Conroy

Thanks for your attention