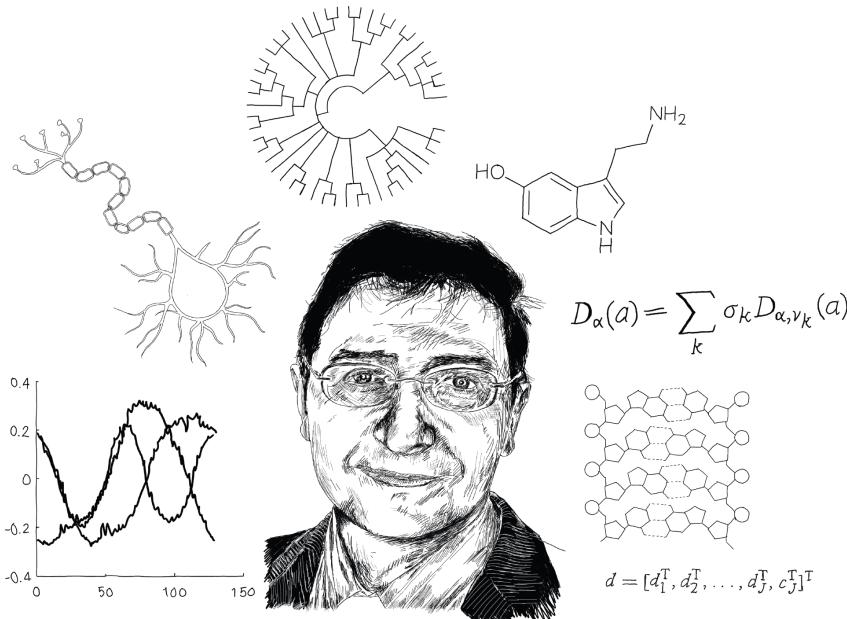


AI AND MEDICINE: GRAPH AND HYPERGRAPH REPRESENTATION LEARNING



Prof. Pietro Lio'

PhD Complex Systems and Non Linear dynamics (Engineering, Firenze)
PhD Genetics (Pavia)

*Dept of Computer Science and Technology
Cambridge Center for AI in Medicine
University of Cambridge
pl219@cam.ac.uk*



UNIVERSITY OF
CAMBRIDGE



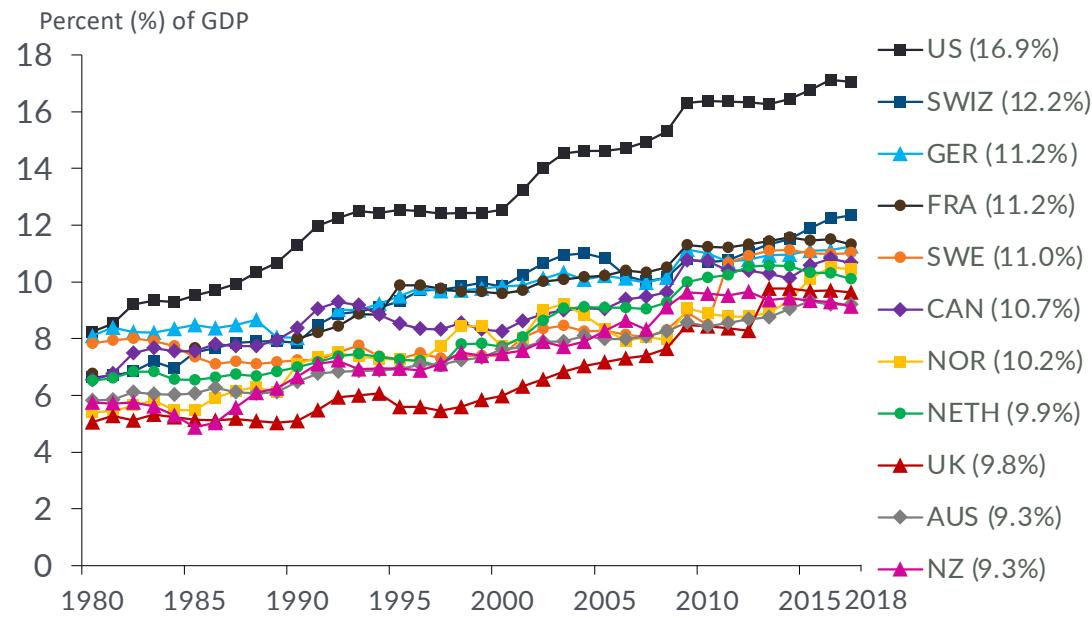
Asia-Pacific Artificial
Intelligence Association



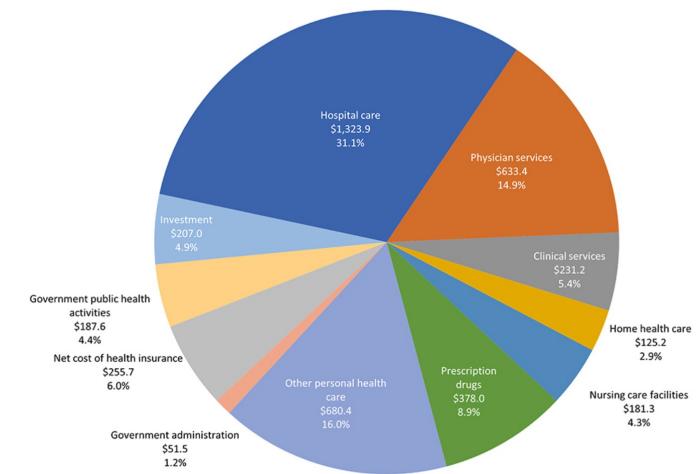
Plan

- AI Challenges in Medicine
- Predicting patient outcome
- The patient digital twin
- The cell and tissues scales
- Complexing hypergraphs
- Helping with logic

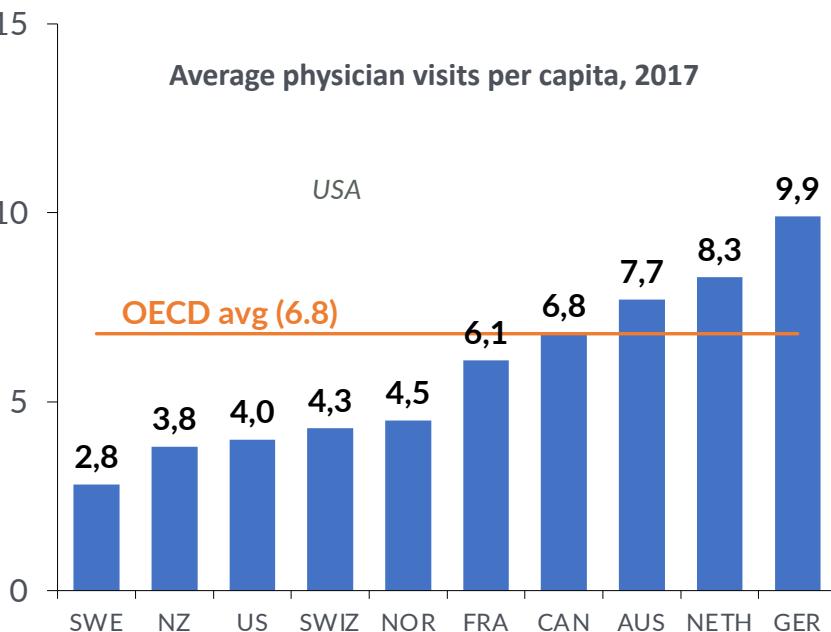
Health Care Spending as a Percent of GDP 1980–2018 *Adjusted for Differences in Cost of Living*



2018*
data:



Average physician visits per capita, 2017



Many challenges , still access to care does not equal access to quality care

Answering the patients (&Drs) key questions

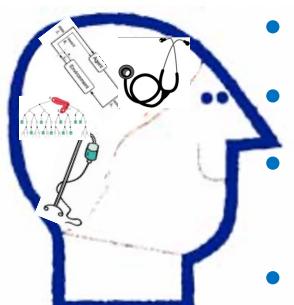


The patient

- Can my cancer be cured (am I going to die)?
- How will I be treated?
- When will I know if my treatment is working?
- Are my family also at risk?
- Will my treatment make me ill?
- Do I need more treatment?



The doctor



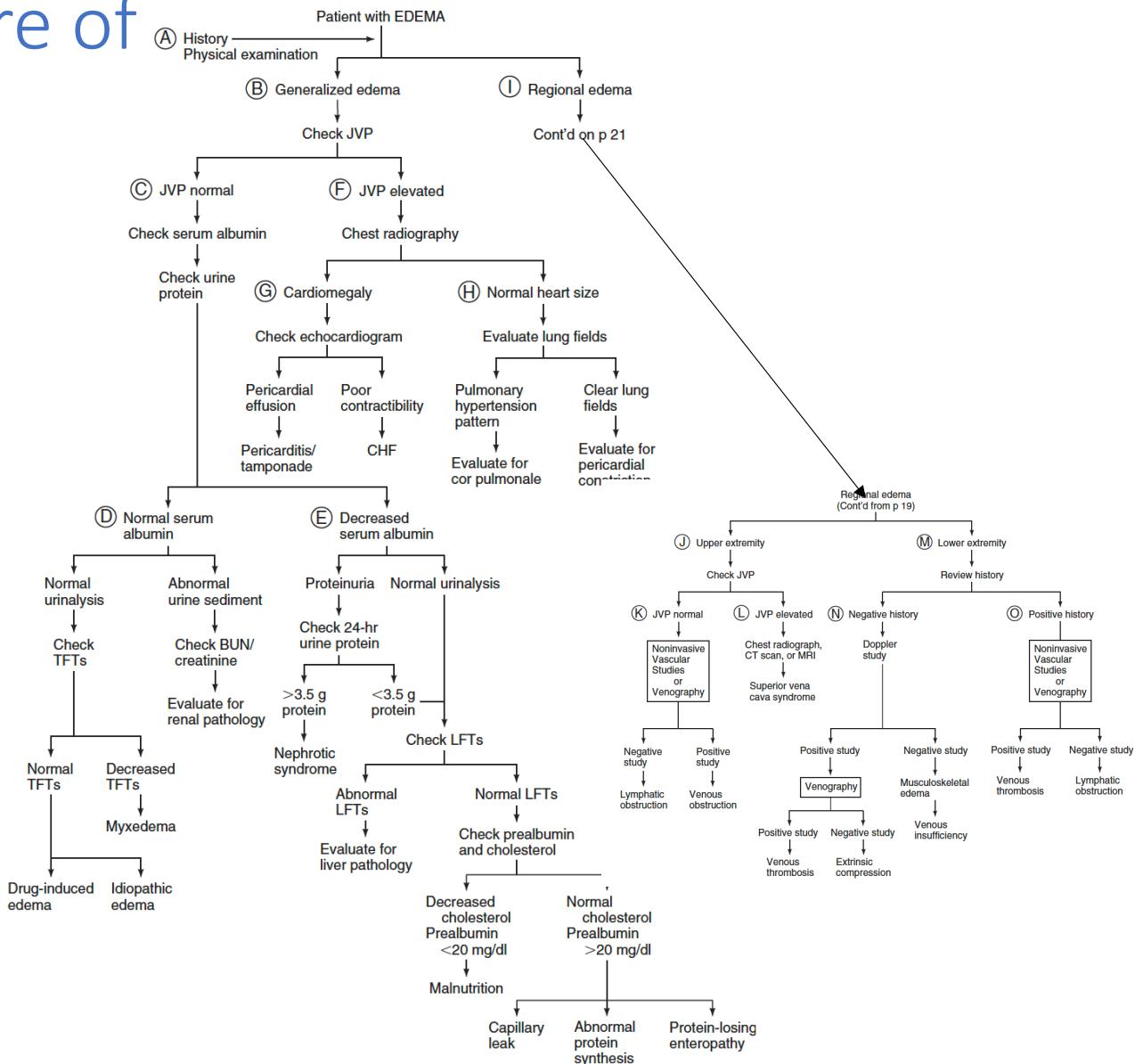
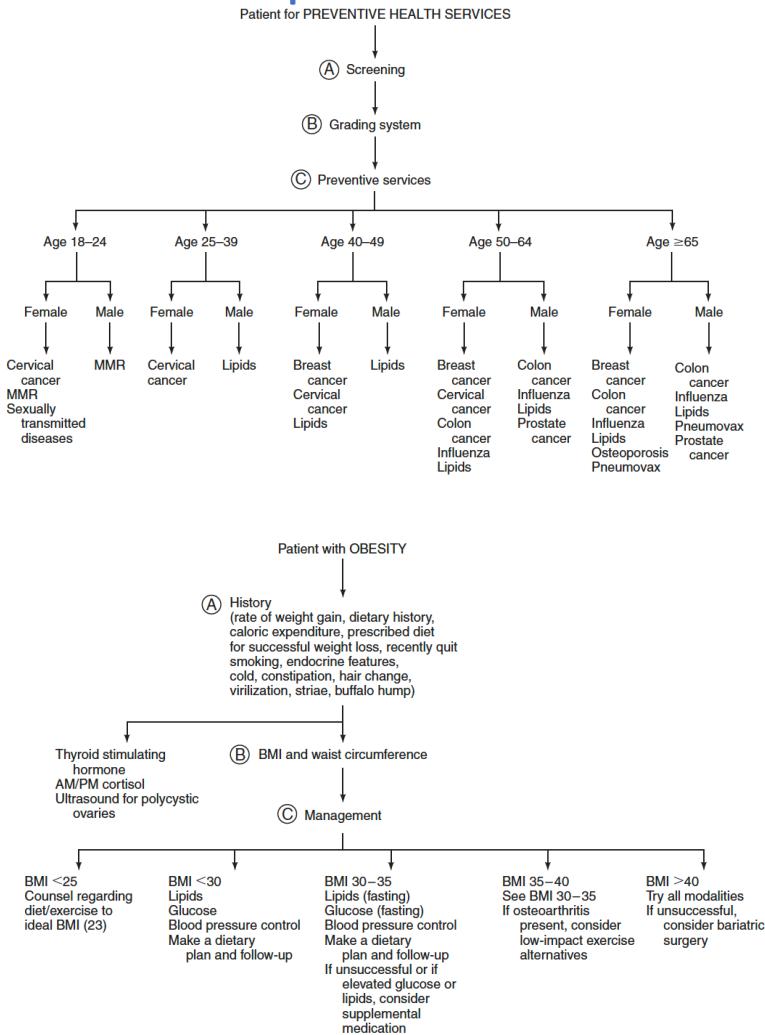
The doctor

- Every cancer responds differently to different treatments
- Treatment is still mostly one-size-fits-all
- We have the data, but we are not exploiting it properly
- Can we make personalised prediction about response to different treatments?
- We need to bring multi-modal AI to the point of care to improve health outcomes.

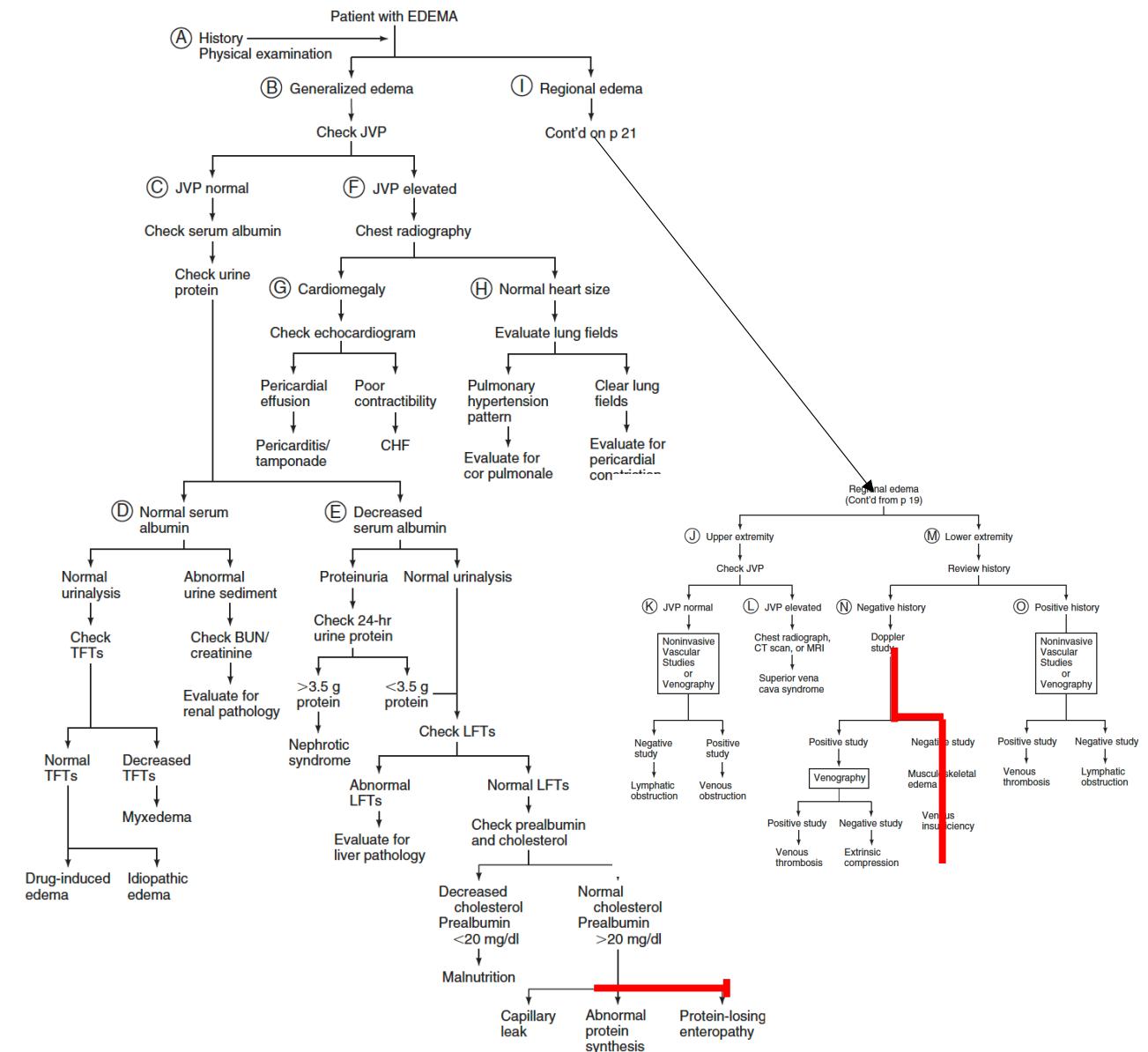
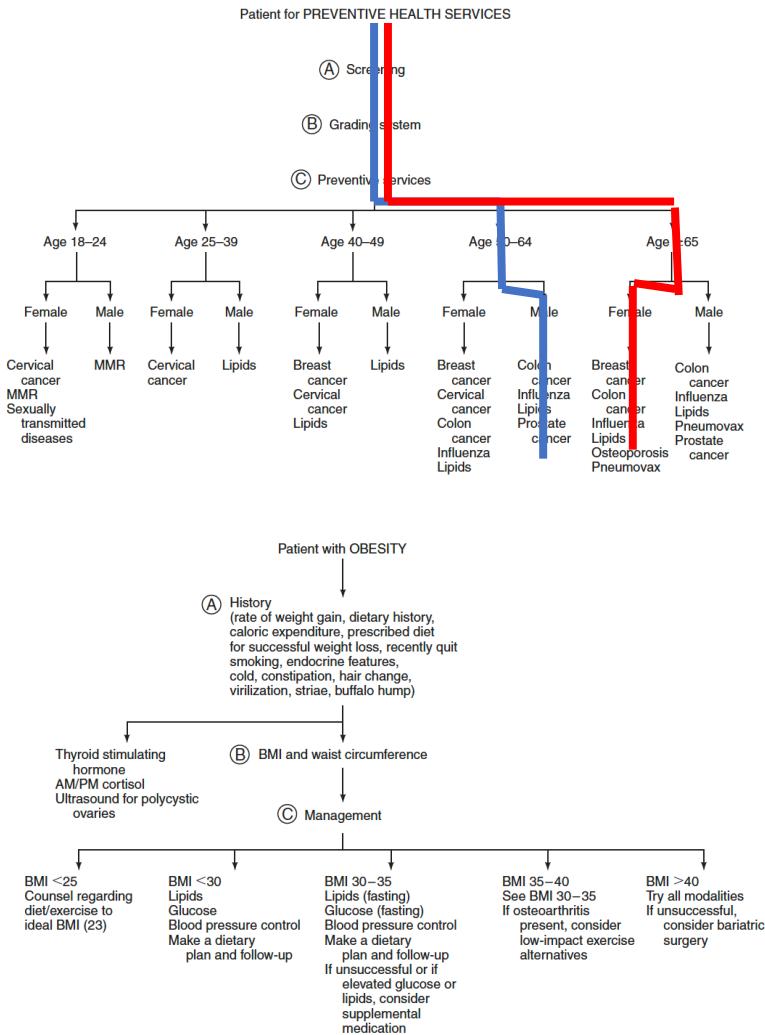


The AI scientist

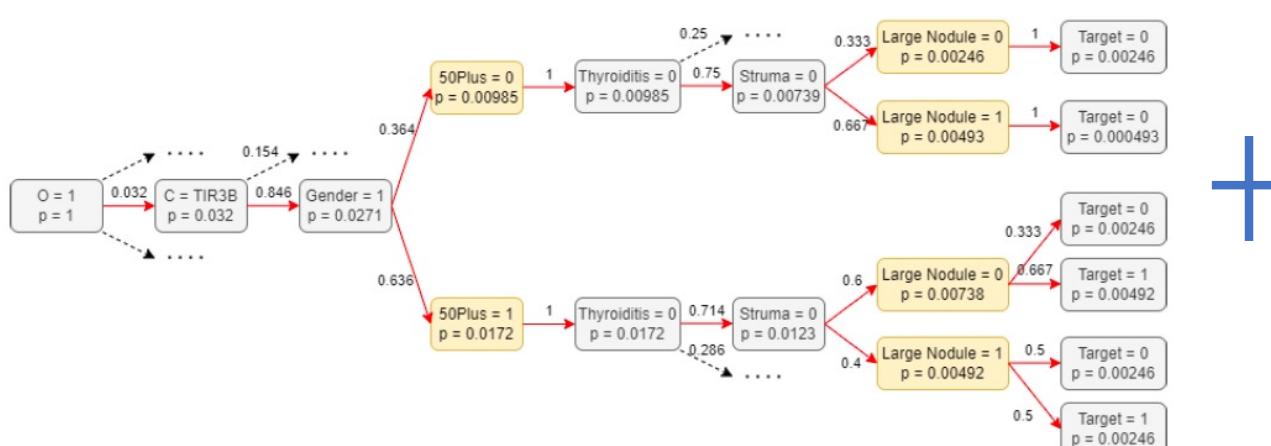
Decision trees at the core of doctor's practice



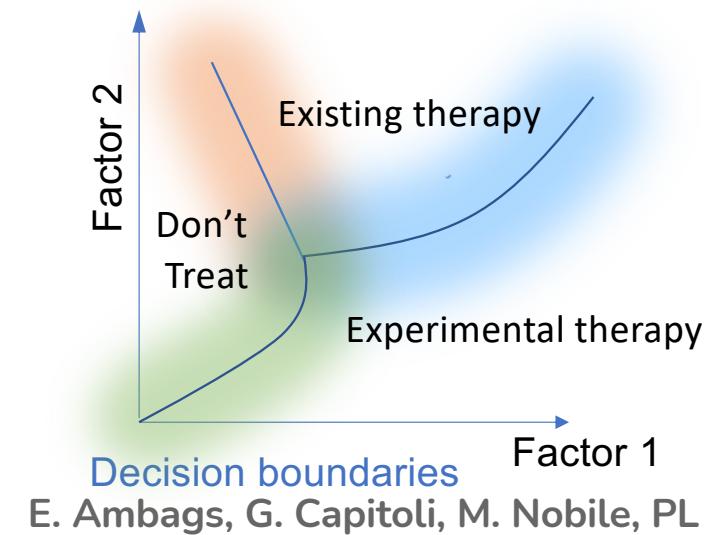
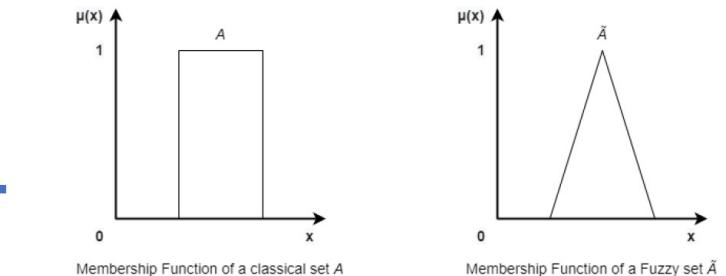
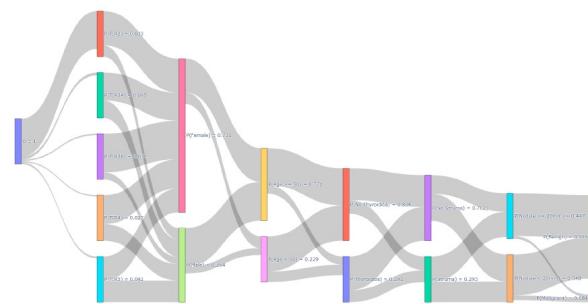
Bias in their usage



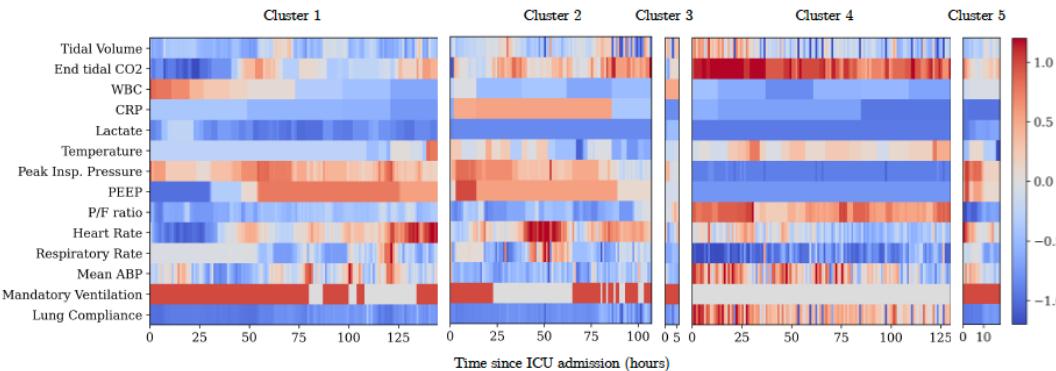
Fuzzy Sets in Probability : considering counterfactuals and bias



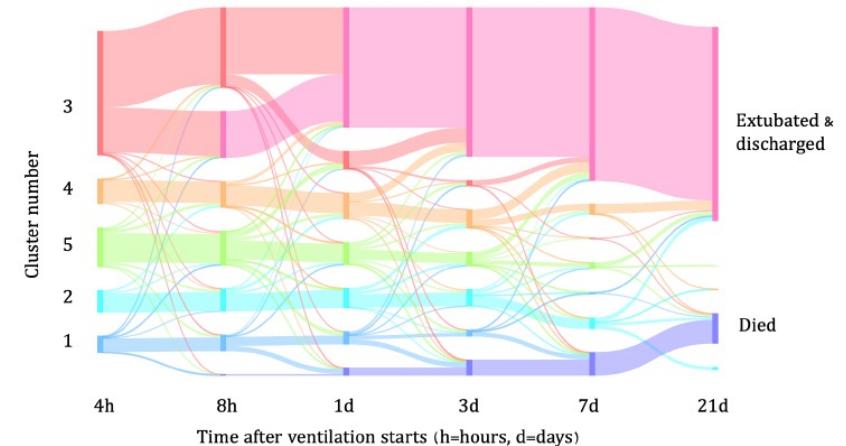
Level	Form	Typical Question
1. Association	$P(\mathcal{A} \mathcal{B})$	What is the probability of event \mathcal{A} given that event \mathcal{B} is true?
2. Intervention	$P(\mathcal{A} do(\mathcal{B}))$	What is the probability of event \mathcal{A} given that event \mathcal{B} was <i>made</i> true?
3. Counterfactuals	$P(\mathcal{A}_c \mathcal{B})$	Given that \mathcal{B} is true, what would the probability of \mathcal{A} be if \mathcal{C} were true?



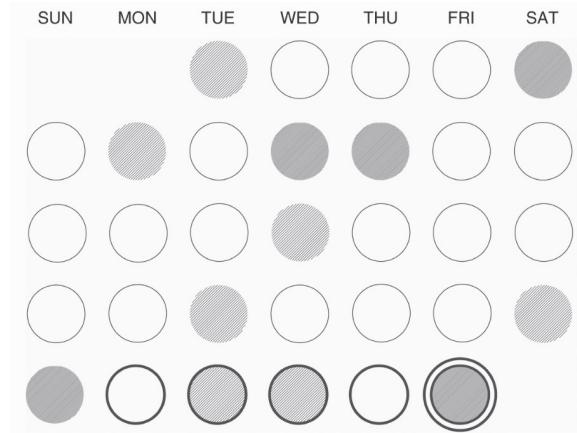
Biomedical data



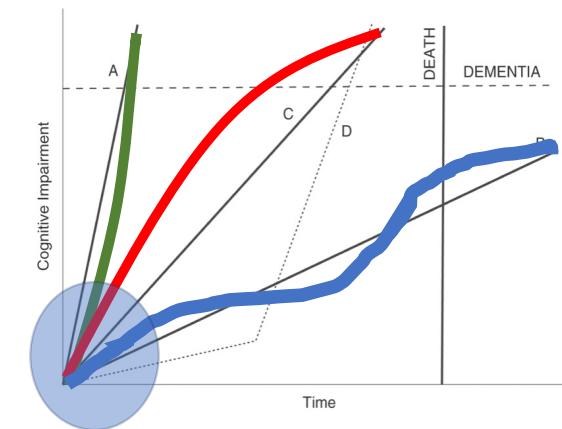
Patient stratification



Cluster trajectories

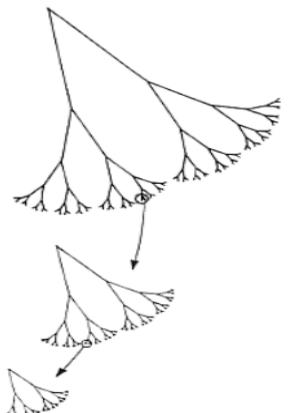


A series of “good days” and “bad days”

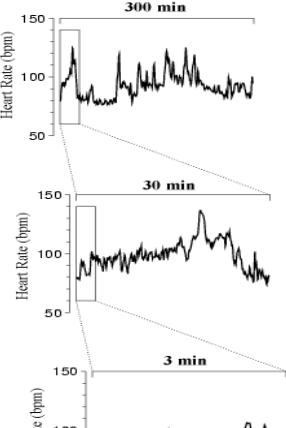


Nonlinear systems are everywhere

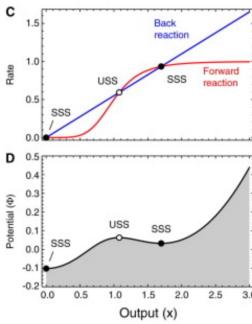
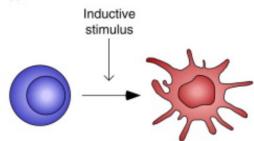
Self-Similar Structure



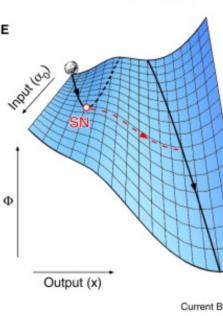
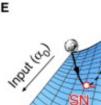
Self-Similar Dynamics



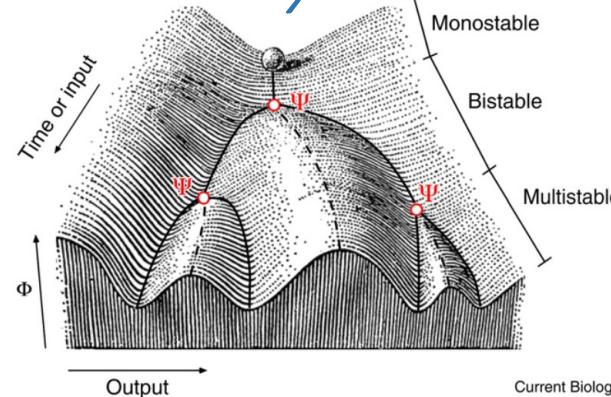
A



C

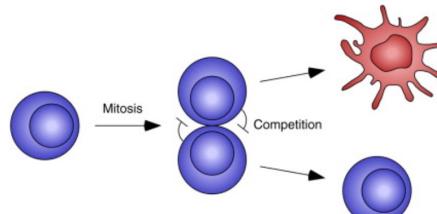


Self-Similar Dynamics

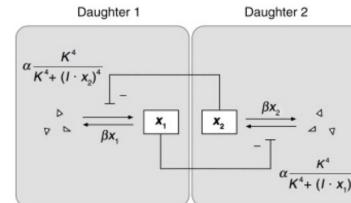


Current Biology

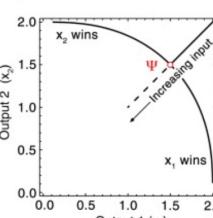
A



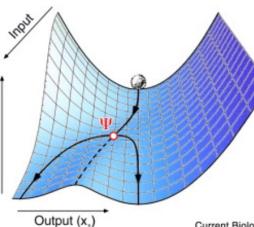
B



C



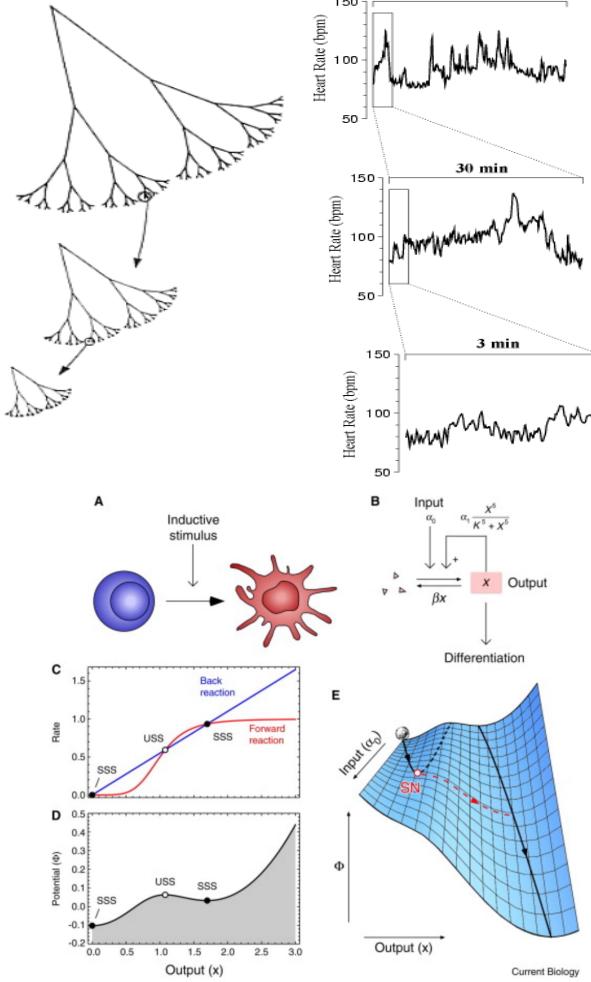
D



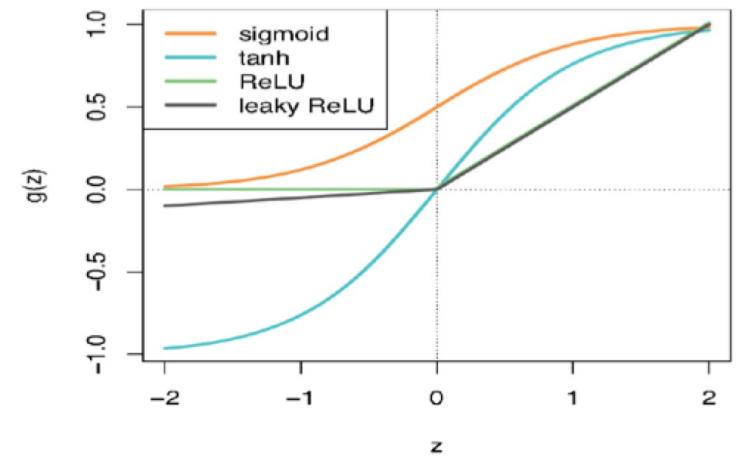
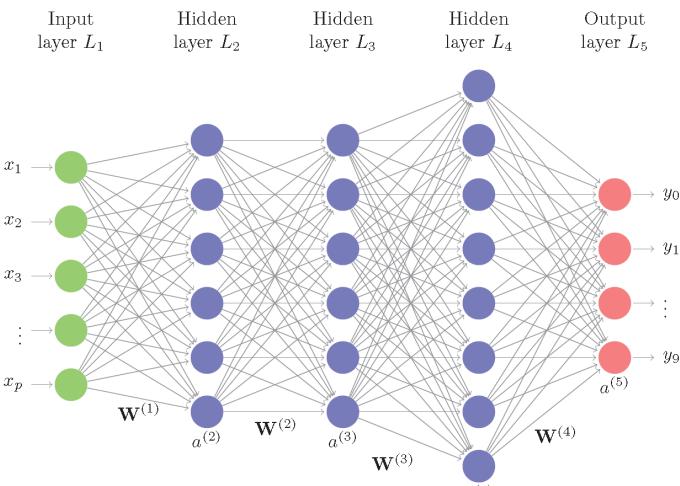
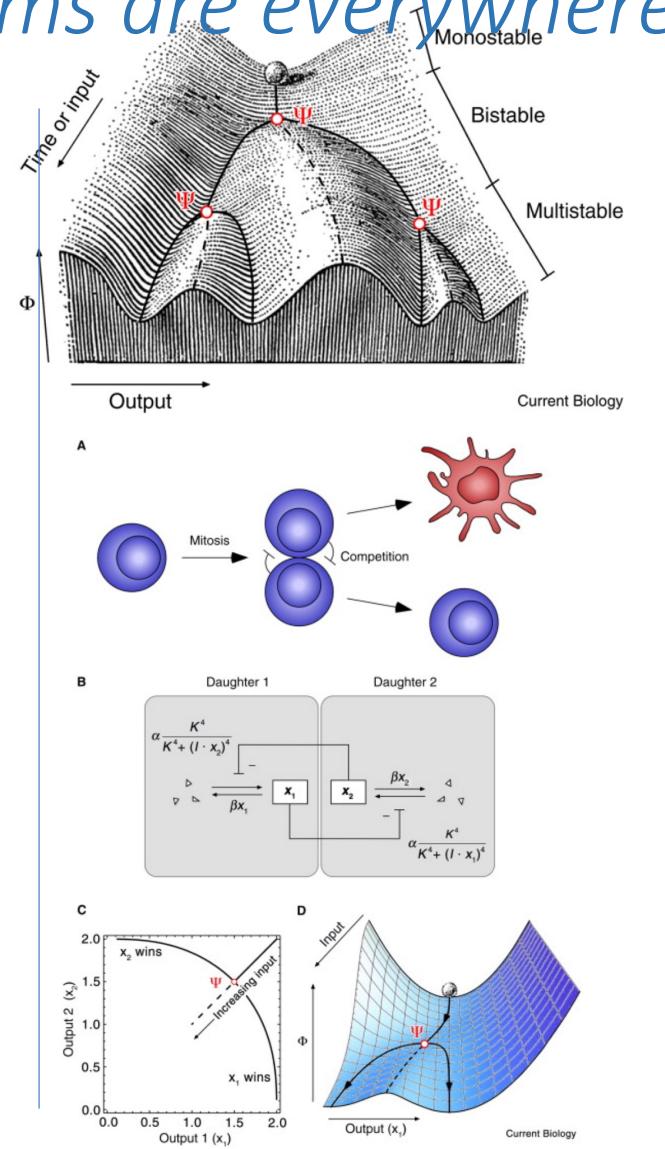
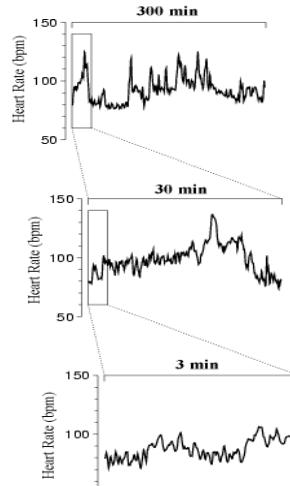
Current Biology

Nonlinear systems are everywhere

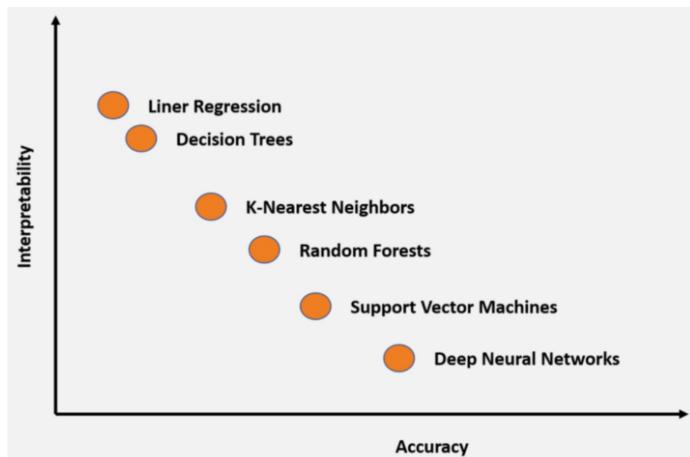
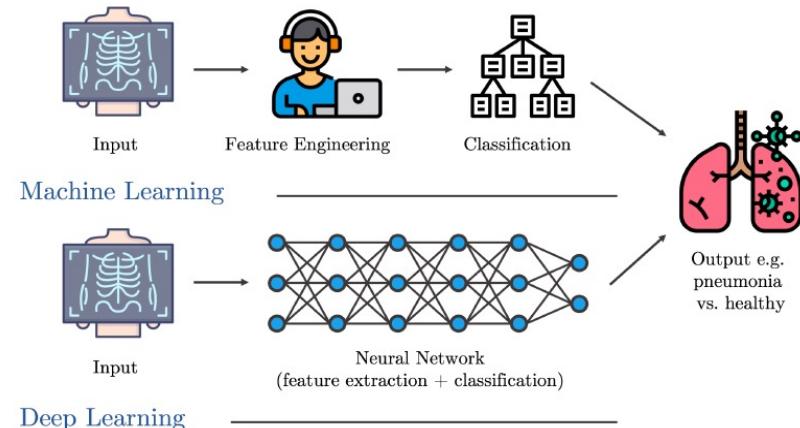
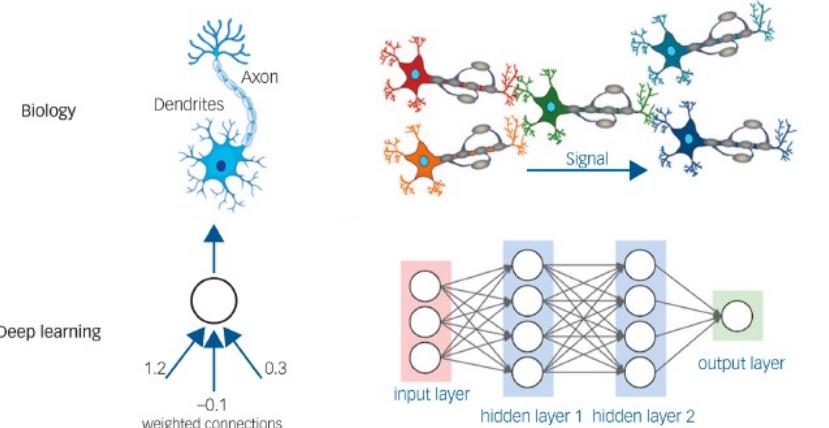
Self-Similar Structure



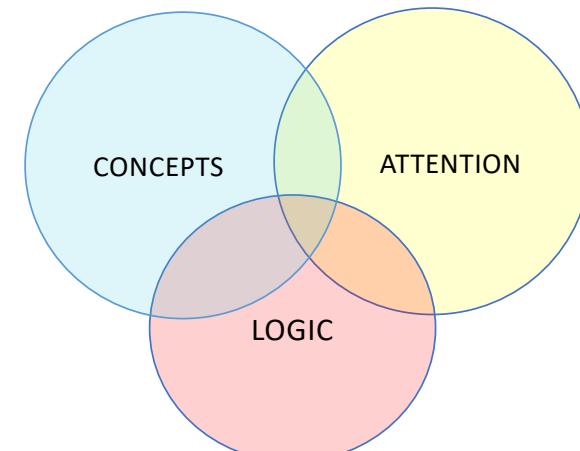
Self-Similar Dynamics



A traditional machine learning/bioinformatics vs deep learning

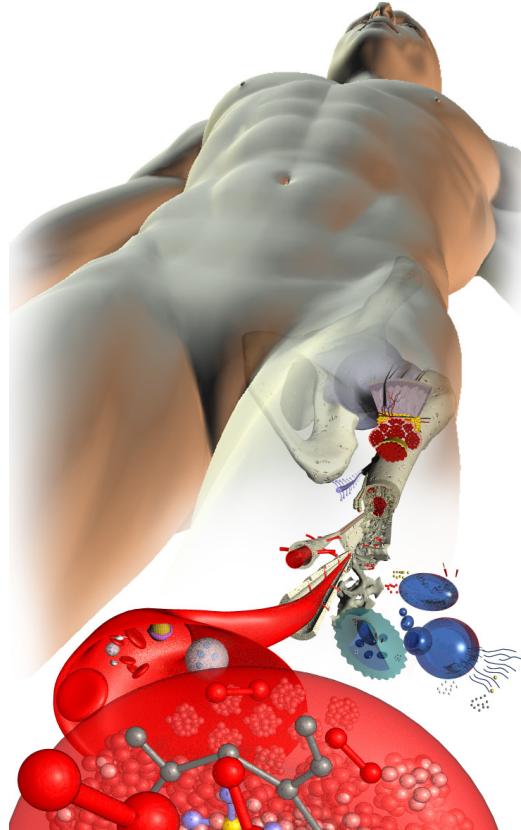


One model is more interpretable than another if it is easier for a human to understand how it makes predictions than the other model.

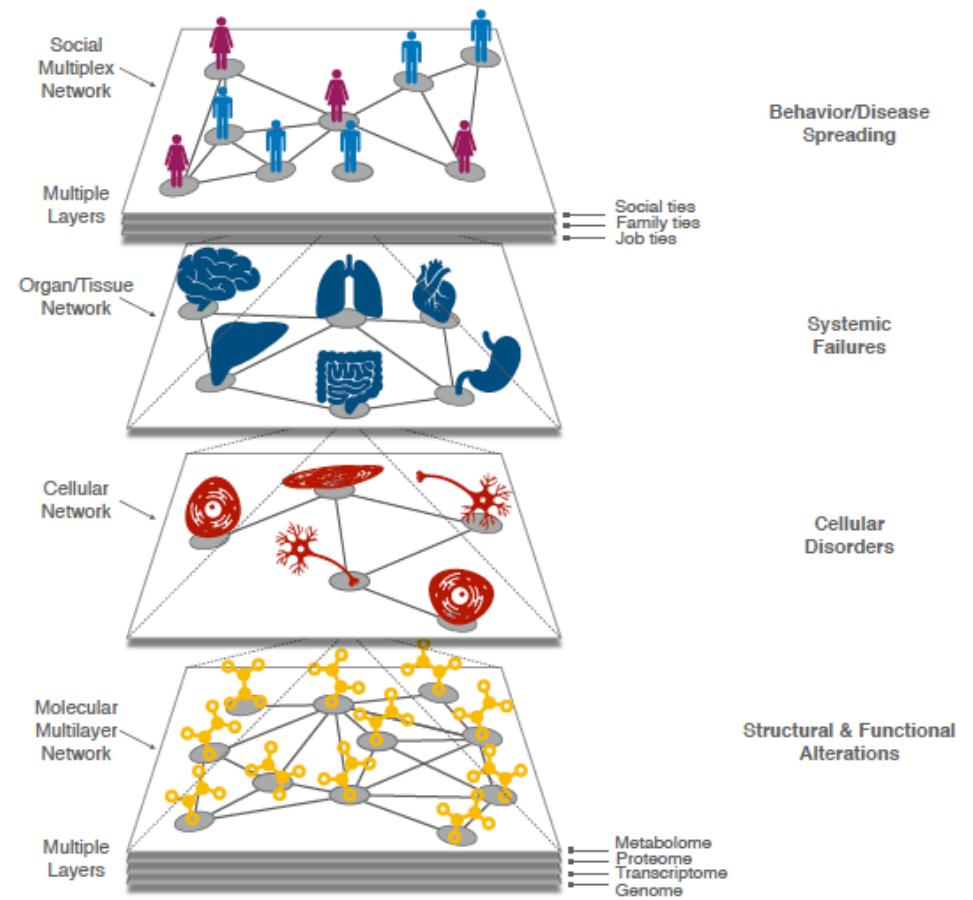


Works by Barbiero, Ciravegna, Giannini, Gori etc

identifying disease markers across scales

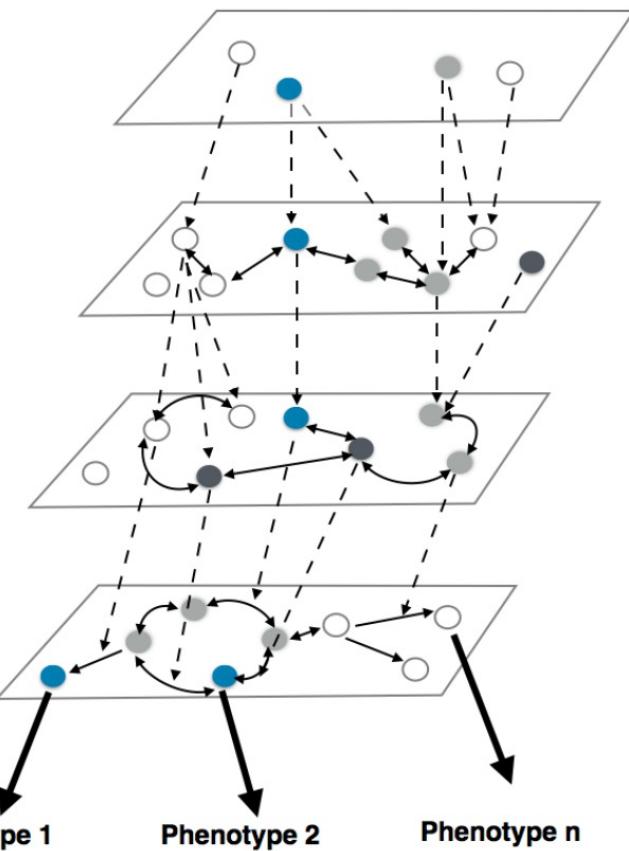
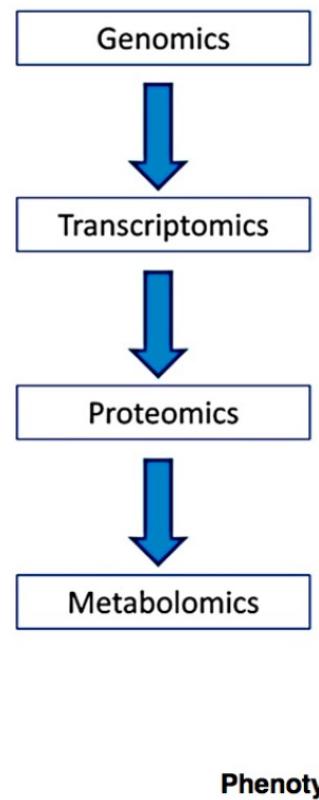
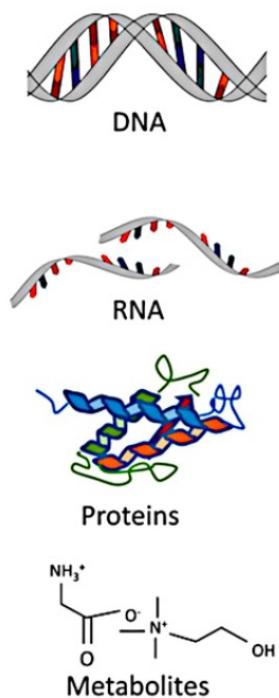


as a graphs system

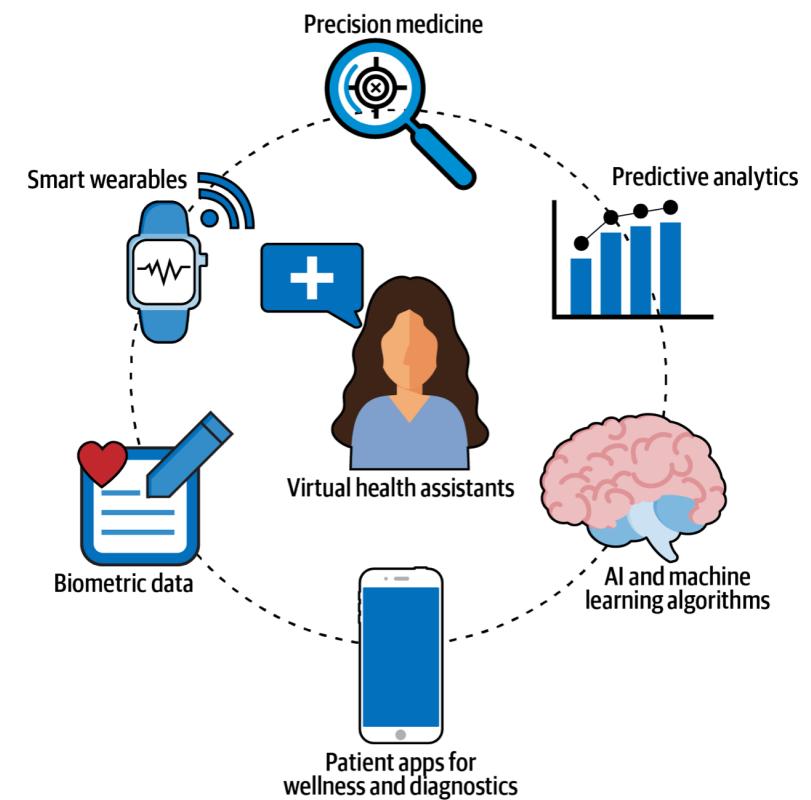


We can observe what happens at almost all scales, from the whole organism down to the molecular level; however, putting things together in order to obtain real understanding is much more difficult and less developed

Benchside

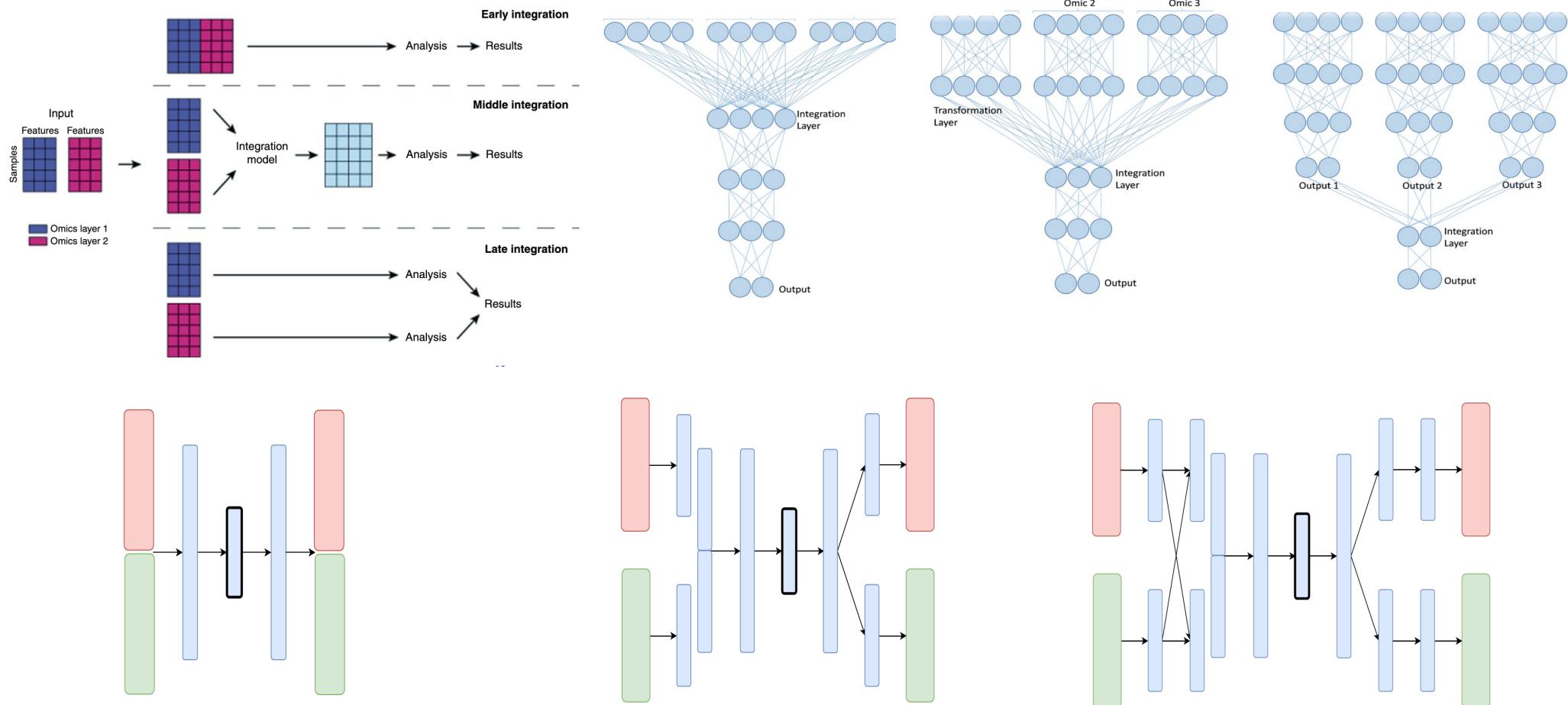


Bedside

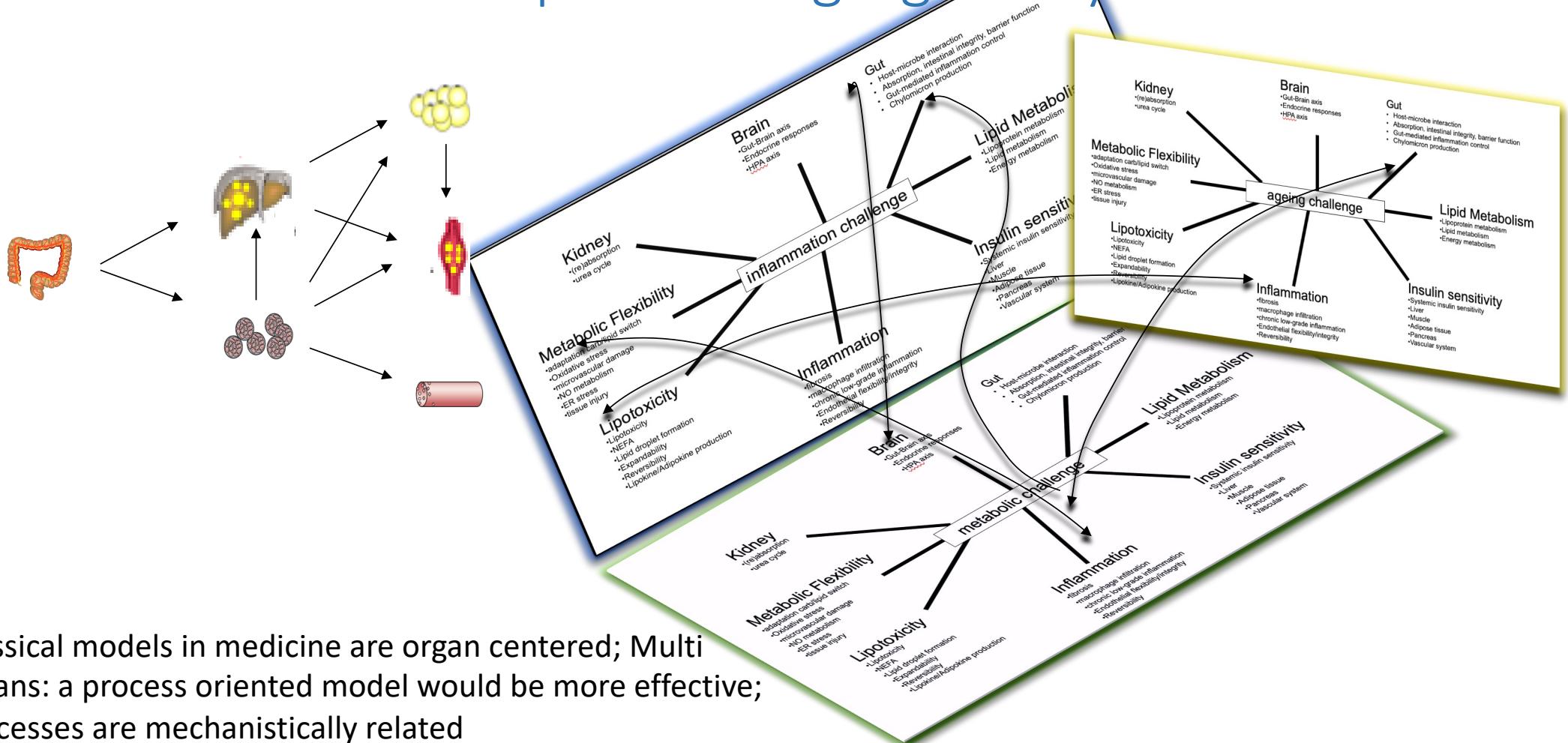


Disease state management platform

Early (left), Middle and Late (right) data integration

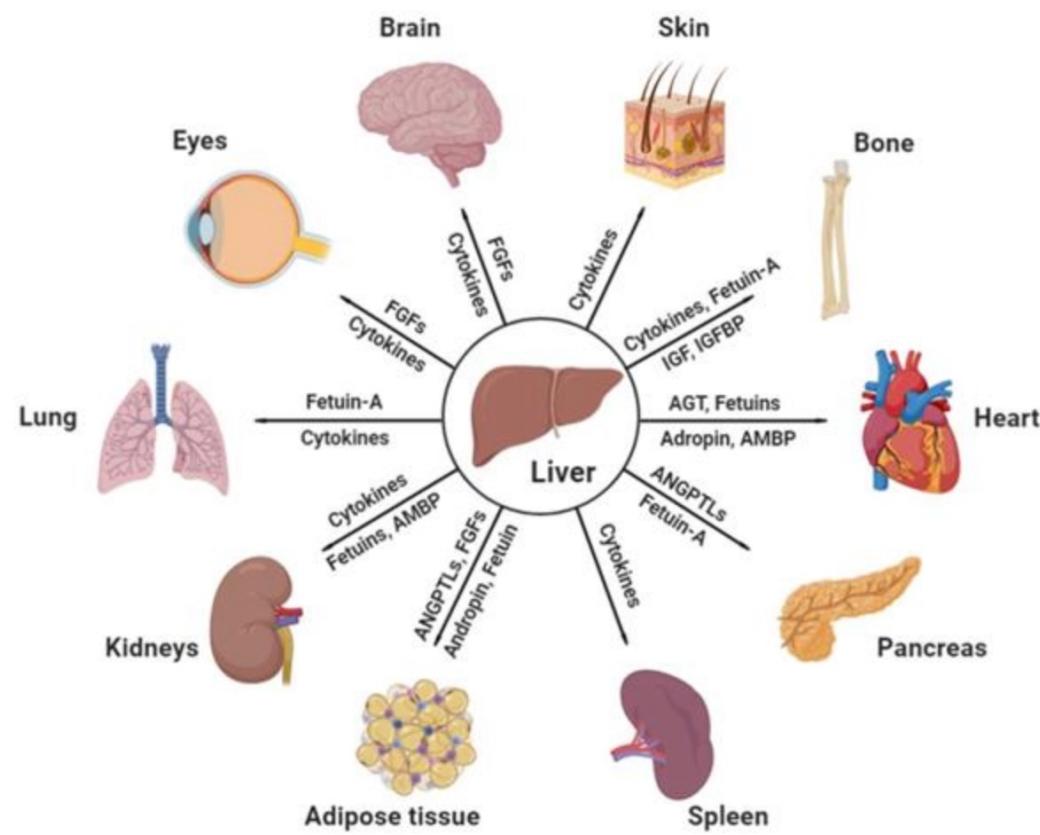


Data-driven could help detecting signals systemic medicine

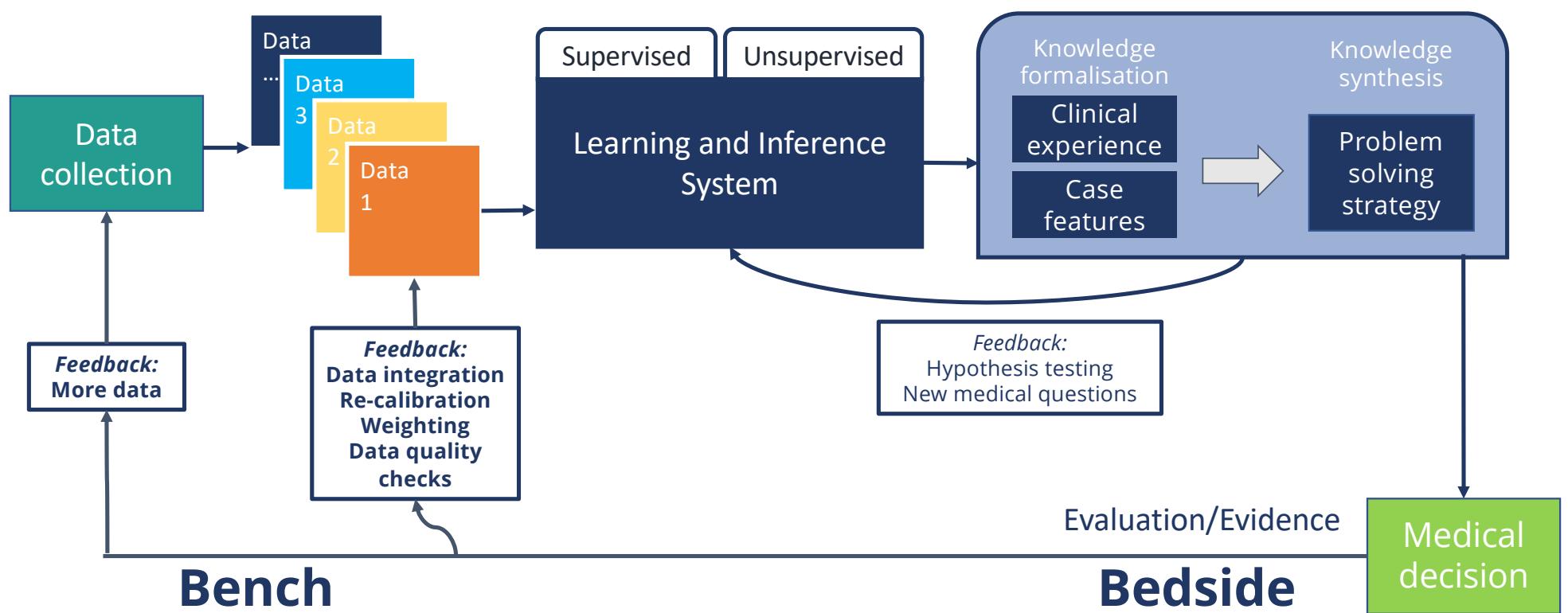


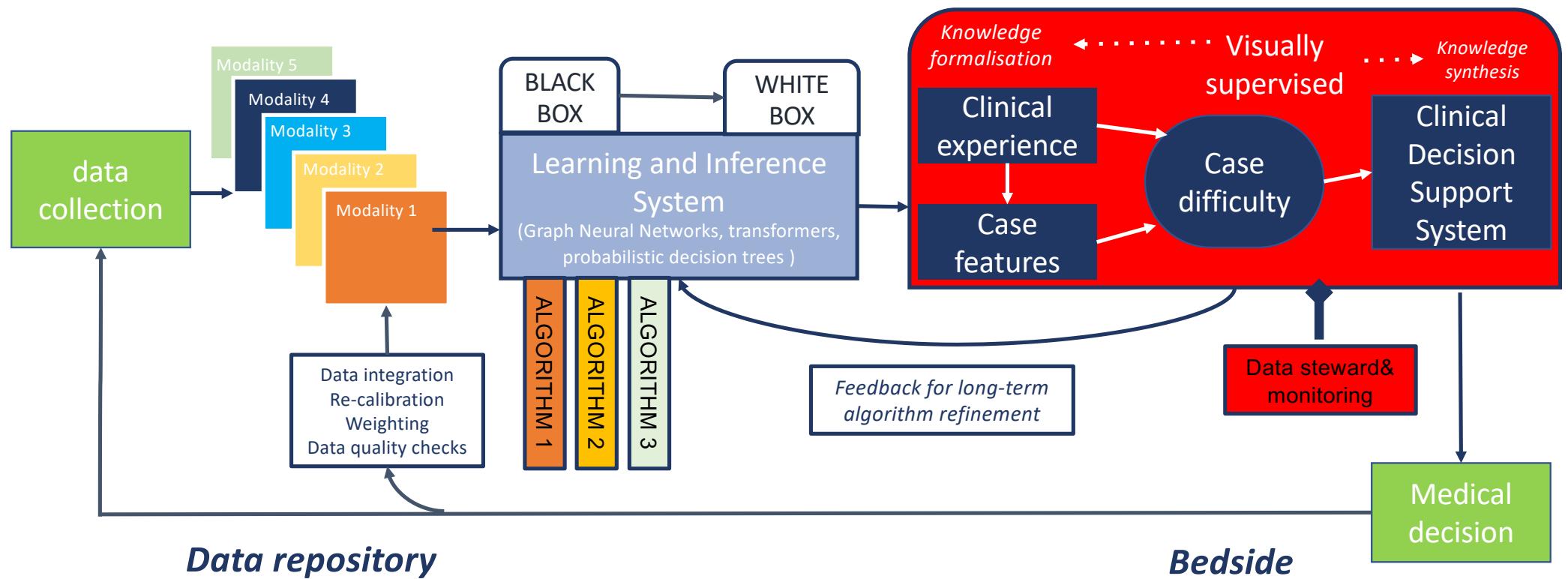
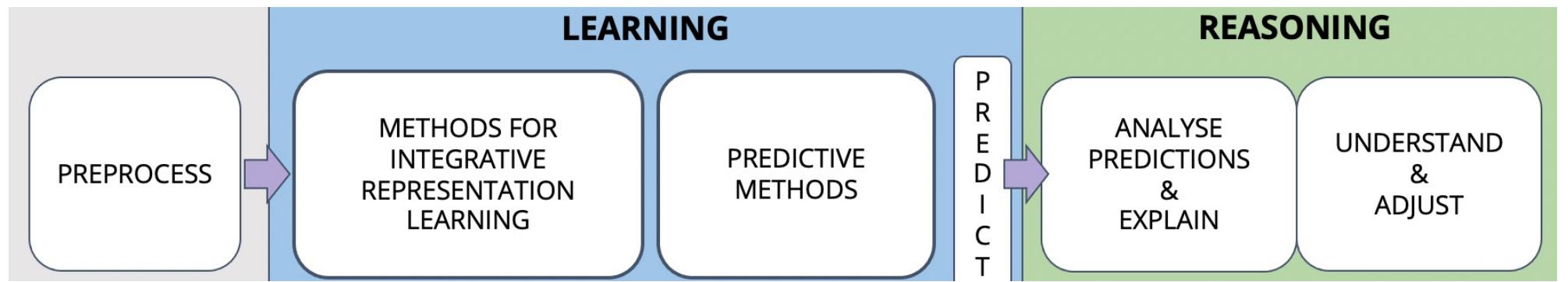
Classical models in medicine are organ centered; Multi organs: a process oriented model would be more effective; Processes are mechanistically related

Spatial-temporal communication



- Long-distance communication is essential for coordination of organ growth and development
- Microbial metabolites produced by the gut microbiome, e.g. short-chain fatty acids (SCFA), have been found to influence lung physiology and injury responses.
- Communicome, for examples hepatokines and cytokines, from the liver to other major distant organs
- highly coordinated, multi-tissue metabolism underlying the body's circadian rhythms





Machine learning makes expert knowledge scalable

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva^{1*}, Brett Kuprel^{1*}, Roberto A. Novoa^{2,3}, Justin Ko², Susan M. Swetter^{2,4}, Helen M. Blau⁵ & Sebastian Thrun⁶

Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents

Joy T. Wu, MBChB, MPH,¹ Ken C. L. Wong, PhD,¹ Yaniv Gur, PhD,¹ Nadeem Ansari, MS,¹ Alexandros Karargyris, PhD,¹ Arjun Sharma, MD,¹ Michael Morris, MD,¹ Babak Saboury, MD,¹ Hassan Ahmad, MD,¹ Orest Boyko, MD, PhD,² Ali Syed, MD,¹ Ashutosh Jadhav, PhD,¹ Hongzhi Wang, PhD,¹ Anup Pillai, PhD,¹ Satyananda Kashyap, PhD,¹ Mehdi Moradi, PhD,¹ and Tanveer Syeda-Mahmood, PhD¹

A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs

Chi-Tung Cheng^{1,7}, Yirui Wang^{1,7}, Huan-Wu Chen³, Po-Meng Hsiao⁴, Chun-Nan Yeh⁵, Chi-Hsun Hsieh¹, Shun Miao², Jing Xiao², Chien-Hung Liao^{1,6✉} & Le Lu^{1,2}

Machine learning will replace human radiologists, pathologists, maybe soon

As artificial intelligence, cognitive computing and machine learning systems become better than humans at medicine and cost less, it might even become unethical not to replace people.

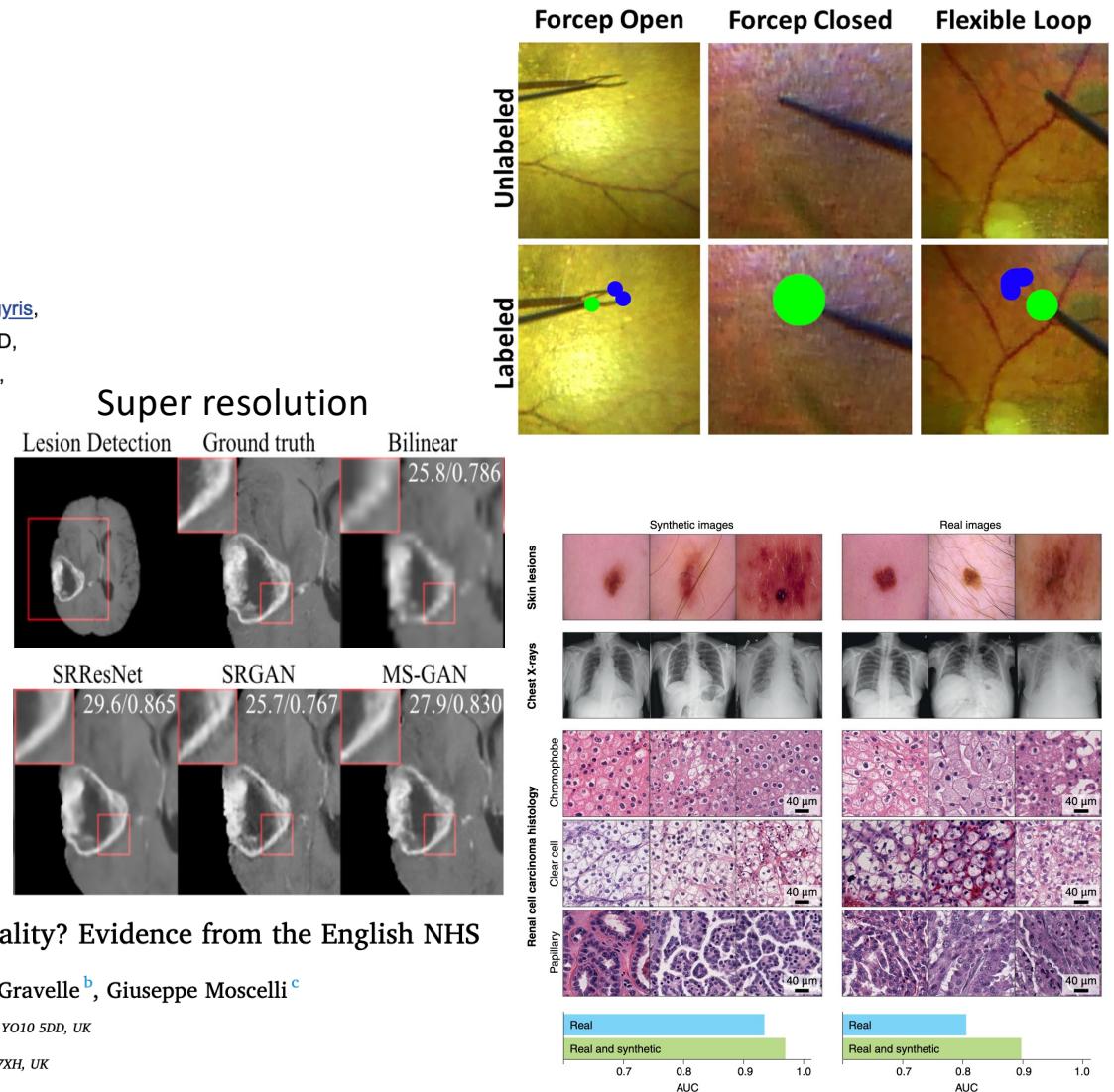
Do small hospitals have lower quality? Evidence from the English NHS

James Gaughan^b, Luigi Siciliani^{a,*}, Hugh Gravelle^b, Giuseppe Moscelli^c

^a Department of Economics and Related Studies, University of York, York, YO10 5DD, UK

^b Centre for Health Economics, University of York, York, YO10 5DD, UK

^c Department of Economics, University of Surrey, Guildford, Surrey, GU2 7XH, UK



Current debate: how do you like AI in Medicine: adagio, andante or allegro or vivace?

Following a music analogy, it is like choosing the pace:

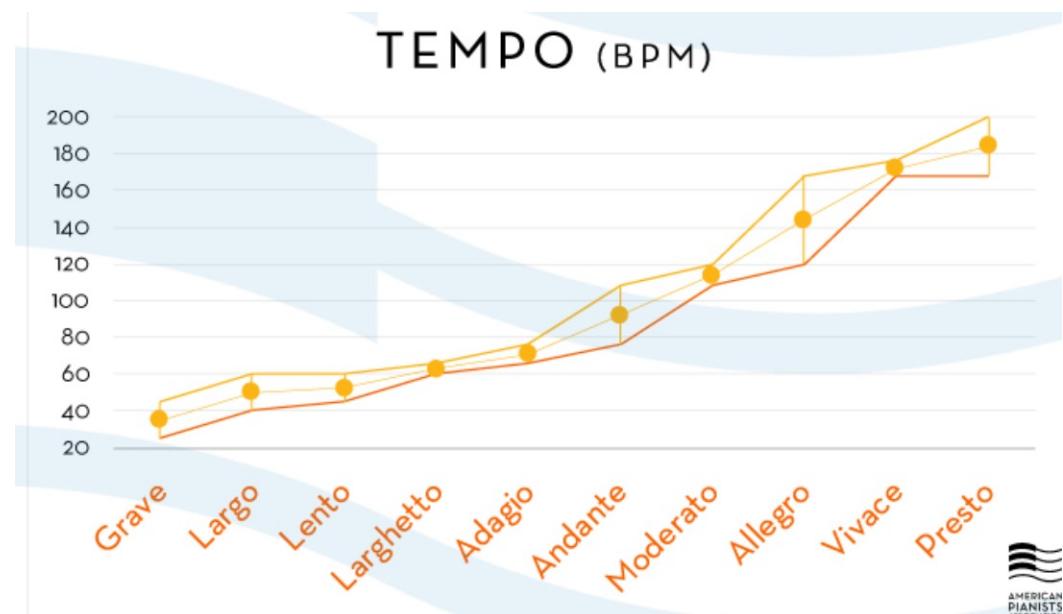
Adagio is a tempo marking in music that means gentle and easy and is not too slow or too fast.

Andante means a walking pace;

Allegro means Quite Quickly

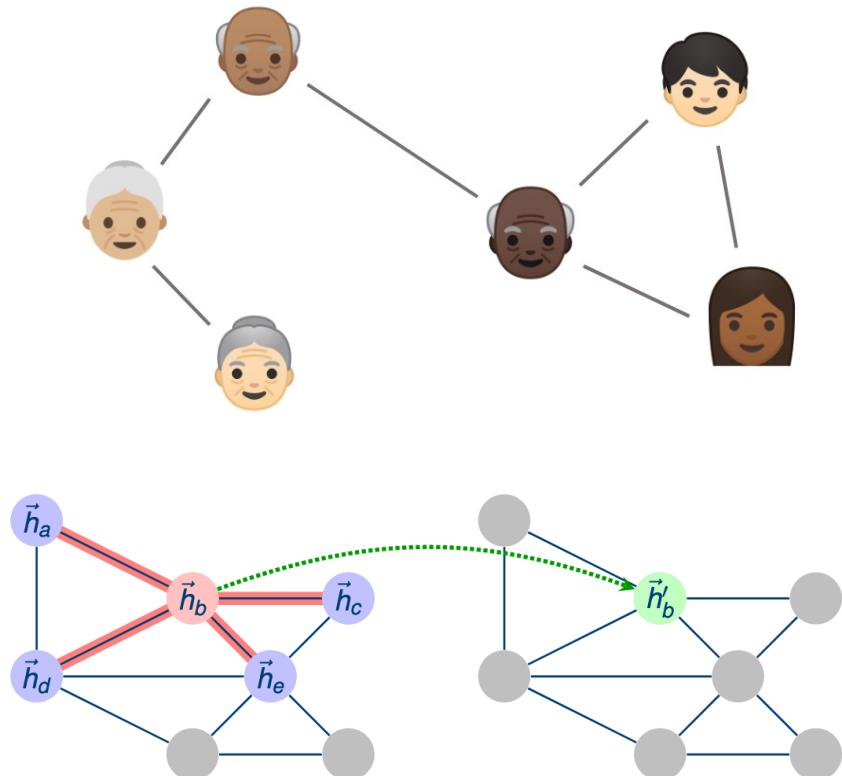
Vivace means lively and fast.

Black box = you know how to build the instruments but You do not have no music score



Predicting Patient Outcomes

“Relatedness”: Grouping Similar Patients



Patient in pink; similar patients in violet

How do doctors work?

Doctors are recalling how they treat rare cases in the past.

The clinician will actively compare how a new patient is similar and different from past experience and how it could inform the view on the current patient.

We emulated this in our architecture: we draw links between patients and allowed the model to “see” data from similar patients (context) as it makes prediction on the current patient.

Credits to Emma Rocheteau

Electronic Health Records in the ICU

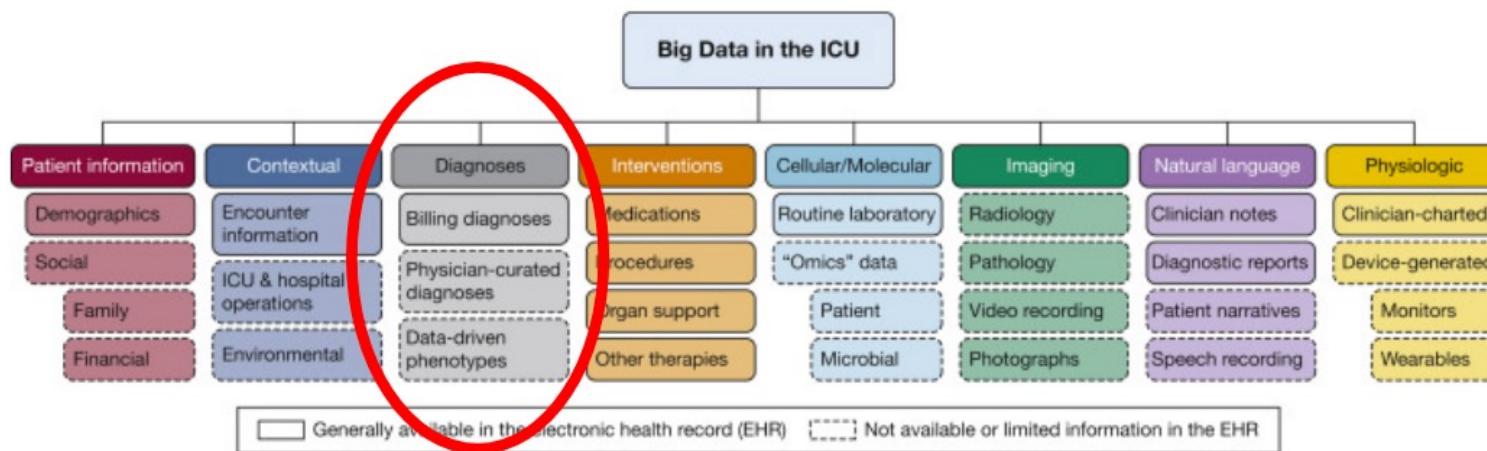
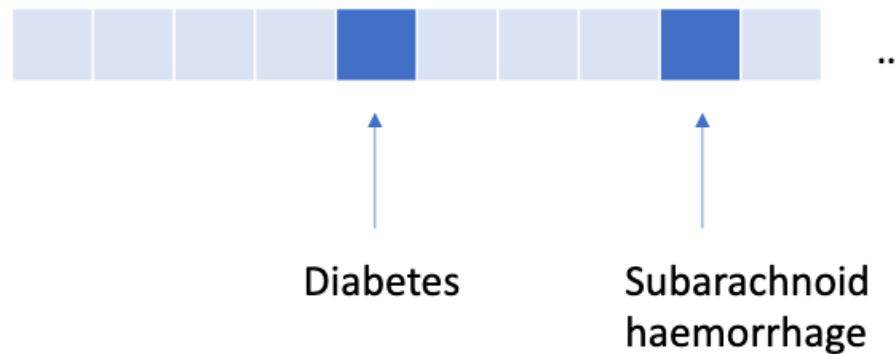


Diagram: Sanchez-Pinto et al. 2018

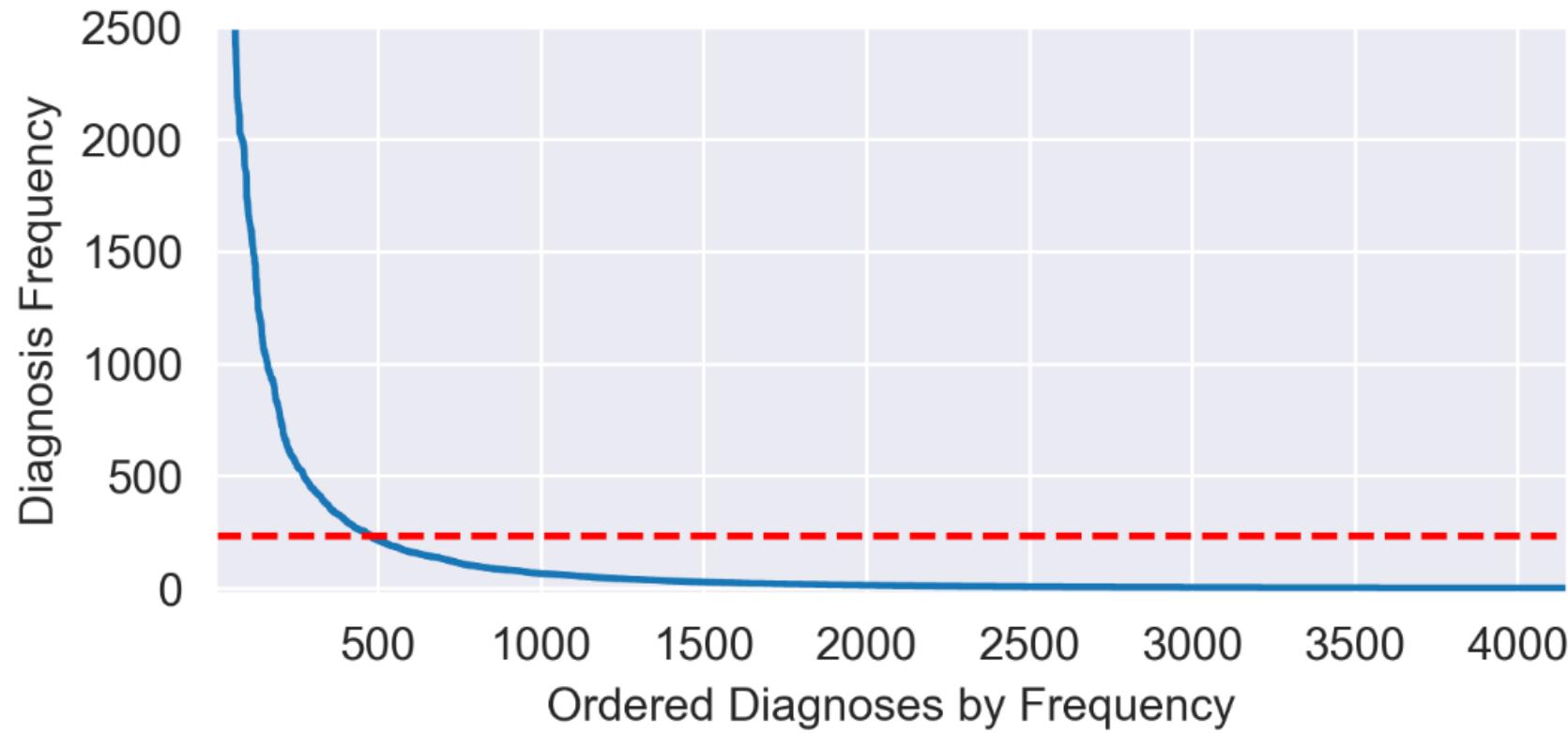
E. Rocheteau*, C. Tong*, P. Velickovic, N. Lane, P. Liò Predicting Patient Outcomes with Graph Representation Learning.

Diagnosis Information is Hard to Use



- ▶ The large number of possibilities make distinguishing between patterns of comorbidity difficult.
- ▶ There is a lack of data for rarer combinations.

Distribution of Diagnoses in the elCU Database



Credits to Emma Rocheteau

Graph Construction

The “relatedness” score between two patients i and j is given by:

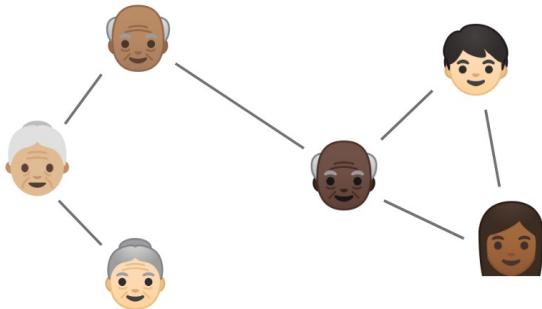
$$\mathcal{M}_{ij} = a \underbrace{\sum_{\mu=1}^m (\mathcal{D}_{i\mu} \mathcal{D}_{j\mu} (d_\mu^{-1} + c))}_{\text{Shared Diagnoses}} - \underbrace{\sum_{\mu=1}^m (\mathcal{D}_{i\mu} + \mathcal{D}_{j\mu})}_{\text{All Diagnoses}} \quad (1)$$

where

- ▶ $\mathcal{D} \in \mathbb{R}^{N \times m}$ is a diagnosis matrix,
- ▶ N is the number of patients,
- ▶ m is the number of unique diagnoses,
- ▶ d_μ is the frequency of diagnosis μ ,
- ▶ a and c are tunable constants.

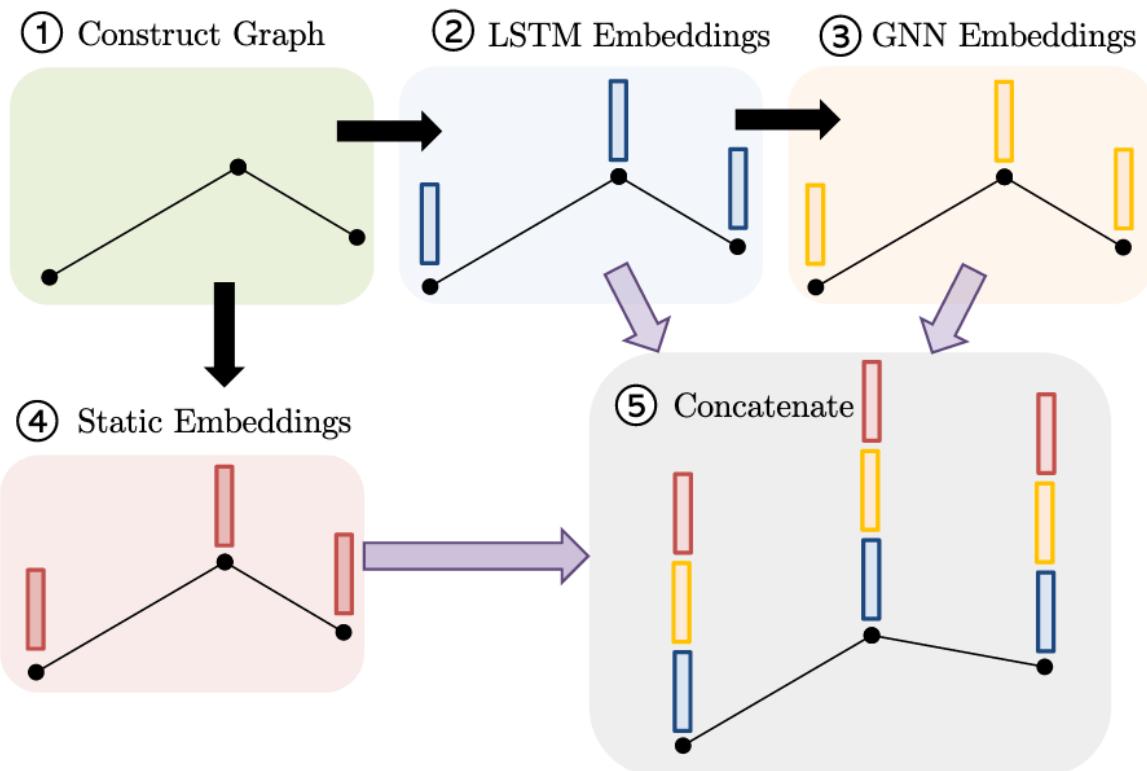
Credits to Emma Rocheteau

Graph Construction



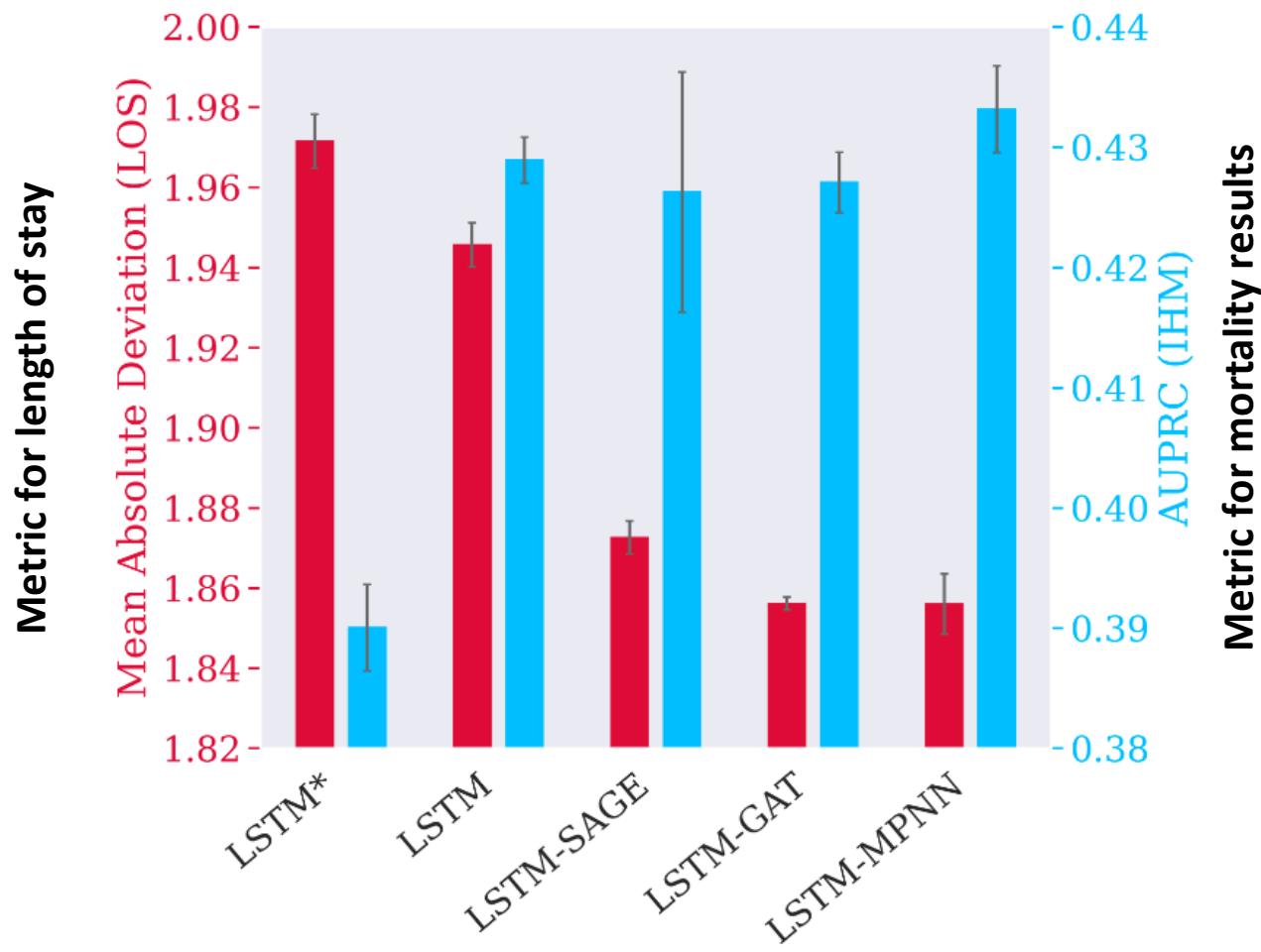
The “relatedness” score between two patients i and j is given by:

$$\mathcal{M}_{ij} = a \sum_{\mu=1}^m \underbrace{\left(\mathcal{D}_{i\mu} \mathcal{D}_{j\mu} (d_\mu^{-1} + c) \right)}_{\text{Shared Diagnoses}} - \sum_{\mu=1}^m \underbrace{\left(\mathcal{D}_{i\mu} + \mathcal{D}_{j\mu} \right)}_{\text{All Diagnoses}}$$



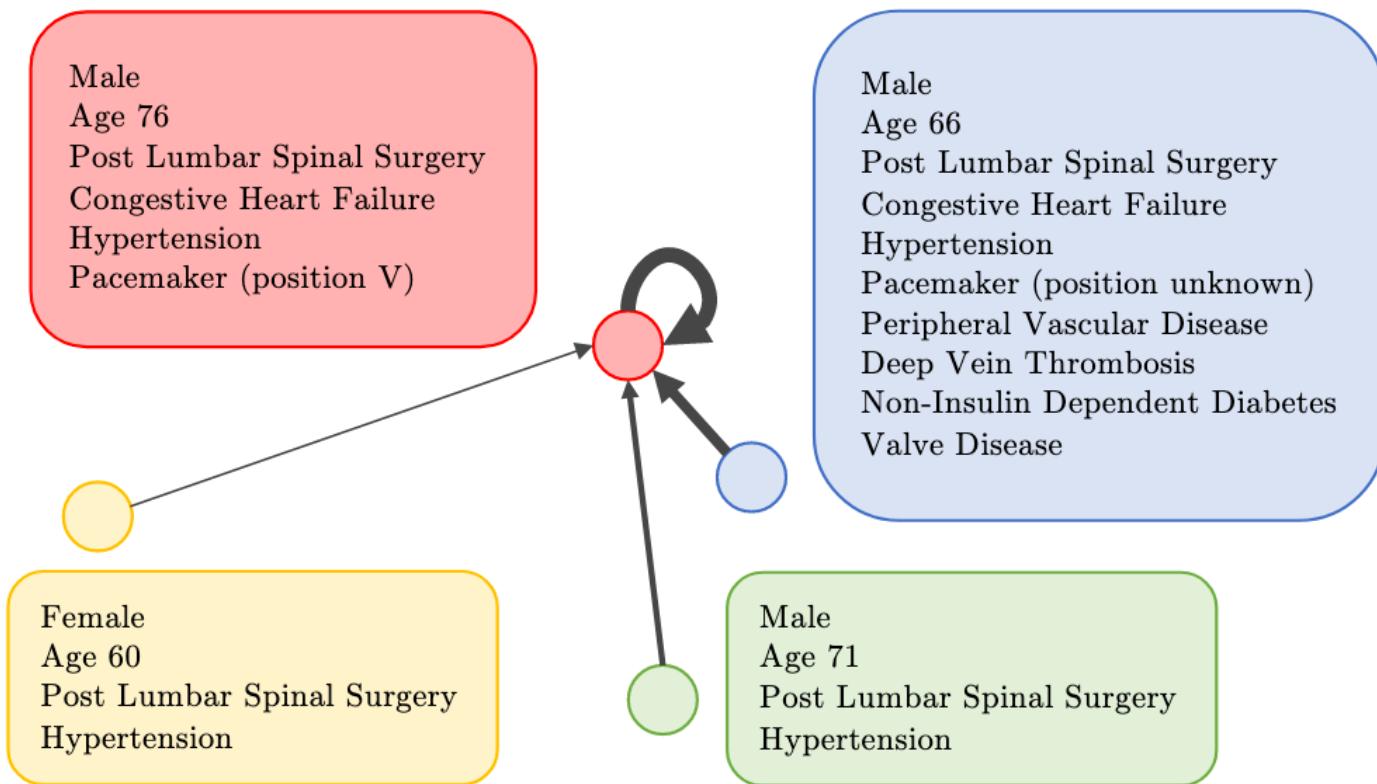
Credits to Emma Rocheteau

Results (small section)



Credits to Emma Rocheteau

Visualisation: LSTM-GAT* attention weights



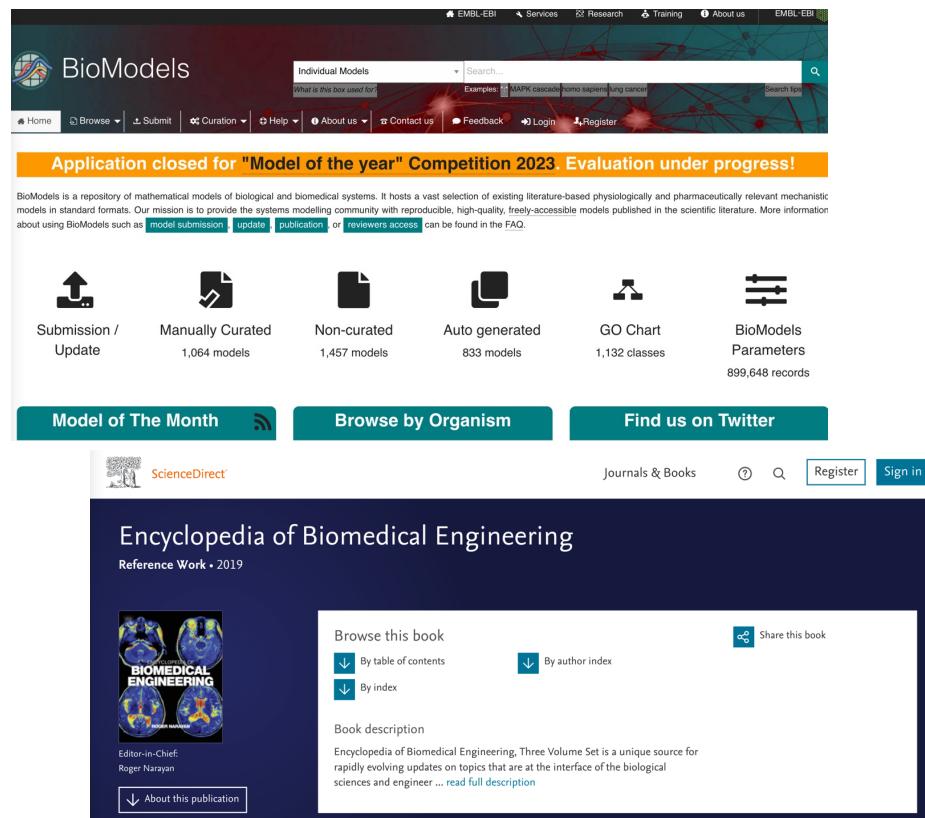
Interpretability benefits;

The patient under observation is the pink; **Attention weights are shown by the thickness of the lines.** The model pays more attention as self attention; then Congestive heart failure counts more (blue patient).

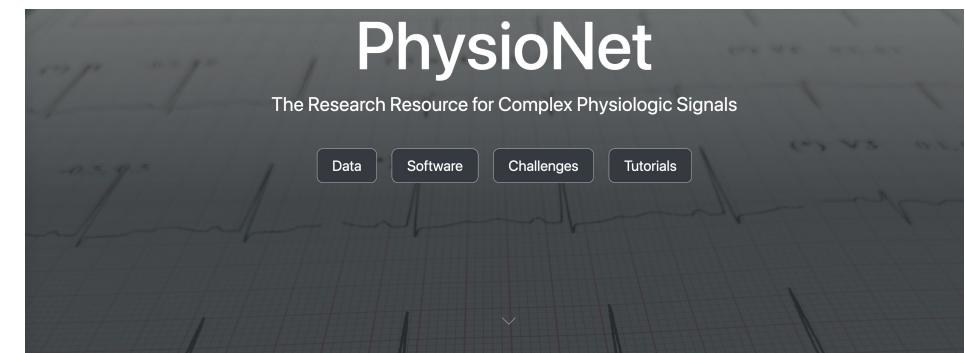
Credits to Emma Rocheteau

The Patient digital Twin

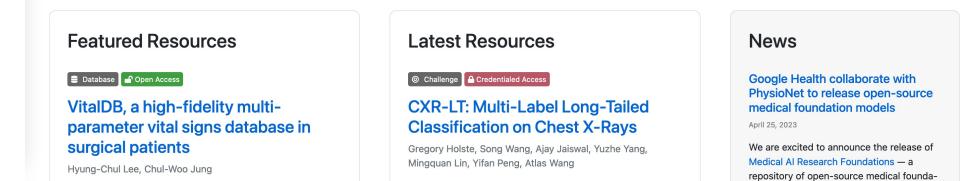
Physiological models



The BioModels homepage features a search bar at the top with a placeholder "Individual Models" and a dropdown menu. Below the search bar is a banner stating "Application closed for 'Model of the year' Competition 2023. Evaluation under progress!". A detailed description follows, mentioning the repository's mission to provide systems modeling community with reproducible, high-quality, freely-accessible models. Below this, there are six icons with corresponding statistics: "Submission / Update" (1,064 models), "Manually Curated" (1,457 models), "Non-curated" (833 models), "Auto generated" (1,132 classes), "GO Chart" (899,648 records), and "BioModels Parameters". At the bottom, there are three buttons: "Model of The Month" (with a RSS icon), "Browse by Organism", and "Find us on Twitter". The footer includes links to ScienceDirect, Journals & Books, a search bar, and user registration/sign-in options.

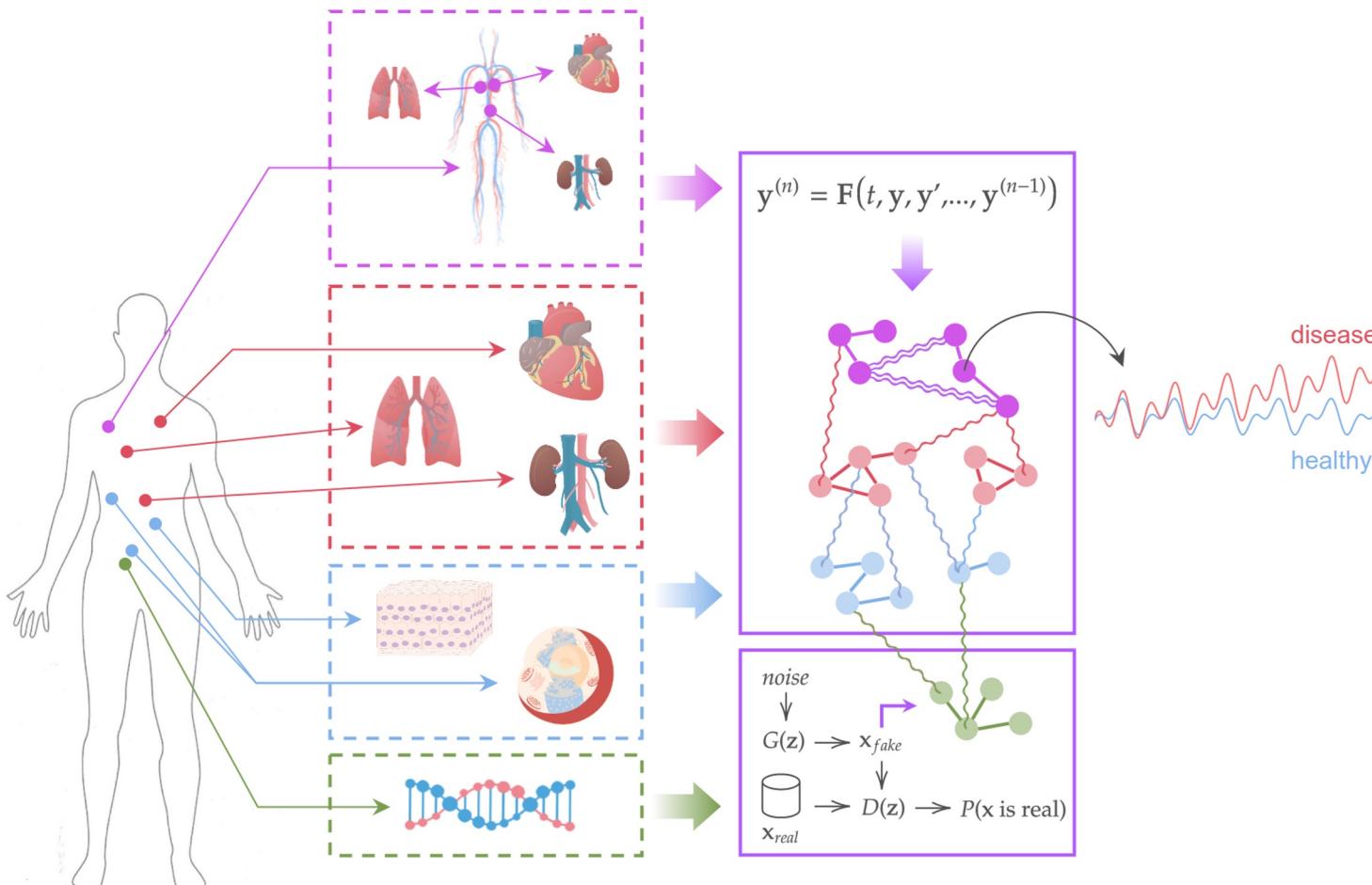


The PhysioNet homepage features a large header with the text "PhysioNet" and "The Research Resource for Complex Physiologic Signals". Below the header is a navigation bar with four buttons: "Data", "Software", "Challenges", and "Tutorials". The main content area shows a grid of ECG traces on a graph paper background.



This section contains three boxes: "Featured Resources" (listing VitalDB as a high-fidelity multi-parameter vital signs database in surgical patients), "Latest Resources" (listing CXR-LT: Multi-Label Long-Tailed Classification on Chest X-Rays), and "News" (announcing Google Health's collaboration with PhysioNet to release open-source medical foundation models).

Digital Patient



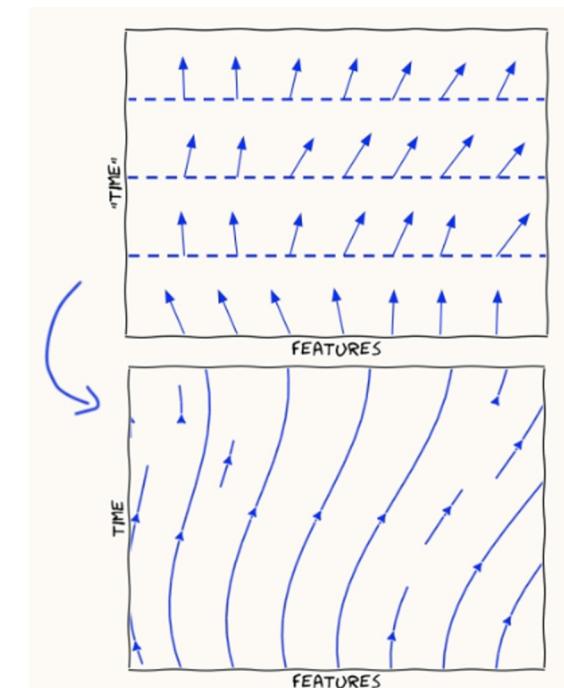
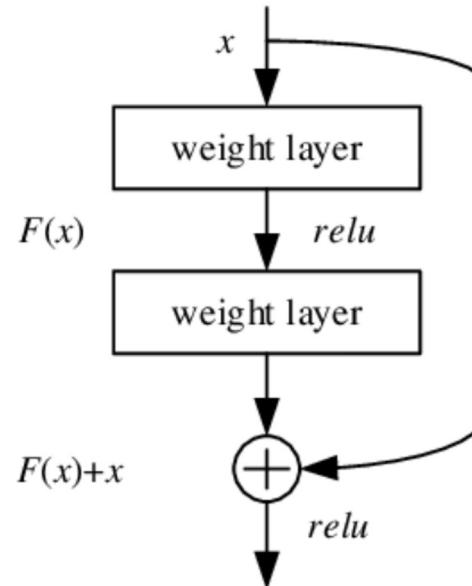
The digital patient is a sort of personal avatar of the patient;

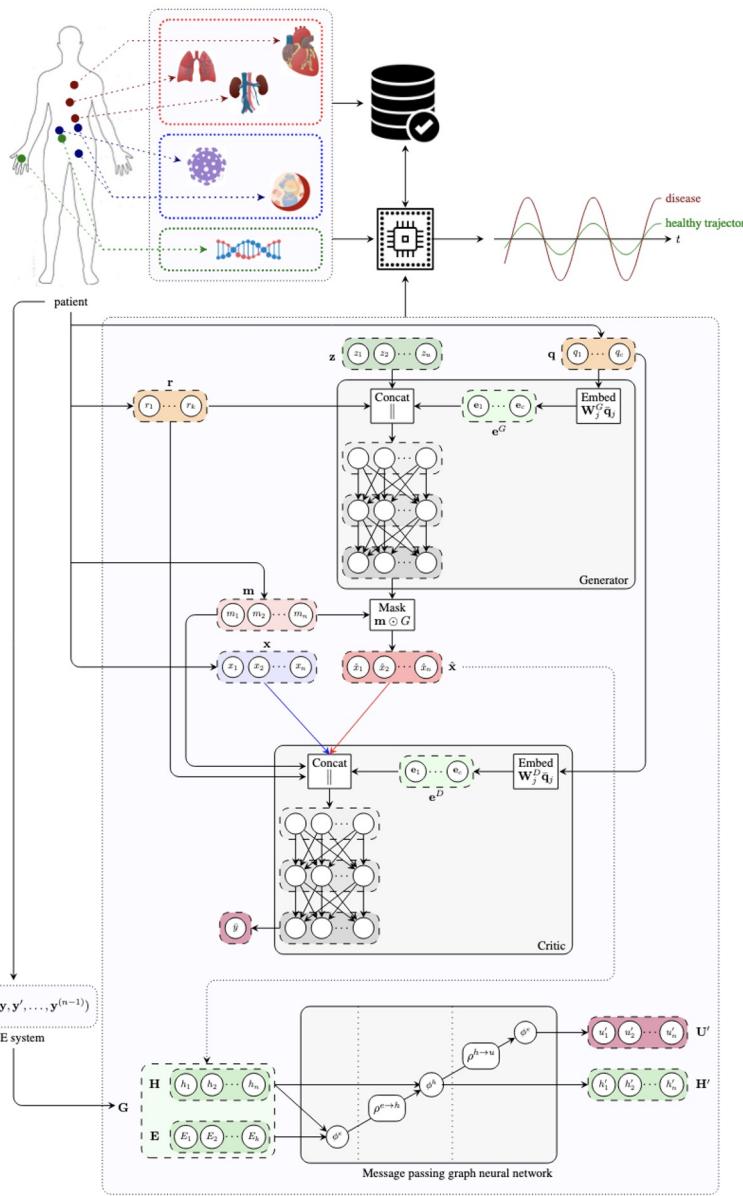
Running experiments on therapies on a digital twin will be cost effective and will provide us best practice to be used on the biological twin.

Here a composite mathematical model could enable the prediction of element functionalities before the insurgence of disease symptoms.

ODEs and deep learning

- A chain of residual blocks in a neural network is basically a solution of the ODE with the Euler method
- The initial condition for the system is “time” 0, which indicates the very first layer of the neural network
- The residual connections are discretised time steps of the Euler method, it means, that we can regulate the depth of the neural network, just with choosing the discretising scheme, hence, making the solution (aka neural network) more or less accurate.



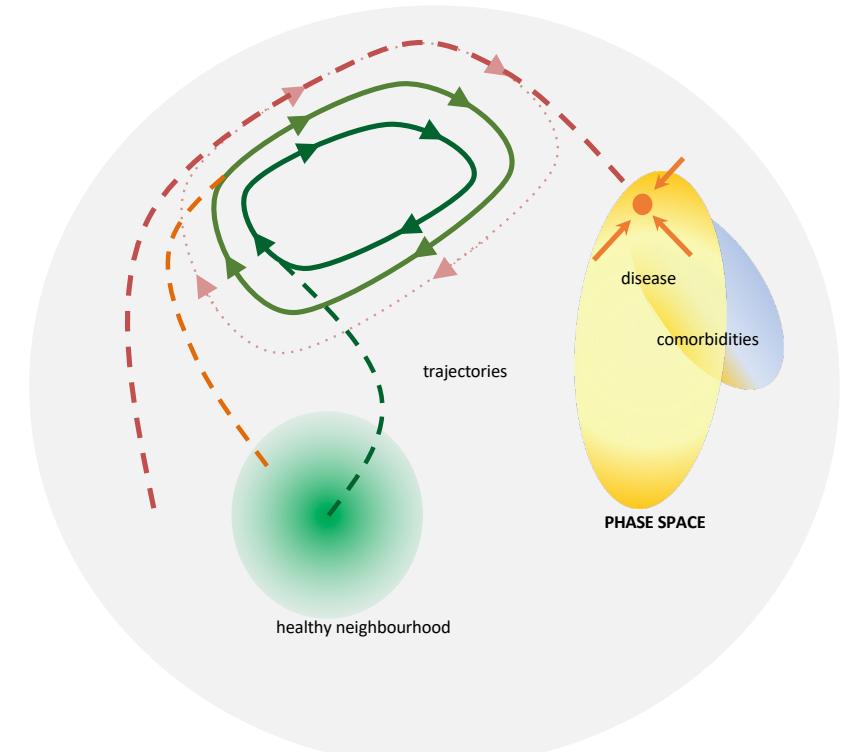
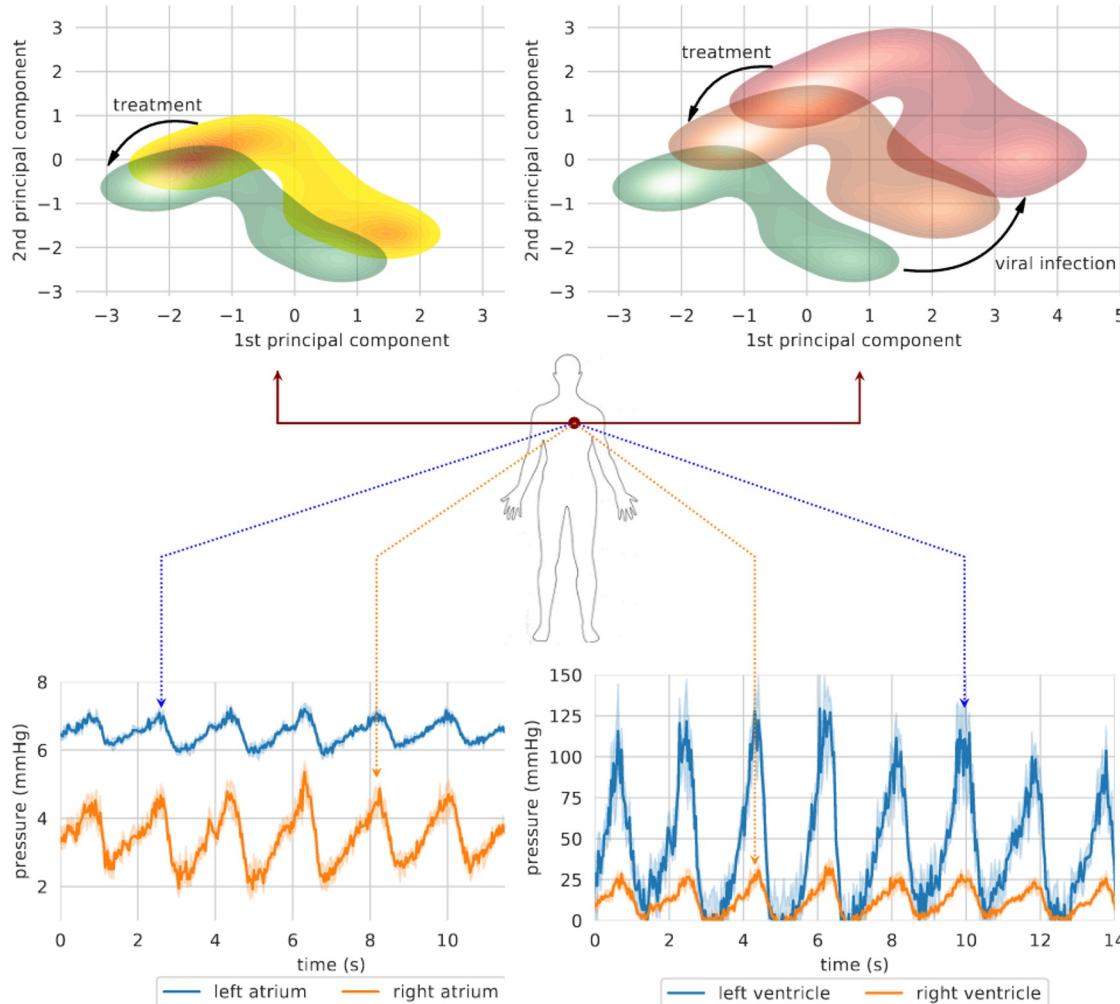


Digital Patient: architecture

The digital patient model is composed of four layers: the transcriptomic layer, the cellular layer, the organ layer (complex networks of cooperating organ), and the exposomic layer (toxicants, dietary regimens, treatments, physical exercise, posture, lifestyle habits, all of them are possible exposures taking part to individual's well-being or disease condition).

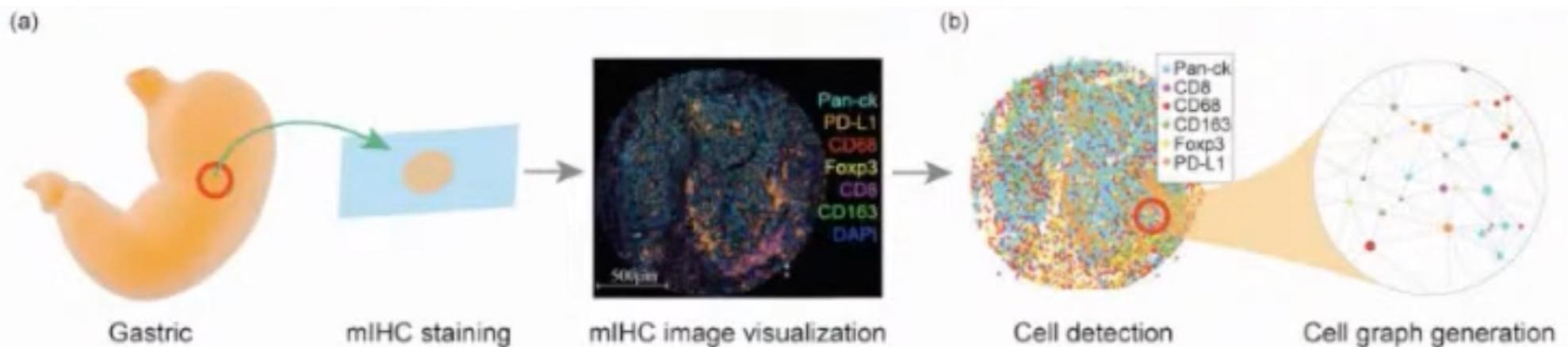
Other layers that bring information on other omics or body sensor network i.e. a collection of networked sensors that can be used to monitor human physiological signals, could be similarly implemented

Digital Patient: results



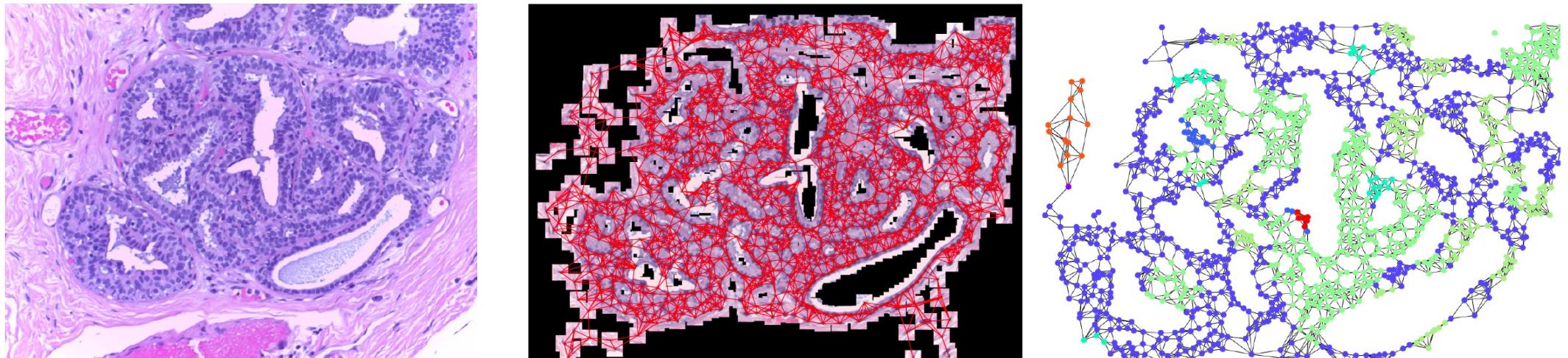
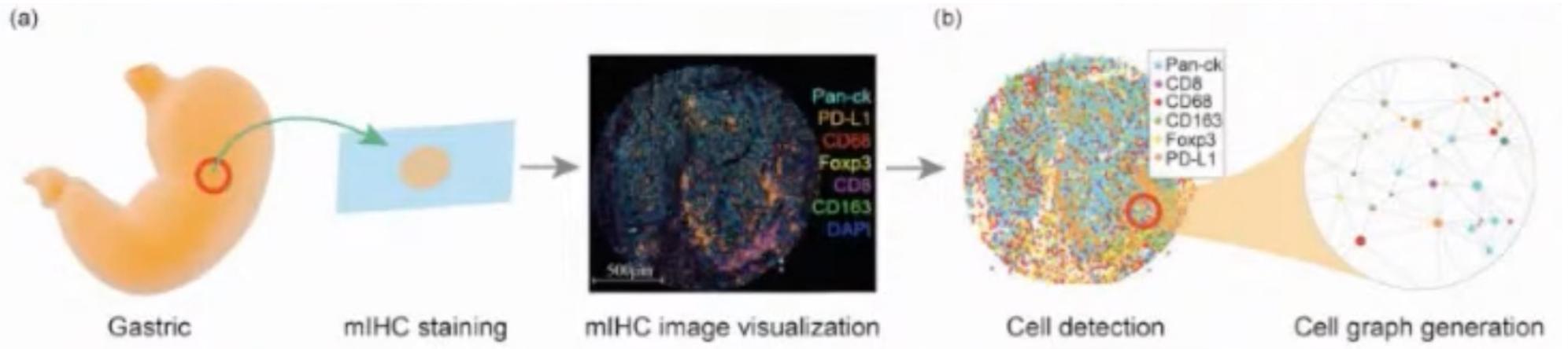
GNN at Cellular and Tissue scales

Tissue level: Cancer Graph network



Wang et al., Cell graph neural networks enable the digital staging of tumor microenvironment and precise prediction of patient survival in gastric cancer, Nature Oncology, 2022

Tissue level: Cancer Graph network



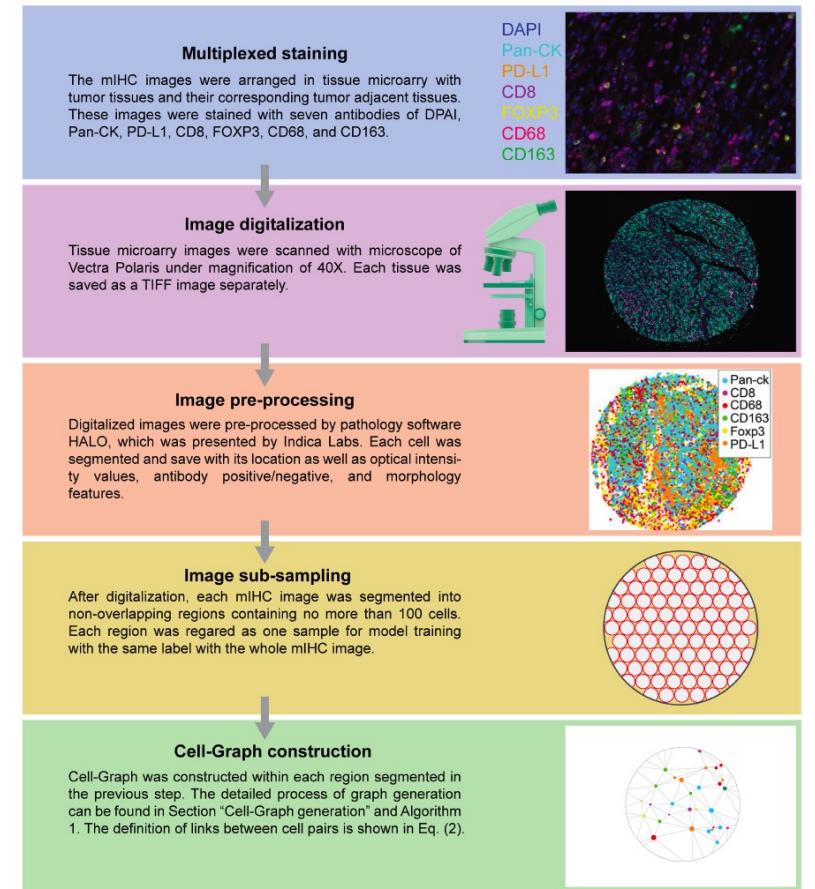
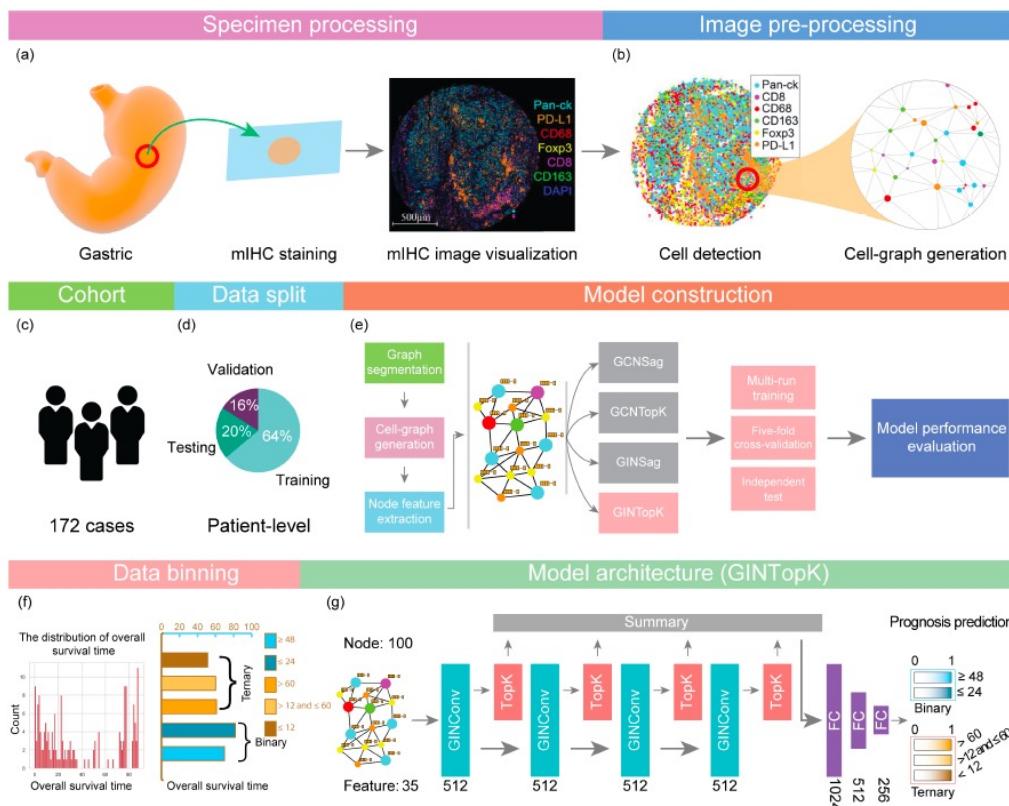
Wang et al., Cell graph neural networks enable the digital staging of tumor microenvironment and precise prediction of patient survival in gastric cancer, Nature Oncology, 2022

Framelet decomposition for cell-graph

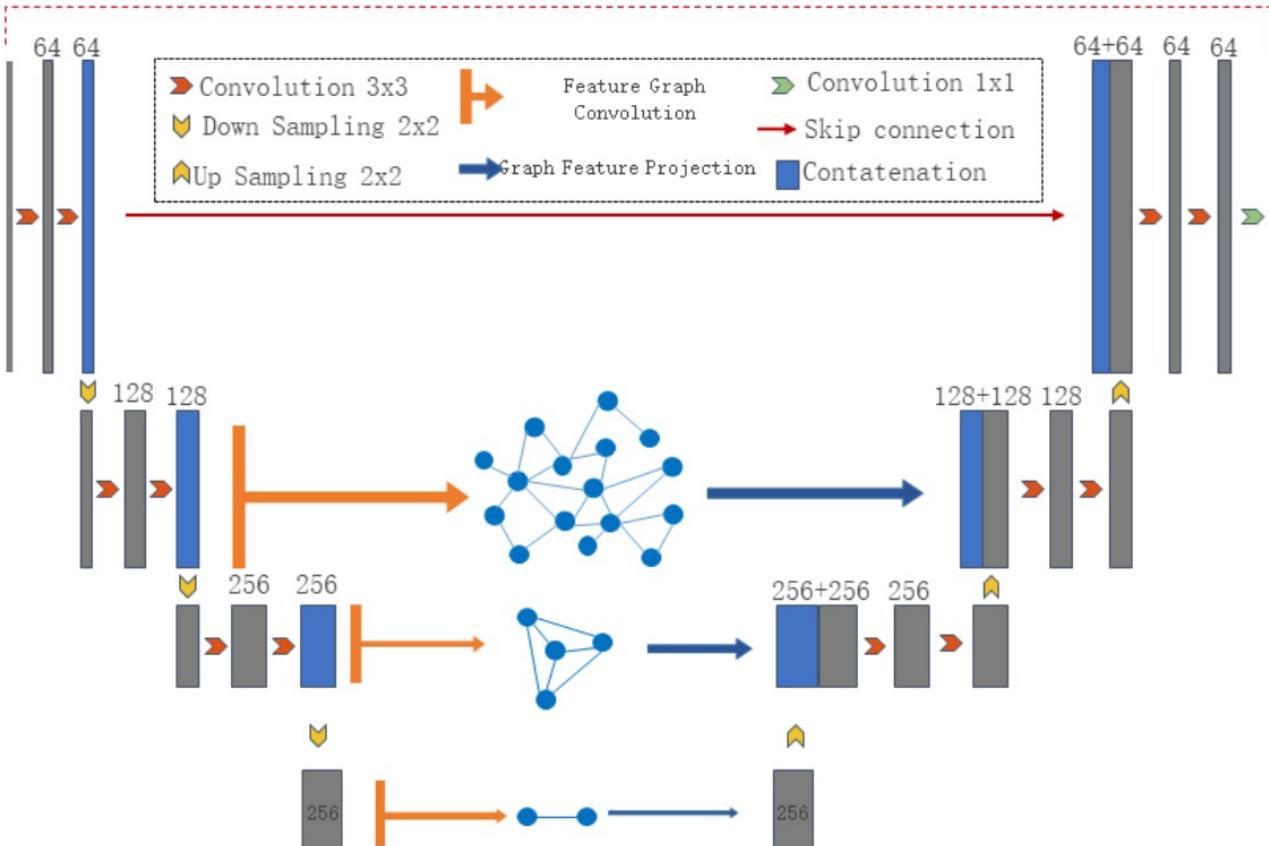
To examine the capacity of Cell-Graph to capture useful spatial features from mIHC images, we conducted a framelet decomposition on the whole mIHC images. The framelet transforms (including framelet decomposition and reconstruction) have proved an important tool for distilling multi-resolution information in low-pass and high-passes from the graph data^{33,34,35,36,37}.

We extracted low-pass and high-pass information of six types of features, corresponding to six different biomarkers DAPI, PAN-CK, CD8, CD68, FOXP3, and PD-L1. Tables S3–S11 show the low-pass and high-pass coefficients of the framelet decomposition on mIHC images of short-term, medium-term, and long-term survivors on the entire mIHC images. For the selected samples, no significant differences were observed from the low-pass channel. However, major differences can be observed from the high-pass channel on the selected samples. More specifically, remarkable signal differences can be seen from the high-pass channel-1 and channel-2 in terms of the features of Cell Area and Nucleus Perimeter (summarized in Supplementary Tables 6–11). These differences highlight the important prognostic value of cell morphological information of the TME, which is consistent with the prognostic value of different types of cell features.

Wang et al: Cell graph neural networks enable digital staging of tumour microenvironment and precisely predict patient survival in gastric cancer. Nature Oncology

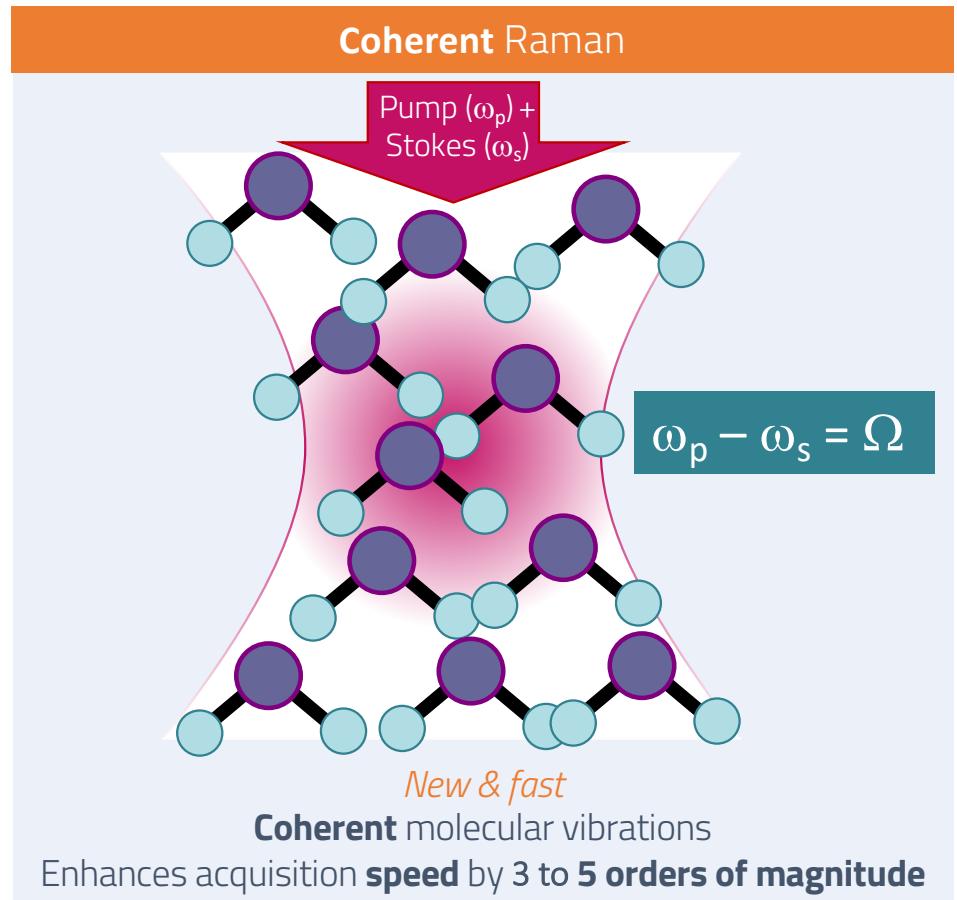
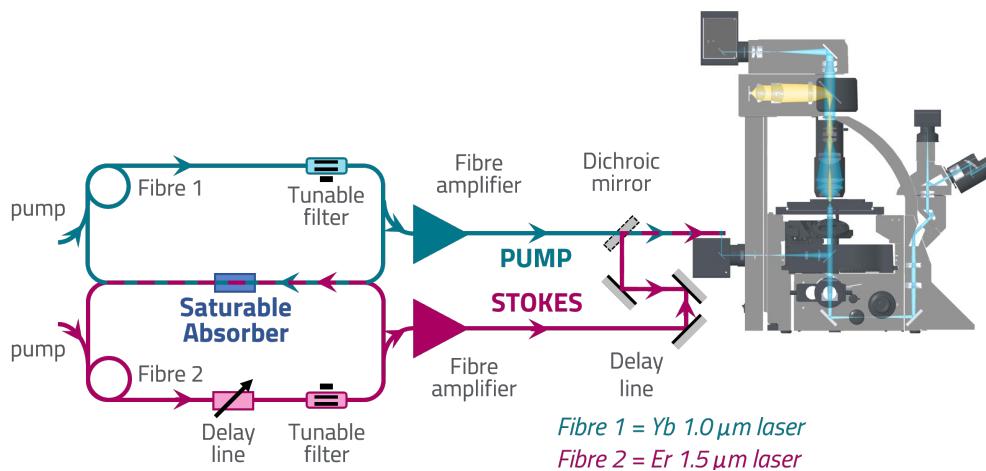
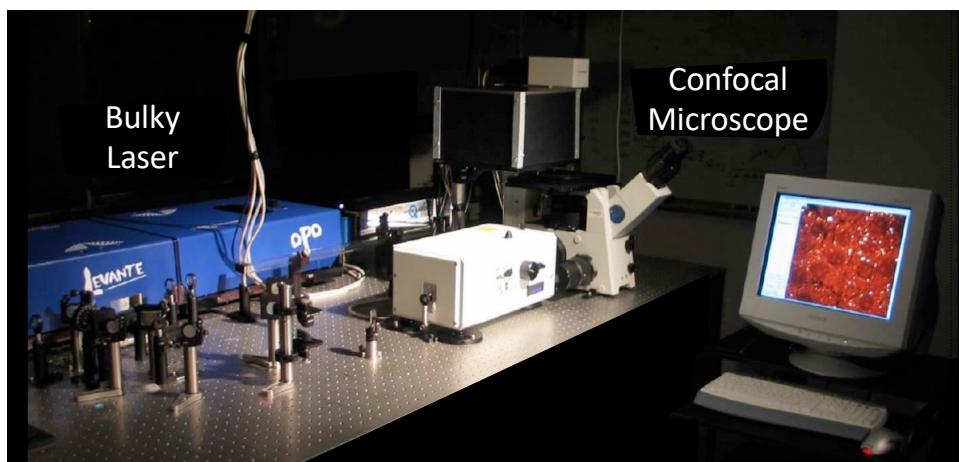


Multi resolution GNN

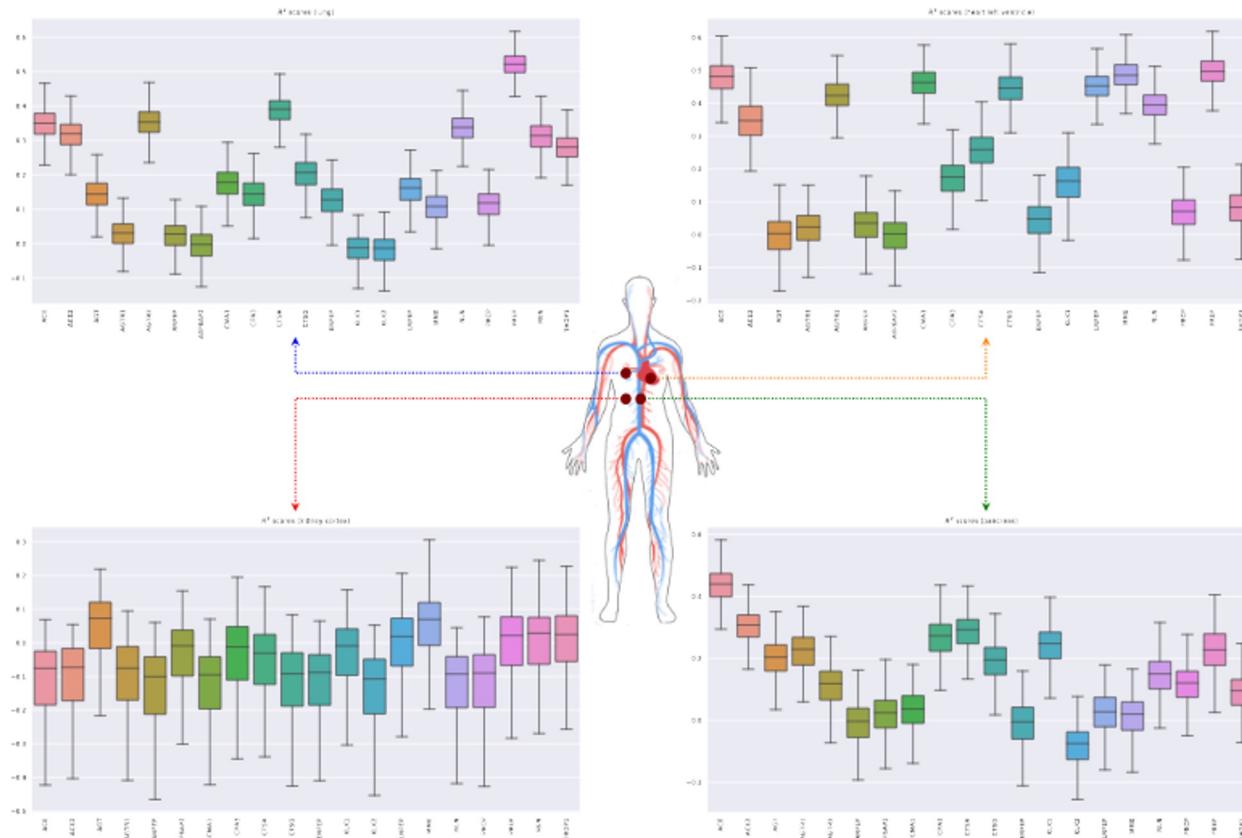


The designed diagram of VGU-Net. The encoder-decoder path is the same as architecture of U-Net. The CNNs in the bottle-neck layer and two skip connections are substituted with two layers of GCNs, performing graph representation learning to take long-range interaction between nodes. The residual connection represented by the dash line is specifically designed for image restoration task

innovation: turnkey, compact, & low cost



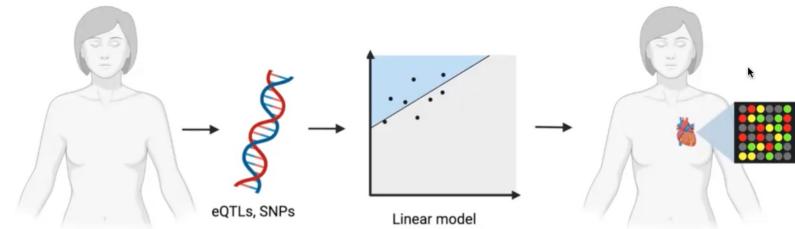
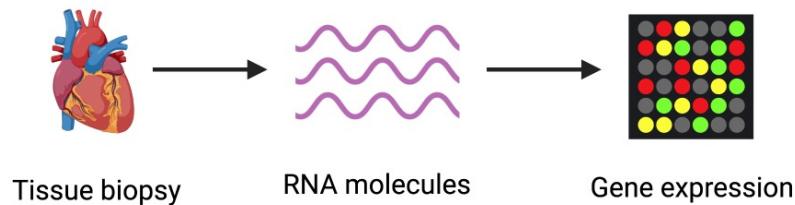
Digital Patient: results



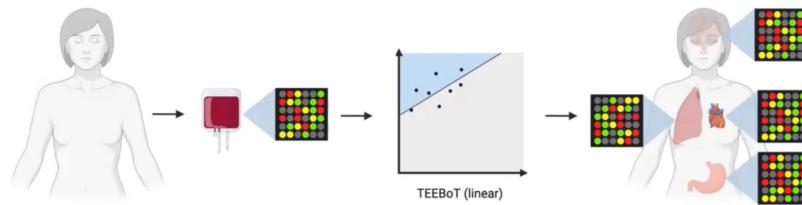
Simulations showing viral infection and ACE-i interventions.

GNN at gene expression

Imputation of gene expression

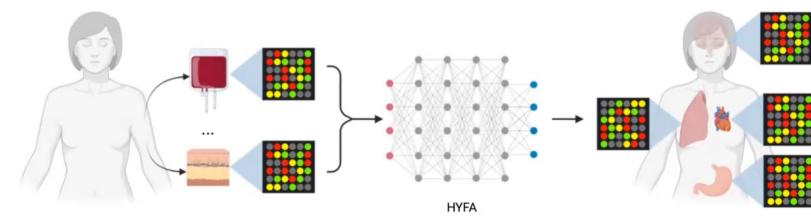


How do we infer gene expression of inaccessible tissues?



Single-tissue imputation of gene expression

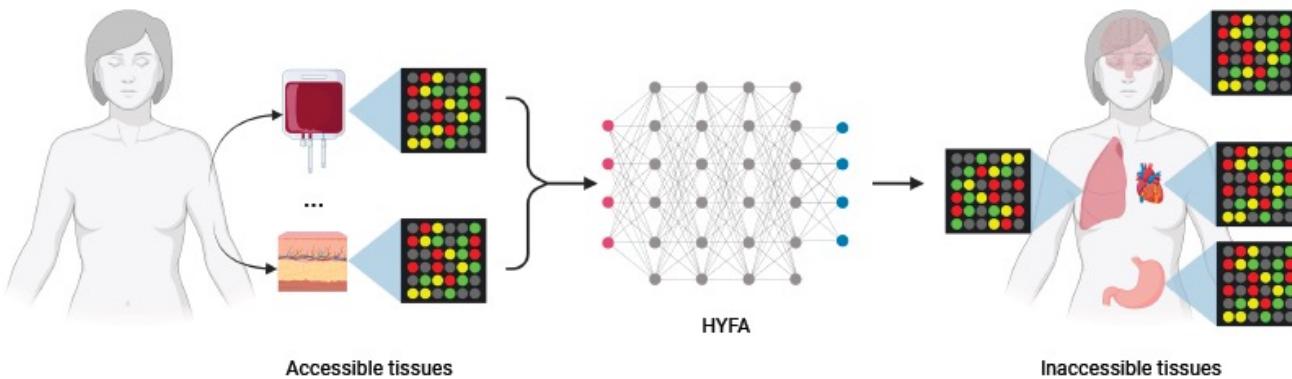
Imputation from genotype-derived features



Multi-tissue imputation of gene expression

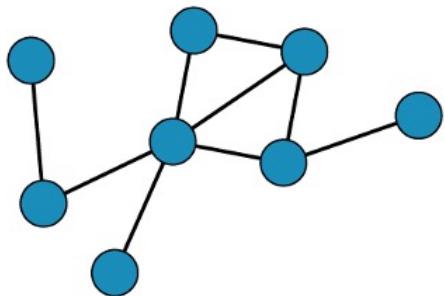
Credits to Ramon Viñas

Hypergraph factorisation

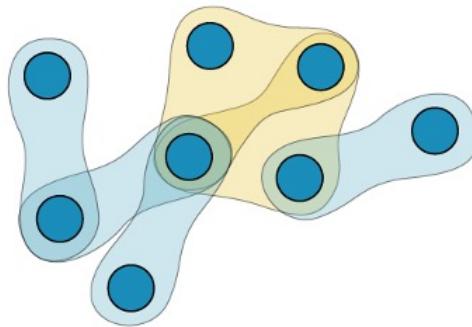


Overview of HYFA. (a) HYFA processes gene expression from a number of collected tissues (e.g. accessible tissues) and infers the transcriptomes of uncollected tissues. (b) Workflow of HYFA..

Network

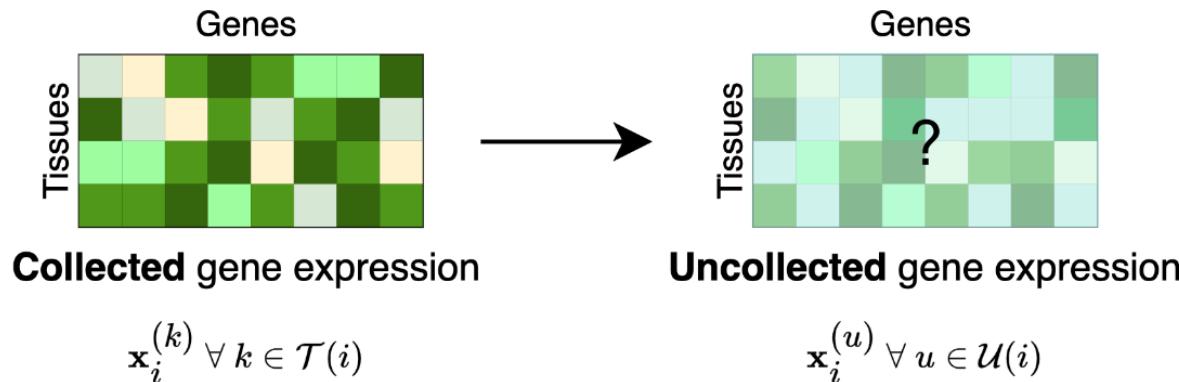


Hypergraph



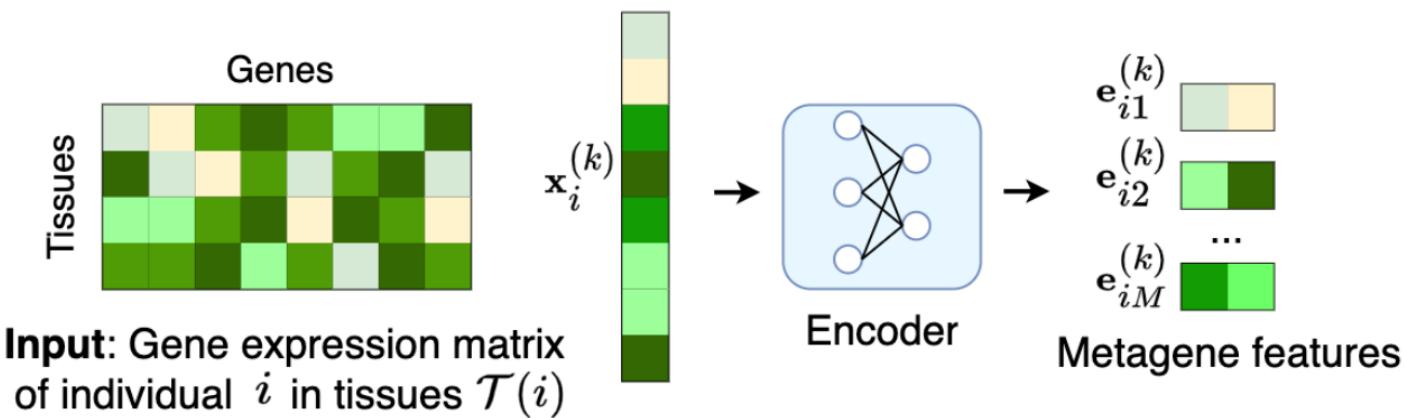
Problem formulation

- ▶ Dataset with N individuals/donors, T tissues, and G genes.
- ▶ For a given individual i :
 - ▶ $\mathcal{T}(i)$ represents the set of *collected* tissues.
 - ▶ $\mathcal{U}(i)$ represents the set of *uncollected* tissues.
- ▶ Denote by $\mathbf{x}_i^{(k)} \in \mathbb{R}^G$ the expression values of donor i measured in tissue k .
- ▶ **Goal:** For individual i , infer the uncollected $\mathbf{x}_i^{(u)} \forall u \in \mathcal{U}(i)$ from the collected $\mathbf{x}_i^{(k)} \forall k \in \mathcal{T}(i)$ gene expression.



Low-dimensional gene expression representation

- ▶ Project gene expression into low-dimensional space



- ▶ *Metagene* features capture abstract gene expression patterns.
- ▶ G genes, M metagenes, $M \ll G$.

Hypergraph representation

Node features

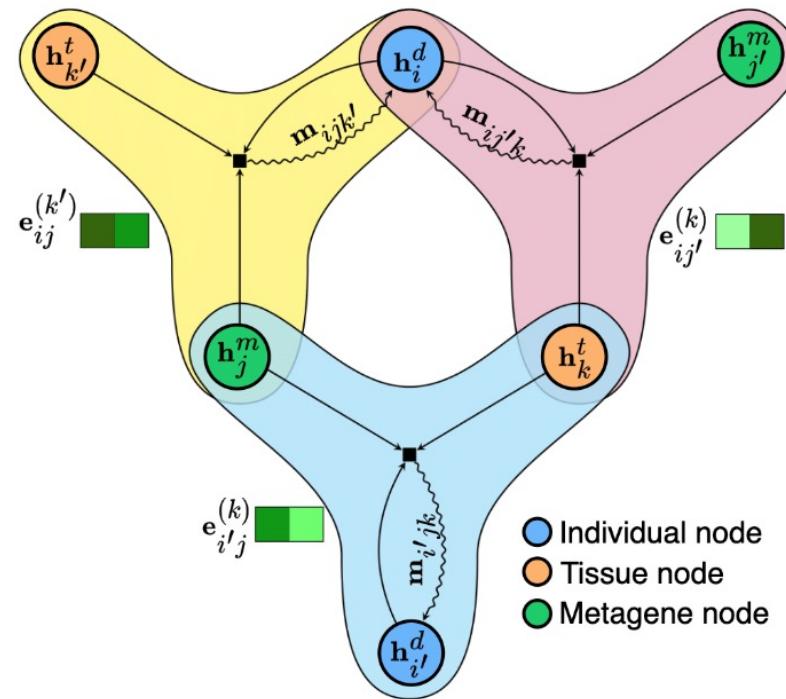
- ▶ Donor: \mathbf{h}_i^d (update)
- ▶ Tissue: \mathbf{h}_k^t (learnable)
- ▶ Metagene: \mathbf{h}_j^m (learnable)

Hyperedges

$$\mathcal{E} = \{(i, j, k, \mathbf{e}_{ij}^{(k)}) \mid \forall (i, j, k), k \in \mathcal{T}(i)\}$$

Hyperedge attributes

- ▶ Expression-derived $\mathbf{e}_{ij}^{(k)}$
 - ▶ individual i
 - ▶ metagene j
 - ▶ tissue k



Message passing updates representations of individual nodes

Message passing on hypergraph

- ▶ Message

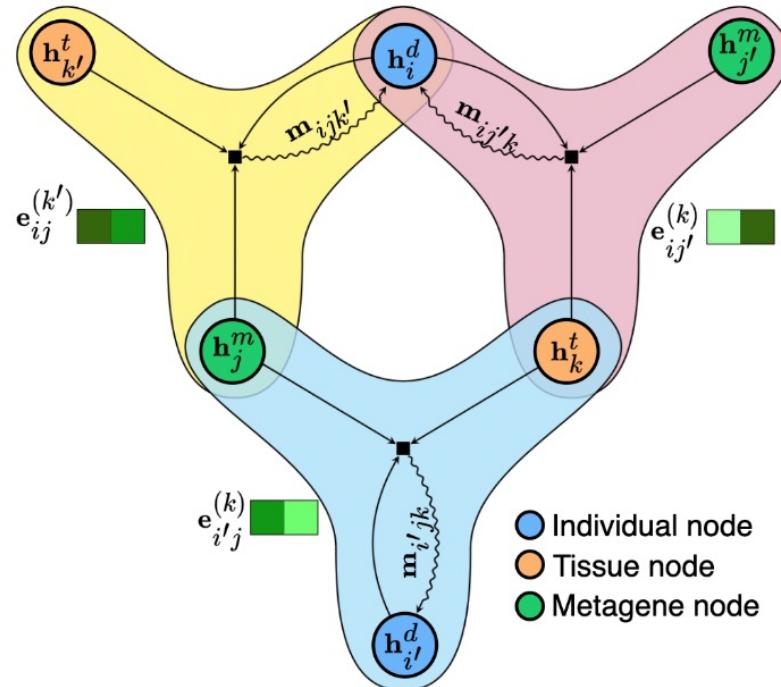
$$\mathbf{m}_{ijk} = \text{MLP}(\mathbf{h}_i^d, \mathbf{h}_j^m, \mathbf{h}_k^t, \mathbf{e}_{ij}^{(k)})$$

- ▶ Aggregate

$$\mathbf{m}_i = \sum_{j=1}^M \sum_{k \in \mathcal{T}(i)} \phi(\mathbf{h}_j^m, \mathbf{h}_k^t, \mathbf{m}_{ijk})$$

- ▶ Update

$$\hat{\mathbf{h}}_i^d = \text{MLP}(\mathbf{h}_i^d, \mathbf{m}_i)$$



Message passing updates representations of individual nodes

Inductive biases to exploit shared regulatory architecture

Message aggregation

► Aggregate

$$\mathbf{m}_i = \sum_{j=1}^M \sum_{k \in \mathcal{T}(i)} \phi(\mathbf{h}_j^m, \mathbf{h}_k^t, \mathbf{m}_{ijk})$$

1. Sum aggregation:

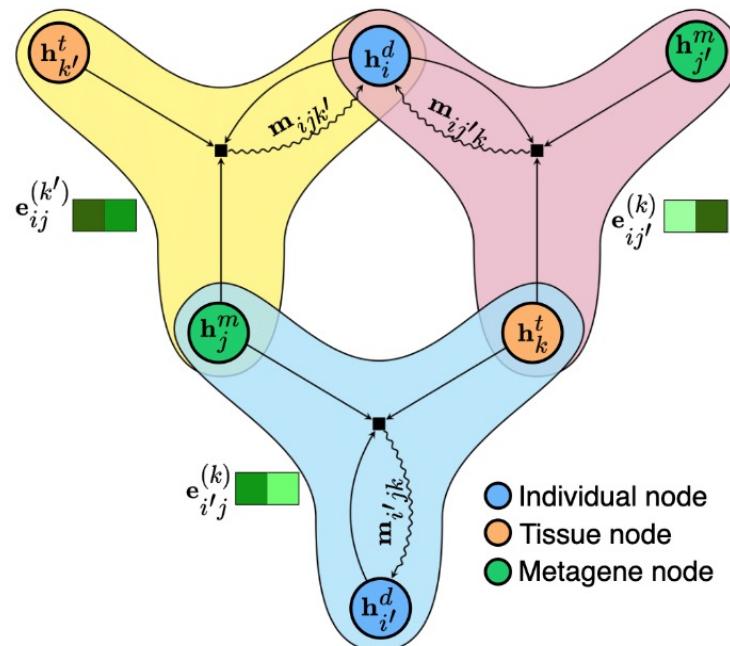
$$\phi(\mathbf{h}_j^m, \mathbf{h}_k^t, \mathbf{m}_{ijk}) = \mathbf{m}_{ijk}$$

2. Attention-based:

$$\phi(\mathbf{h}_j^m, \mathbf{h}_k^t, \mathbf{m}_{ijk}) = \alpha_{jk} \mathbf{m}_{ijk}$$

$$\alpha_{jk} = \frac{\exp e(\mathbf{h}_j^m, \mathbf{h}_k^t)}{\sum_v \exp e(\mathbf{h}_v^m, \mathbf{h}_k^t)}$$

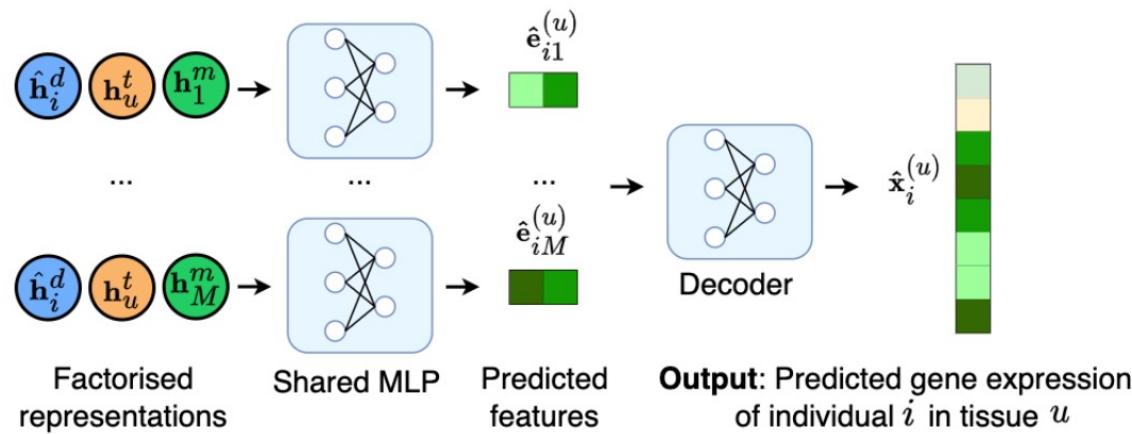
$$e(\mathbf{h}_j^m, \mathbf{h}_k^t) = \mathbf{a}^\top \text{LeakyReLU}(\mathbf{W}[\mathbf{h}_j^m || \mathbf{h}_k^t])$$



Message passing updates representations of individual nodes

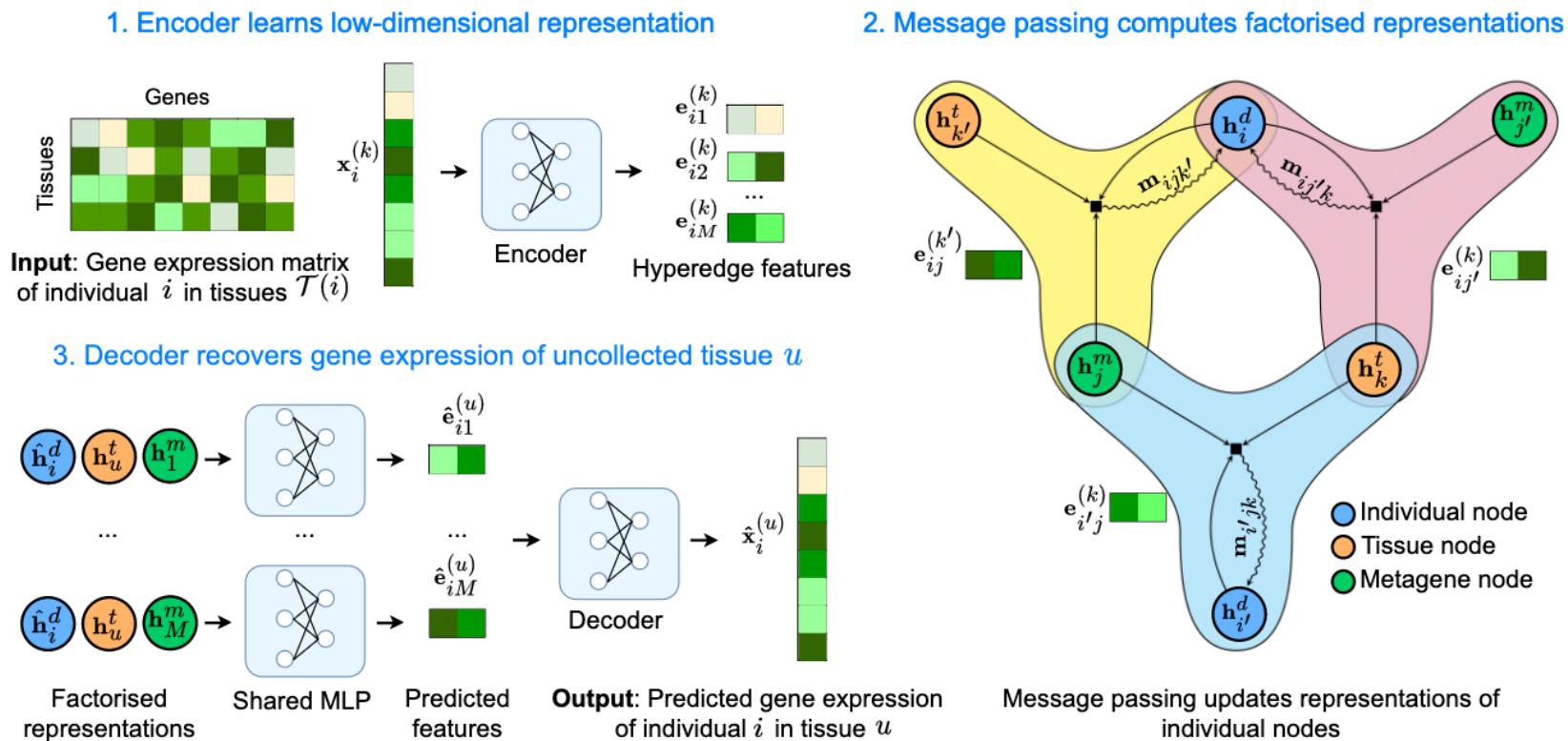
Gene expression imputation for uncollected tissue

► Uncollected tissue u



► Hyperedge-level prediction task

Hypergraph factorisation (HYFA)



Credits to Ramon Viñas

Hypergraph factorisation

Advantages

- ▶ Integrates **multi-tissue** gene expression information
- ▶ Variable number of reference tissues per individual
- ▶ Inductive biases to exploit shared regulatory architecture
- ▶ Parameter-efficient approach
- ▶ Genotype-agnostic

Limitations

- ▶ Assumes same set of genes per tissue
- ▶ Does not support technical batches (yet)

The Genotype-Tissue Expression project (GTEx)

Statistics

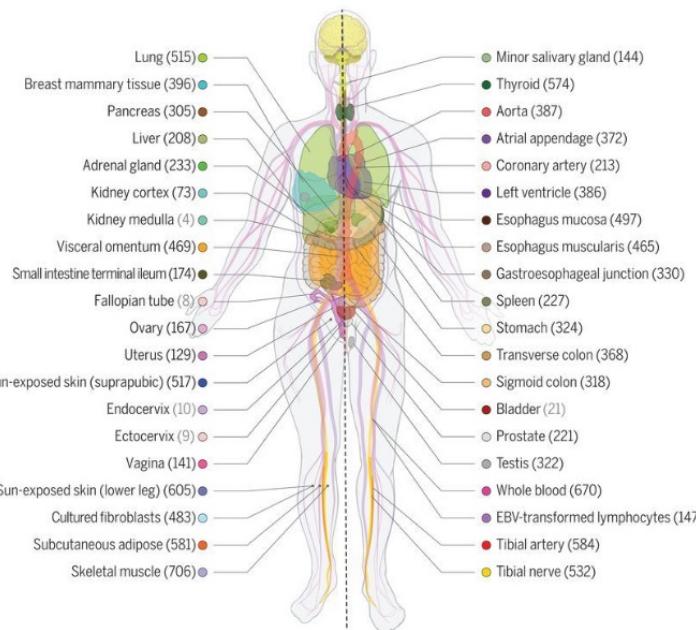
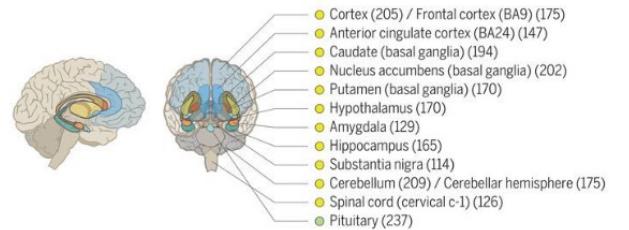
- ▶ 54 tissues
- ▶ 838 individuals

Modalities

- ▶ Gene expression
- ▶ Genotype
- ▶ Histology
- ▶ ...

Applications

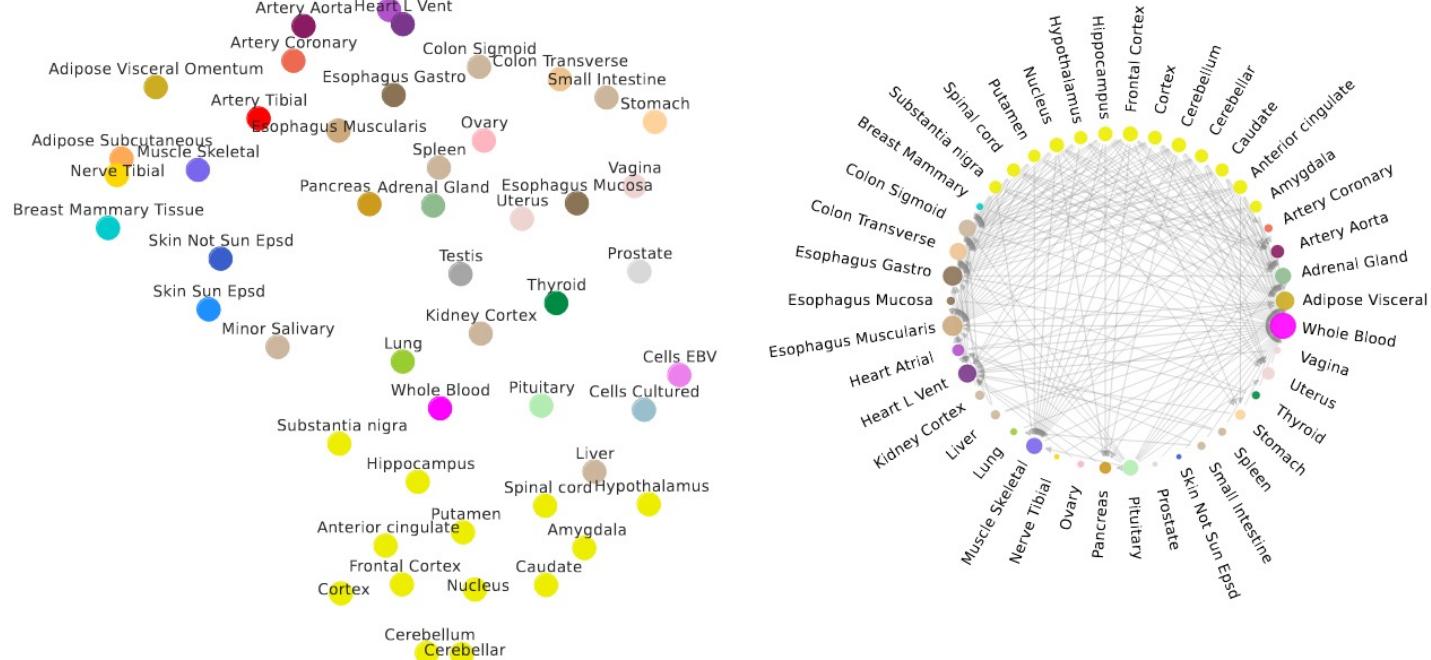
- ▶ eQTL studies
- ▶ Population genetics
- ▶ Genomic-histology associations
- ▶ Cell-type composition
- ▶ ...



The GTEx Consortium atlas of genetic regulatory effects across human tissues, *Science* 2020

Analysis of cross-tissue relationships

- ▶ Elucidate coordinated gene regulation



(a) UMAP representation of the learnt tissue embeddings

(b) Predictability of target tissues measured by the average per-sample Pearson ρ correlation coefficients ($\rho > 0.5$)

Credits to Ramon Viñas

Imputation with whole blood as reference

- ▶ Performance using a single reference tissue (whole blood)

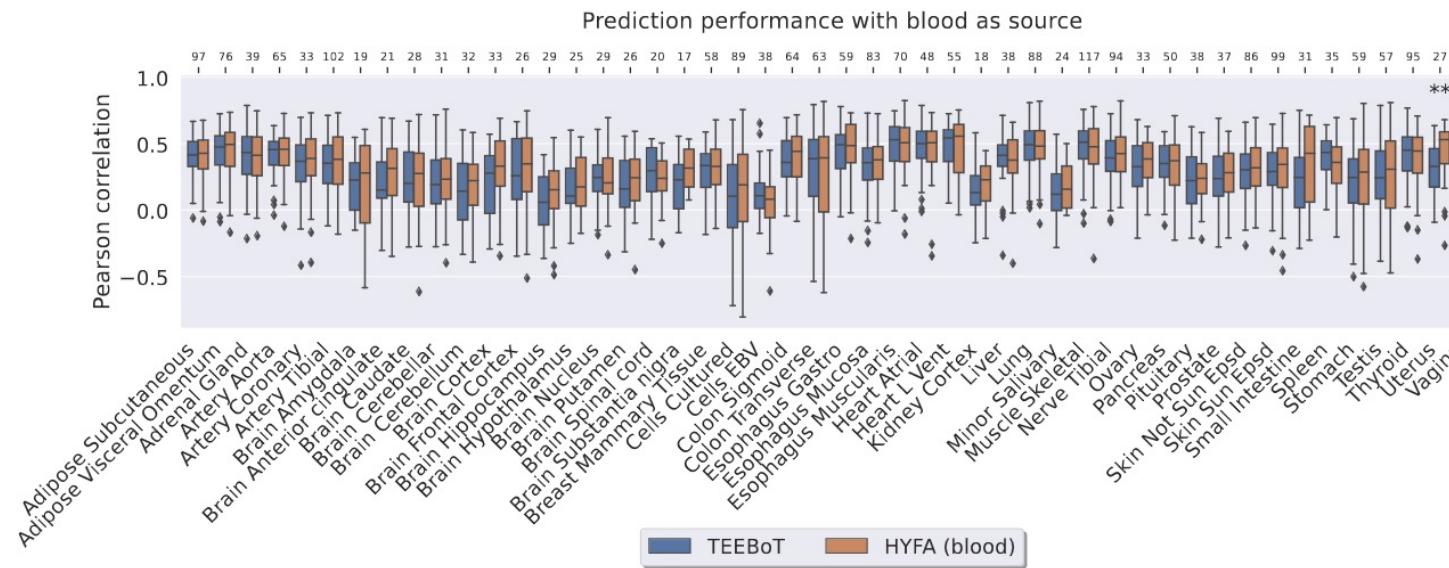


Figure: Performance comparison across gene expression imputation methods.
Per-tissue comparison between HYFA and TEEBoT when using whole-blood and as reference.

Credits to Ramon Viñas

Imputation with accessible tissues as reference

- ▶ Multiple reference tissues (accessible tissues)
- ▶ HYFA outperformed TEEBoT in 38/38 tissues (statistically significant in 26/38 tissues)

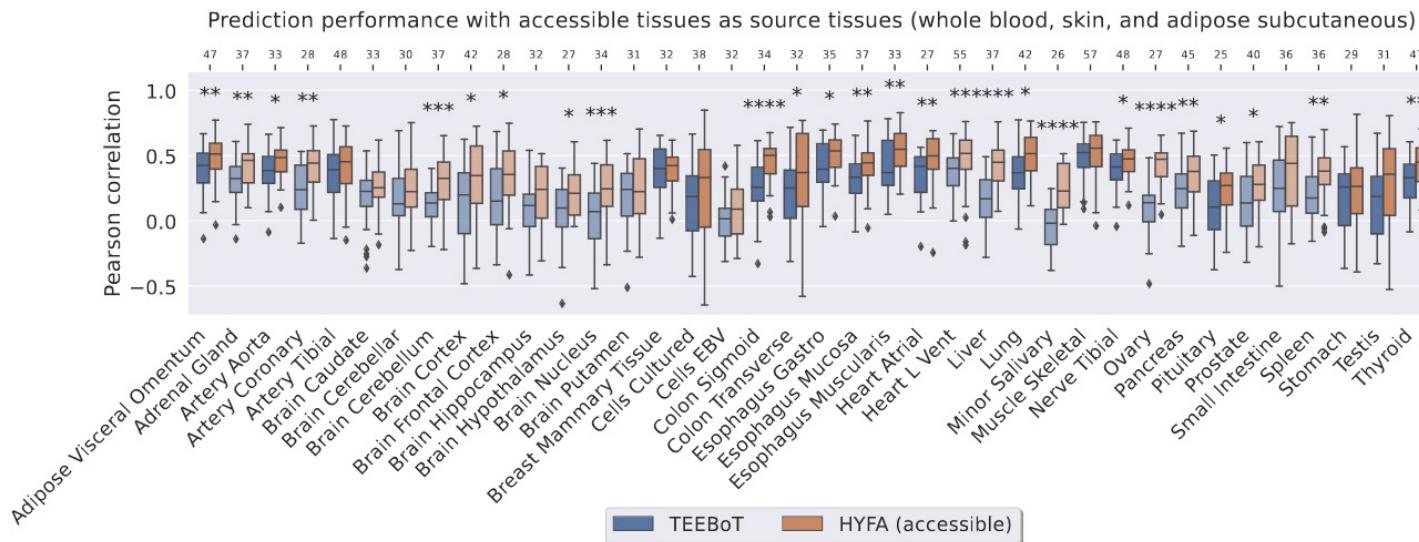


Figure: Per-tissue comparison between HYFA and TEEBoT when using all accessible tissues (whole blood, skin sun exposed, skin not sun exposed, and adipose subcutaneous) as reference. Translucent bars indicate test & validation individuals.

Credits to Ramon Viñas

Background: eQTL discovery

- ▶ Statistical method used to identify genetic variants associated with gene expression levels.

Cis-eQTL

SNP X has an effect on local Gene A



Westra et al., "From genome to function by studying eQTLs", Elsevier 2014

- ▶ SNP: genomic variant at a single base position in the DNA.
- ▶ eQTL: genomic variant associated with variation of gene expression.

Credits to Ramon Viñas

HYFA improves eQTL discovery

- ▶ eQTL datasets are characterised by small sample sizes.
- ▶ Imputed expression of all uncollected tissues with HYFA.

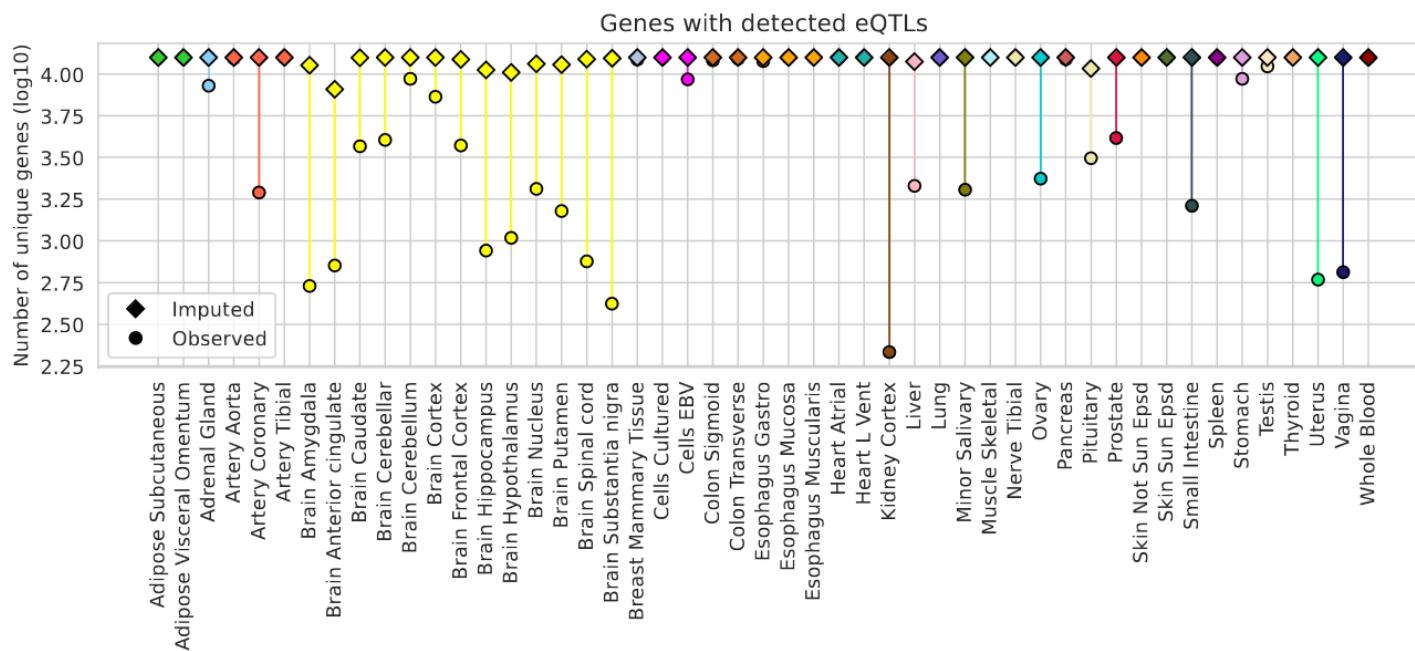


Figure: Number of unique genes with detected eQTLs ($FDR < 0.1$) on observed (circle) and full (observed plus imputed; rhombus) GTEx data. Note logarithmic scale of y-axis.

HYFA improves eQTL discovery

- Detected eQTLs increased for small sample size tissues.

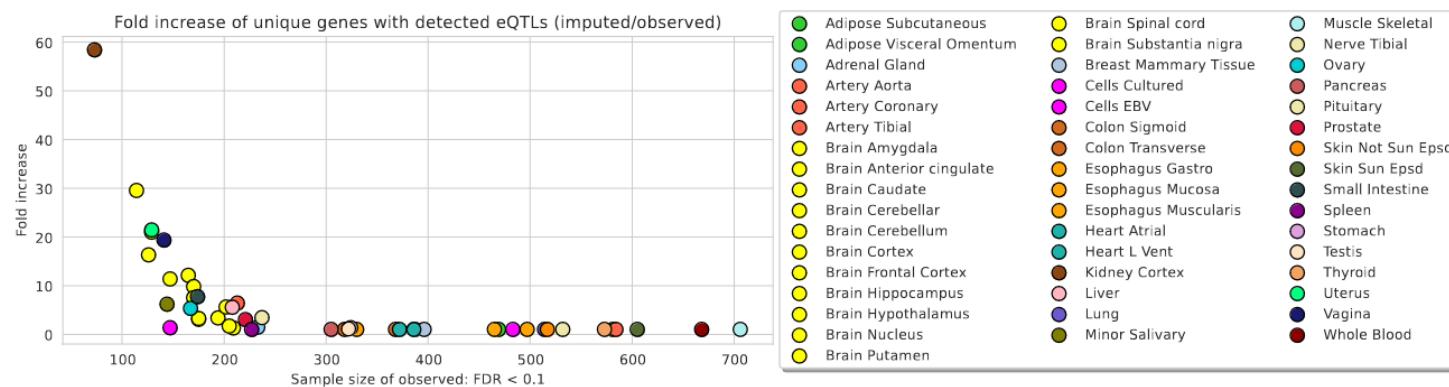


Figure: Fold increase in number of unique genes with mapped eQTLs (y-axis) versus observed sample size (x-axis).

- Results highlight effectiveness of HYFA's inductive biases.

HYFA improves eQTL discovery

- ▶ Enrichment of experimentally validated eQTLs.

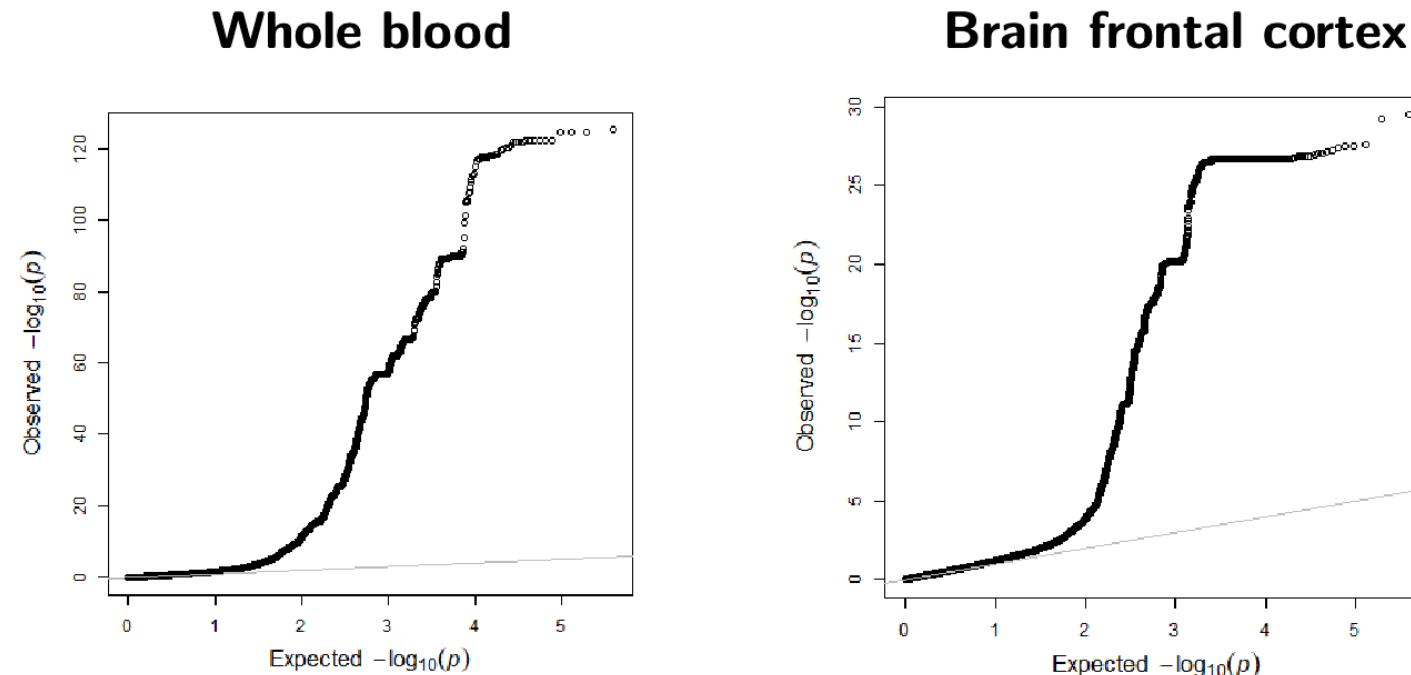


Figure: HYFA recovers experimentally validated expression Quantitative Trait Loci (eQTLs). Quantile-quantile plot showing the causal variants' association with gene expression in blood (**left**) and brain frontal cortex (**right**) in the HYFA-derived dataset using experimentally validated causal variant data from the application of Massively Parallel Reporter Assay [1].

Credits to Ramon Viñas

HYFA improves eQTL discovery

- Detected eQTLs increased for low sample size tissues.

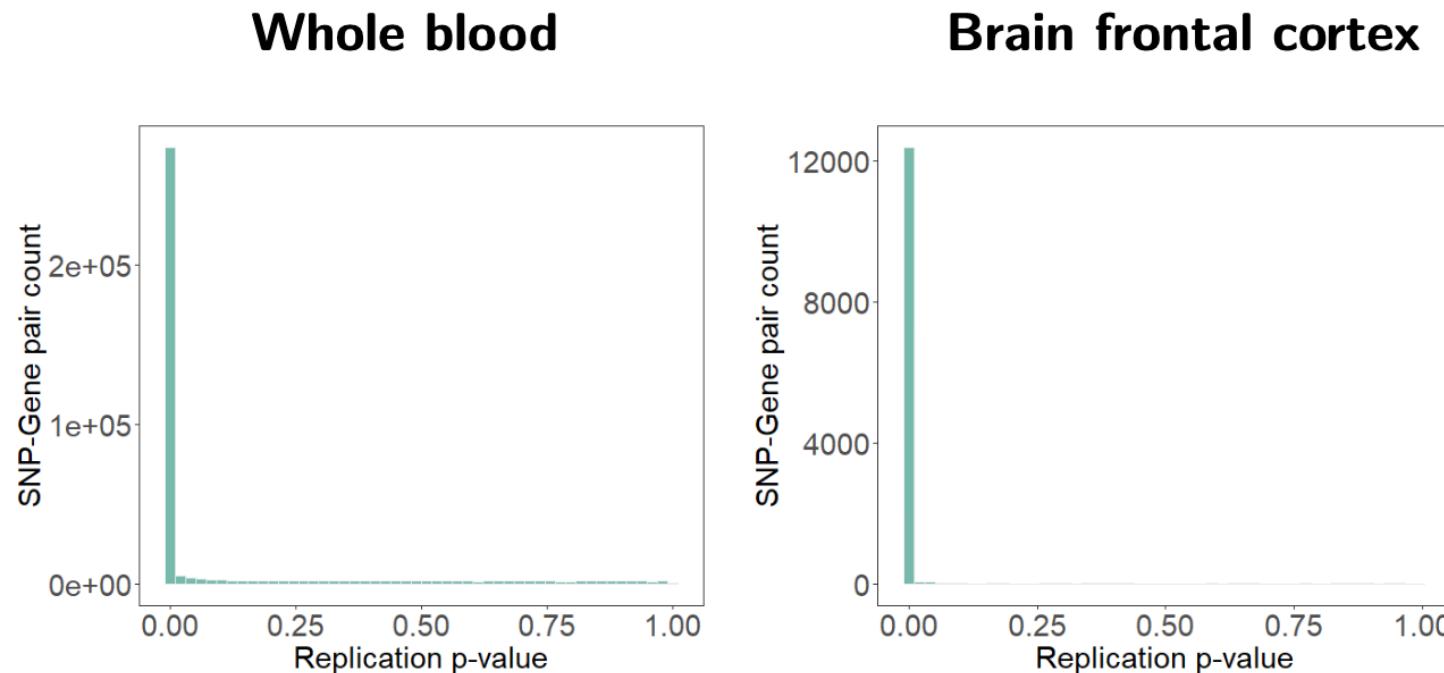


Figure: HYFA recovers replicable expression Quantitative Trait Loci (eQTLs).
Histogram of replication p-values among the HYFA-identified cis-eQTLs for whole blood (**left**) and brain frontal cortex (**right**).

Credits to Ramon Viñas

Metagene-factor enrichment analyses

- ▶ The most enriched types of terms were signaling pathways.

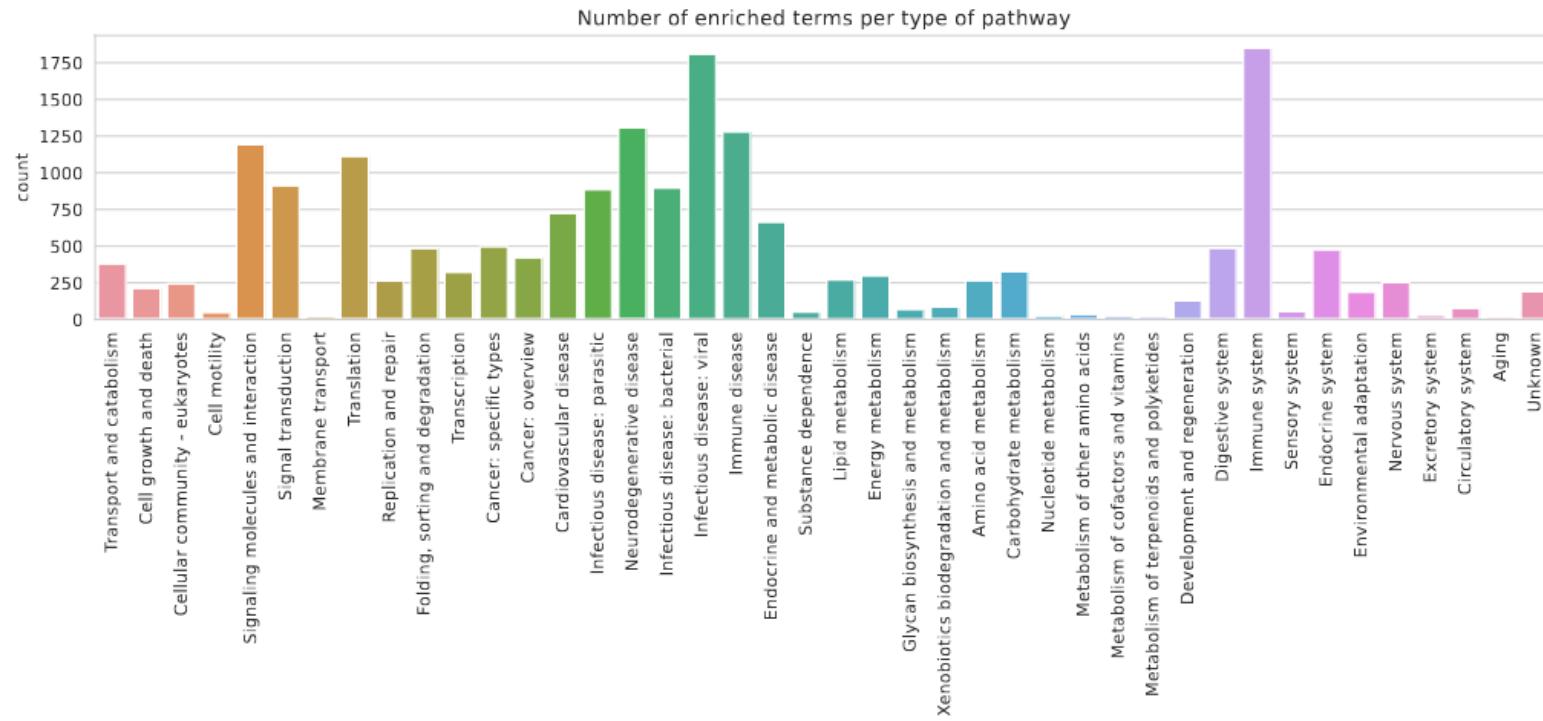


Figure: Total number of enriched terms for each type of pathway.

Metagene-factor enrichment analyses

- We identified metagene-factors related to pathways of neurodegeneration.

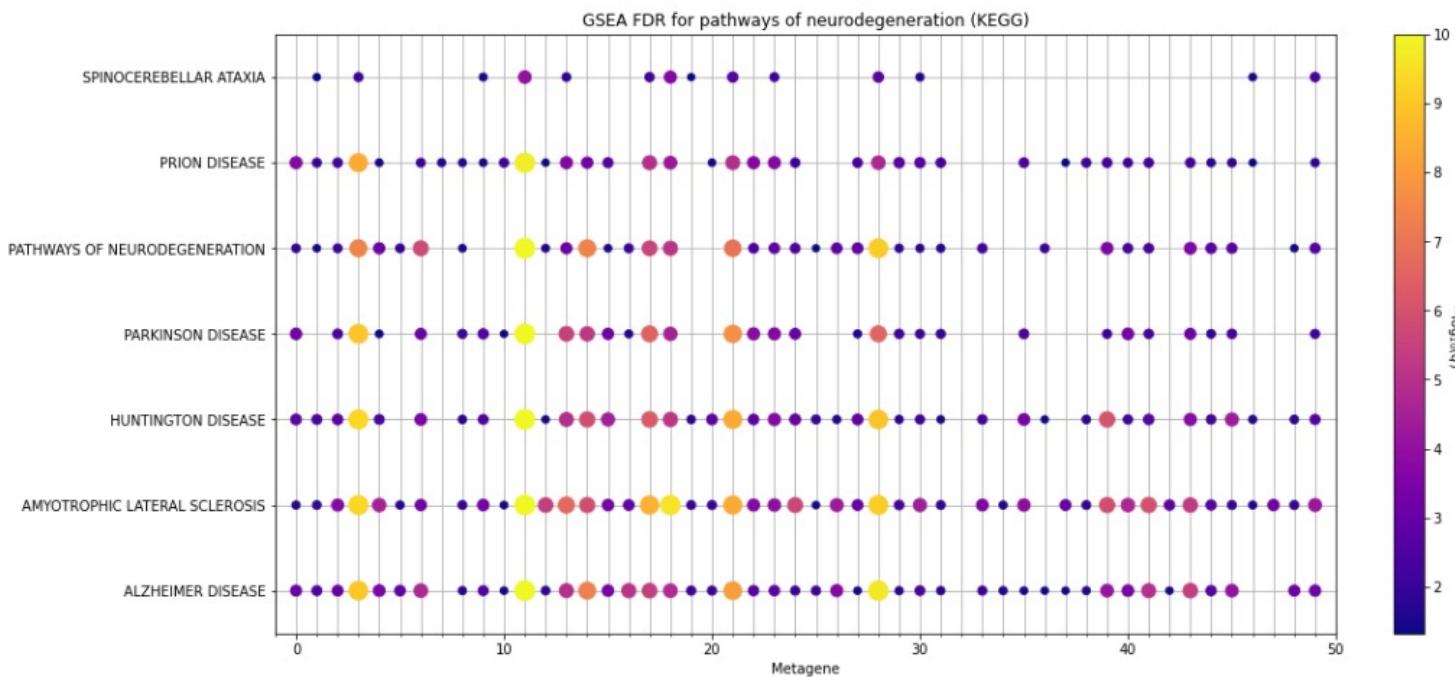


Figure: FDR for pathways of neurodegeneration. For every pathway and metagene, we selected the factor with lowest FDR and depicted statistically significant values (FDR < 0.05). Point sizes are proportional to $-\log$ FDR values.

Metagene-factor enrichment analyses

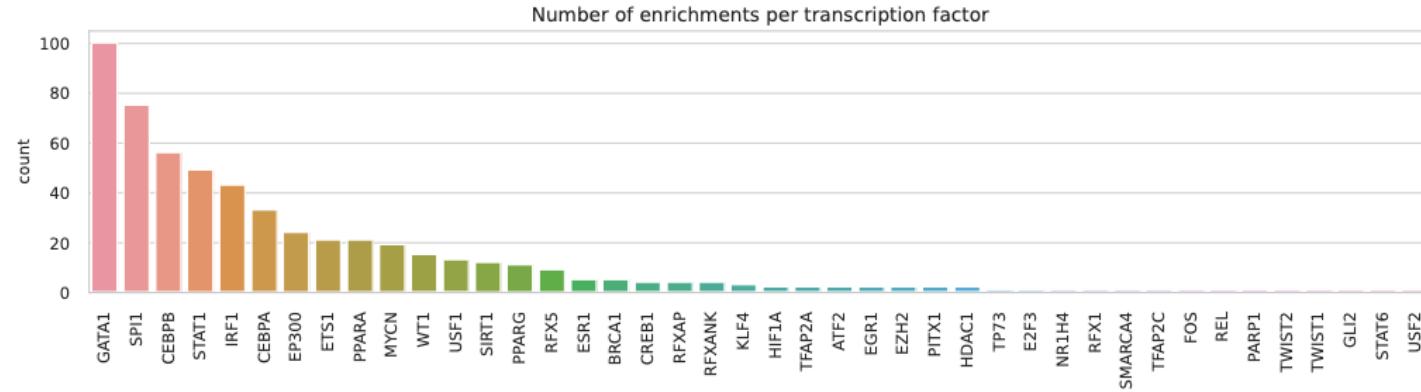


Figure: Top enriched transcription factors (TFs), ranked by the total number of metagene factors in which the TFs were enriched (FDR < 0.05).

Top enriched transcription factors

- ▶ GATA1: known to regulate the development of red blood cells
- ▶ SPI1: controls hematopoietic cell fate
- ▶ CEBPs: play an important role in the control of tissue-specific gene expression
- ▶ STAT1: drives the expression of *many* target genes

Metagene-factor enrichment analyses

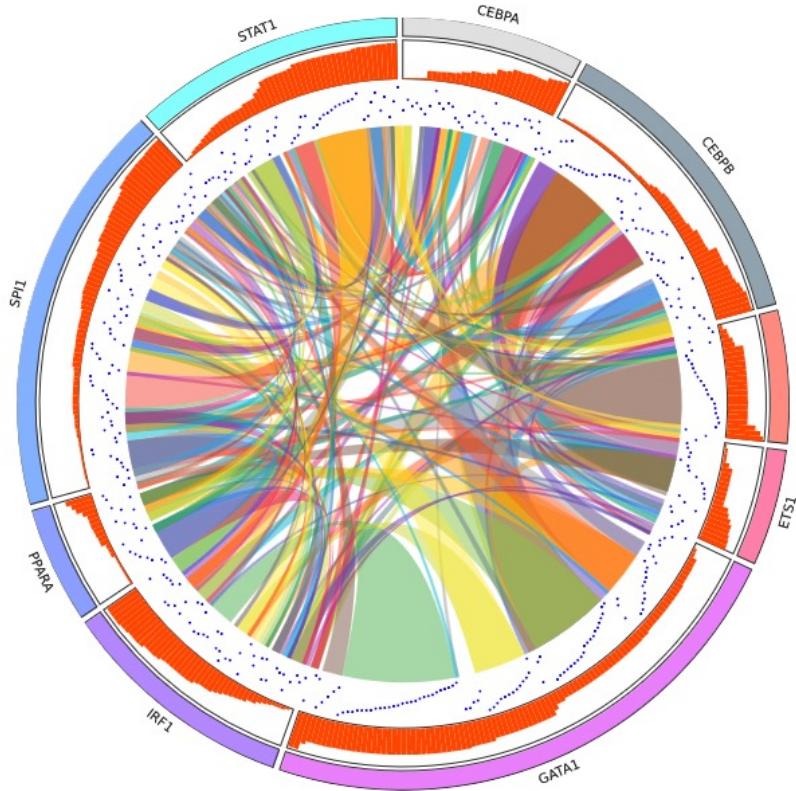


Figure: Circos plot of the top 9 enriched TFs (outer layer). The angular size is proportional to the number of enrichments. The second layer (bar plot) depicts the factor IDs where the TF was enriched, ranging from 0 (lowest bar) to 98 (higher bar). The third layer shows the corresponding metagene IDs (blue dots) of the enriched metagene-factors, increasing monotonically within the same factor. The edges in the middle connect TFs whenever they are both enriched in the same factor ($FDR < 0.05$).

Metagene-factor enrichment analyses

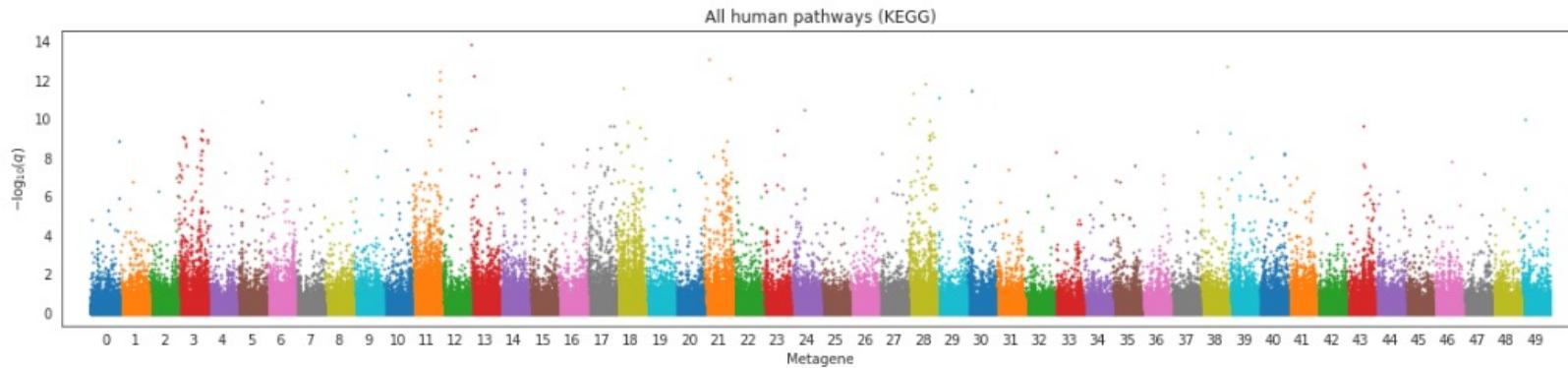


Figure: Pathway enrichment analysis of metagene factors. Manhattan plot of the GSEA results on the metagenes ($n=50$) and factors ($n=98$) learned by HYFA. The x-axis represents metagenes (colored bins) and every offset within the bin corresponds to a different factor. The y-axis is the $-\log q$ -value (FDR) from the GSEA permutation test, corrected for multiple testing via the Benjamini-Hochberg procedure.

Credits to Ramon Viñas

Conclusions

Key takeaways

- ▶ Proposed an approach for joint multi-tissue and cell-type gene expression imputation.
- ▶ Similar performance to TEEBoT for single-reference tissue scenario.
- ▶ Outperforms existing approaches when multiple reference tissues are available.
- ▶ Results uncover a large number of previously undetected tissue-specific eQTLs.
- ▶ HYFA's learned metagene-factors are amenable to biological interpretation.

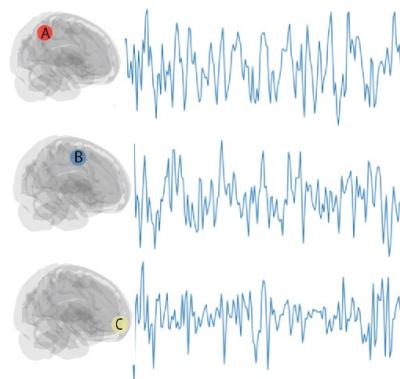
Future work

- ▶ Applications in other fields beyond computational genomics?

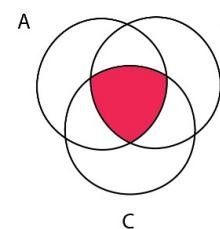
Complexing Hypergraphs using Sheaf

Hypergraph Neural Networks seem appropriate modelling for many natural phenomena

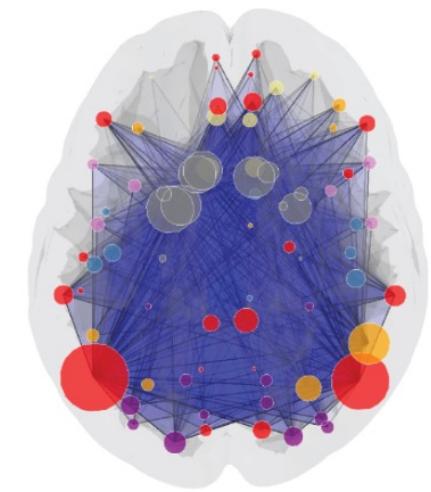
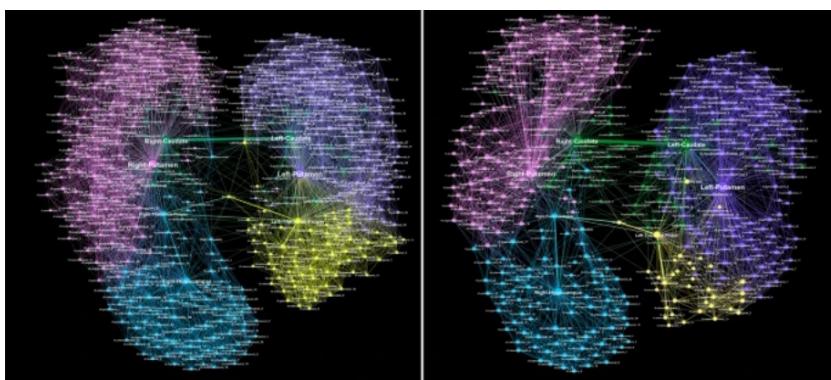
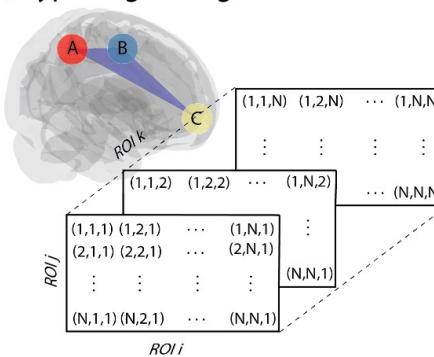
i) multivariate time series



ii) multivariate similarity metric



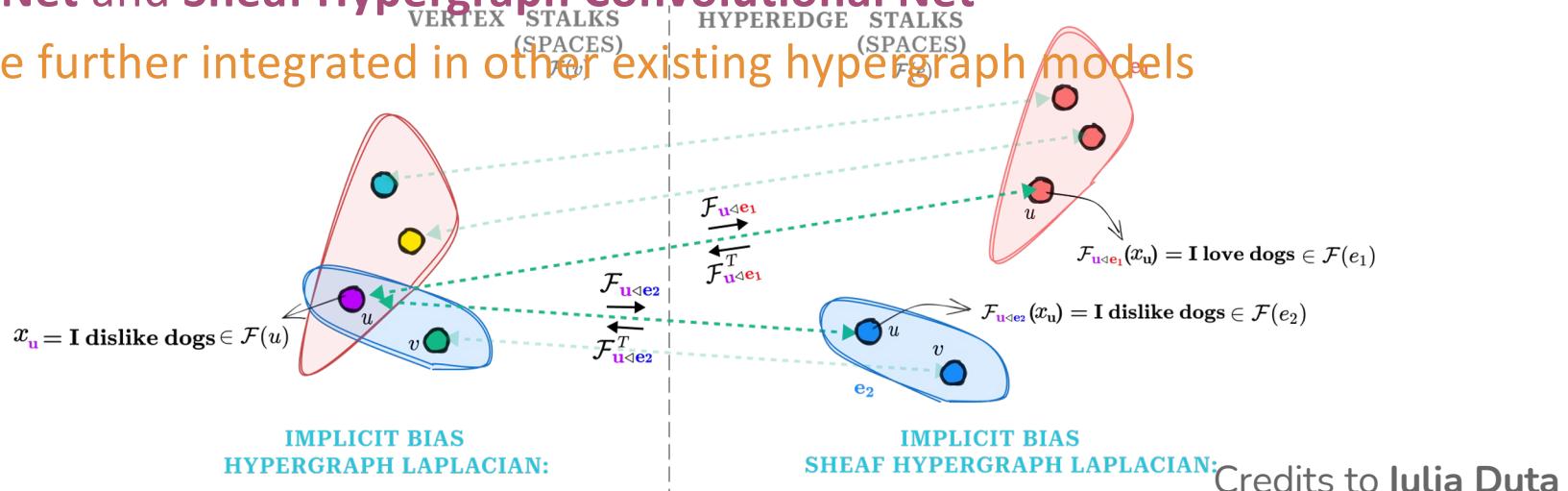
iii) hyperedges weights



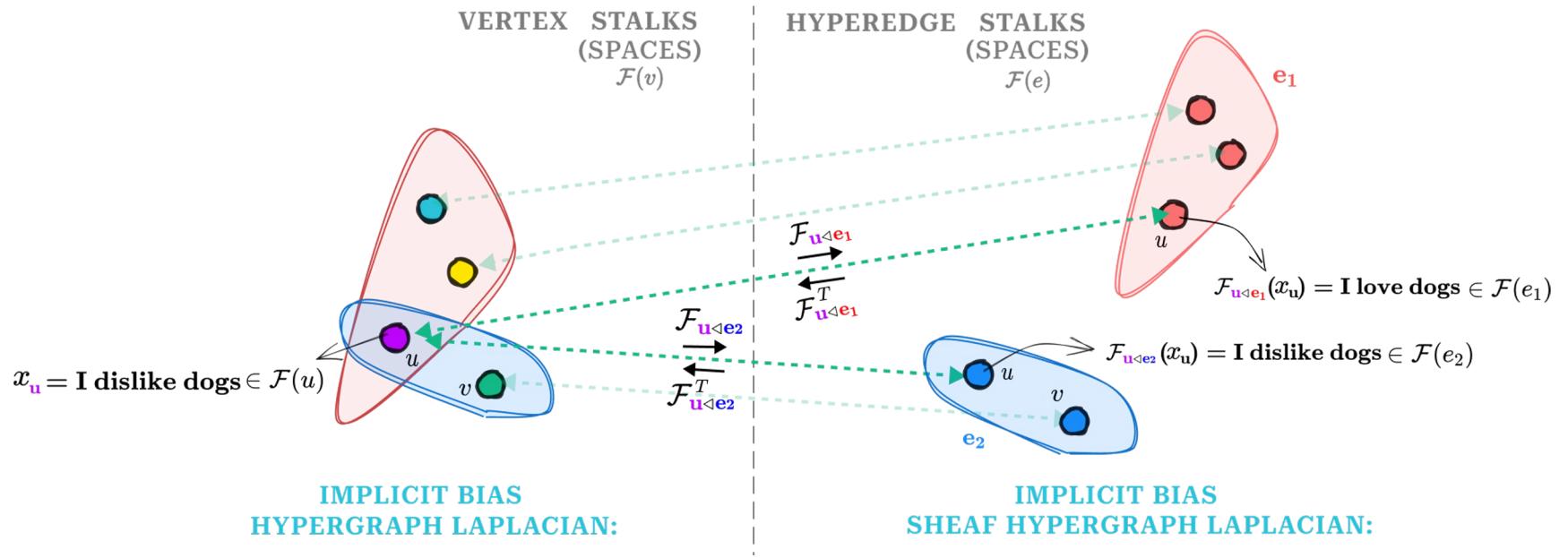
Santos et al. Emergence of High-Order Functional Hubs in the Human Brain, ted February 12, 2023. 10.1101/2023.02.10.528083
Andjelković et al. The topology of higher-order complexes associated with brain hubs in human connectomes

Sheaf Hypergraph Neural Network

- introduce **cellular sheaf for hypergraphs**
- propose both a **linear and a non-linear sheaf hypergraph Laplacian**
- theoretical characterize the **inductive biases generated by the diffusion processes** of these Laplacians
- propose two novel architectures for hypergraph processing: **Sheaf Hypergraph Neural Net** and **Sheaf Hypergraph Convolutional Net**
- could be further integrated in other existing hypergraph models



Cellular Sheaf for Hypergraph

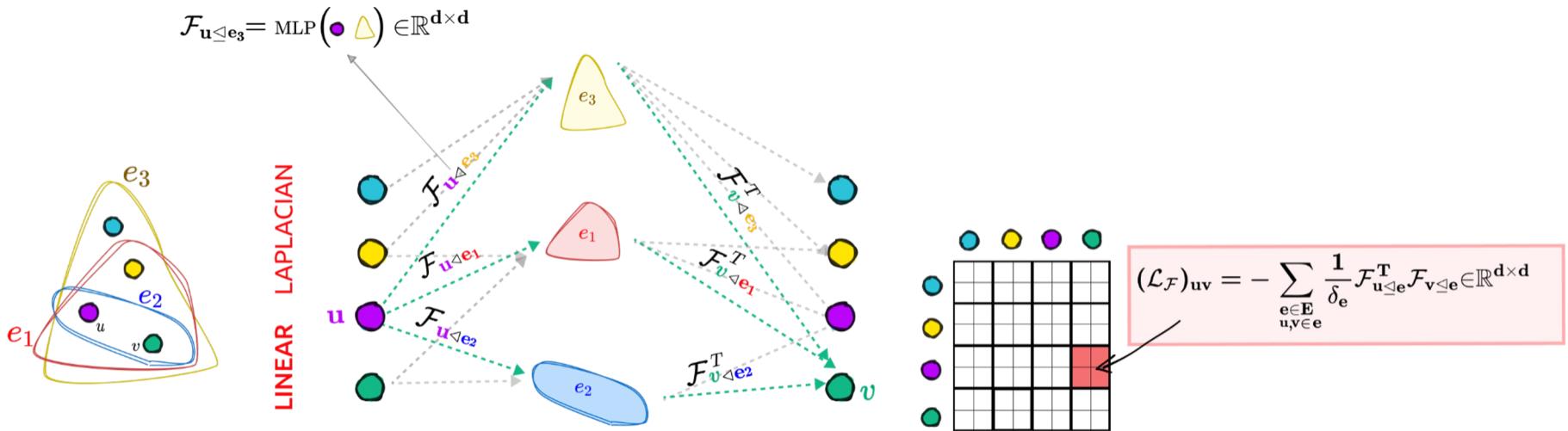


A cellular sheaf associated with a hypergraph constitutes of

- 1) **vertex stalks**: spaces associated with each vertex
- 2) **edge stalks**: spaces associated with each hyperedge
- 3) **restriction maps**: linear projection between (node, hyperedge) incident pairs

Credits to Iulia Duta

Linear Sheaf Hypergraph Laplacian



Theorem 1: Diffusion process using Linear Sheaf Laplacian encourages a form of hyperedge agreement in a more complex space, while preventing the nodes to become uniform.

IMPLICIT BIAS
HYPERGRAPH LAPLACIAN:

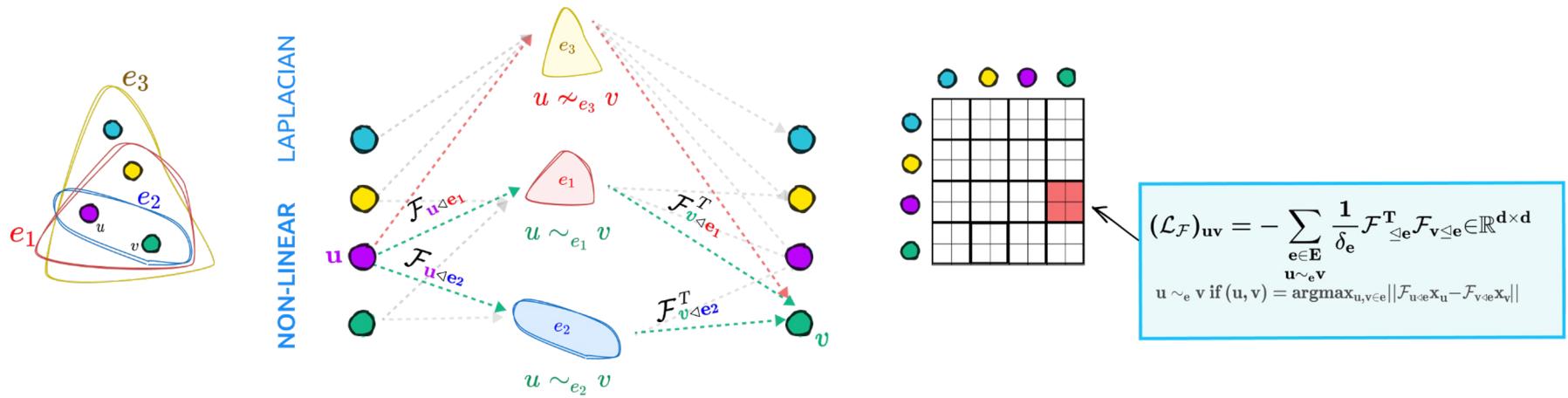
$$E_{L_2}(x) = \frac{1}{2} \sum_e \frac{1}{\delta_e} \sum_{u, v \in e} \|D_v^{-\frac{1}{2}} x_v - D_u^{-\frac{1}{2}} x_u\|_2^2$$

IMPLICIT BIAS
SHEAF HYPERGRAPH LAPLACIAN:

$$E_{L_2}^{\mathcal{F}}(x) = \frac{1}{2} \sum_e \frac{1}{\delta_e} \sum_{u, v \in e} \|\mathcal{F}_{v \triangle e} D_v^{-\frac{1}{2}} x_v - \mathcal{F}_{u \triangle e} D_u^{-\frac{1}{2}} x_u\|_2^2$$

Credits to Iulia Duta

Non-Linear Sheaf Hypergraph Laplacian



Theorem 2: Diffusion using Non-Linear Sheaf Laplacian minimize the total variance in the stalk space associated with each hyperedge, preventing the features to become uniform.

IMPLICIT BIAS
HYPERGRAPH LAPLACIAN:

$$\bar{E}_{TV}(x) = \frac{1}{2} \sum_e \frac{1}{\delta_e} \max_{u, v \in e} \|D_v^{-\frac{1}{2}} x_v - D_u^{-\frac{1}{2}} x_u\|_2^2$$

IMPLICIT BIAS
SHEAF HYPERGRAPH LAPLACIAN:

$$\bar{E}_F(x) = \frac{1}{2} \sum_e \frac{1}{\delta_e} \max_{u, v \in e} \|F_{v \leq e} D_v^{-\frac{1}{2}} x_v - F_{u \leq e} D_u^{-\frac{1}{2}} x_u\|_2^2$$

Credits to Iulia Duta

Sheaf Hypergraph Networks

The two Laplacians leads to **two architectures:**
Sheaf HyperGNN and Sheaf HyperGCN respectively

$$Y = \sigma((I_{nd} - \overset{\bullet}{\Delta})(I_n \otimes W_1)\tilde{X}W_2)$$

Task: semi-supervised node classification on various hypergraph benchmarks

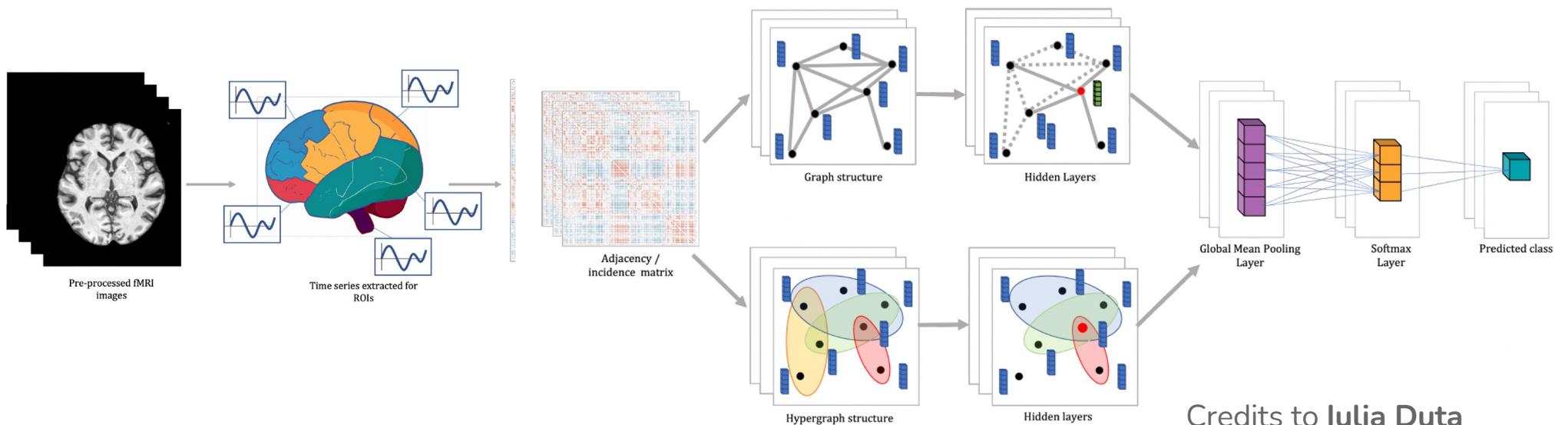
Table 1: **Performance on a collection of hypergraph benchmarks.** Our models using sheaf hypergraph Laplacians demonstrate a clear advantage over their counterparts using classical Laplacians (HyperGNN and HyperGCN). Compared to other recent methods, SheafHyperGNN and SheafHyperGCN achieve competitive performance and attain state-of-the-art results in five of the datasets.

Name	Cora	Citeseer	Pubmed	Cora_CA	DBLP_CA	Senate	House	Congress
HCHA	79.14 \pm 1.02	72.42 \pm 1.42	86.41 \pm 0.36	82.55 \pm 0.97	90.92 \pm 0.22	48.62 \pm 4.41	61.36 \pm 2.53	90.43 \pm 1.20
HNHN	76.36 \pm 1.92	72.64 \pm 1.57	86.90 \pm 0.30	77.19 \pm 1.49	86.78 \pm 0.29	50.93 \pm 6.33	67.8 \pm 2.59	53.35 \pm 1.45
AllDeepSets	76.88 \pm 1.80	70.83 \pm 1.63	88.75 \pm 0.33	81.97 \pm 1.50	91.27 \pm 0.27	48.17 \pm 5.67	67.82 \pm 2.40	91.80 \pm 1.53
AllSetTransformers	78.58 \pm 1.47	73.08 \pm 1.20	88.72 \pm 0.37	83.63 \pm 1.47	91.53 \pm 0.23	51.83 \pm 5.22	69.33 \pm 2.20	92.16 \pm 1.05
UniGCNII	78.81 \pm 1.05	73.05 \pm 2.21	88.25 \pm 0.33	83.60 \pm 1.14	91.69 \pm 0.19	49.30 \pm 4.25	67.25 \pm 2.57	94.81 \pm 0.81
HyperND	79.20 \pm 1.14	72.62 \pm 1.49	86.68 \pm 0.43	80.62 \pm 1.32	90.35 \pm 0.26	52.82 \pm 3.20	51.70 \pm 3.37	74.63 \pm 3.62
ED-HNN	80.31 \pm 1.35	73.70 \pm 1.38	89.03 \pm 0.53	83.97 \pm 1.55	91.90 \pm 0.19	64.79 \pm 5.14	72.45 \pm 2.28	95.00 \pm 0.99
HyperGCN ^[2]	78.36 \pm 2.01	71.01 \pm 2.21	80.81 \pm 12.4	79.50 \pm 2.11	89.42 \pm 0.16*	51.13 \pm 4.15	69.29 \pm 2.05	89.67 \pm 1.22
SheafHyperGCN	80.06 \pm 1.12	73.27 \pm 0.50	87.09 \pm 0.71	83.26 \pm 1.20	90.83 \pm 0.23	66.33 \pm 4.58	72.66 \pm 2.26	90.37 \pm 1.52
HyperGNN	79.39 \pm 1.36	72.45 \pm 1.16	86.44 \pm 0.44	82.64 \pm 1.65	91.03 \pm 0.20	48.59 \pm 4.52	61.39 \pm 2.96	91.26 \pm 1.15
SheafHyperGNN	81.30 \pm 1.70	74.71 \pm 1.23	87.68 \pm 0.60	85.52 \pm 1.28	91.59 \pm 0.24	68.73 \pm 4.68	73.84 \pm 2.30	91.81 \pm 1.60

Credits to Iulia Duta

Alzheimer's Disease Classification using Hypergraph Neural Networks

- Exploring various ways of extracting higher-order structure based on fMRI
- Process the resulting structure using Hypergraph Network to predict different stages of Alzheimer's Disease



Adding Logic in search of explainability

Automatic decision?

Guidance

Ethics, Transparency and Accountability Framework for Automated Decision-Making

Published 13 May 2021

Contents

- [What the framework is for](#)
- [Understanding automated and algorithmic decision-making](#)
- [Before you use this framework](#)
- [How to use this framework](#)
 - 1. Test to avoid any unintended outcomes or consequences
 - 2. Deliver fair services for all of our users and citizens
 - 3. Be clear who is responsible
 - 4. Handle data safely and protect citizens' interests
 - 5. Help users and citizens understand how it impacts ..

What the framework is for

Context

The ethical considerations of artificial intelligence and automated systems is at the centre of technological advancement.

According to a recent [EU survey](#) and a [British Computer Society survey](#) in the UK, there is a distinct distrust in the regulation of advanced technology. A [review by the Committee on Standards in Public Life](#) found that the government should produce clearer guidance on using artificial intelligence ethically in the public sector.



EN English

Search

Home > ... > Law > Reform > Rules for business and organisations > Dealing with citizens > Are there restrictions on the use of automated decision-making?

Are there restrictions on the use of automated decision-making?

PAGE CONTENTS

- [Answer](#)
- [Example](#)
- [References](#)

Answer

Yes, individuals should not be subject to a decision that is based solely on automated processing (such as algorithms) and that is legally binding or which significantly affects them.

A decision may be considered as producing legal effects when the individual's legal rights or legal status are impacted (such as their right to vote for example). In addition, processing can significantly affect an individual if it influences their personal circumstances, their behaviour or their choices (for example an automatic processing may lead to the refusal of an online credit application).

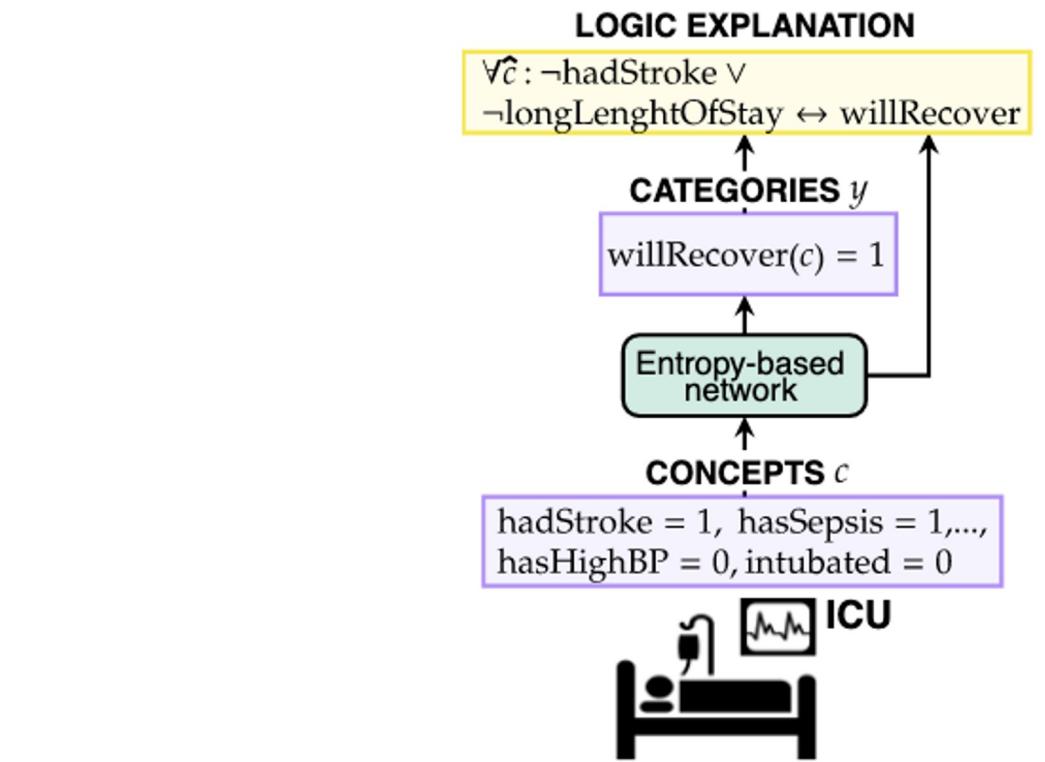
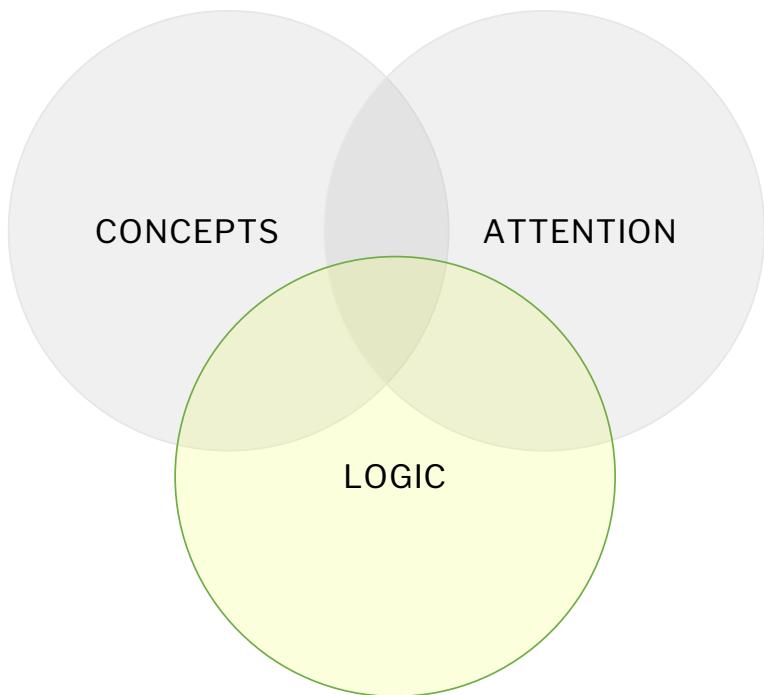
The use of automated processing for decision-making is authorised only in the following cases:

- the decision based on the algorithm is necessary (i.e. there must be no other way to achieve the same goal) to enter into or to perform a contract with the individual whose data your company/organisation processed via the algorithm (for example an online loan application)

Explainability

- Most techniques explaining black boxes focus on finding or ranking the most relevant features used by the black box to make predictions.
- Such “feature-scoring” methods are very efficient and widely used, but they cannot explain how neural networks compose such features to make predictions.
- Concept-based approaches have become increasingly popular as they provide explanations in terms of human-understandable categories (i.e. the “concepts”) rather than raw features

Clinicians-in-the-loop: using logic

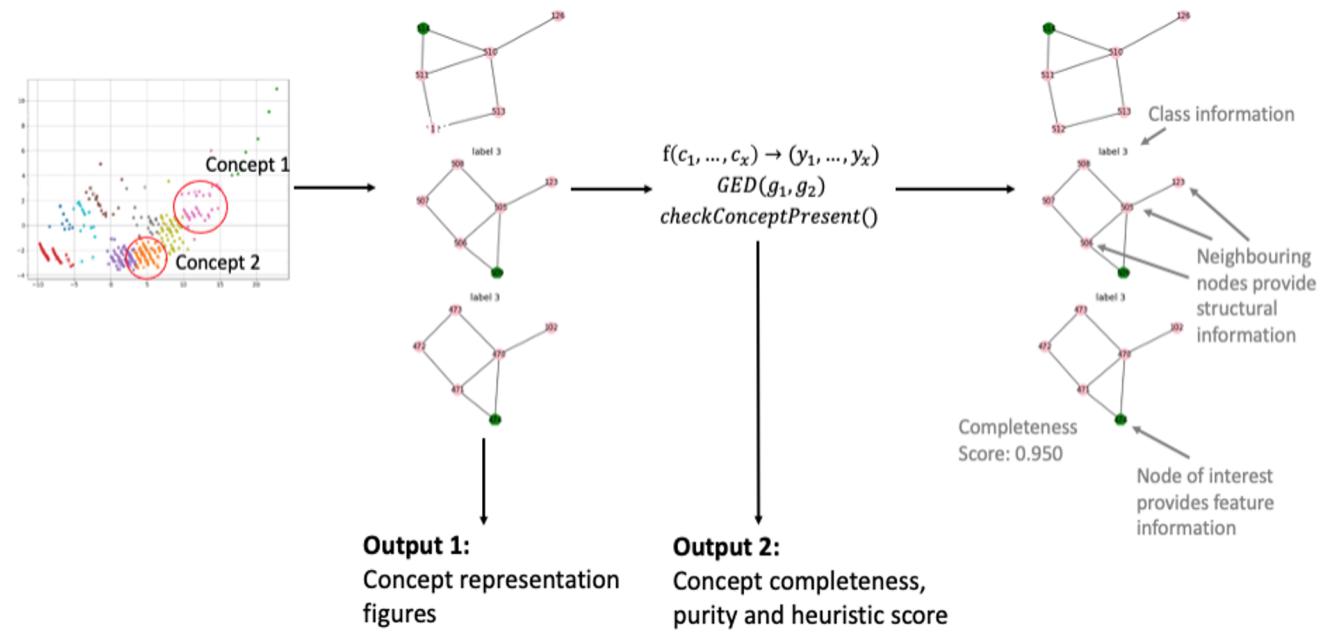


Credits to Pietro Barbiero

Ciravegna, Gabriele, et al. "Human-Driven FOL Explanations of Deep Learning." IJCAI. 2020.

GCEExplainer

1. Cluster activation space with k-Means
 - Cluster = concept
2. Visualise best representations of these clusters
 - Results show nodes with similar neighbourhood structure and features clustered together
3. Evaluate using concept completeness and purity
4. User uses these metrics and visualisations to reason about concept



Credits to Charlotte Magister

Magister, Lucie Charlotte, et al. "GCEExplainer: Human-in-the-Loop Concept-based Explanations for Graph Neural Networks." *arXiv preprint arXiv:2107.11889* (2021).

Logic Explained Networks (LENs)

With Sheaf and hypergraph the deep learning has reached a good level of black box complexity

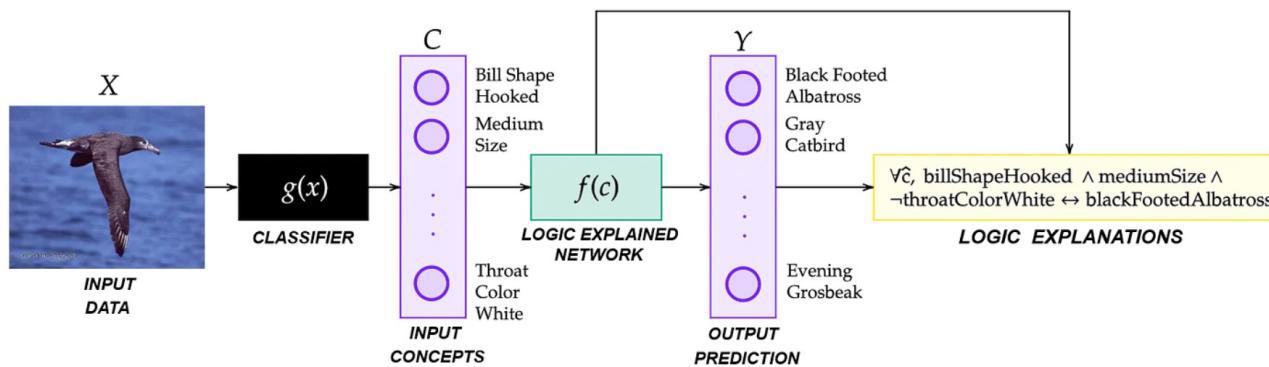
Neural networks cannot explain *how* they arrive to a prediction, hence their deployment in safety-critical applications is discouraged.

Logic Explained Networks are novel “*explainable-by-design*” deep learning models, providing logic explanations of their predictions.

Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

Logic Explained Networks (LENs)

- Logic Explained Networks (or LENs) are a special family of concept-based neural networks providing first-order logic (FOL) explanations for their decisions.



Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

Logic Explained Networks (LENs)

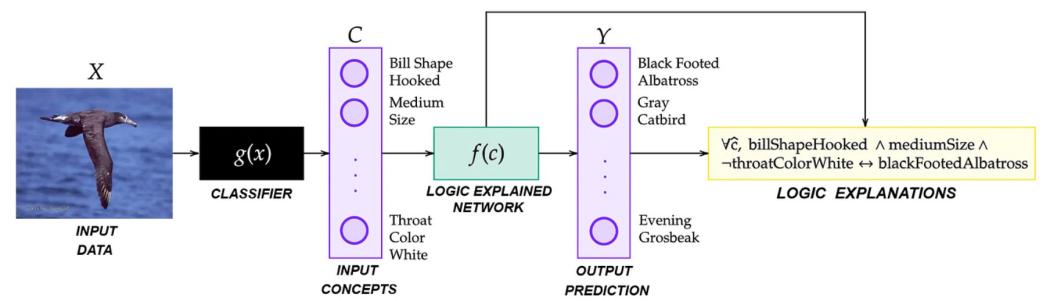
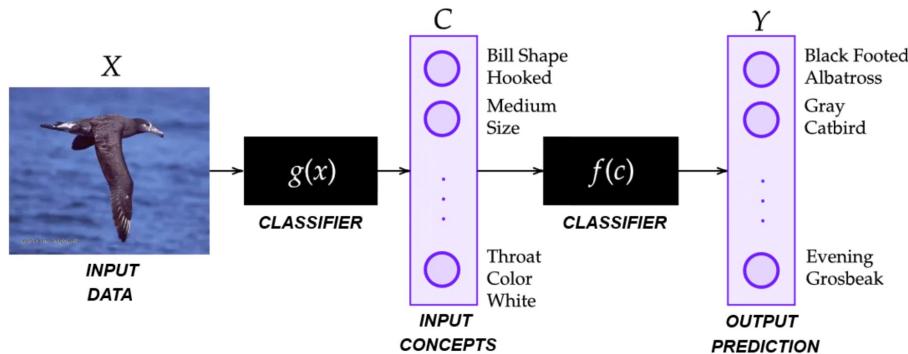
there is a clear distinction between post-hoc methods and the LEN paradigm which is explainable by design.

Post-hoc methods generally do not impose constraints on the classifier: the model itself is free from any constraints related to the explanation method. This is why this class of methods is called post hoc: After the usual training is completed, the XAI method kicks in.

The LENs paradigm instead is explainable by design as it embeds additional constraints both in the architecture and in the loss function, making the network self explainable.

This is why we say that LENs are explainable by design: The classifier itself is constrained to learn in a way that makes explanations emerge automatically.

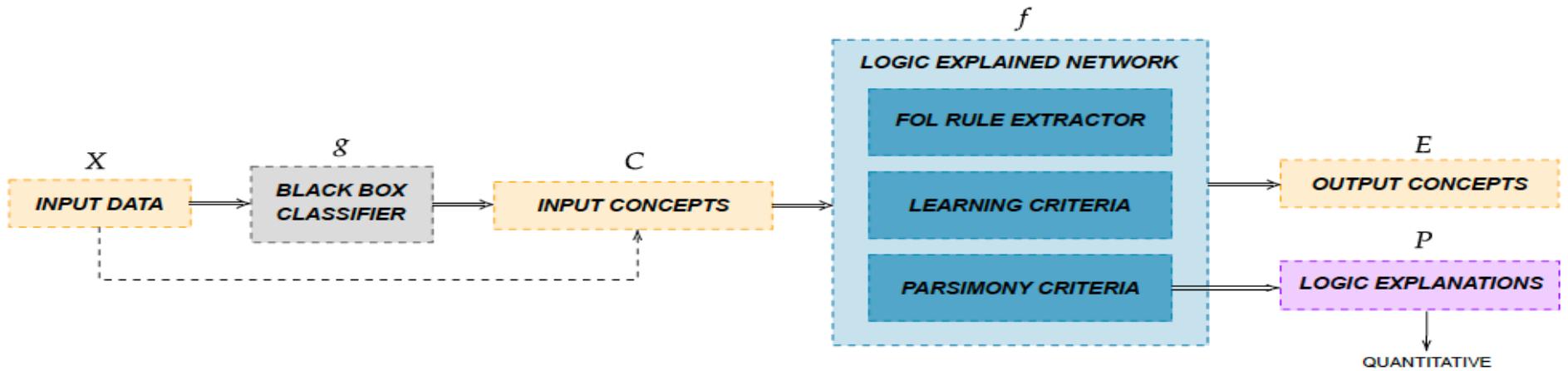
Logic Explained Networks (LENs)



- The classifier g performs the same action (i.e. predicting concepts from images);
- However, the classifier f is now a ***Logic Explained Network***, providing both the predictions for the target classes + logic formulas explaining how the network f leverages the input concepts to arrive to a decision!

Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

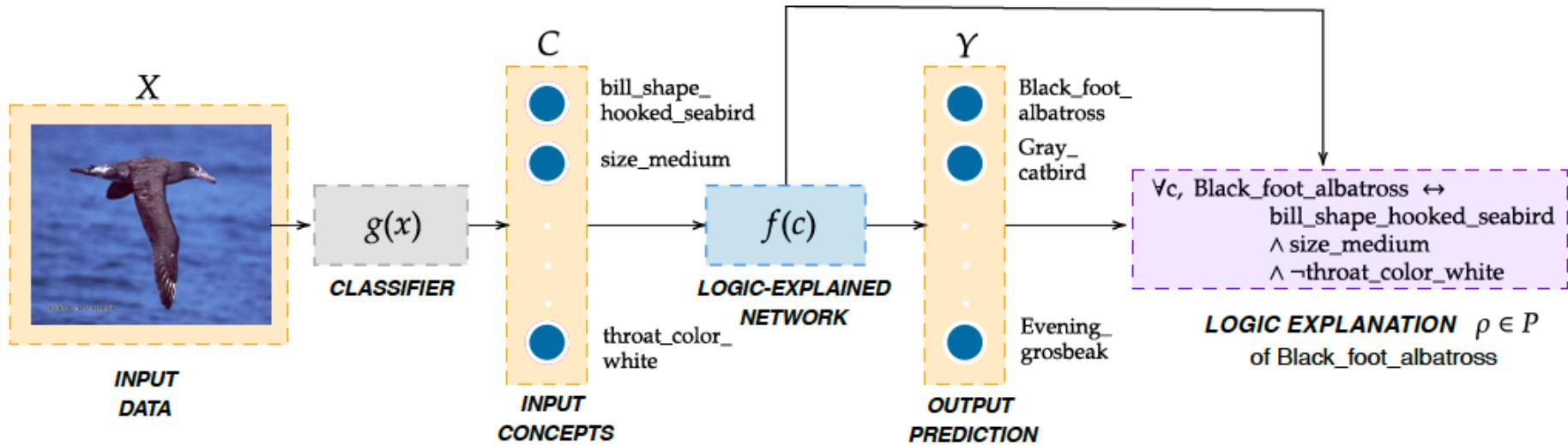
Logic Explained Networks (LENs)



Logic Explained Networks (LENs, f) are neural networks capable of making predictions of a set of output concepts (activation scores belonging to E) and providing First-Order Logic explanations (belonging to P) in function of the LEN inputs. Inputs might be other concepts (activation scores belonging to C) either computed by a neural network (g) or directly provided within the available data (each data sample belongs to X). There are several different ways of instantiating this generic model into real-world problems (Section 3). Within the blue box, the key components of LENs are listed (Section 4).

Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

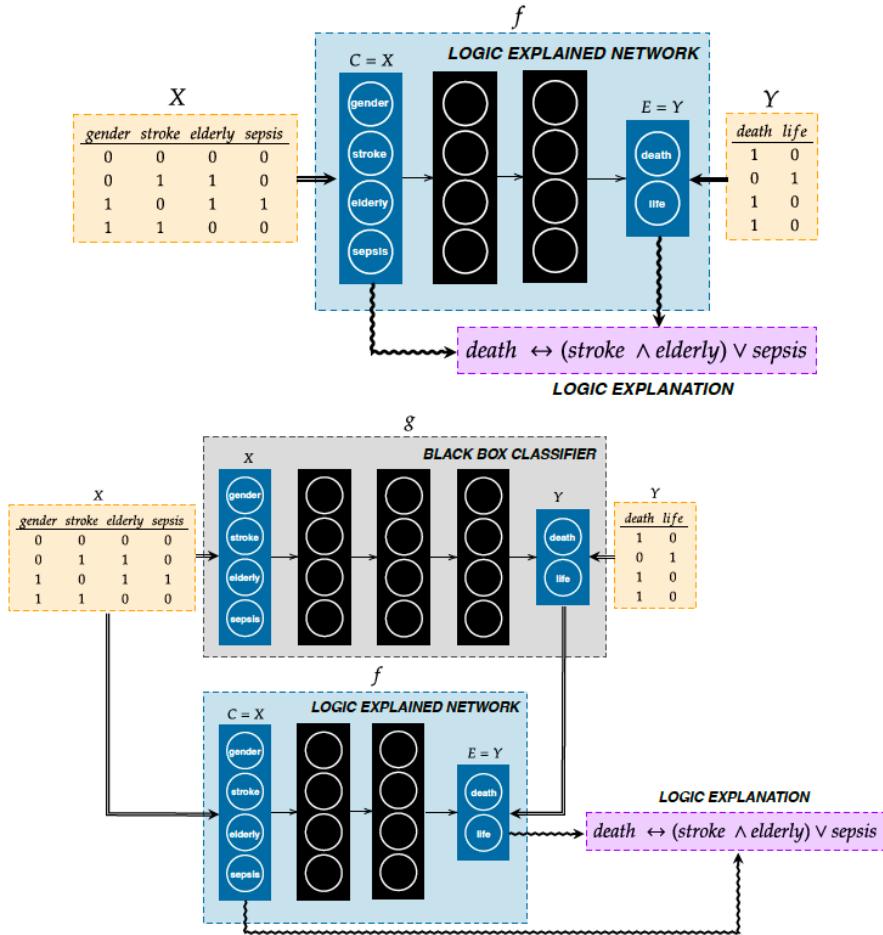
Logic Explained Networks (LENs)



A LEN is placed on top of a convolutional neural network $g(\cdot)$ in order to (i) classify the species of the bird in input and (ii) provide an explanation on why it belongs to this class. The logic explanation in the example showcases the predicted output class (all the output concepts can be explained), dropping the argument of the predicates for compactness

Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

Logic Explained Networks (LENs)



End-to-end (E2E) LENs directly work on input data that are interpretable per se and treated as concepts ($C = X$), while the output concepts are the activation scores of the classes of the dataset .

Bottom: the LEN provides explanations of a black-box classifier. The universal quantifier and the argument of the predicates have been dropped for simplicity.

Ciravegna, Gabriele, et al.“Logic Explained Networks.” *arXiv preprint arXiv:2108.05149* (2021).

Thanks to
Ramon Viñas Torné , Felix Opolka, Charlotte
Magister, Dobrik Georgiev, Pietro Barbiero, Iulia
Duta, Chaitanya Joshi, Gabriele Ciravegna, Petar
Velickovic, Michael Bronstein, Cris Bodnar,
Francesco Prinzi, Marco Nobile, Giulia Capitoli,
Emma Ambags, Yin-Cong Zhi, Emma Rocheteau, Yu
Guang Wang, Charlie Harris, Simon Mathis, Bianca
Dumitrascu, Eric Gamazon, Marco Gori, Xiaowen
Dong, Mihaela van der Schaar