What is a “good” visual explanation for AI?

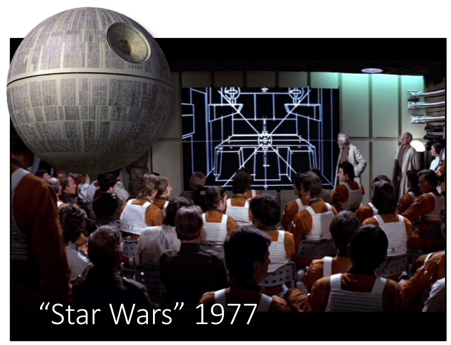
Liz Marai

Electronic Visualization Laboratory
Department of Computer Science
University of Illinois at Chicago

evl electronic visualization laboratory

EVL: Pioneering technology for interdisciplinary, collaborative work

- Established 1973, Tom DeFanti & Dan Sandin
- At the forefront of VR research since 1992
- Introduced CAVE, first projection-based VR system in the world
- Wide range of immersive technologies for data analysis


"Star Wars" 1977

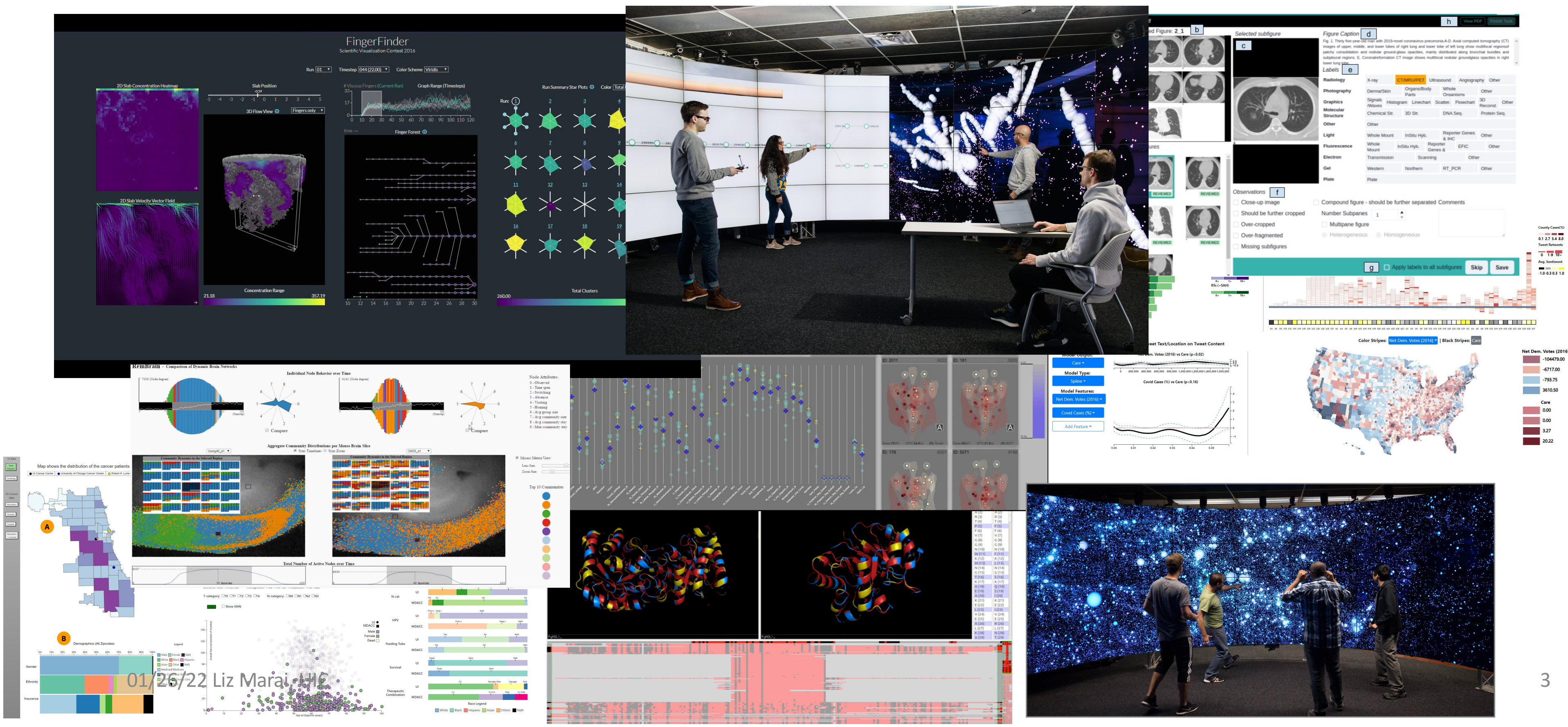

CAVE 1992


TacTile 2008

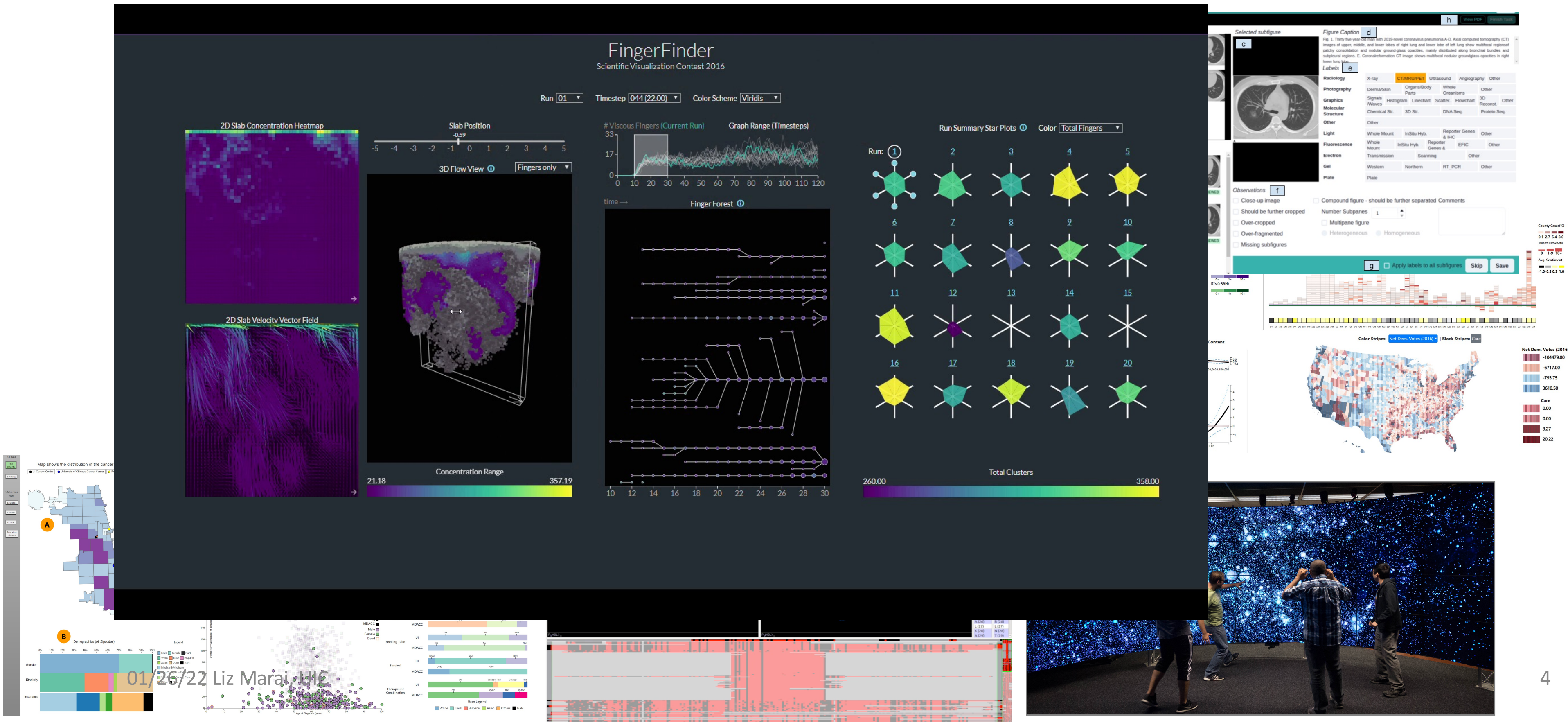

SAGE (2004-2014) and SAGE2 (2014-present)


CAVE2 2012

Problem-Oriented and Theoretical Work



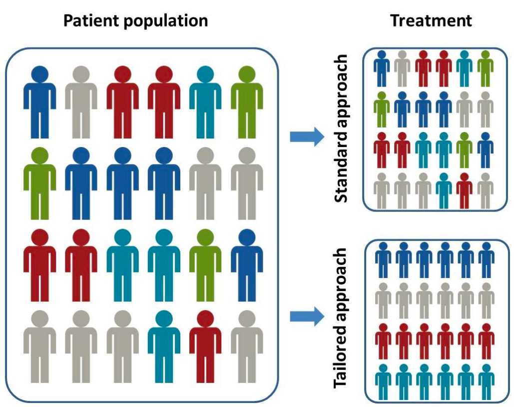
Problem-Oriented and Theoretical Work



A Computational Oncology Endeavor

Joint UIC work (medical imaging, visual computing, ML) w/ G. Canahuate (data mining), D. Fuller (radiation oncology), and 25+ other people at four sites (UIC, MD Anderson, U Iowa, UMN)

Precision Medicine in Oncology



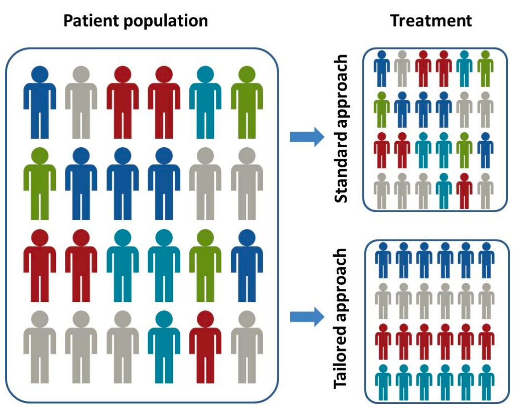
Big data from cohorts can enhance:

- Clinical decision-making
- Care-delivery during/after treatment

at the individual patient level.

Image from https://www.labkey.com/

Precision Medicine in Oncology



Big data from cohorts can enhance:

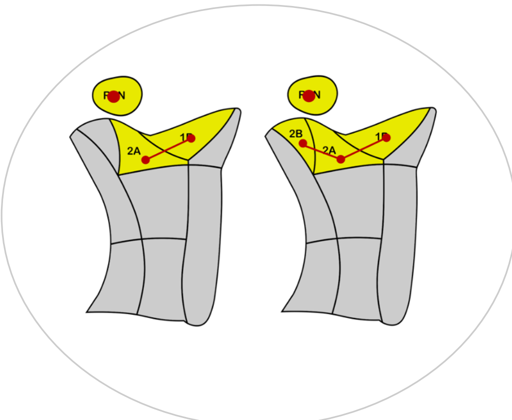
- Clinical decision-making
- Care-delivery during/after treatment

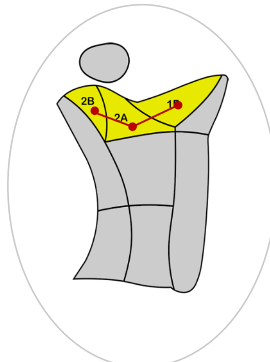
at the individual patient level.

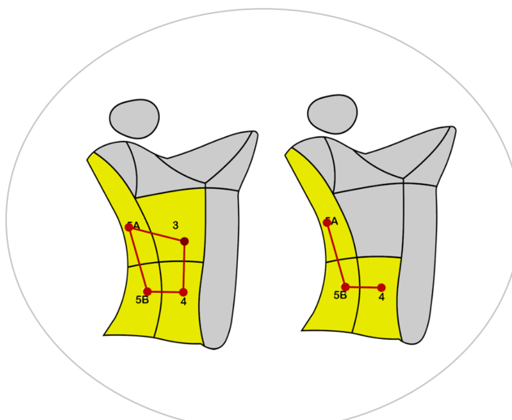
by clustering spatially similar patients

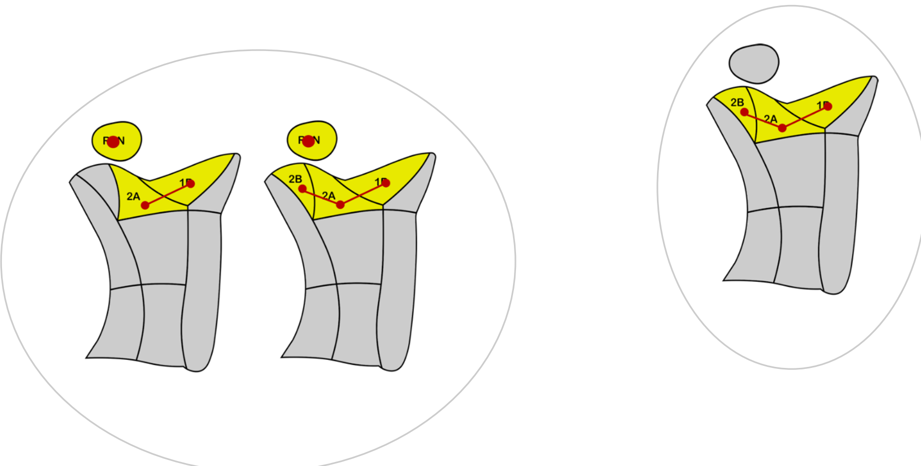
Image from https://www.labkey.com/

Head and Neck Cancer Research: Group patients based on disease spread patterns

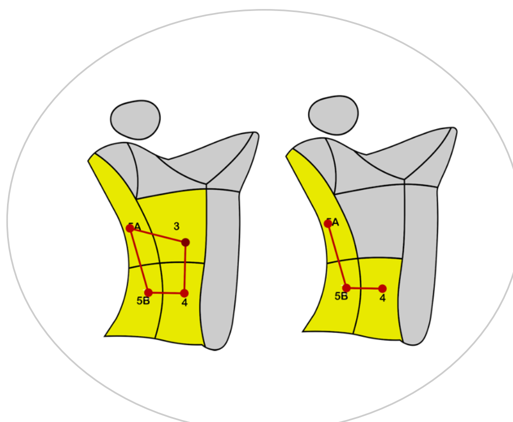


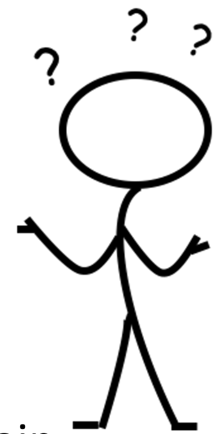






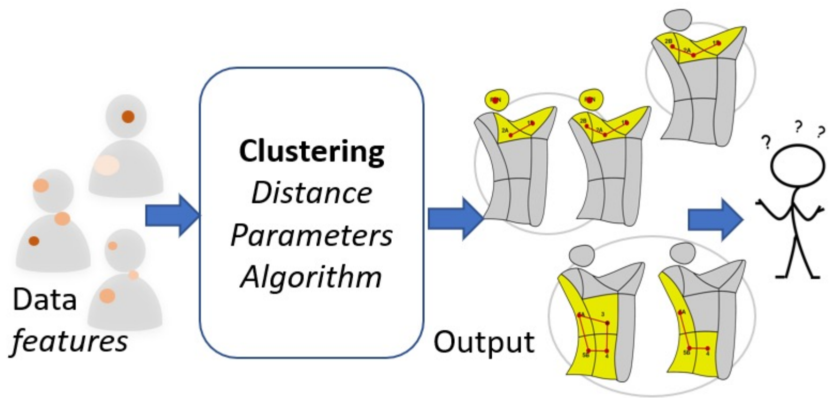
New clustering
methodology using
spatial data



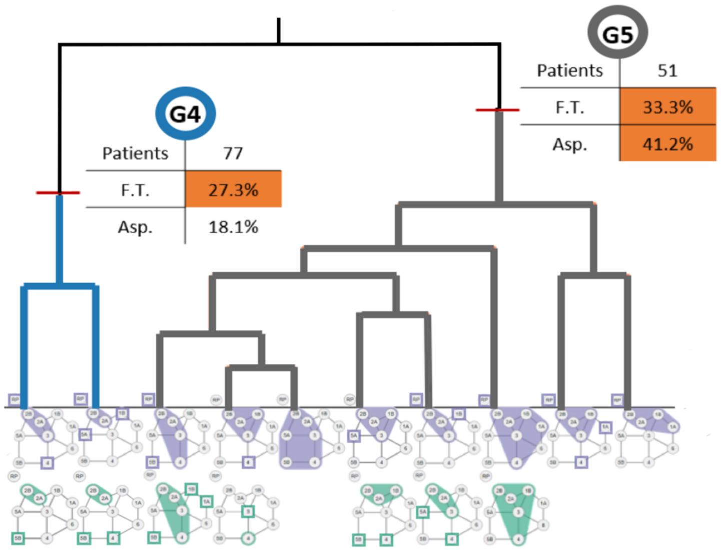


How to explain
spatial clustering to non-experts?

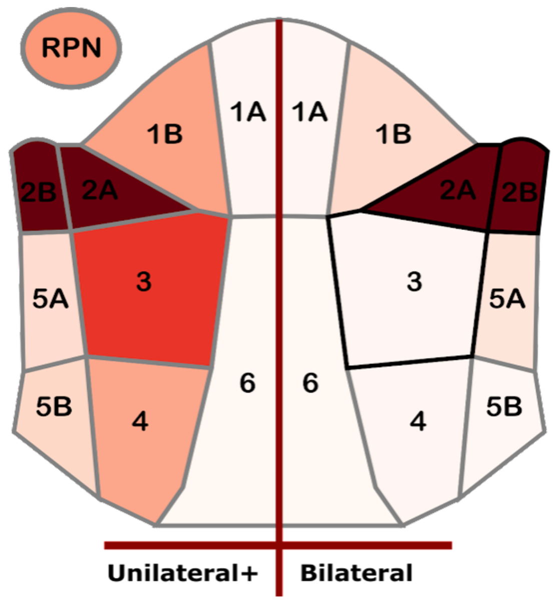
Clustering with Spatial Data



From Model Building to Dissemination

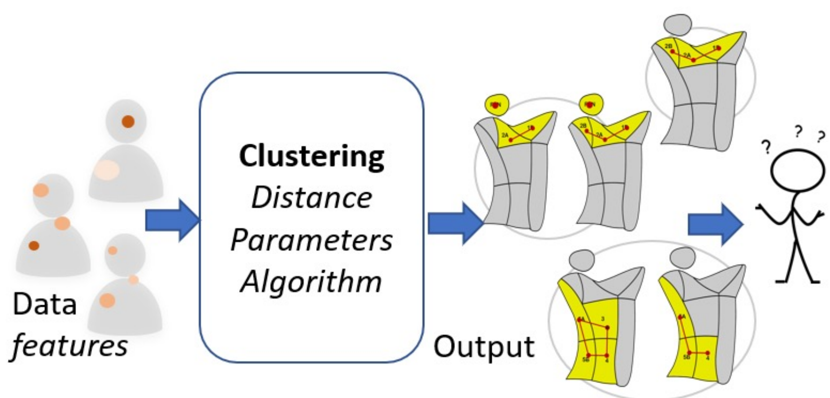


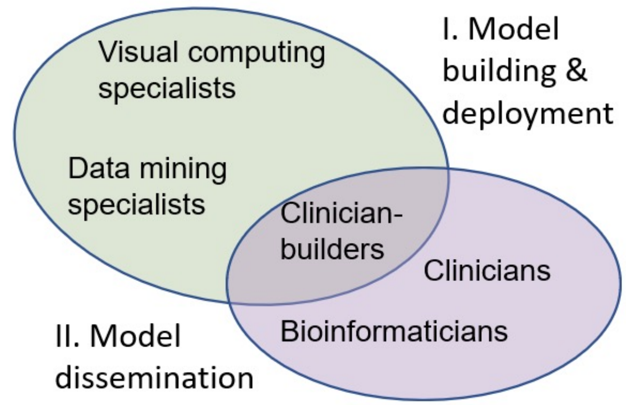
Model Building



Model Dissemination

Collaboration: Participatory to Broader





- EVL and Computational Oncology
- XAI Considerations
- Vis in XAI
- Clustering with Spatial Data: RT
- Clustering with Spatial Data: LN
- A “Good” Visual Explanation for AI

XAI Considerations

5 lenses: layman, machine learning, social science, public policy, and healthcare.

Benefits of XAI.

XAI (eXplainable Artificial Intelligence)

- Expert systems research in the ‘70s
- Resurgence since 2010-2020, grant solicitations, popular press
  - Many AI applications not adopted by the intended audience
- Hypothesis: by building more explainable systems, their audience will understand & adopt the AI

“Explain” (Merriam-Webster Dictionary)

1: to make (something) clear or easy to understand

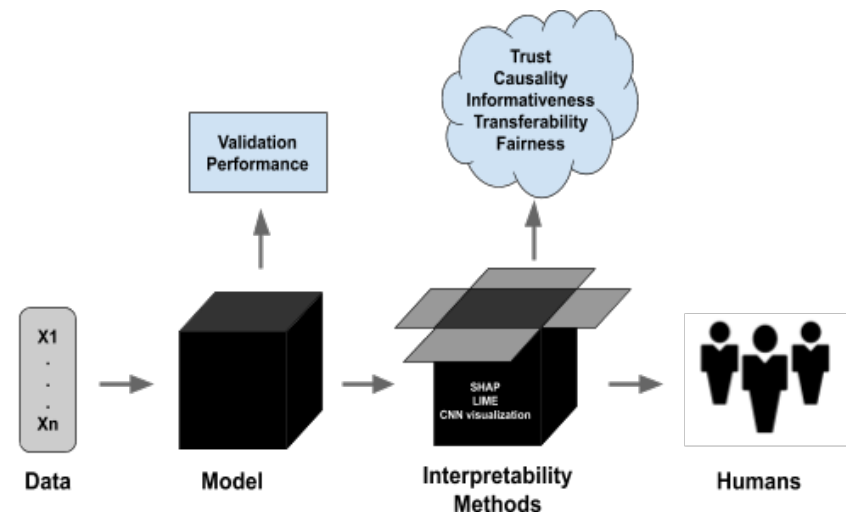
// The professor *explained* the poem to the class.

2: to tell, show, or be the reason for or cause of something

// Scientists could not *explain* the strange lights in the sky.

(in healthcare, 2: is not ”causality”)

XAI as Interpretability in Machine Learning

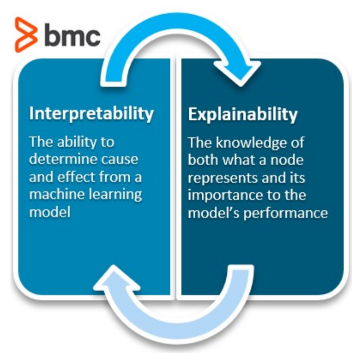


https://blog.ml.cmu.edu/2020/08/31/6-interpretability/

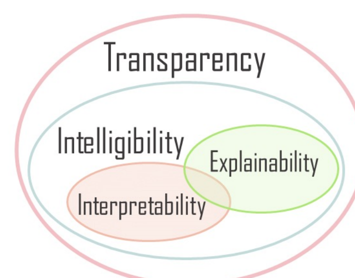
Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31-57.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

- *Interpretability* allows us to understand what exactly a model is learning, what other information the model has to offer, and the *justifications* behind the model decisions, and evaluate all of these



https://www.bmc.com/blogs/machine-learning-interpretability-vs-explainability/



A Survey of Explainable AI Terminology, Clinciu & Hastie'19

XAI in Social Sciences

arXiv:1706.07269v3 [cs.AI] 15 Aug 2018

Explanation in Artificial Intelligence:
Insights from the Social Sciences

Tim Miller

School of Computing and Information Systems
University of Melbourne, Melbourne, Australia
tmiller@unimelb.edu.au

Abstract

There has been a recent resurgence in the area of explainable artificial intelligence as researchers and practitioners seek to make their algorithms more understandable. Much of this research is focused on explicitly explaining decisions or actions to a human observer, and it should not be controversial to say that looking at how humans explain to each other can serve as a useful starting point for explanation in artificial intelligence. However, it is fair to say that most work in explainable artificial intelligence uses only the researchers' intuition of what constitutes a 'good' explanation. There exists vast and valuable bodies of research in philosophy, psychology, and cognitive science of how people define, generate, select, evaluate, and present explanations, which argues that people employ certain cognitive biases and social expectations towards the explanation process. This paper argues that the field of explainable artificial intelligence should build on this existing research, and reviews relevant papers from philosophy, cognitive psychology/science, and social psychology, which study these topics. It draws out some important findings, and discusses ways that these can be infused with work on explainable artificial intelligence.

Keywords: Explanation, Explainability, Interpretability, Explainable AI, Transparency

Contents

Preprint submitted to Journal Name *August 16, 2018*

- XAI should build on research in philosophy, psychology, social sciences, social psychology
- Not about the model, but about the human interacting with the AI

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

XAI in Social Sciences

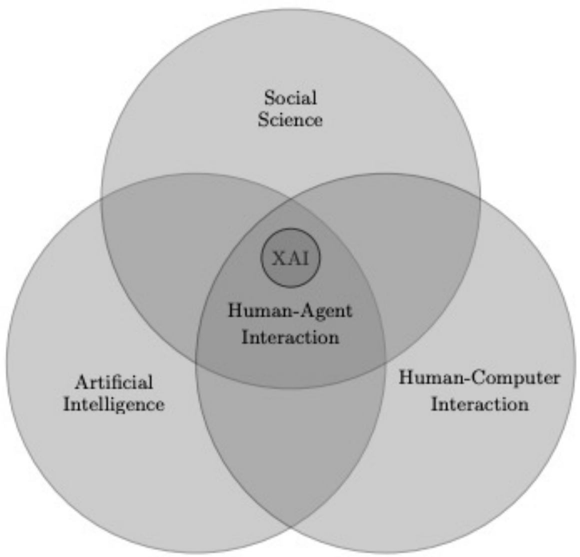


Figure 1: Scope of Explainable Artificial Intelligence

- What are the characteristics of an explanation?
  - Contrastive; they are sought in response to particular counterfactual cases
  - Social & Selective; transfer of knowledge & humans select one or two causes
  - Probabilities "probably" don't matter; referring to probabilities is not as effective as referring to causes

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1-38.

electronic visualization laboratory

XAI in Public Policy

THE ROYAL SOCIETY

Explainable AI: the basics

POLICY BRIEFING

Explainable AI: the basics Policy briefing Issued: November 2019 DES6051

ISBN: 978-1-78252-433-5

This report can be viewed online at: royalsociety.org/ai-interpretability

- Characteristics of AI:
  - *interpretable*, implying some sense of understanding how the technology works
  - *explainable*, implying that a wider range of users can understand why or how a conclusion was reached
  - *transparent*, implying some level of accessibility to the data or algorithm

XAI in Healthcare

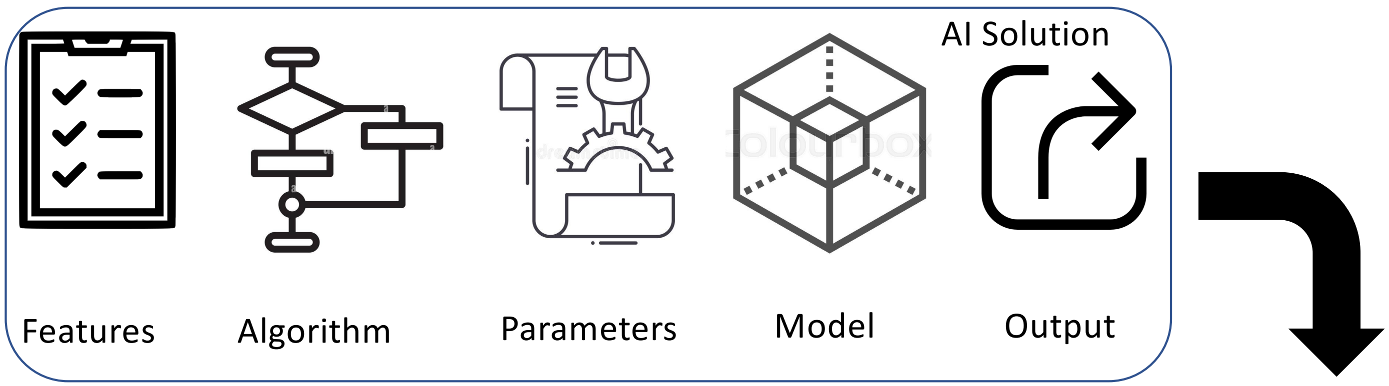
- Early expert systems in healthcare as XAI
  - e.g., MYCIN, early '70s AI, written in Lisp

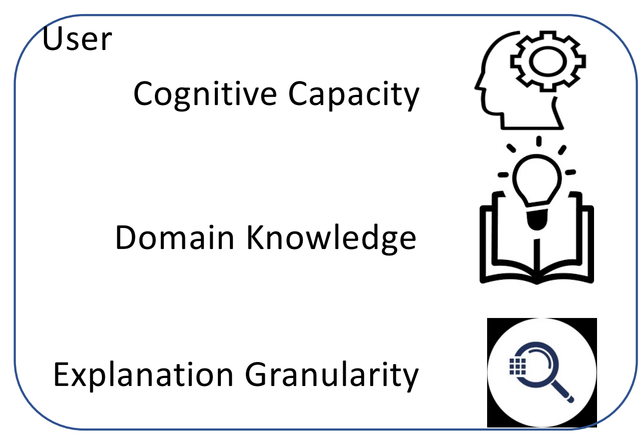
Today

- 18% rule-based
- 2% ML
- 80% heuristics

Interpretable machine learning in healthcare
Ahmed at al., 2018 BCBHI

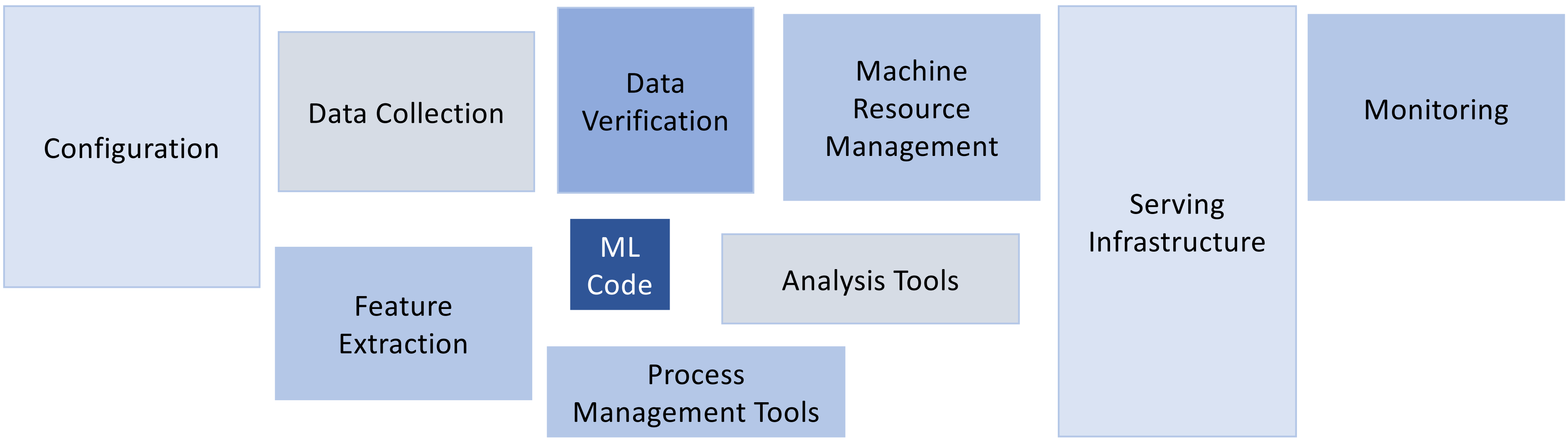
XAI in Healthcare: Wholistic View





Interpretable machine learning in healthcare
Ahmed at al., 2018 BCBHI

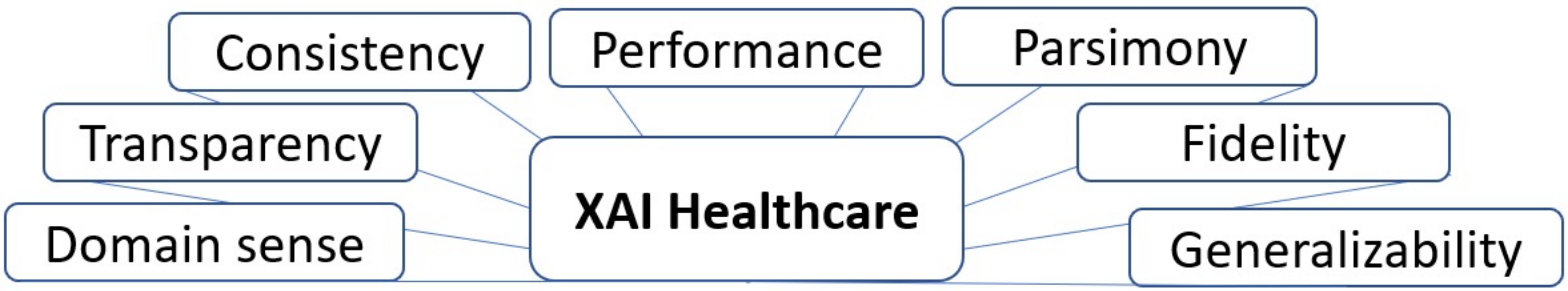
Only a Small Fraction of ML Systems is ML Code



Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in neural information processing systems*, *28*, 2503-2511.

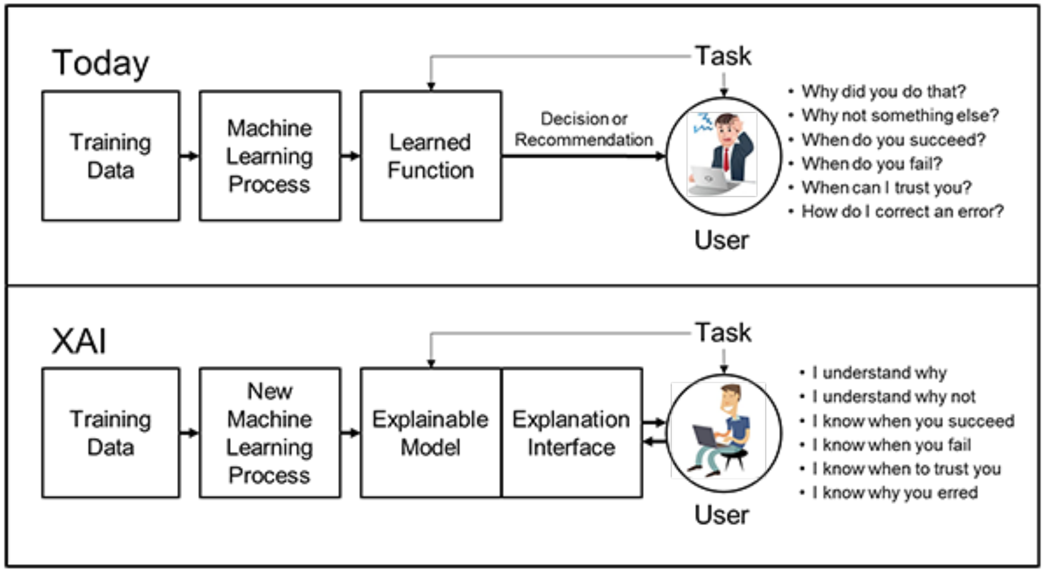
XAI in Healthcare

- What is a “good” XAI?



Interpretable machine learning in healthcare
Ahmed at al., 2018 BCBHI

XAI: Why Bother?



https://www.darpa.mil/program/explainable-artificial-intelligence

- When fairness is critical
- When consequences are far-reaching
- When a new/unknown hypothesis is drawn by the AI
  - E.g., “Pneumonia patients with asthma have lower death risk”

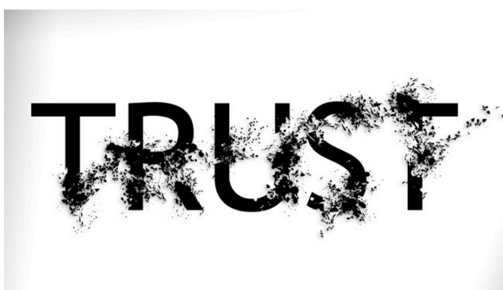
Brief Recap of Terminology

- Explain (layman): to make something or a cause easy to understand
- Interpret (ML): understand the model workings
- XAI (social sci): consider the human (e.g., counterfactuals)
- XAI (public policy): consider the model, the human, the data/algo
- XAI (healthcare): consider the model, the human (cognitive, domain knowledge, needs), the data/algo, the features/params

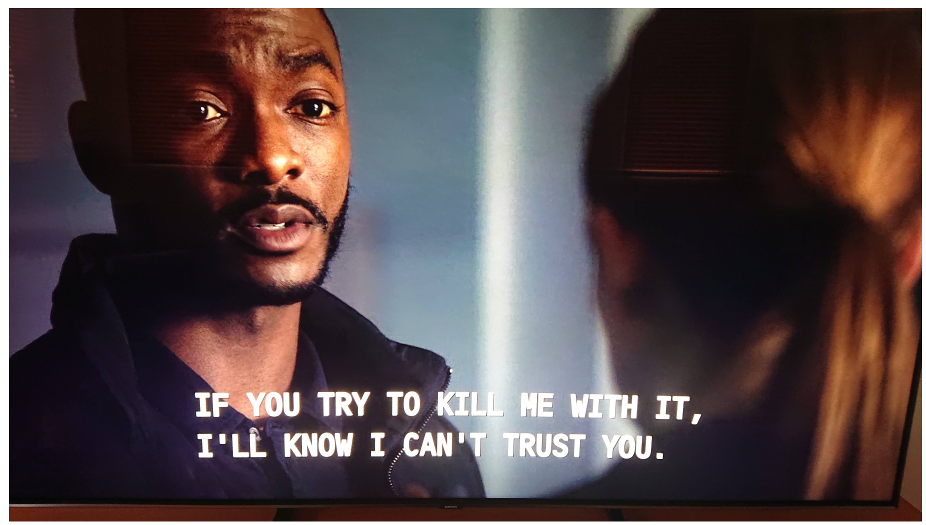
Trust

- “Hard won and easily lost”

NB: Explanations can also confuse or mislead.



larrybroughton.net



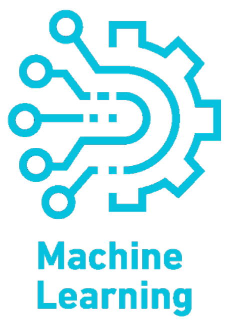
ABC’s “Agents of SHIELD”

- EVL and Computational Oncology
- XAI Considerations
- Vis in XAI
- Clustering with Spatial Data: RT
- Clustering with Spatial Data: LN
- A “Good” Visual Explanation for AI

Visual Explanations in AI

Vis in Human-Machine Analysis

AI



Mixed human-machine analysis

- can leverage and balance computational and human effort in data analysis

- AI < IA (Intelligence Amplification)

Brooks, F.P. 1996, **The computer scientist as toolsmith Part II**

IA

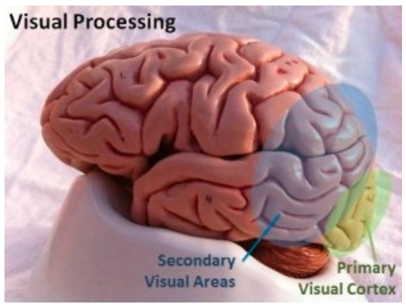
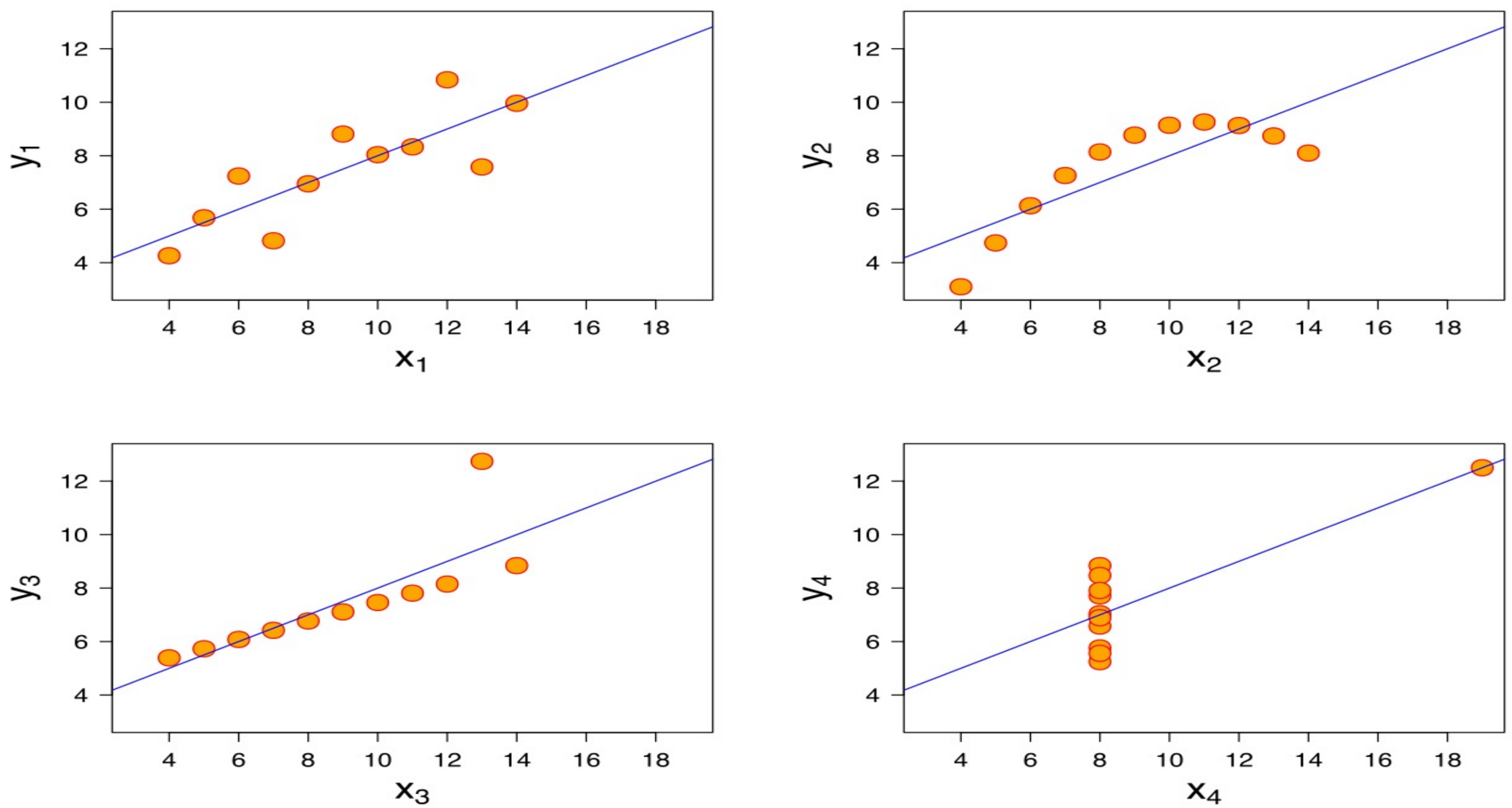


Image from https://www.flaticon.com/

Anscombe's Quartet

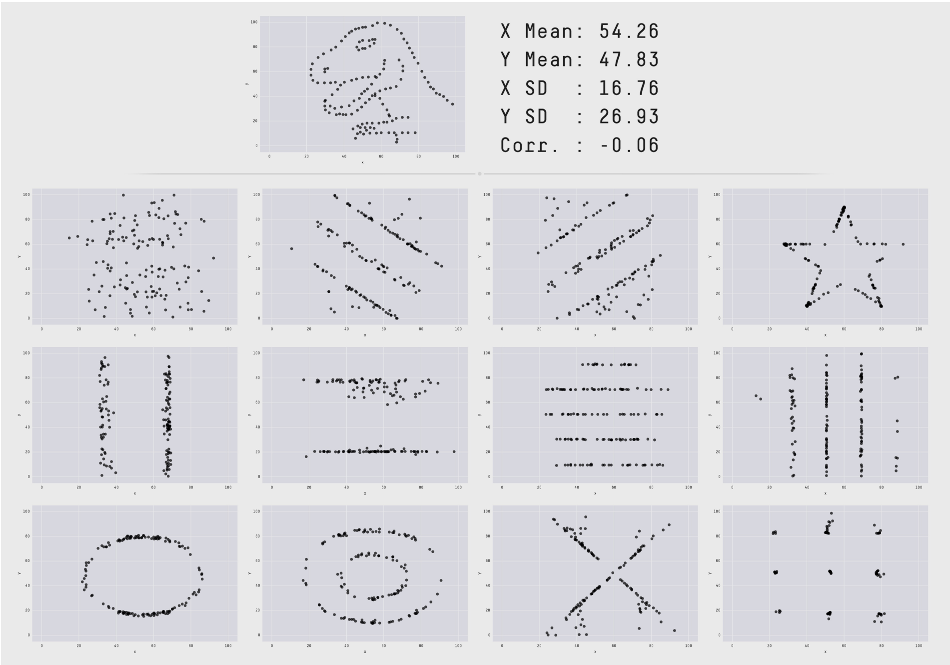
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

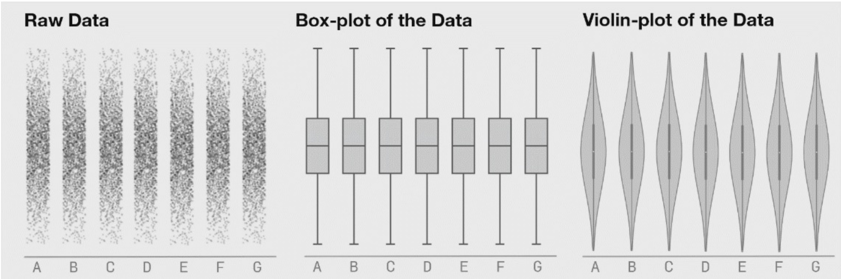
Visualizing Anscombe's Quartet



Spatial Structure in Data Science

http://www.autodeskresearch.com/publications/samestats



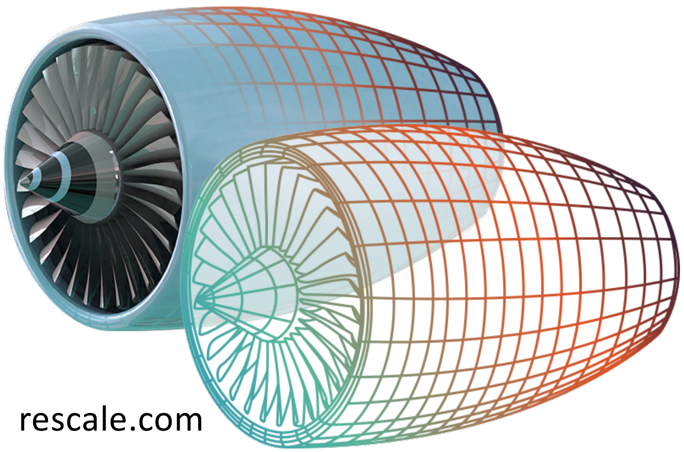


Do We Have to Use Visualization?

Not necessarily. E.g., post-hoc explanations (as opposed to model transparency)

- Text explanations
- Local explanations, e.g., saliency maps
- Explanation by example, e.g., KNN
- Other basic data visualizations like t-SNE
- …

Example: Digital Twin Dyad in HNC



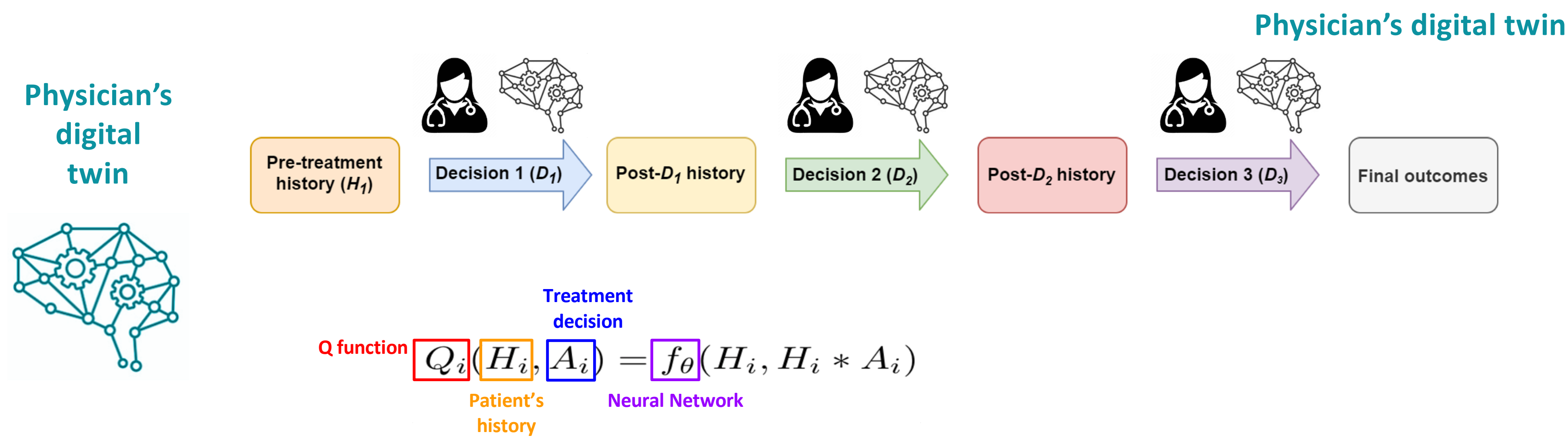
rescale.com

- Digital twin: a concept from manufacturing, where a digital replica is built and simulated
- Concept adopted by NIH; hope to replicate digitally biological systems

Tardini, E., Zhang, X., Canahuate, G., Wentzel, A., Mohamed, A. S., Van Dijk, L., ... & Marai, G. E. (2021). Optimal policy determination in sequential systemic and locoregional therapy of oropharyngeal squamous carcinomas: A patient-physician digital twin dyad with deep Q-learning for treatment selection. *medRxiv*.

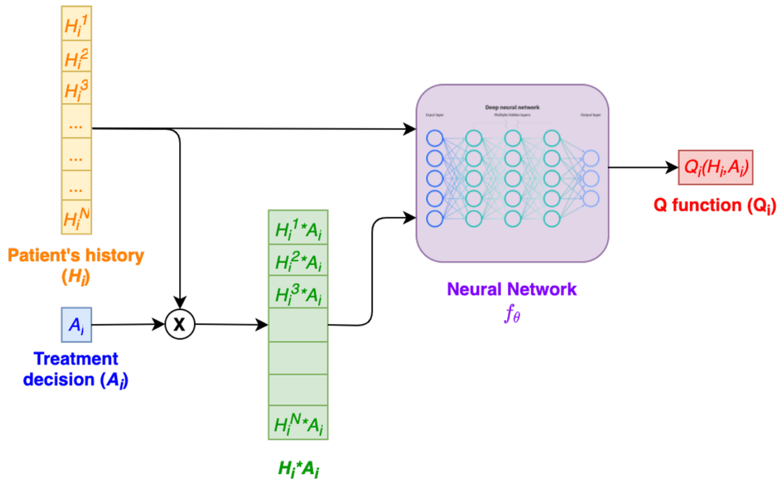
Digital Twin Dyad in HNC Treatment

- Multi-stage treatment: chemotherapy, radiotherapy, surgery
- Conflicting outcomes: efficacy vs quality of life
- Reinforcement q-learning via deep learners

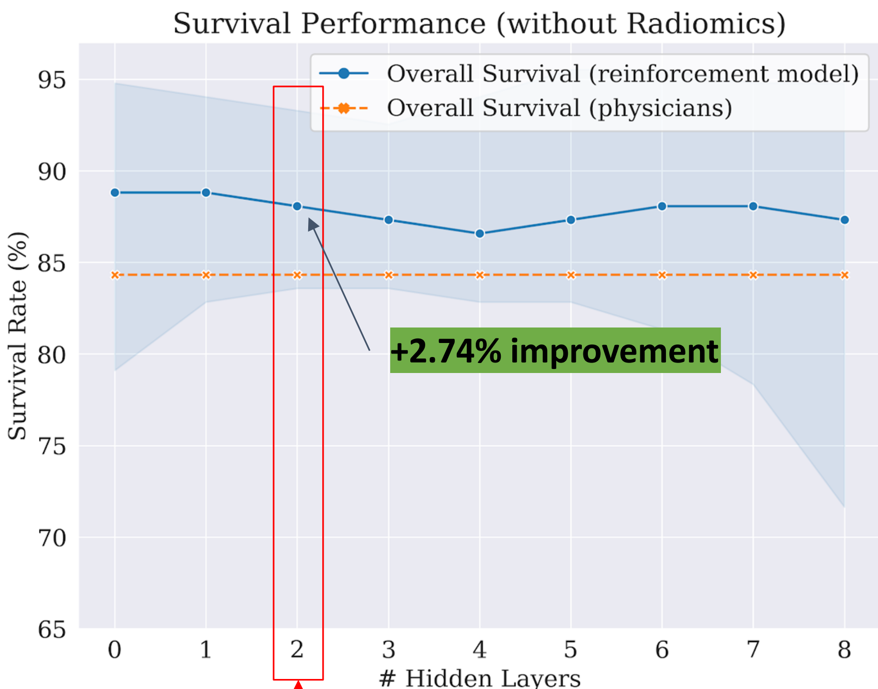


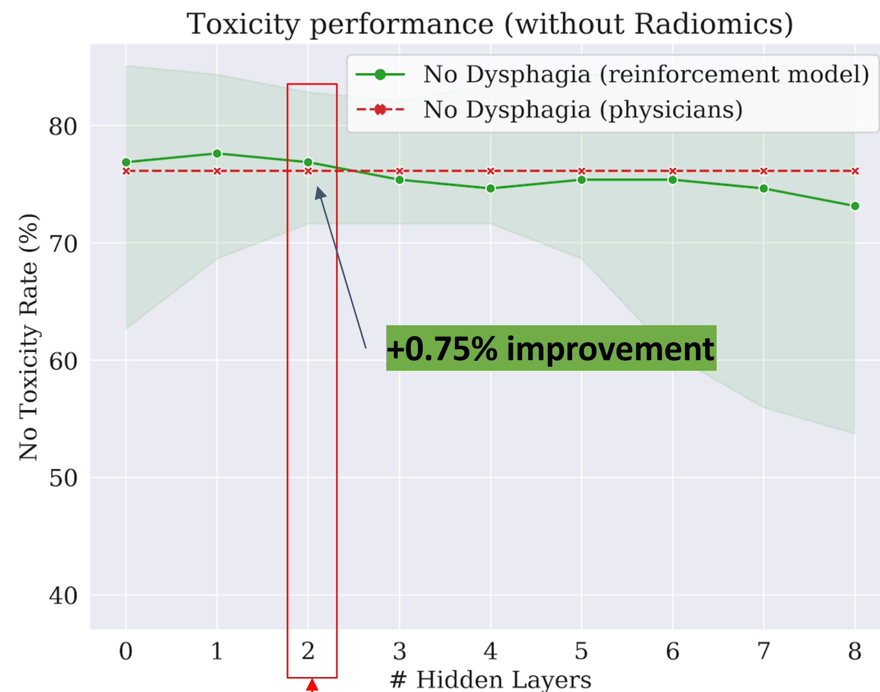
Digital Twin Dyad in HNC Treatment

- Essentially, an AI system that learns from past medical experience
- Treatment simulator for patient outcomes



Physician digital twin (simulated outcomes)





Best
model

Textual Explanation

- Clinician team prescribed: D1: **not** IC, D2: CC, D3: not ND
- DQL sequence prescribed: D1: IC, D2: CC, D3: not ND.
- Medical records examination: patient had only one functioning kidney, thus no chemotherapy prescribed, as precaution to prevent renal injury.
- Dyad system performed well given the input specifications of this case

Textual Explanation

- Clinician team prescribed: D1: not IC, D2: CC, D3: not ND
- DQL sequence prescribed: D1: IC, D2: CC, D3: not ND.
- Medical records examination: patient had only one functioning kidney, thus no chemotherapy prescribed, as precaution to prevent renal injury.
- Dyad system performed well given the input specifications of this case

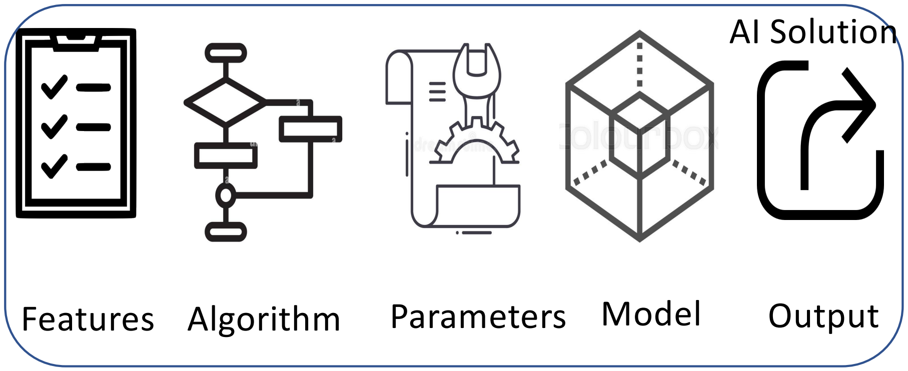
- OK for proof of concept, but what when we add spatial information?

Vis in XAI

- Vis can enable XAI
- Chatzimparmpas et al 2020, https://trustmlvis.lnu.se/

- Representing outcomes
- Looking "inside" the models
- Depicting conditional variation
- Exploring what-if scenarios, steering

- Whenever spatial data is involved



What is a “Good” Visual XAI?

- 45 scientists (bioinformatics, biology, visual computing)

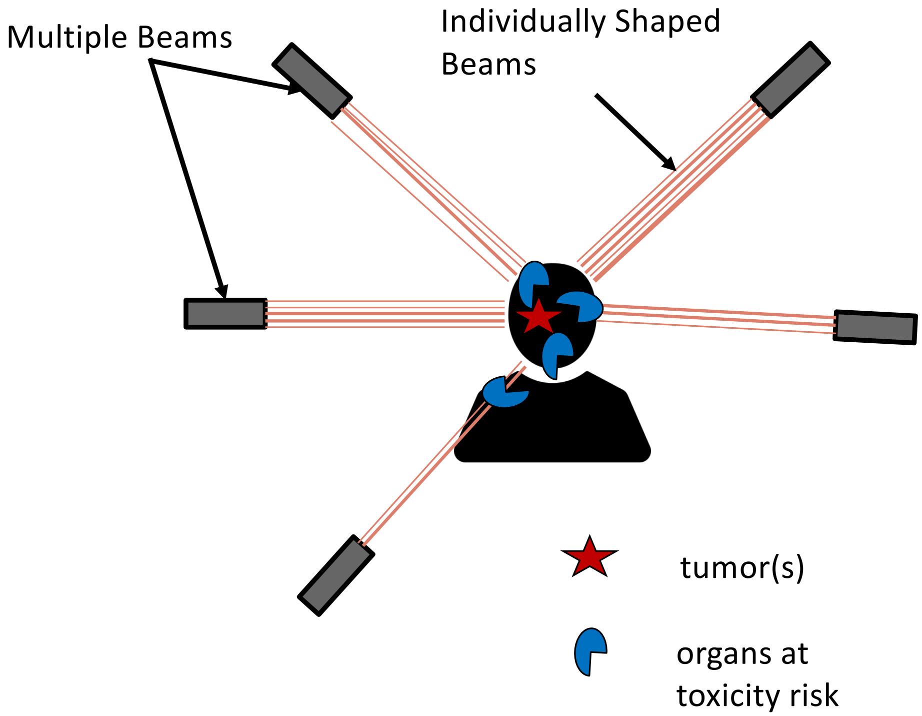
- … ahem?
- *“Does it help me understand?”*
- *“Does it help me understand faster?”*

- EVL and Computational Oncology
- XAI Considerations
- Vis in XAI
- Clustering with Spatial Data: RT
- Clustering with Spatial Data: LN
- A “Good” Visual Explanation for AI

Clustering with Spatial Data: Radiation Therapy

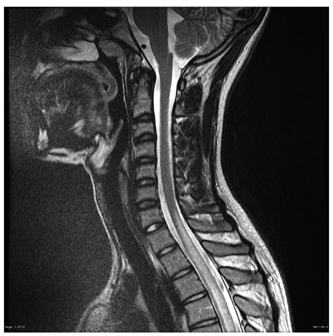
Radiation Therapy Planning is Complex

- Modern RT plans are complex and require human expertise
  - >1 week to create
- Radiation to surrounding organs also causes toxicity (side effects) such as permanent dry mouth or loss of vision
- Idea: try to predict distribution of radiation, then predict toxicity

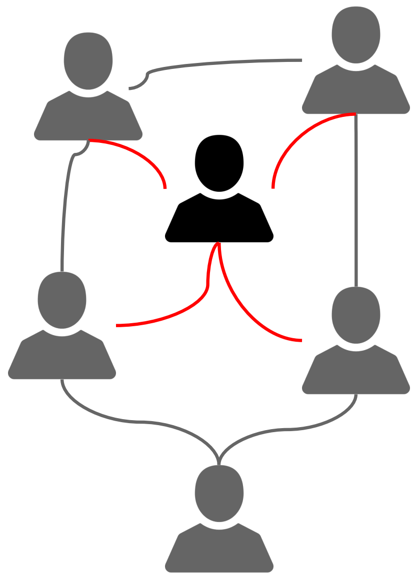


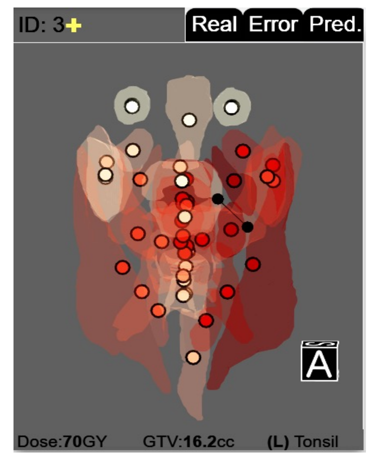
electronic
visualization
laboratory

Using Patient Repositories to Enable Similarity-Based Prediction



Case courtesy of A.Prof Frank
Gaillard, Radiopaedia.org, rID: 19649





Cohort Data
(CT volume images)

Most Similar Patients
w.r.t. Tumor Location

Predicted Radiation
From Similar Examples

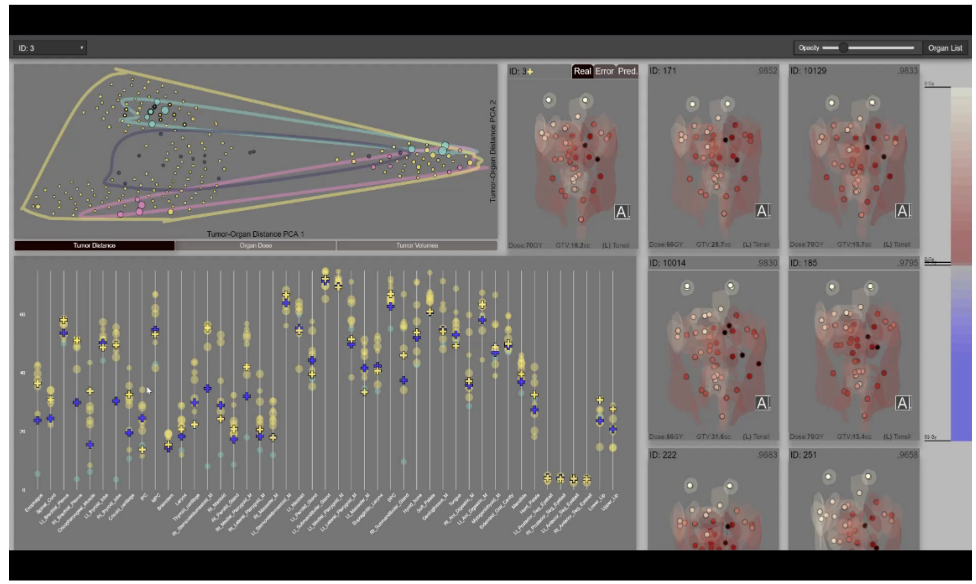
Computational Challenge

- No existing spatial similarity methodology
- Current human-based approach:
  - medical image inspection
  - prior knowledge
  - physician/institutional memory
- Current approach not scalable

Image: K. Reed, MSgt, http://www.af.mil

Visual Computing Approach

- Computing w/ images & 3D models
- Novel spatial similarity measure (T-SSIM)
- Predictive algorithm for RT dose-distribution
- Application of visual steering to precision radiation oncology



Wentzel, A., Hanula, P., Luciani, T., Elgohari, B., Elhalawani, H., Canahuate, G., ... & Marai, G. E. (2019). Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. *IEEE transactions on visualization and computer graphics*, *26*(1), 949-959.

Related Work

- Spatial Similarity
  - Graph [Sun 2011], Shape [Iyer 2005], and Deep-learning [Nguyen 2018]
  - Can't handle large groups of organs
- Biomedical + Nonspatial data visualization
  - MRI Images + Statistical views [Nunes 2014], RT cohort + uncertainty in Bladders [Raidou 2018]
  - No spatial similarity, or RT to surrounding organs
- Visual Steering for Model Development
  - Clustering analysis [Kwon, 2018], RT dose-response modeling [Naqa 2006], Environmental Modeling [Poco 2014]
  - Different problem space than ours

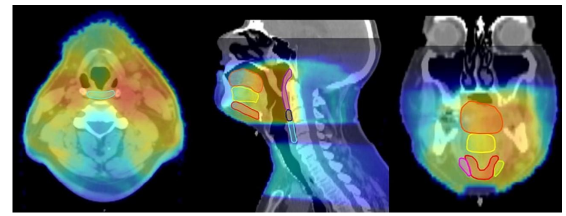
Data and Goals

Data

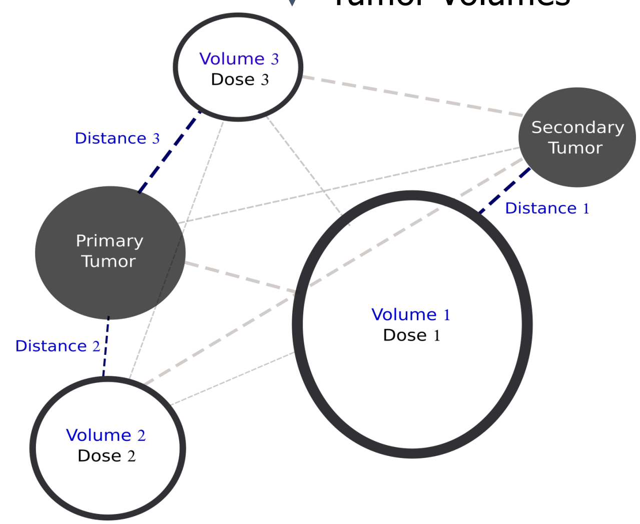
- 165 Head/Neck cancer patients
- CT scans and RT plans
  - 45 Surrounding Organs + Tumor(s)
  - Positions/distances, volumes, dose
- Known demographics & toxicity

Goals

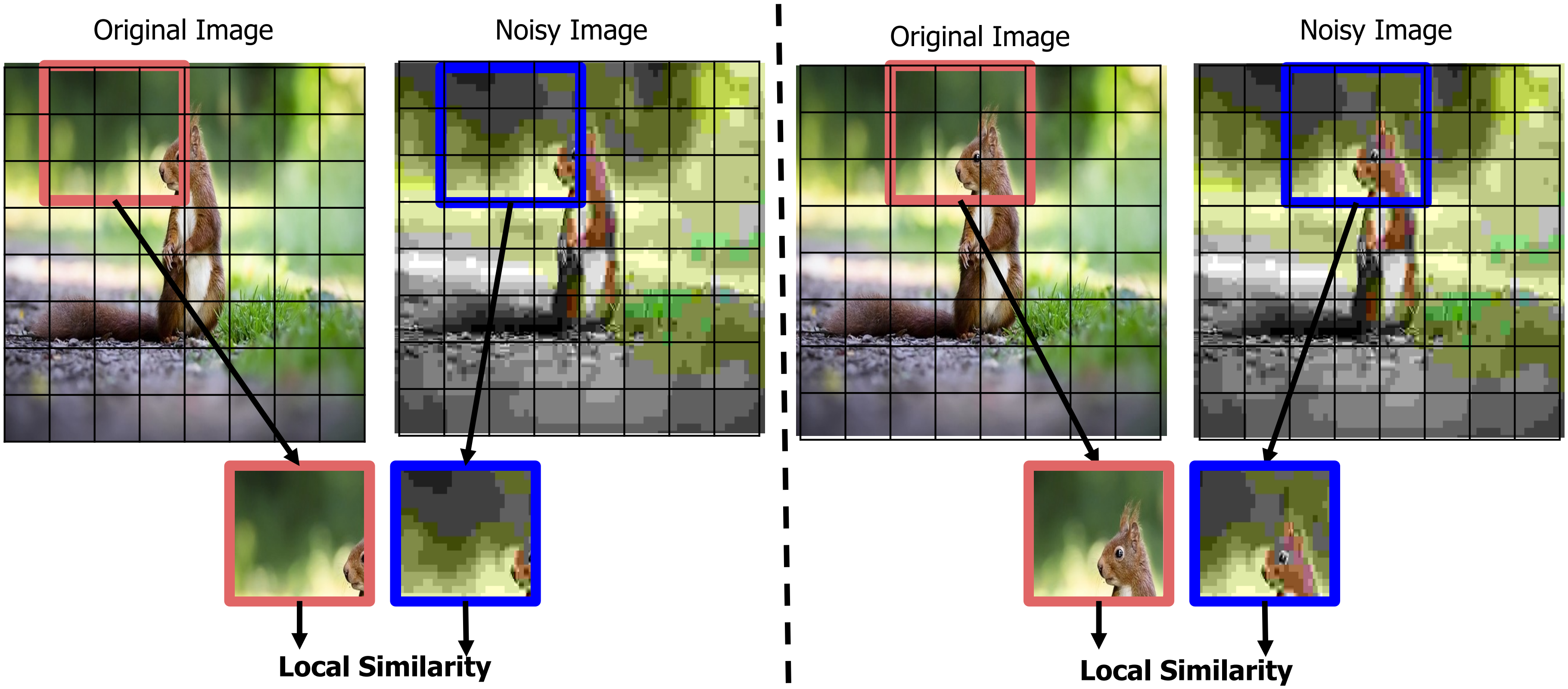
- Measure similarity based on spatial data
- Estimate delivered dose distribution
- Analyze patterns in RT plans



Segment Organ & Tumor Volumes



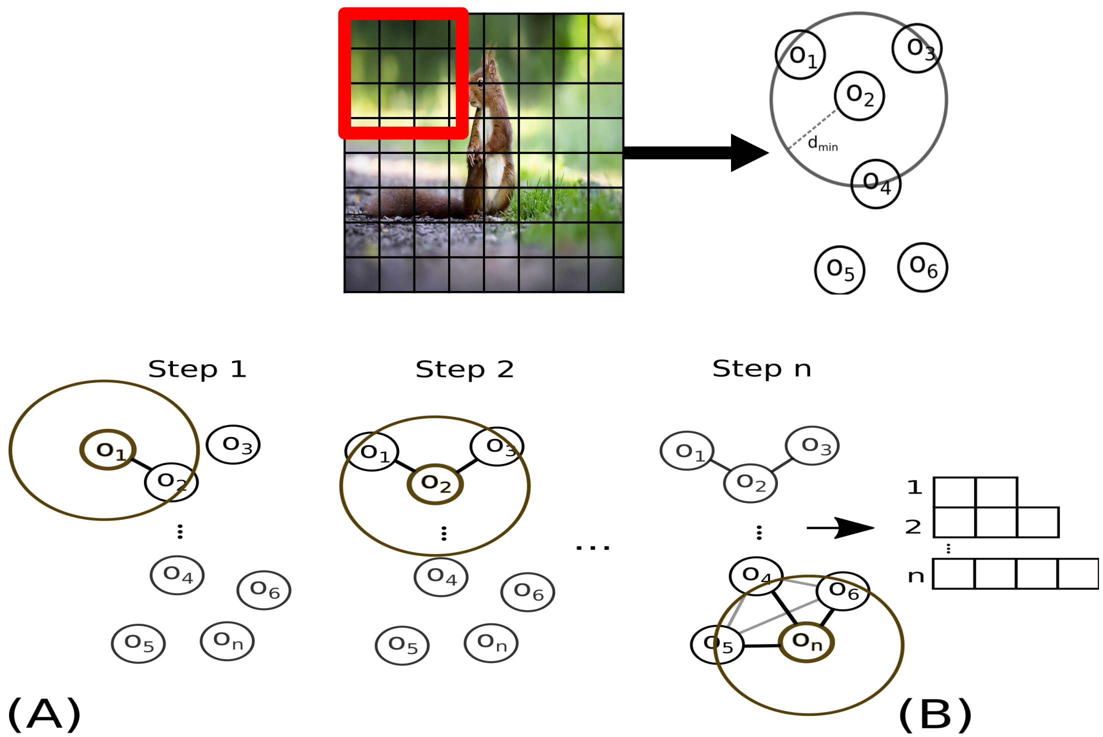
Inspiration: Image Similarity with SSIM



From Image Similarity to Topological Similarity

From 2D images to organ 3D topology:

- 2D Image Window -> Spherical 3D Window
- Pixel sliding -> Organ-Center sliding



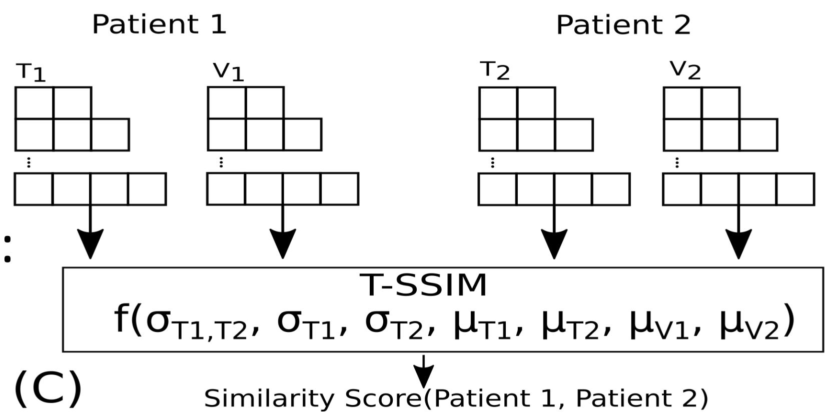
Topological Similarity Measure

SSIM formula uses 3 pixel-value channels:

- Structure
- Intensity
- Contrast

T-SSIM formula uses distance and volumes:

- Structure -> Distance to tumor
- Intensity -> Organ Volume
- Contrast -> not used here

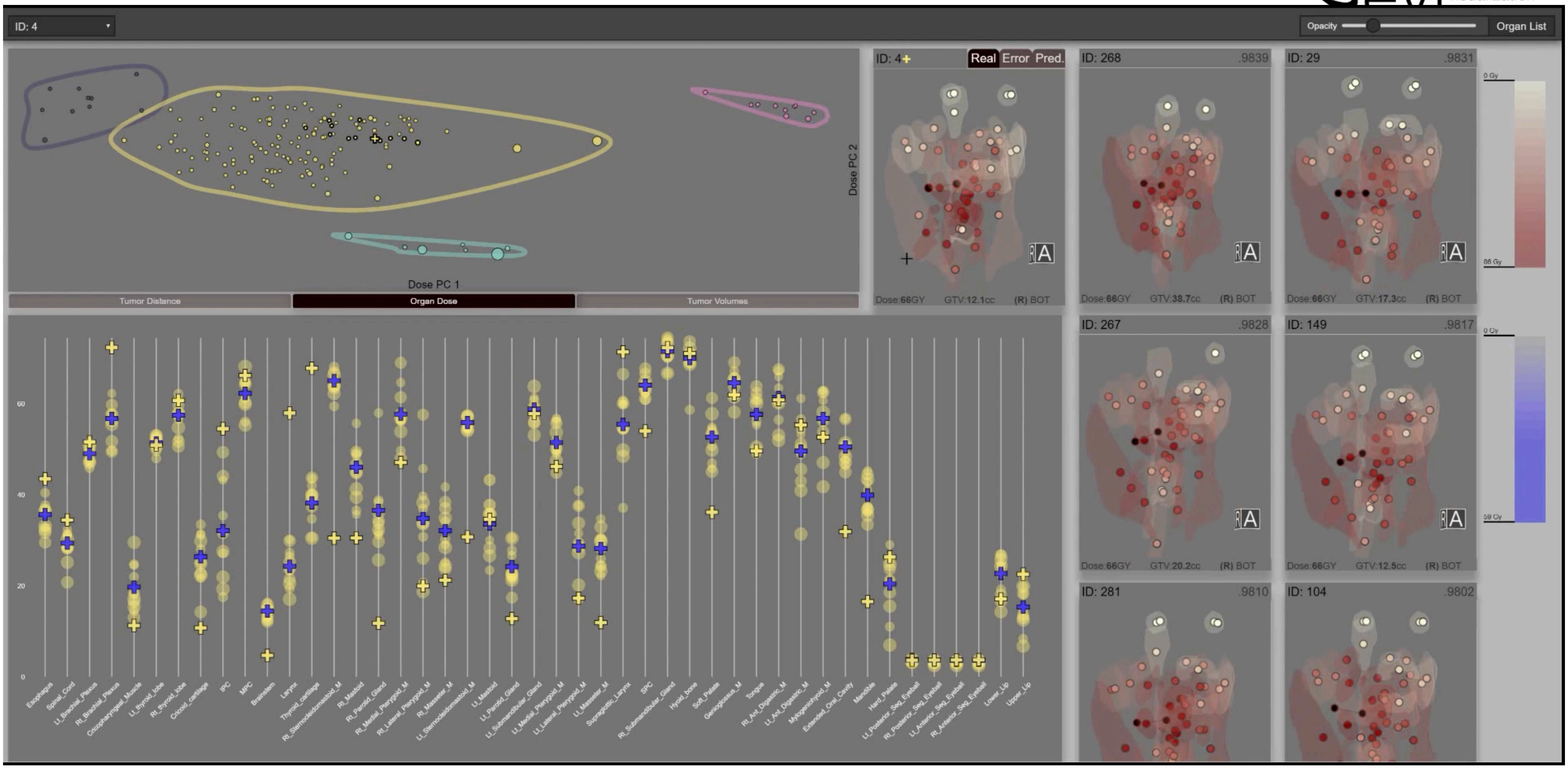


Similarity-based Predictive Algorithm

- New Patient:
  - Match k most similar patients using T-SSIM
  - k determined via line search = (cluster size)½
- Calculate per-organ weighted average of doses of similar patients:

$$Rad_{predicted} = \frac{\sum_{n \in Neighbors} T_n * Rad_n}{\sum_{n \in Neighbors} T_n}$$

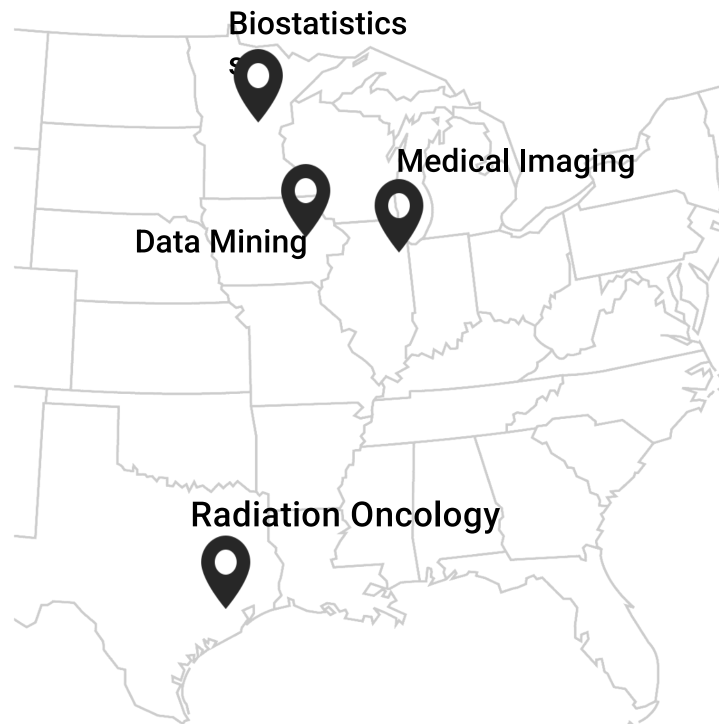
- Report the error between the predicted dose distribution and actual RT dose distribution


ID: 4
Opacity
Organ List
Dose PC 1
Dose PC 2
Tumor Distance
Organ Dose
Tumor Volumes
Real
Error
Pred.
ID: 268
.9839
ID: 29
.9831
ID: 267
.9828
ID: 149
.9817
ID: 281
.9810
ID: 104
.9802

Hybrid Evaluation

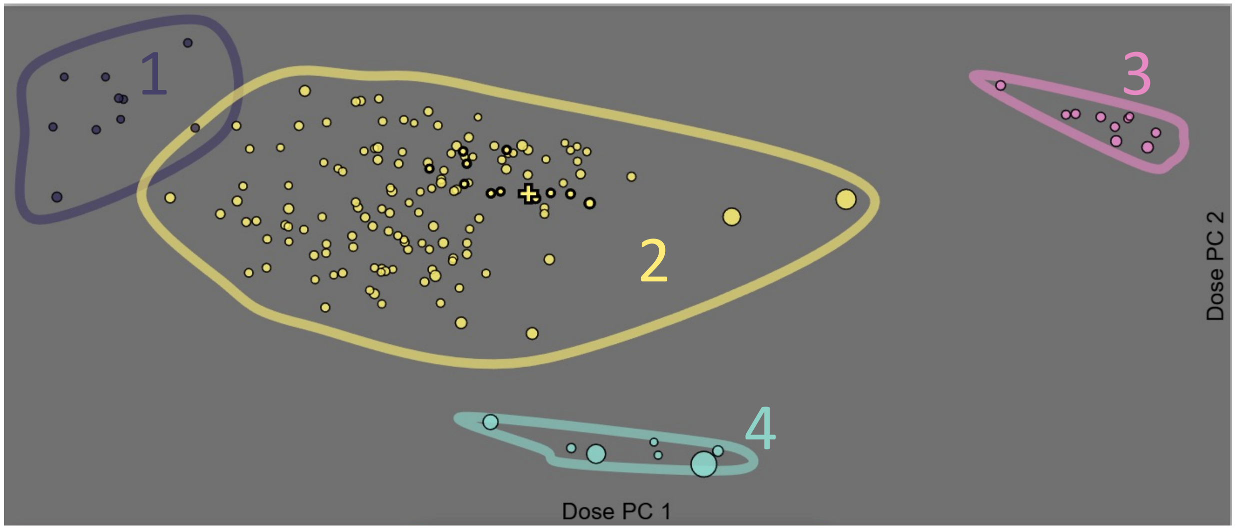
- Qualitative:
  - Two case studies w/ 4 Domain Experts
  - Visual steering using results from one expert's clustering
  - Analyzing + troubleshooting prediction

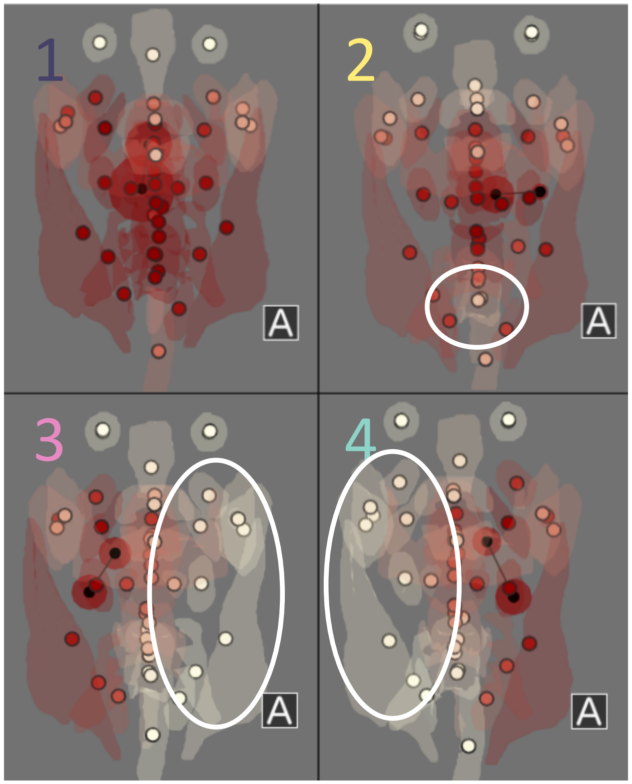
- Quantitative
  - Leave-One-Out cross-validation
  - Mean, absolute percent error

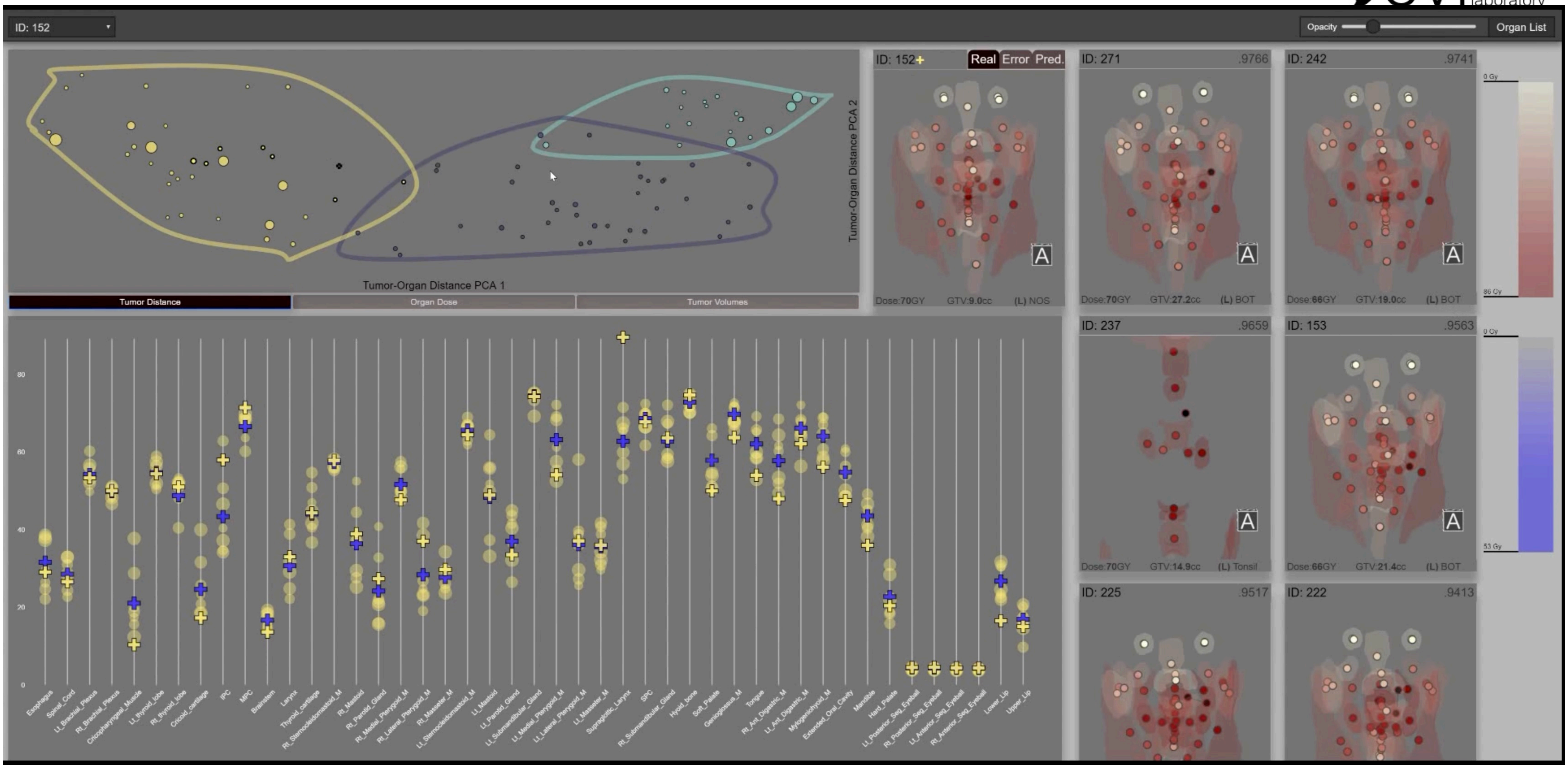


Visual Steering Results

- 165 patient dataset: successfully retrieved patients with similar tumor location
- Identified 4 archetype RT plans in repository
- Enabled 4 domain experts to synergize their efforts

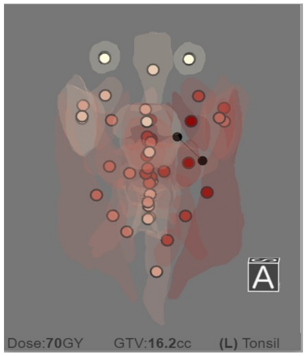




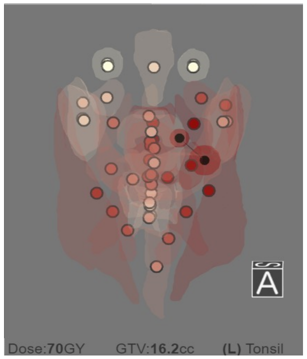

ID: 152
Opacity
Organ List
Tumor-Organ Distance PCA 1
Tumor-Organ Distance PCA 2
Tumor Distance
Organ Dose
Tumor Volumes
ID: 152
Real
Error
Pred.
Dose:70GY GTV:9.0cc (L) NOS
ID: 271 .9766
Dose:70GY GTV:27.2cc (L) BOT
ID: 242 .9741
Dose:66GY GTV:19.0cc (L) BOT
ID: 237 .9659
Dose:70GY GTV:14.9cc (L) Tonsil
ID: 153 .9563
Dose:66GY GTV:21.4cc (L) BOT
ID: 225 .9517
ID: 222 .9413

Quantitative Analysis Results

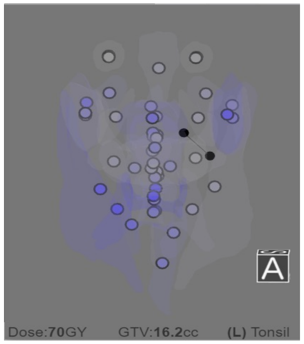
- Mean absolute percent error
  - Archetype-unaware: 16.7% (6.2 Gy) ± 9.3%
  - Archetype-aware: 12.3% (4.7 Gy) ± 4.4%
- Running times on 8GB DDR4 RAM and Intel i5 2.5GHz processor:
  - Processing: 100.5s
  - Prediction: 476.5s
  - <10 min total on a laptop, compared to 1+ week with a medical expert



Predicted Plan



Actual Plan

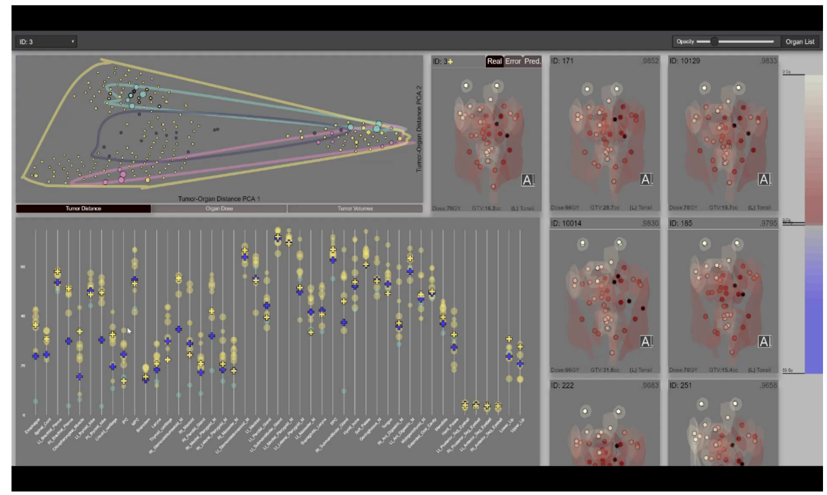


Prediction Error

evl electronic visualization laboratory

Location: Tumor-to-OAR Distance Similarity

- Similarity metric based on 3D structure
- Correlates with RT dosage
  - can identify similar RT plans
  - can identify outliers
  - may predict RT
- Spatially-aware clustering
  - Improves AJCC8 survival prediction quality (beyond radiomics)
- It also correlates w/ groups at risk for specific toxicity outcomes



Tumor-to-OAR Correlates with Toxicity

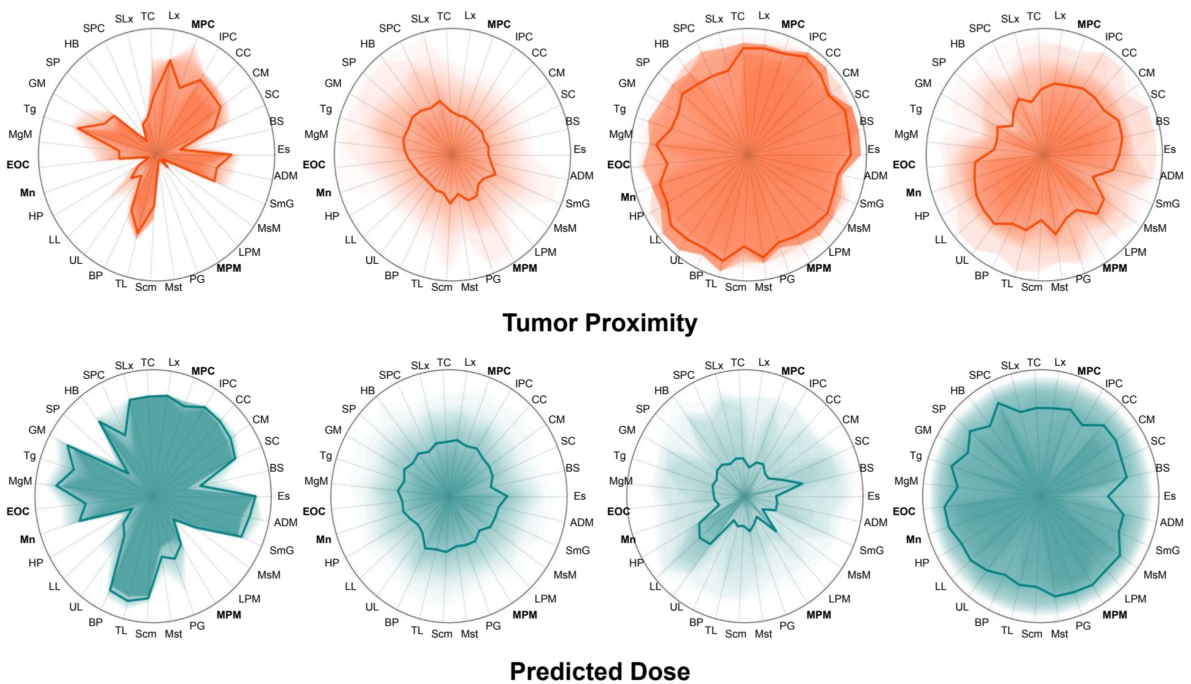
- 200 patients

Spatial Cluster	Count	% RAD	% Feeding Tube	% Aspiration	Extended Oral Cavity Predicted Dose (Gy)	Mandible Predicted Dose (Gy)	Average Medial Pterygoid Muscle Predicted Dose (Gy)	Mandible-Tumor Distance (mm)	Medial Pharyngeal Constrictor-Tumor Distance (mm)
Spatial Cluster 1	3	0	0	0	51.96 (1.97)	39.92 (2.16)	37.94 (0.5)	13.35 (1.44)	0.29 (1.64)
Spatial Cluster 2	114	5.3	3.5	2.6	50.84 (1.37)	38.21 (1.02)	38.03 (1.1)	7.18 (3.82)	12.23 (4.55)
Spatial Cluster 3	35	11.4	8.6	5.7	48.66 (3.20)	36.33 (2.82)	35.91 (2.1)	0.04 (1.81)	0.28 (0.60)
Spatial Cluster 4	48	50	31.3	29.2	57.11 (2.93)	44.84 (3.69)	42.03 (1.9)	1.64 (2.28)	5.86 (4.23)

Wentzel, A., Hanula, P., van Dijk, L. V., Elgohari, B., Mohamed, A. S., Cardenas, C. E., ... & Marai, G. E. (2020). Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy. *Radiotherapy and Oncology*, 148, 245-251.

electronic visualization laboratory

Tumor-to-OAR Location and Toxicity

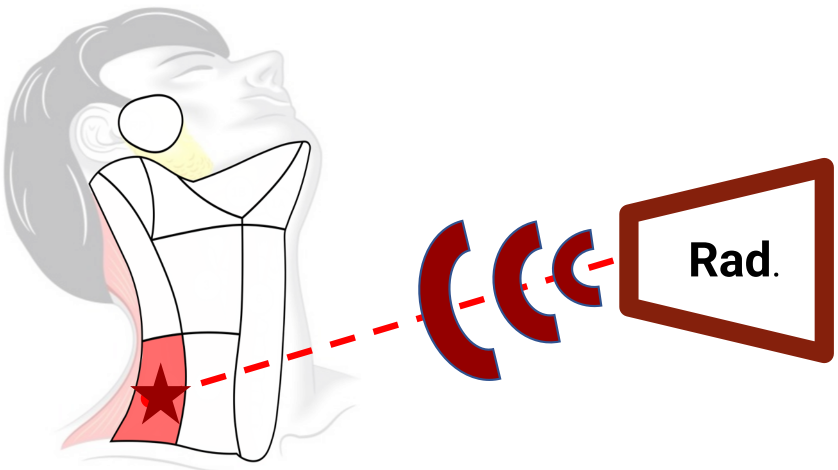


Abbreviation Legend:
UL: Upper Lip, LL: Lower Lip, HP: Hard Palate, Mn: Mandible, EOC: Extended Oral Cavity, MgM: Mylo Geniohyoid Muscle, Tg: Tongue, GM: Genioglossus Muscle, SP: Soft Palate, HB: Hyoid Bone, SPC: Superior Pharyngeal Constrictor, SLx: Supraglottic Larynx, TC: Thyroid Cartilage, Lx: Larynx, MPC: Medial Pharyngeal Constrictor, IPC: Inferior Pharyngeal Constrictor, CC: Cricoid Cartilage, CM: Cricopharyngeal Muscle, SC: Spinal Cord, BS: Brainstem, Es: Esophagus, ADM: Anterior Digastric Muscle, SmG: Submandibular Gland, MsM: Masseter Muscle, LPM: Lateral Pterygoid Muscle, MPM: Medial Pterygoid Muscle, PG: Parotid Gland, Mst: Mastoid, Scm: Sternocleidomastoid, TL: Thyroid Lobe, BP: Brachial Plexus.

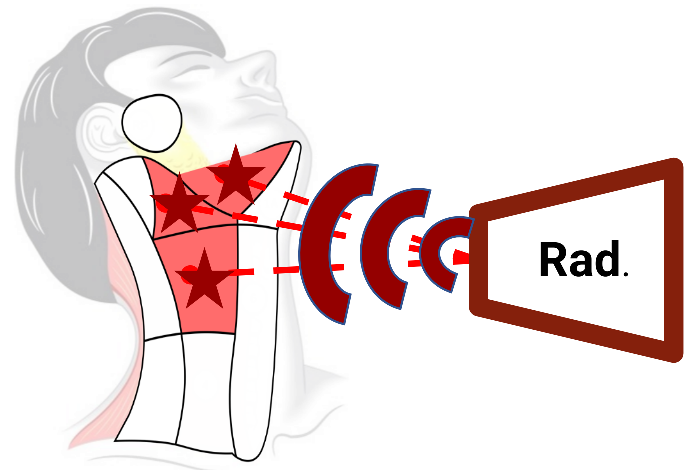
- EVL and Computational Oncology
- XAI Considerations
- Vis in XAI
- Clustering with Spatial Data: RT
- Clustering with Spatial Data: LN
- A “Good” Visual Explanation for AI

Clustering with Spatial Data: Disease Spread via Lymph Nodes

Head & Neck Cancer Therapy Depends on the Disease Spread



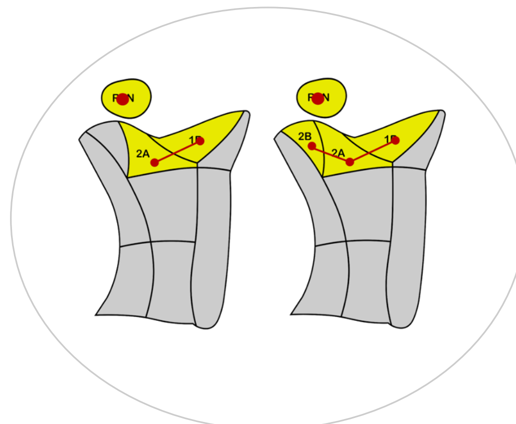
Low risk of toxicity

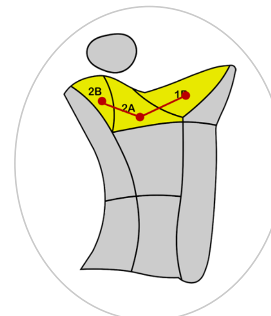


High risk of toxicity

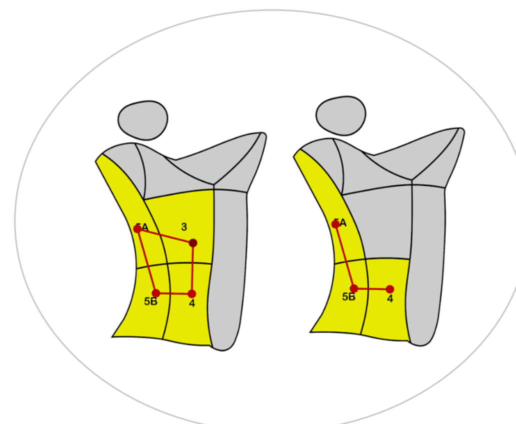
Group Patients Based on Disease Spread Patterns

Head & Neck Cancer Patients

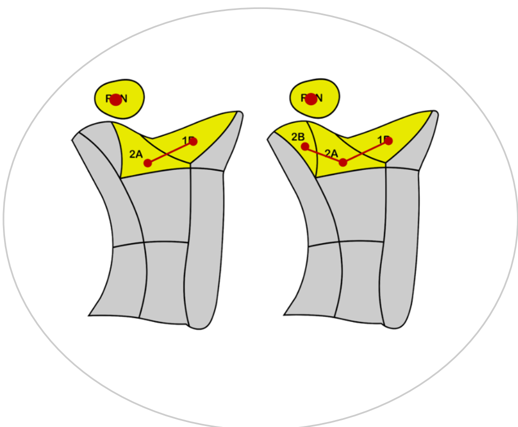




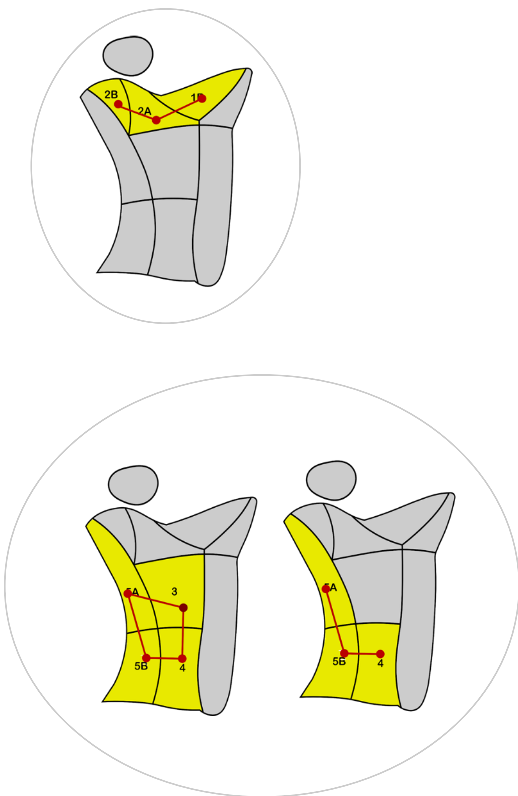
New clustering methodology using spatial data

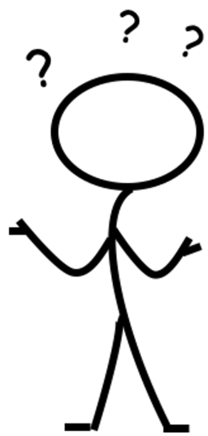


How Do We Explain Spatial Clustering?



New clustering methodology using spatial data





How to explain spatial clustering to non-experts?

Challenges

- 582 HNC patients treated at MD Anderson Cancer Center
- Important underlying anatomical structure
- No spatial clustering methodology

- Many unique spread patterns (63)
- Symmetry to be leveraged
- Analyzing # clusters & cluster membership

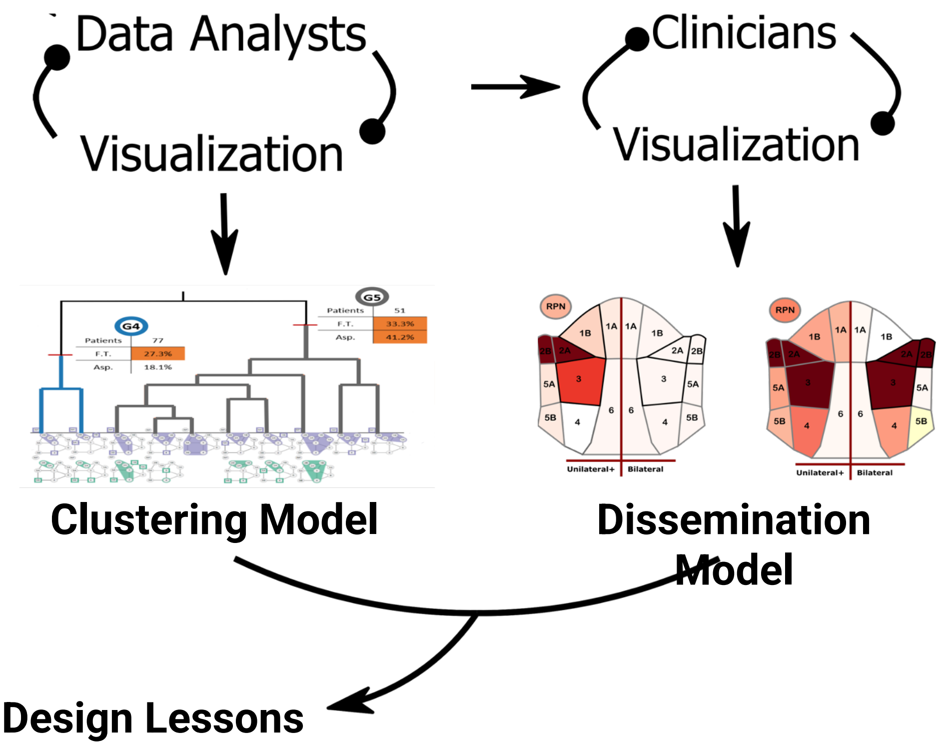
- Participatory development vs Broader dissemination
- Multi-year, multi-site project, evolving requirements

Explaining Spatial Clustering

1. Participatory Design w/ cancer researchers

1. Vis for explainable Spatial Clusters

1. Design Lessons



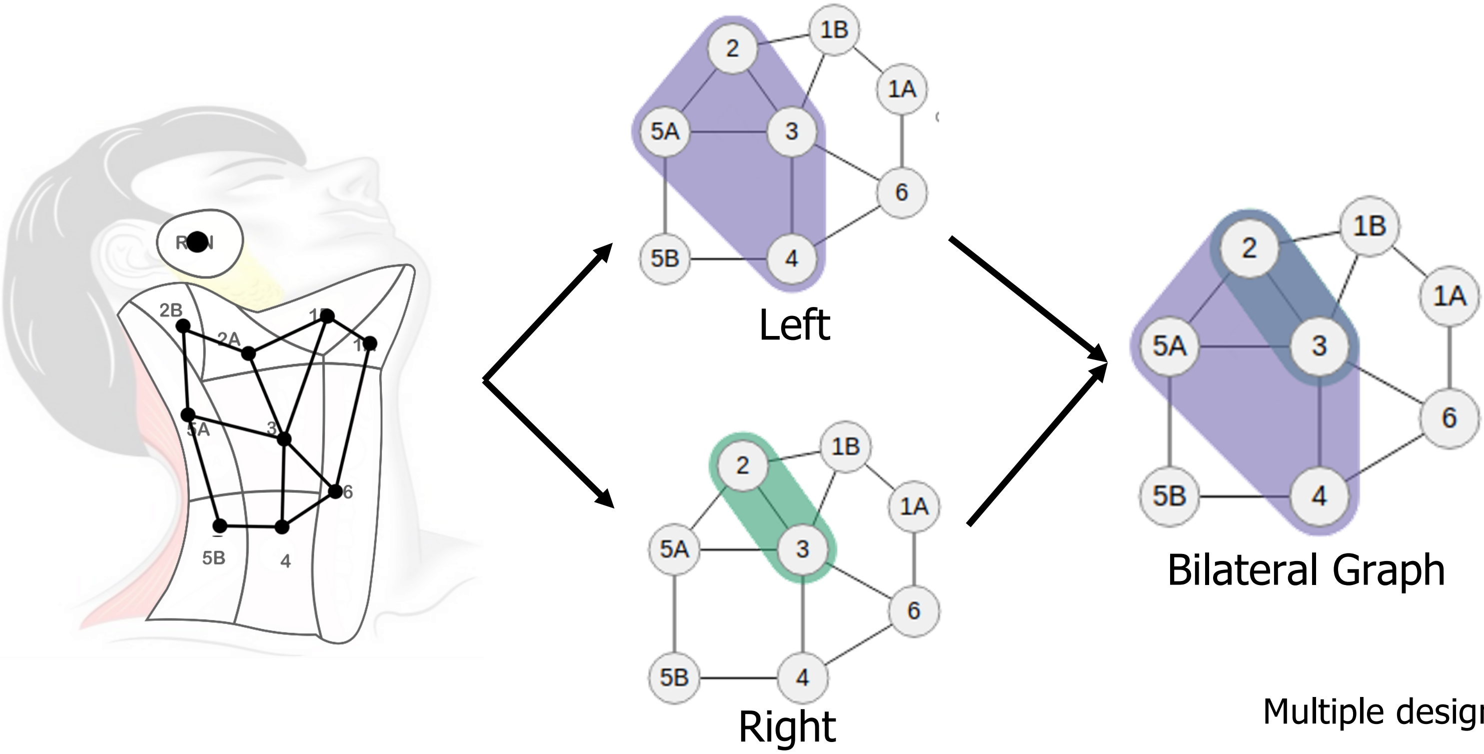
Wentzel, A., Canahuate, G., Van Dijk, L. V., Mohamed, A. S., Fuller, C. D., & Marai, G. E. (2020, October). Explainable Spatial Clustering: Leveraging Spatial Data in Radiation Oncology. In *2020 IEEE Visualization Conference (VIS)* (pp. 281-285). IEEE.

Related Work

- Cluster Visualization
  - Low dimensional embeddings [Alper 2011, Cava 2017]
  - Parallel Coordinate Plots [Chou 1999]
  - Specialized Glyphs [Cao 2011]
    - No work in spatial cluster visualization
- Healthcare/Cohort Analytics
  - Steering Regression Models [Dingen 2018]
    - Only considers linear effects
  - Bladder Shape Analysis [Grossman 2019, Raidou 2020]
    - Focuses on shape variation rather than clusters and outcomes

same spatial structure, yet different clusters

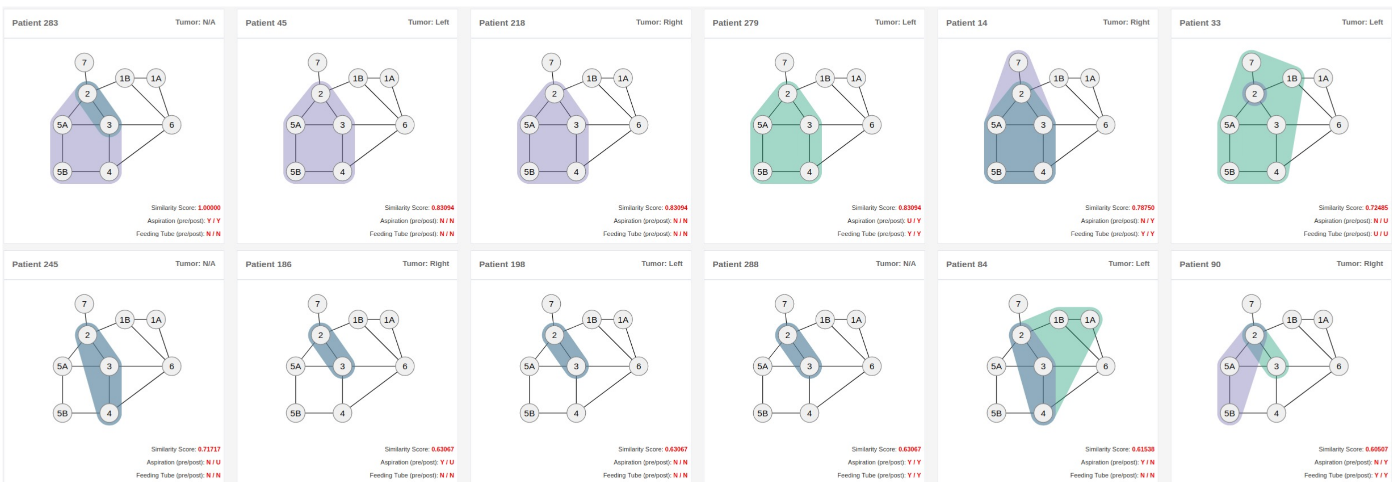
Clustering Model: Graph-based



Multiple design iterations.

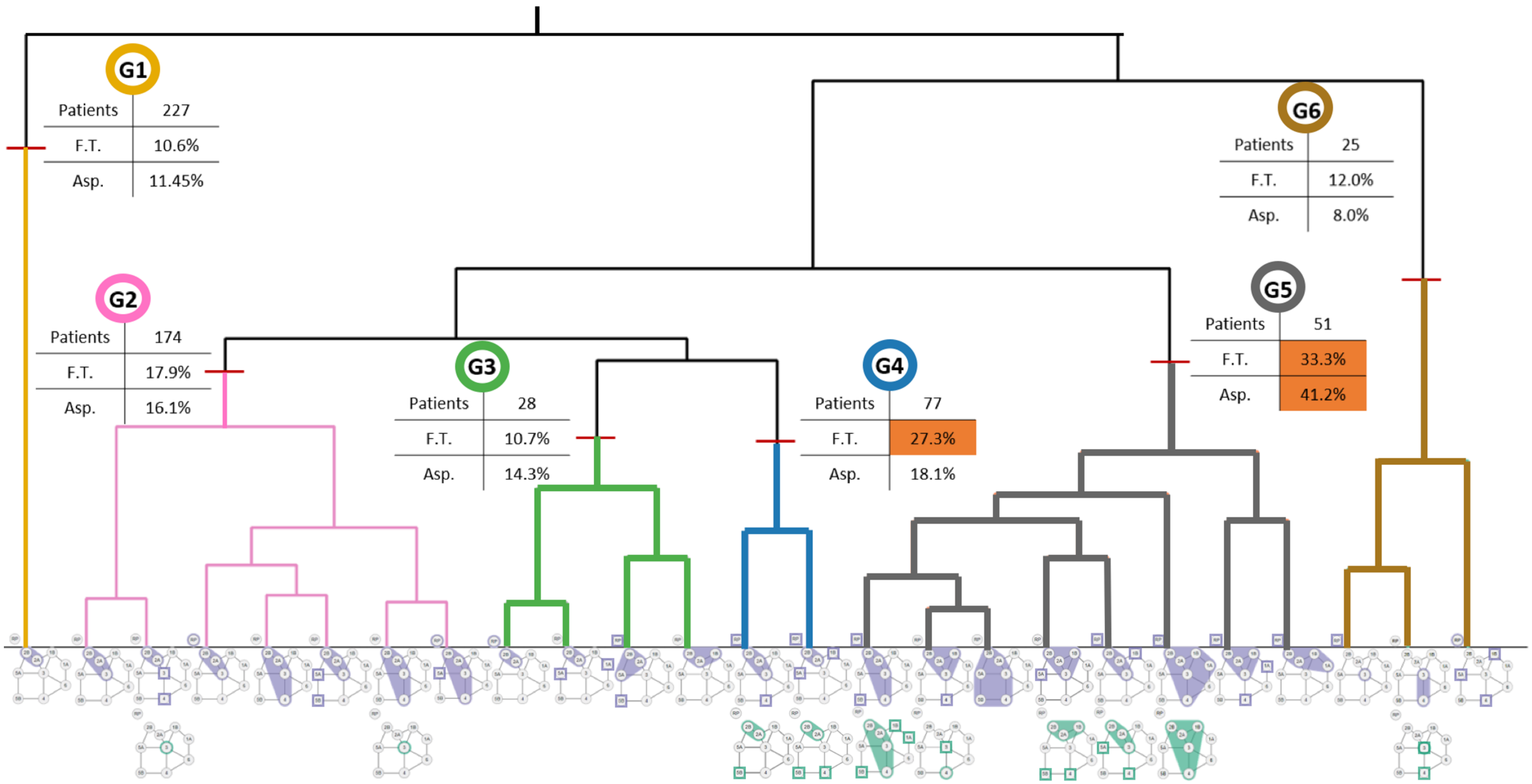
Model Development

- Note: some patients have no exact match in the repository



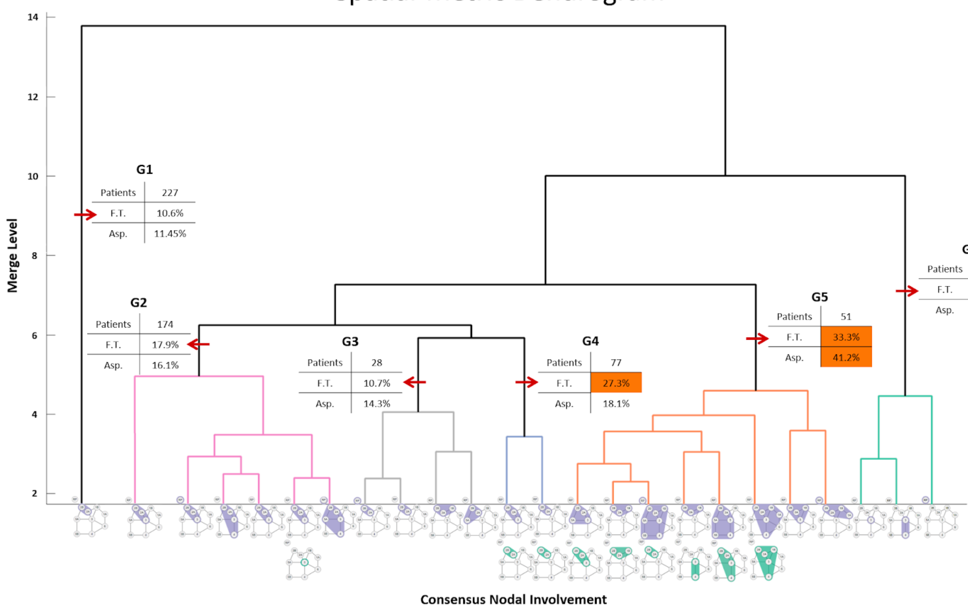
Luciani, T., Wentzel, A., Elgohari, B., Elhalawani, H., Mohamed, A., Canahuate, G., ... & Marai, G. E. (2020). A spatial neighborhood methodology for computing and analyzing lymph node carcinoma similarity in precision medicine. *Journal of Biomedical Informatics*, *5*, 100067.

Clustering Model: Cluster Membership

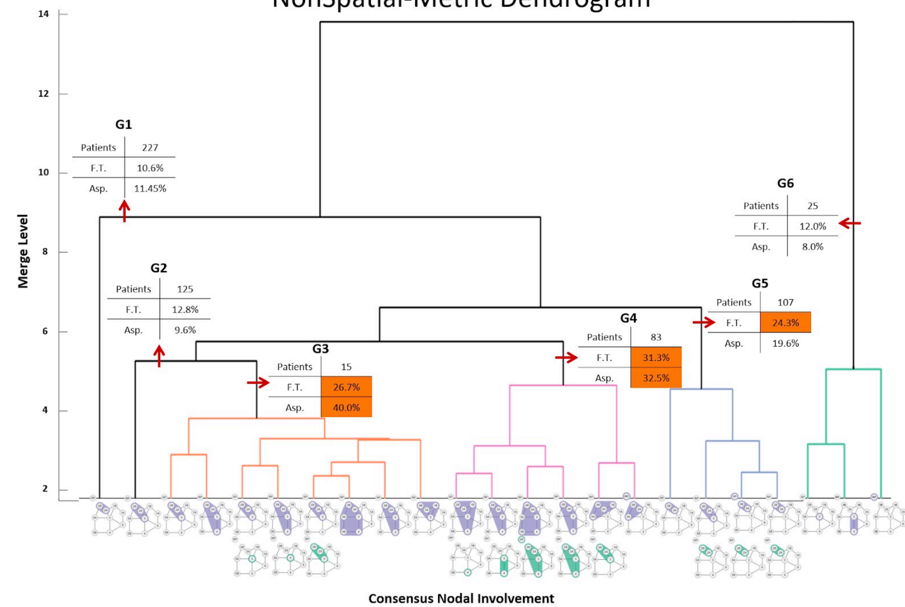


Spatial vs. Categorical Metric

Spatial-Metric Dendrogram

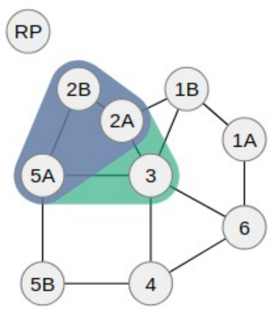


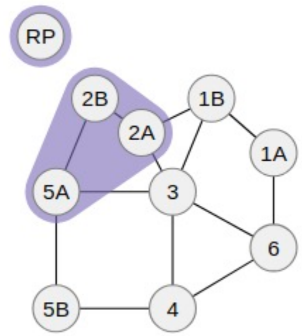
NonSpatial-Metric Dendrogram

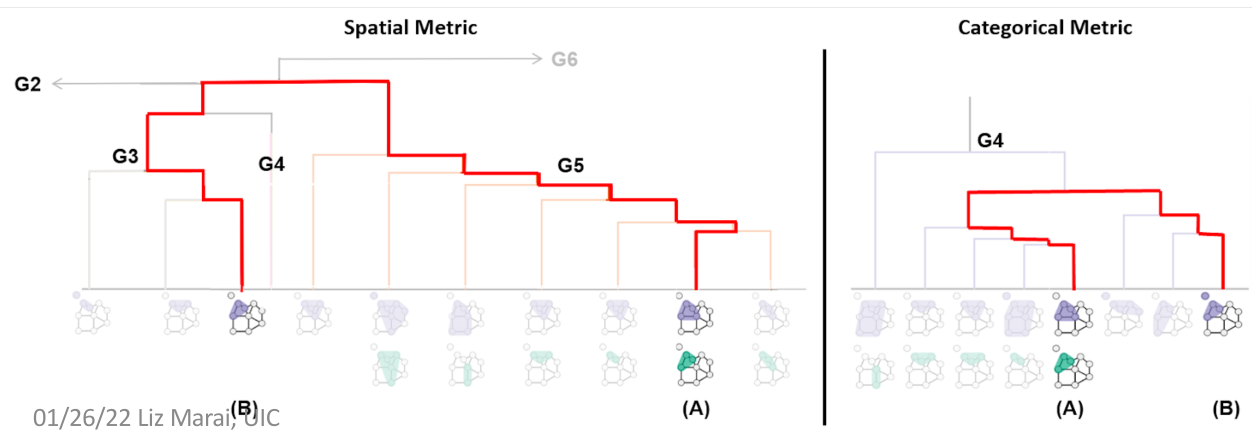


227 patients (out of ~600) have simple spread, to one node only. Both metrics separate those. For the rest, they disagree about 50%.

Spatial vs Categorical Metric







Clustering Model: Pros and Cons

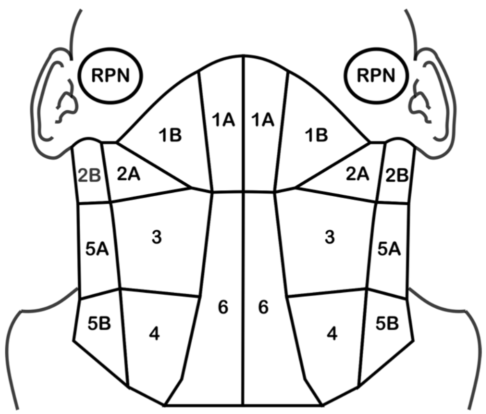
Pros:

- compact visual representation
- accounts for symmetry
- graph-theory aligned
- supports comparison and clustering

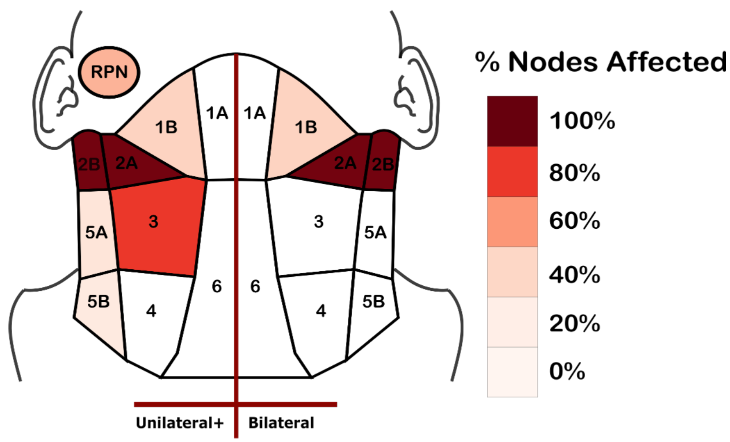
Cons:

- abstract & complex, difficult to interpret by others
- less emphasis on toxicity outcome correlates

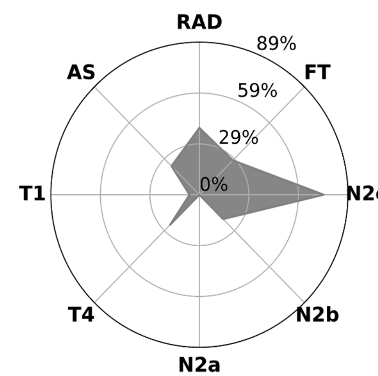
Dissemination Model: Anatomy-based



Anatomical Map



Heatmap Cluster Representation

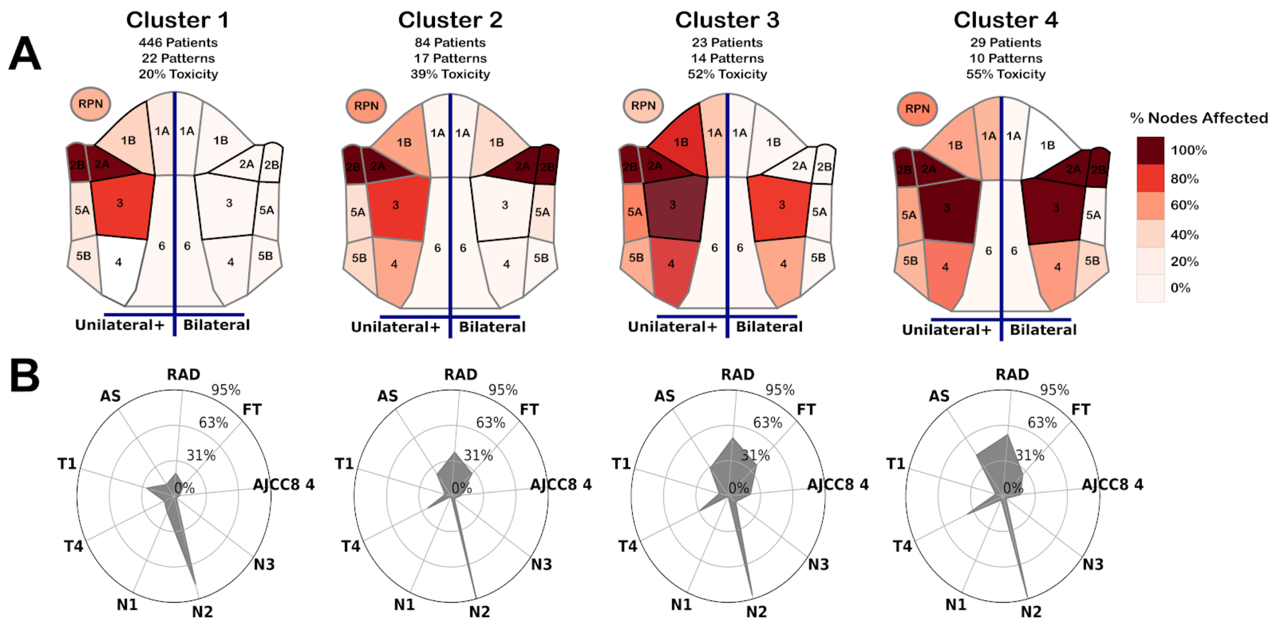


Clinical Correlates

Multiple design iterations.

Wentzel, A., Luciani, T., van Dijk, L. V., Taku, N., Elgohari, B., Mohamed, A. S., ... & Marai, G. E. (2021). Precision association of lymphatic disease spread with radiation-associated toxicity in oropharyngeal squamous carcinomas. *Radiotherapy and Oncology*.

Final Model for Dissemination



(A) Heat map of nodal involvements within each cluster. Left side indicates % of patients in the cluster with at least one involved node on each level, while the right side encodes the percentage of patients with bilateral involvement within each node. Regions outlined in black denote regions that are most discriminative cluster membership and can be used to determine if 99% of patients are within a given cluster. (B) Radar chart showing the % of patients in each cluster with a given toxicity or inclusion in a high-risk clinical staging category. FT: Feeding Tube, AS: Aspiration, RAD: Radiation-induced dysphagia, T1/T4: T-category 1/4, AJCC 4: AJCC clinical (8th edition) stage IV, N1/N2/N3: N-category 1/2/3.

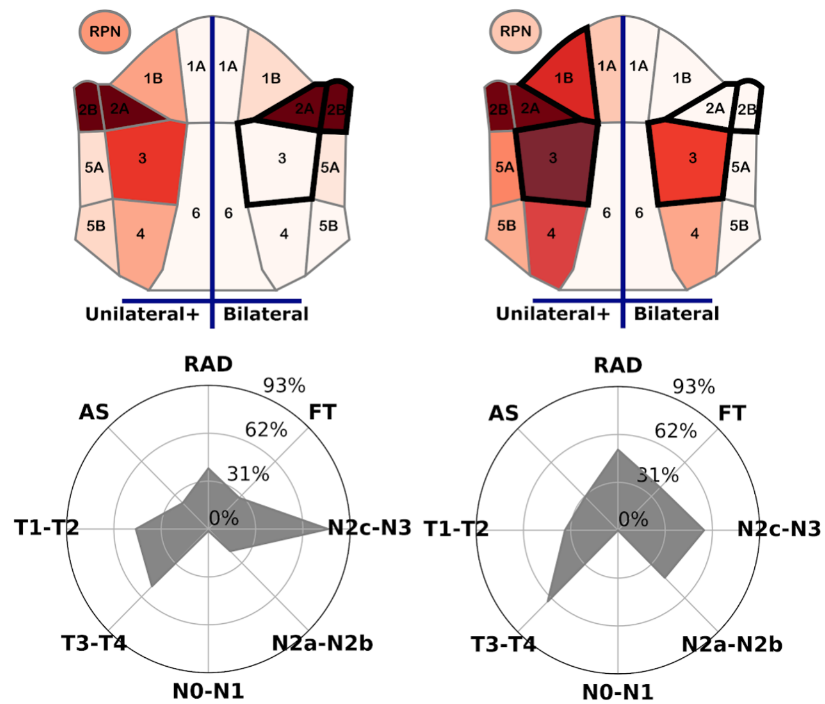
Dissemination Model: Cluster Comparison

Pros:

- easier to distribute in print
- anatomy-based
- easier to interpret
- shows correlates

Cons:

- too complex for model development



- EVL and Computational Oncology
- XAI Considerations
- Vis in XAI
- Clustering with Spatial Data: RT
- Clustering with Spatial Data: LN
- A “Good” Visual Explanation for AI

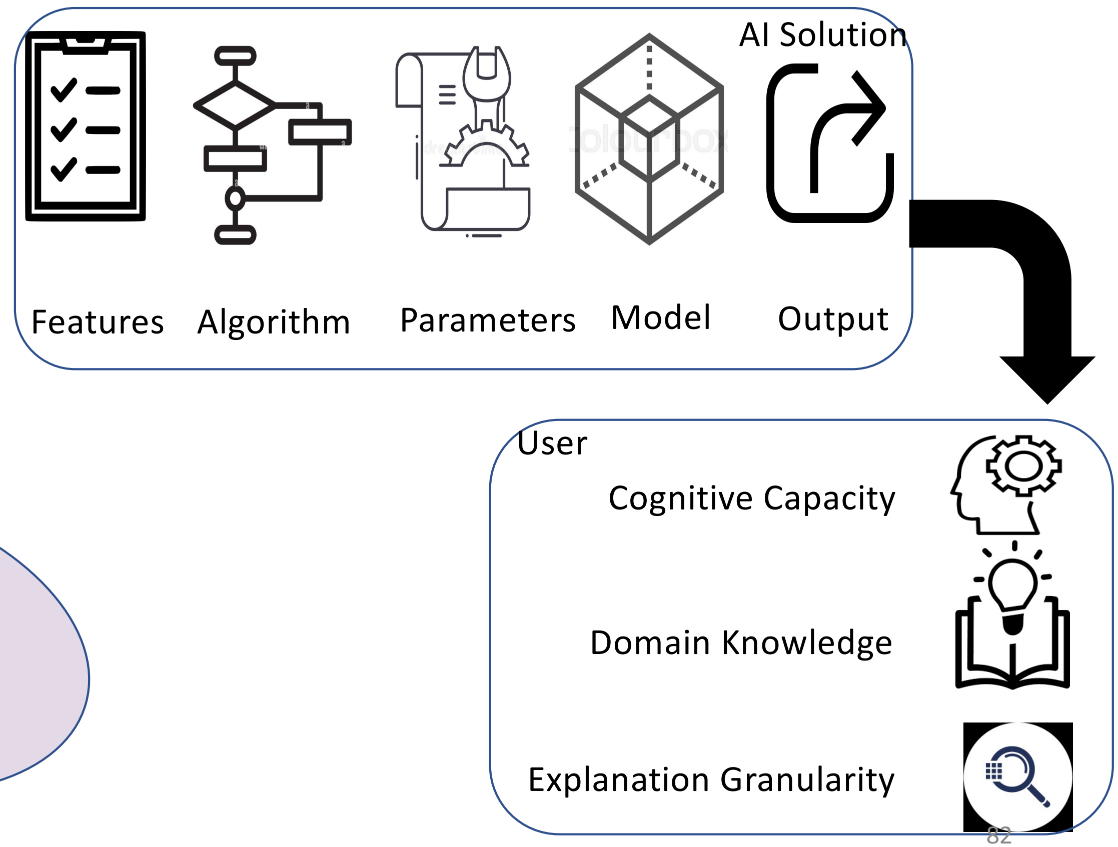
What is a “Good” Visual Explanation for AI?

“Us” and “Them”

Visual computing specialists

Data mining specialists

Model Builders

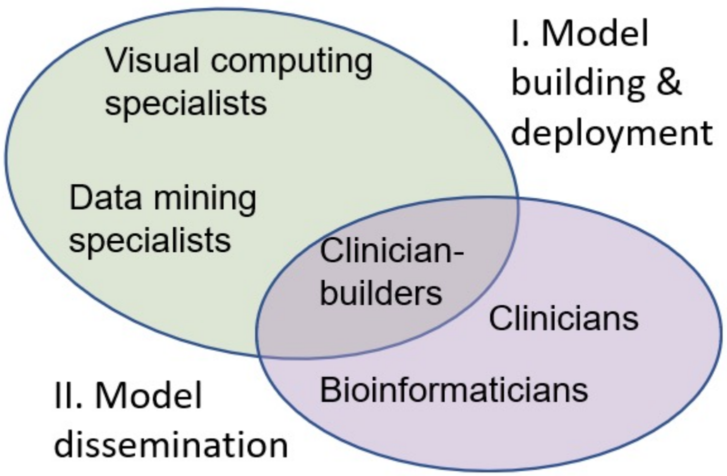


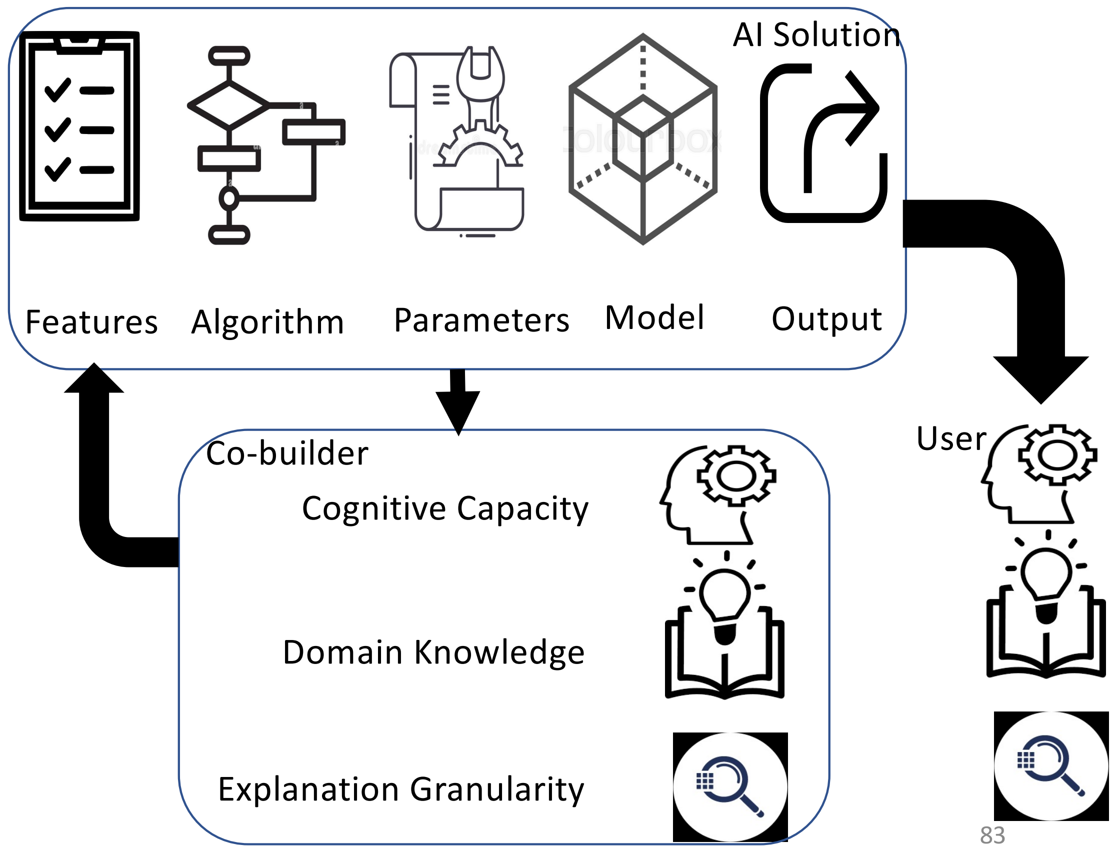
Clinicians

Bioinformaticians

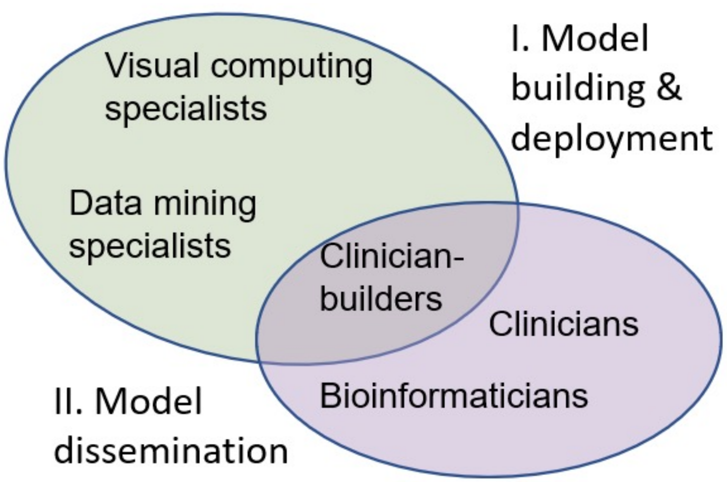
Model Users

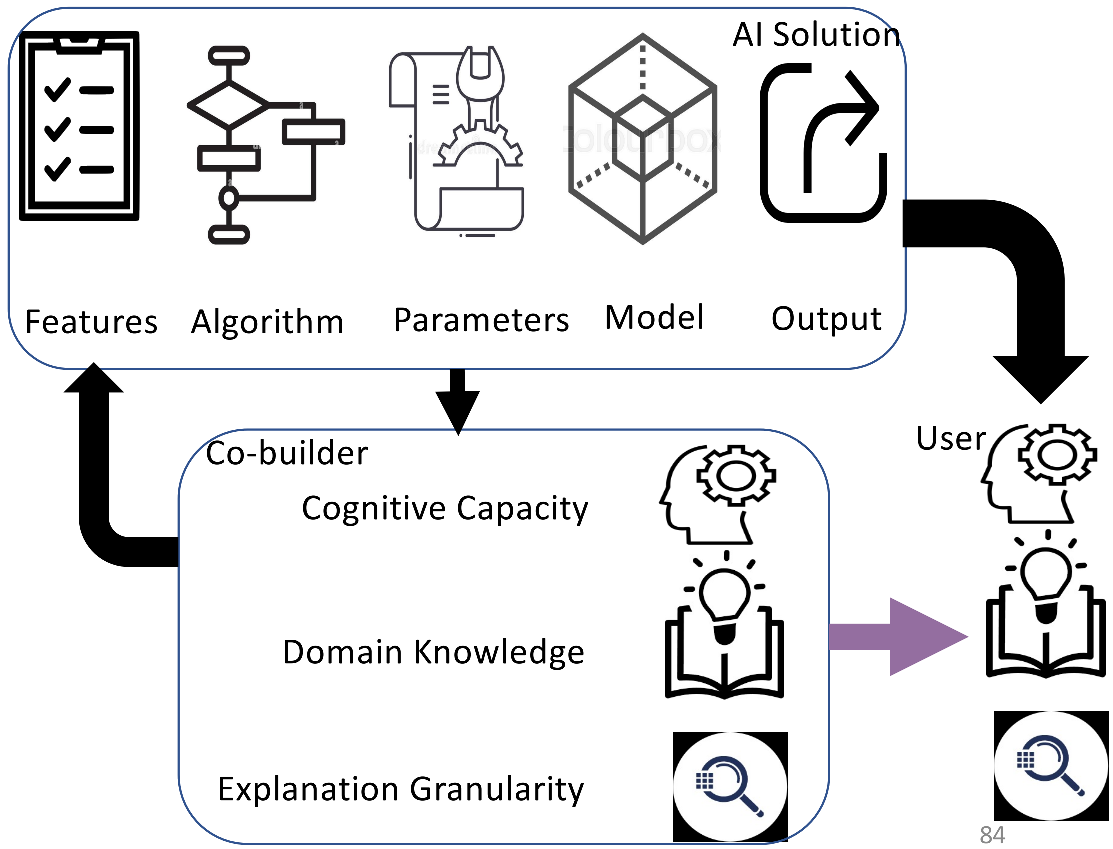
“Us” and “Them”


I. Model
building &
deployment
Visual computing
specialists
Data mining
specialists
Clinician-
builders
Clinicians
Bioinformaticians
II. Model
dissemination

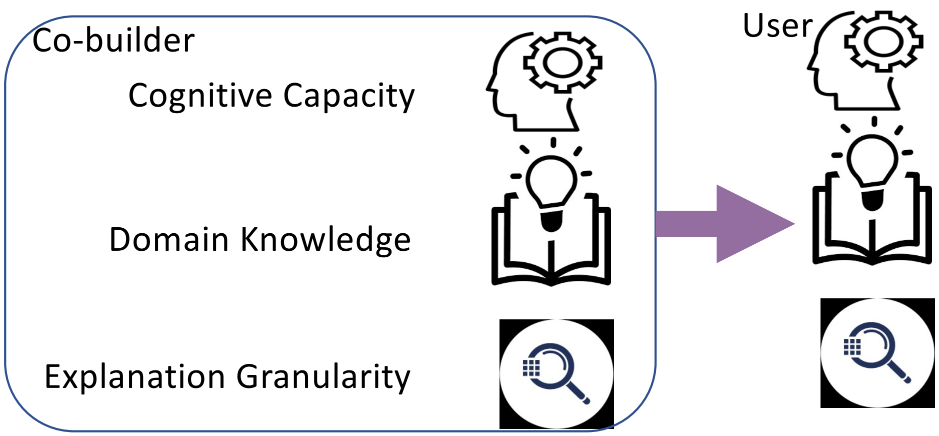

AI Solution
Features
Algorithm
Parameters
Model
Output
Co-builder
Cognitive Capacity
Domain Knowledge
Explanation Granularity
User

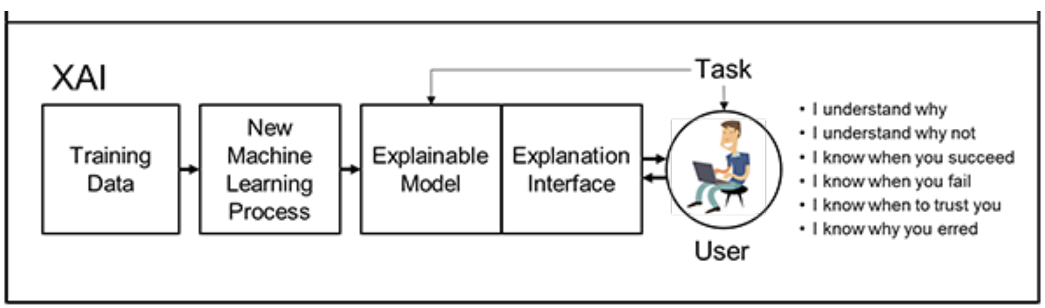
“Us” and “Them”: Evangelizers





Us and Them: Policymakers





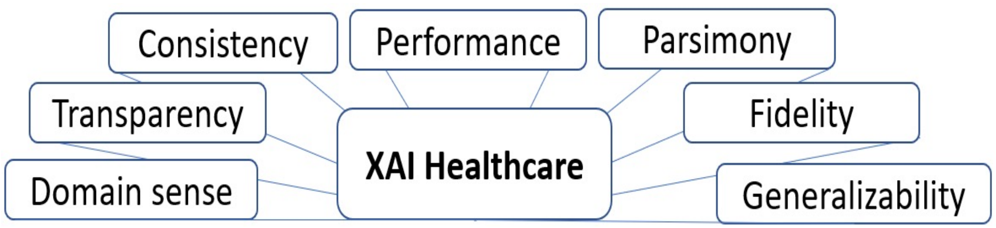
- It's possible the "users" have indeed very small XAI needs
- Yet, co-builders are "influencers" or "evangelizers" or "policymakers", and they have significant XAI needs and tremendous influence on the 95% "users"

https://www.darpa.mil/program/explainable-artificial-intelligence

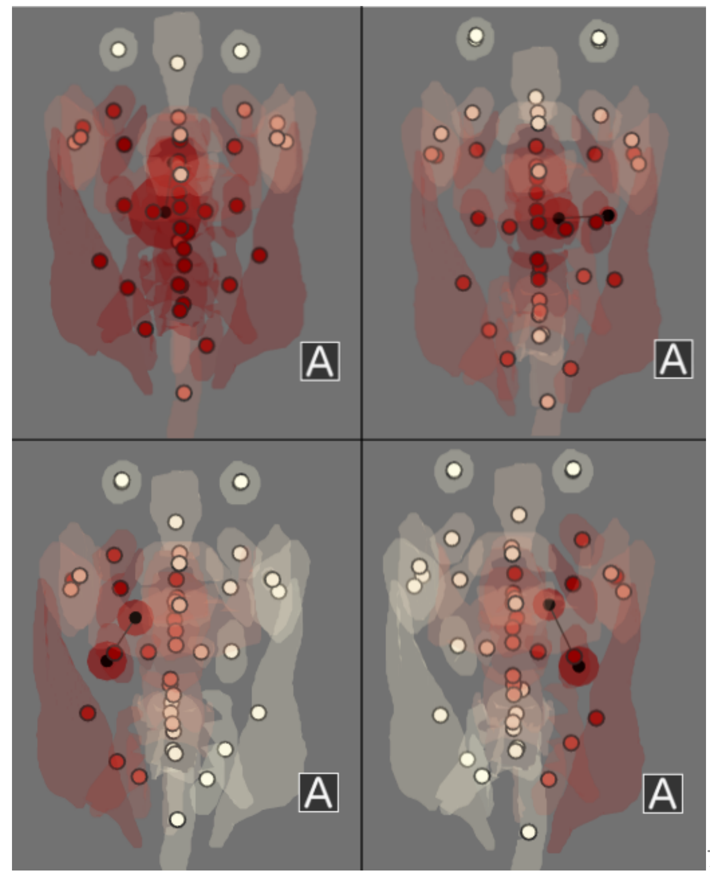
Goodness:
Domain Sense, Actionability, Transparency

Domain Sense

The visual explanations need to make sense in the application domain

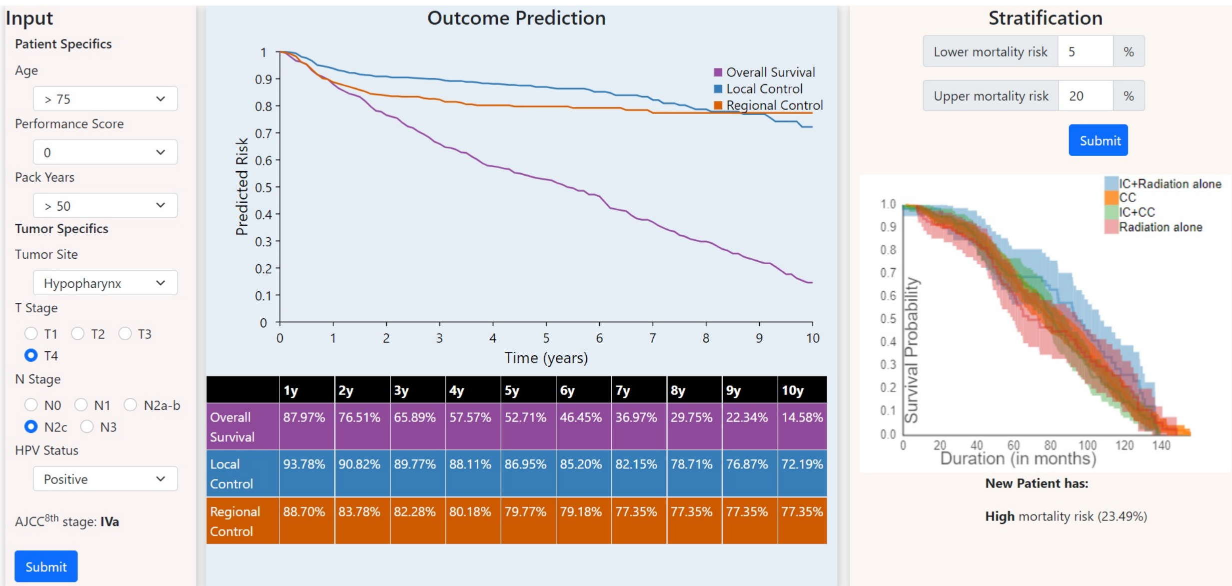


Interpretable machine learning in healthcare
Ahmed at al., 2018 BCBHI



Wentzel, Andrew, et al. "Explainable Spatial Clustering: Leveraging Spatial Data in Radiation Oncology." *2020 IEEE Visualization Conference (VIS)*. IEEE, 2020.

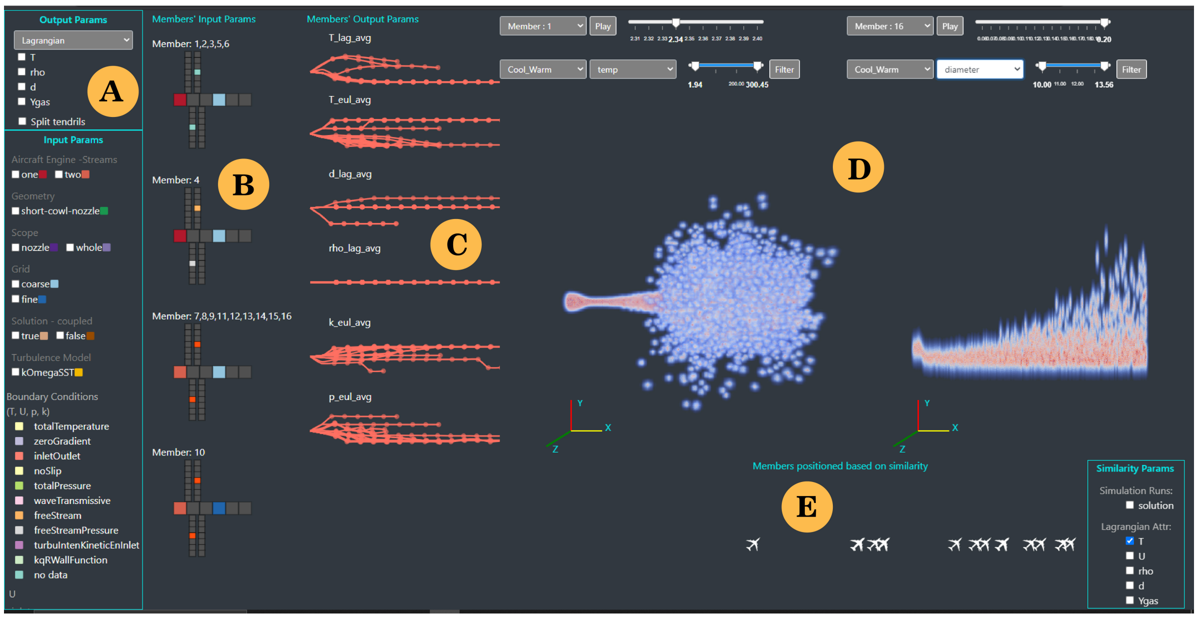
Actionability



Floricel, Carla, et al. "Thalis: Human-machine analysis of longitudinal symptoms in cancer therapy." *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021): 151-161.

evl electronic visualization laboratory

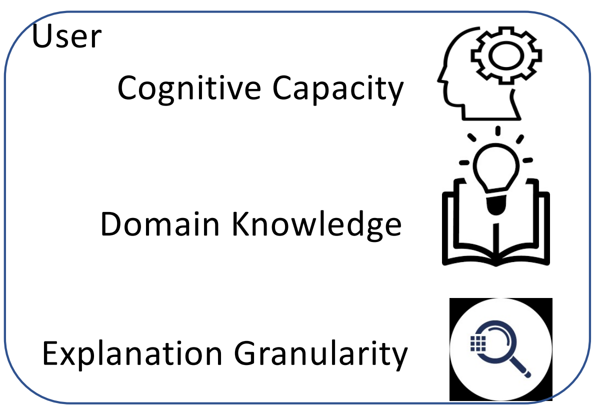
Actionability

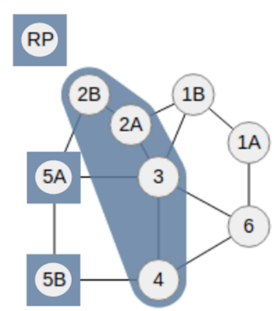


Floricel, Carla, et al. "Thalis: Human-machine analysis of longitudinal symptoms in cancer therapy." *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021): 151-161.

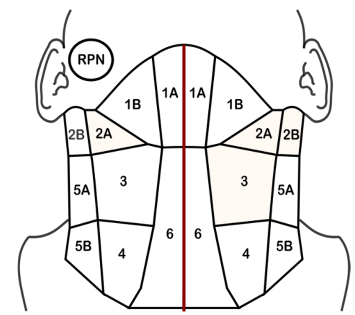
Transparency

- Who needs an explanation?
- How much AI/ML knowledge do they have? How much domain knowledge do they have?
- Do they care about the model? Do they only care about features or output?





Background = Graph Theory



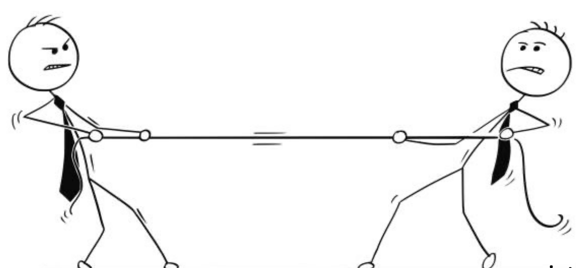
Background = Anatomy

Goodness:
Parsimony, Fidelity, Consistency, Performance

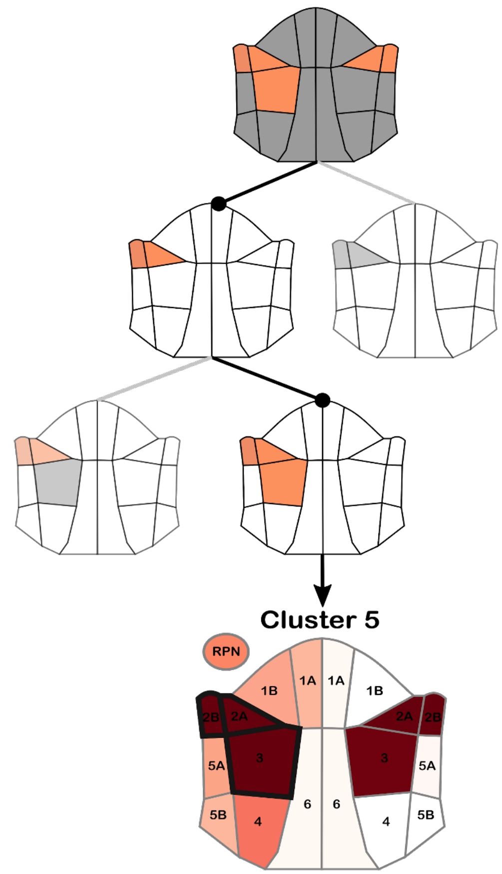
Parsimony and Fidelity

- Parsimony: the simplest explanation
- Fidelity: the most faithful explanation

- Again, context and purpose matter
- Parsimony vs Fidelity tension

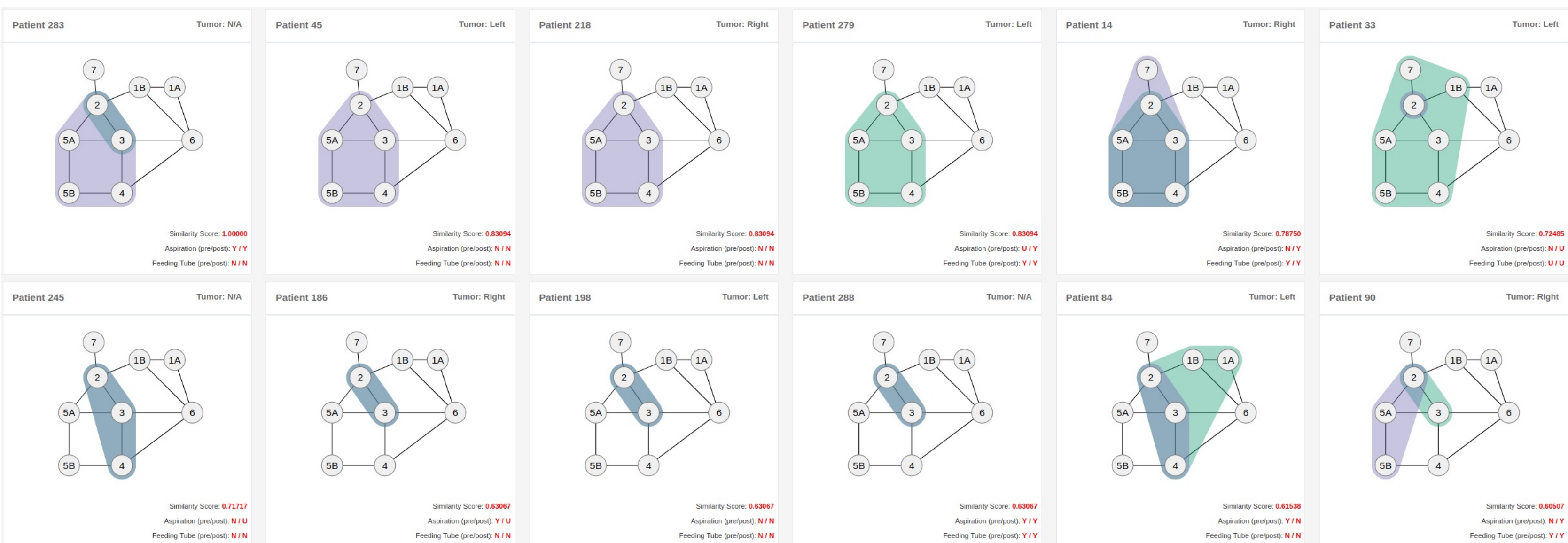


istock.com



Generalizability

- Breast cancer researchers saw immediately a connection based on the graph representation, but not when looking at the anatomical one



Consistency

- One of the more confusing criteria:
  - Consistency means the methods for producing explanations should be consistent across different models and across different runs of the model

Interpretable machine learning in healthcare
Ahmed at al., 2018 BCBHI

  - Consistency means the explanations are stable across models and runs?
    - Local Linear Explanation methods (e.g., LIME and SHAP) are plagued by many defects including unstable explanations, divergence of actual implementations from the promised theoretical properties, and explanations for the wrong label.

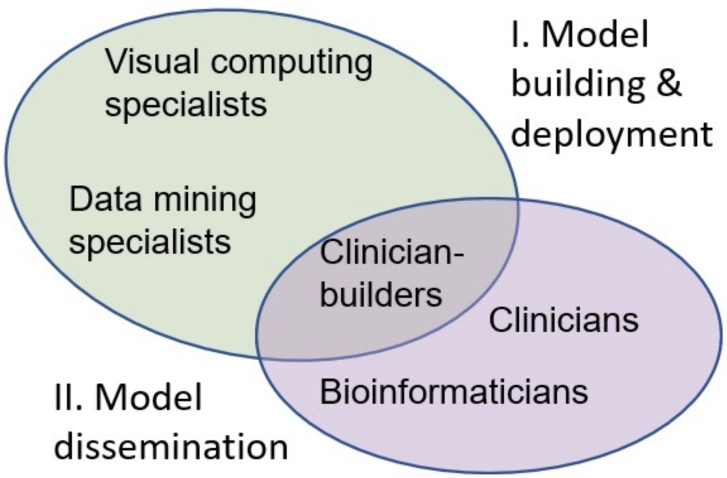
To trust or not to trust an explanation: using LEAF to evaluate local linear XAI methods
Amparore at al., 2021 PeerJ

Performance

What are the time constraints?

- need to act quickly (pathology, computational steering)
- need to act deliberately and fairly (tumor board, policy making)

A Multi-Dimensional, Multi-Phase Model



transparency	
fidelity	
generalizability	…
actionability	actionability
domain sense	domain sense
…	parsimony
consistency	consistency
	performance
model building	model dissemination

Recap and Conclusion

- XAI is beneficial, esp. to policy makers and co-builders
- XAI goes way beyond black box interpretability
- Vis can enable XAI, in particular when spatial data is involved
- Goodness of visual XAI is multidimensional and multi-phase
- Goodness criteria shift between development and dissemination
- Goodness criteria go beyond transparency
  - At least in healthcare, *domain sense, actionability, and parsimony* matter

evl electronic visualization laboratory

Thanks

- www.evl.uic.edu
- EVL members & collaborators

- NSF CDS&E-1854815, CNS-1828265, CNS-1625941, DMS-1557559, IIS-2031095
- NIH NCI-R01-CA214825, NCI-R01-CA225190, NLM-R01-LM012527, NCI-R01-CA258827

- We’re hiring! Multiple faculty positions including CG
- Fastest growing CS department in the USA
- Downtown Chicago