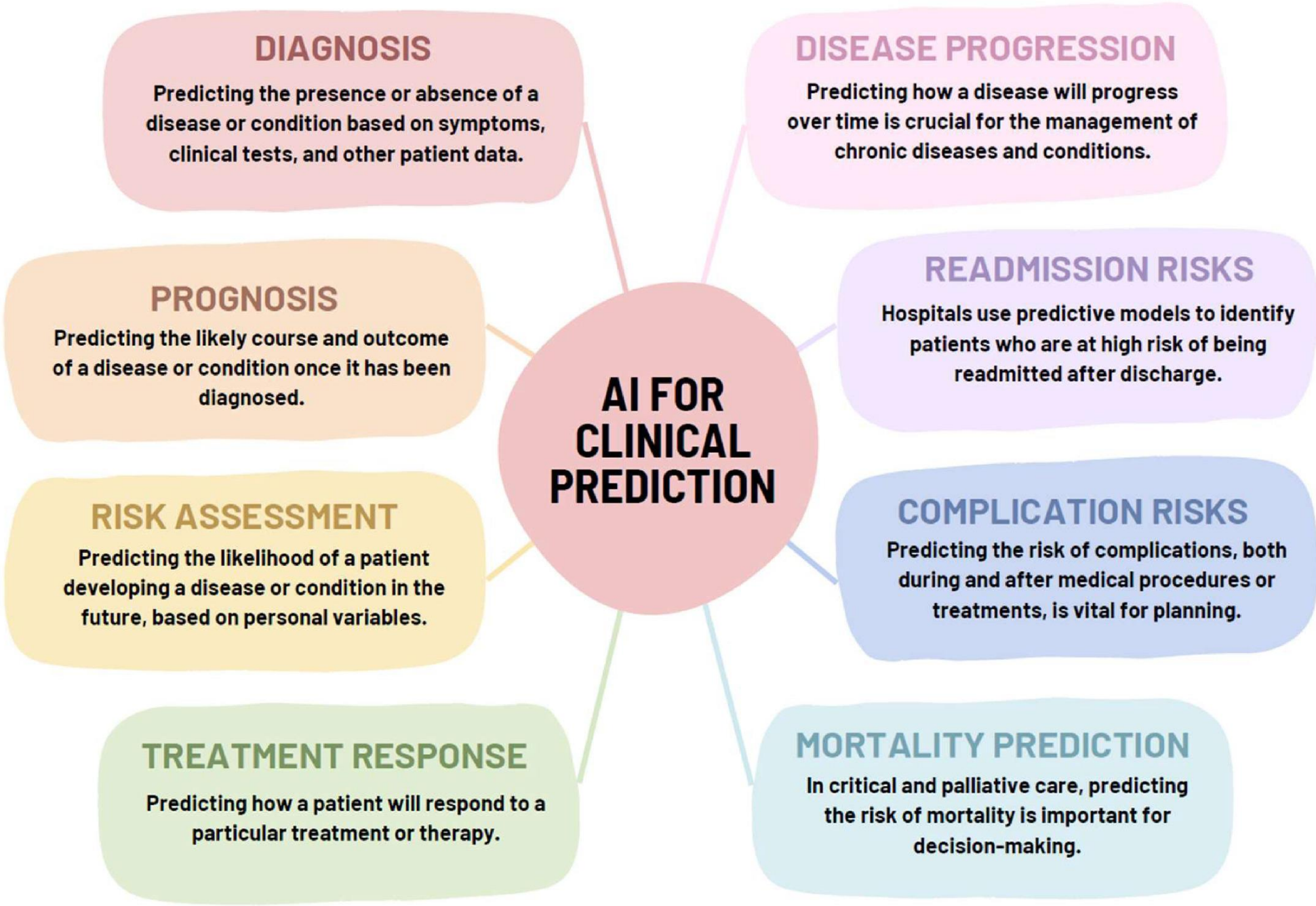
What, Why, & How of Clinical Risk Prediction Using Electronic Health Records

Sandhya Tripathi
Department of Anesthesiology
Washington University School of Medicine in St Louis

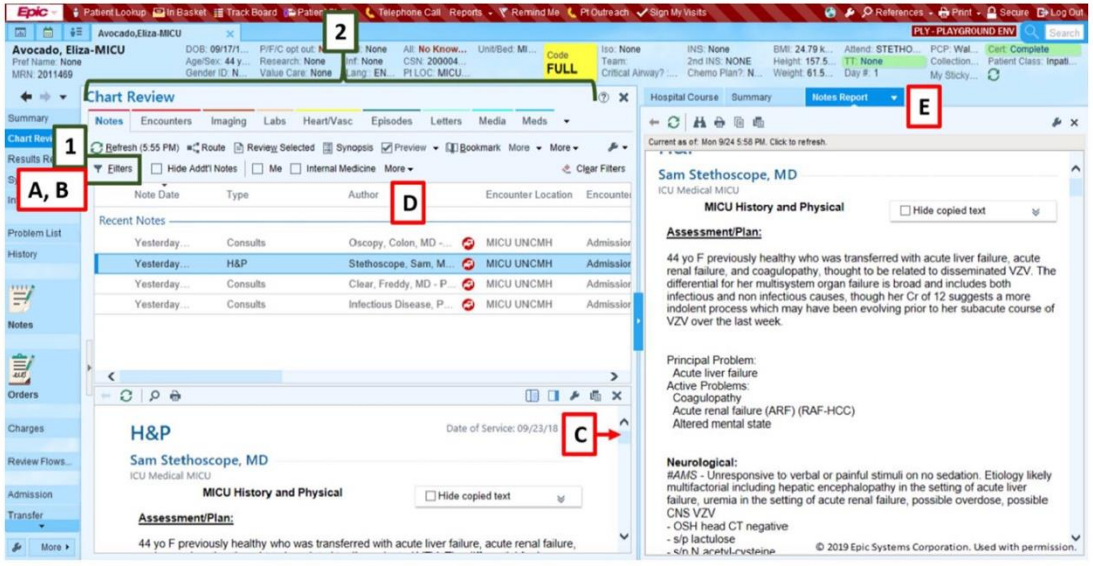
Outline

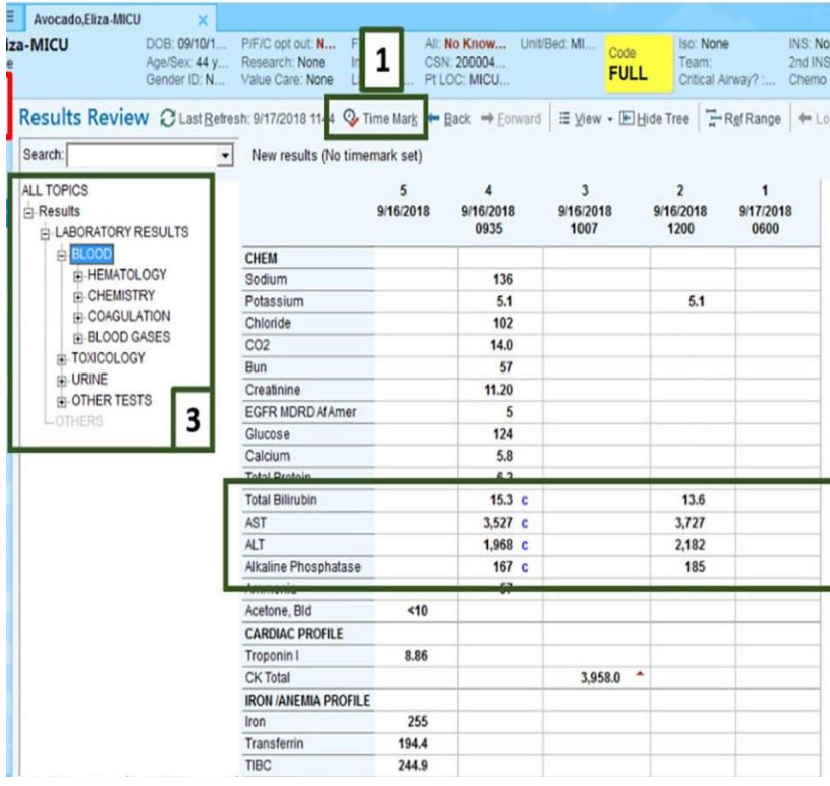
- EHR for clinical risk prediction
- The big picture in perioperative space
- Contrastive learning overview
- What to do with so many EHR systems?
- Extracting maximum information from EHR
- What to be cautious of ?



Source: Khalifa, M. and Albadawy, M., 2024. Artificial intelligence for clinical prediction: exploring key domains and essential functions. *Computer Methods and Programs in Biomedicine Update*, p.100148.

Electronic Health Records





Source: Khairat, S., Coleman, C., Teal, R., Rezk, S., Rand, V., Bice, T. and Carson, S.S., 2021. Physician experiences of screen-level features in a prominent electronic health record: Design recommendations from a qualitative study. *Health Informatics Journal*, *27*(1), p.1460458221997914.

Examples: EHR for Clinical Risk Prediction

Development and Validation of a Machine Learning Model to Identify Patients Before Surgery at High Risk for Postoperative Adverse Events

Aman Mahajan, MD, PhD[1]; Stephen Esper, MD, MBA[1]; Thien Htay Oo, PhD[2]; et al

» Author Affiliations | Article Information

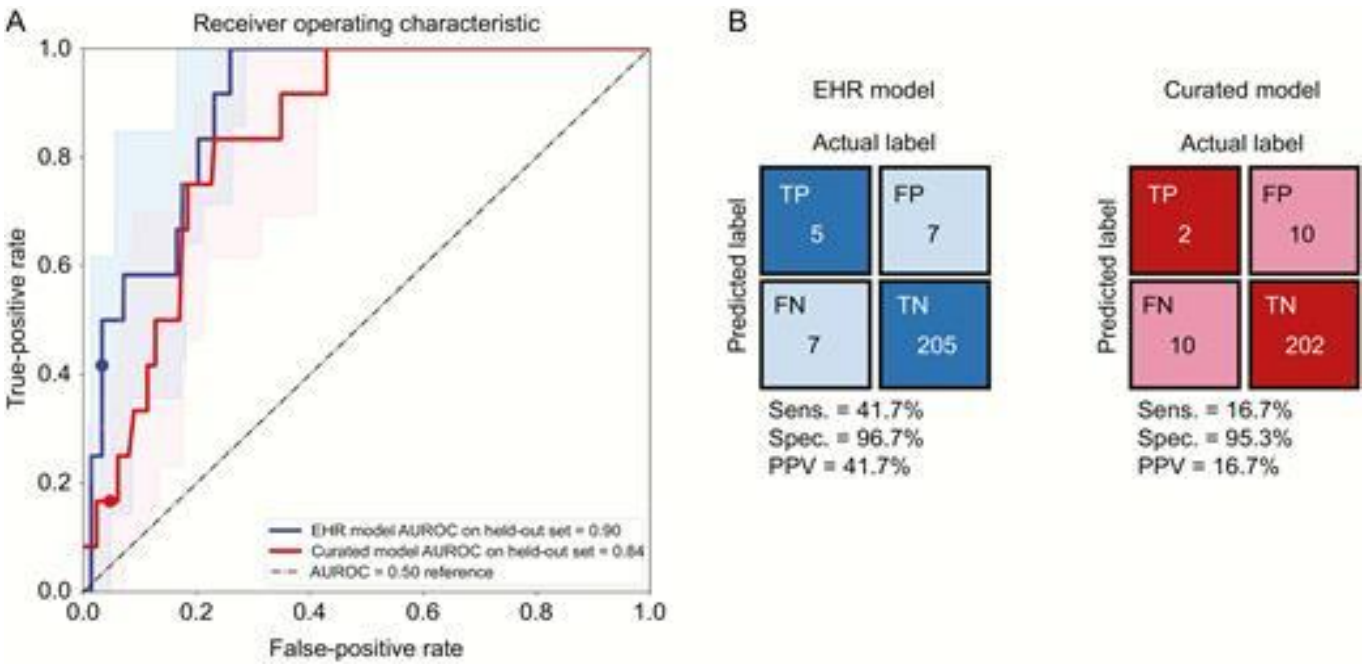
JAMA Netw Open. 2023;6(7):e2322285. doi:10.1001/jamanetworkopen.2023.22285

Key Points

Question Can an automated machine learning model accurately identify patients at high risk of adverse outcomes and predict surgical mortality or major complications using only preoperative variables?

Findings This prognostic study of 1477561 patients undergoing surgical procedures found that an automated machine learning model had an area under the receiver operating characteristic curve (AUROC) for mortality of 0.972, 0.946, and 0.956 for the training, test, and prospective set, respectively, and an AUROC for major adverse cardiac and cerebrovascular events or mortality of 0.923 and 0.899 on training and test sets, respectively. The model outperformed the National Surgical Quality Improvement Program Surgical Risk Calculator by 0.048 AUROC.

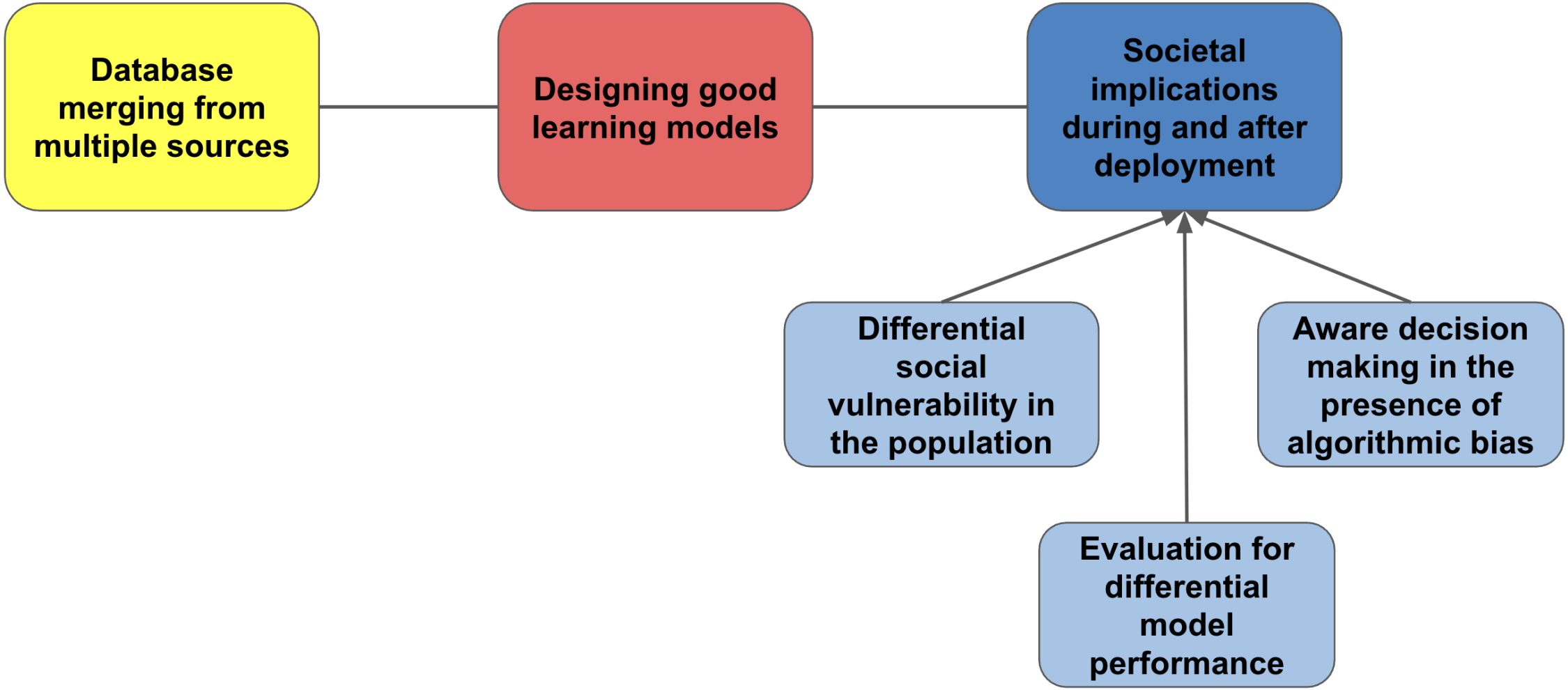
Meaning These findings suggest that a machine learning model may be used to optimize perioperative care.



Predicting Clostridium difficile

Source: Li, B.Y., Oh, J., Young, V.B., Rao, K. and Wiens, J., 2019, May. Using machine learning and the electronic health record to predict complicated Clostridium difficile infection. In *Open forum infectious diseases* (Vol. 6, No. 5, p. ofz186). US: Oxford University Press.

The Big Picture: Perioperative Risk Prediction

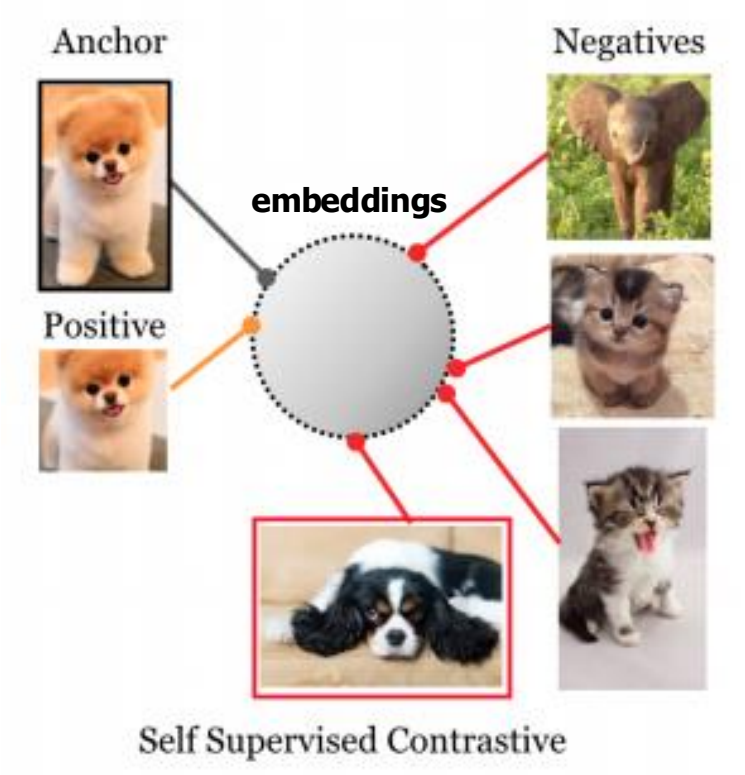


Contrastive learning

Background

What is contrastive learning?

Learn higher level feature representation where the data itself provides supervision via comparison
Similar data points close, dissimilar ones are far apart



$\mathbf{x}$ anchor point

$\mathbf{x}^+$ positive

$\mathbf{x}^-$ negative

f encoder

$$sim(f(\mathbf{x}, \mathbf{x}^+)) >> sim(f(\mathbf{x}, \mathbf{x}^-))$$

Source: Khosla et. al. 2021

Contrastive learning components

Input data + Models + Optimization over an objective

What ??

Different views

Different modalities

How ??

Augmentations

Different sources of information

Encoders

Projection heads

What ??

Loss maximizing similarity between representations

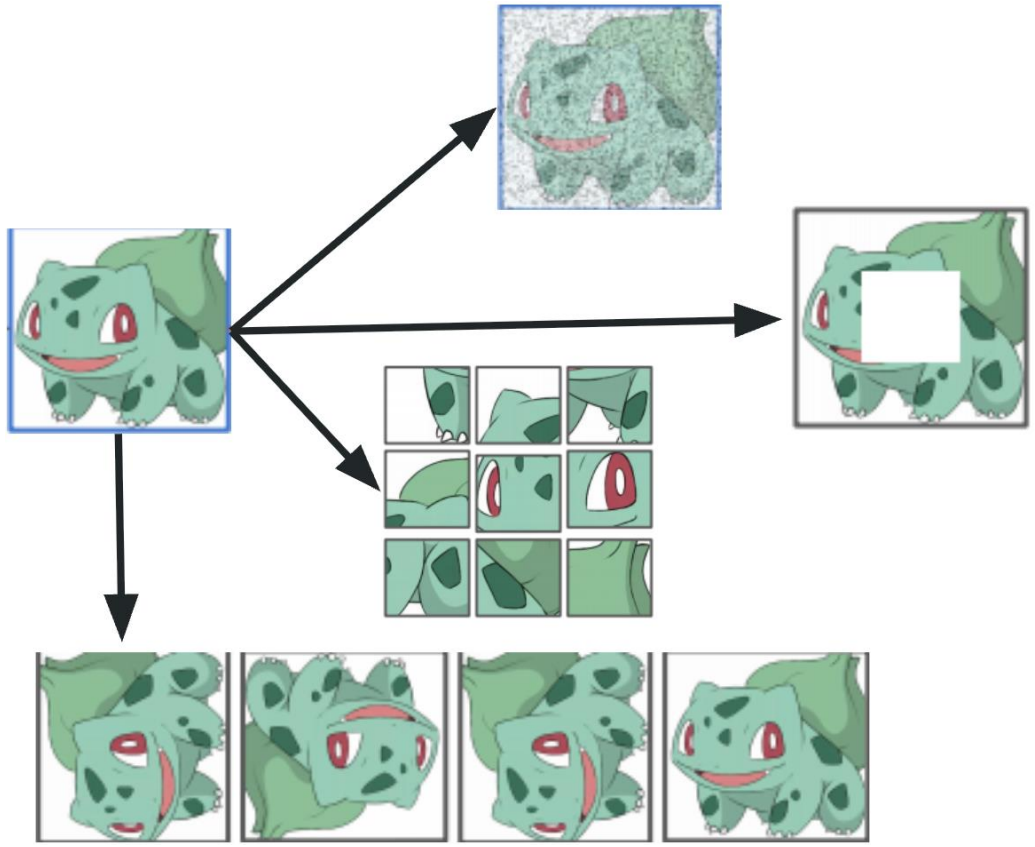
How ??

Training and batch construction (choice of negative pairs) strategies

Examples of data augmentations

Label preserving: A good set of views are those that share the minimal information necessary to perform well at the downstream task.

Vision: Random crop, Rotation, Adding noise, masking, color jitter



Text: lexical editing, back translation

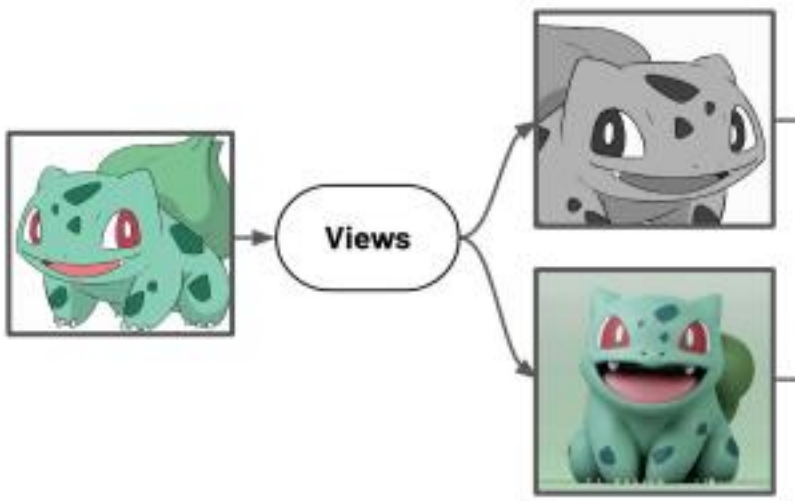
Operation	Sentence
None	A sad, superior human comedy played out on the back roads of life.
SR	A ***lamentable***, superior human comedy played out on the ***backward*** road of life.
RI	A sad, superior human comedy played out on ***funniness*** the back roads of life.
RS	A sad, superior human comedy played out on ***roads*** back ***the*** of life.
RD	A sad, superior human out on the roads of life.

EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks.

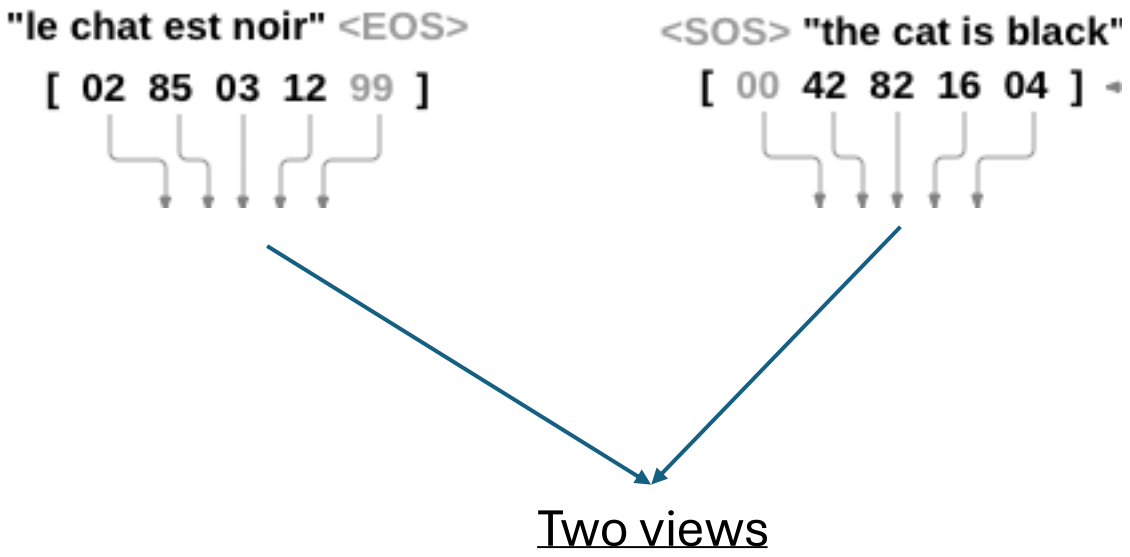
What makes for good views for contrastive learning?

Multiple views of the same data

Images

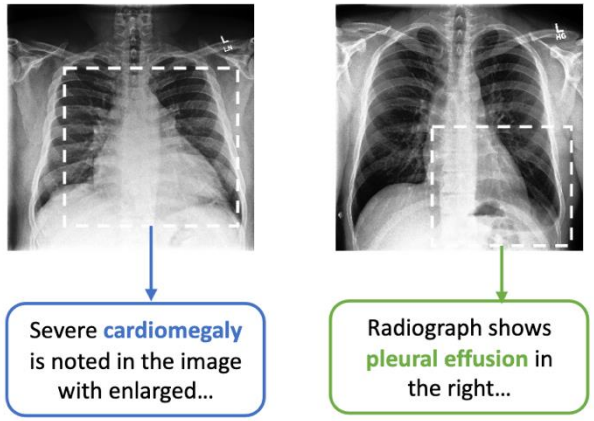


Text

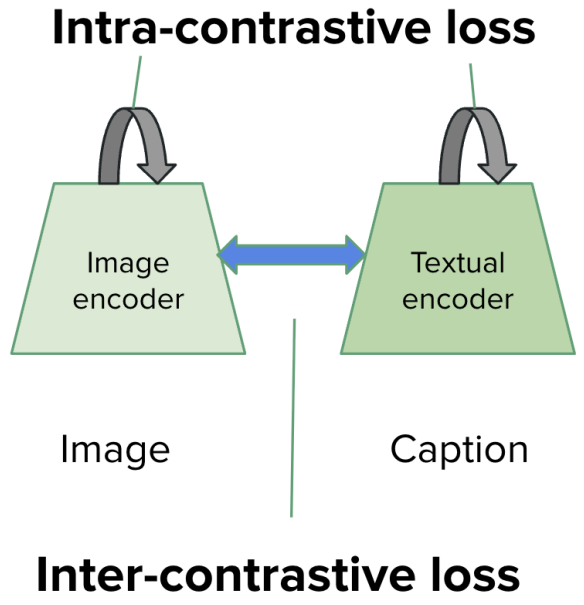


More than one modality or data types?

- Availability of additional source of information
  - Captions of images
  - Subtitles in videos
- Issues with unimodal self supervised learning
  - High inter-class similarity in some domains
- Can be used in
  - Zero shot learning by using one set of embeddings and matching
  - Combining representation from different modality sources
    - Demographics + Vitals + medications for readmission prediction



Solution class: ConVIRT



Contrastive loss objectives

Contrastive loss (2005)

Face verification

Min: embedding distance if x1,x2 genuine pair
Max: if x1, x2 imposter pair

Noise contrastive estimation 2010

Learning by comparison of target (positive) and noise (negative) distribution

Relates to the log-likelihood in a logistic regression model for discrimination

Noise contrastive estimation based losses

InfoNCE loss 2018

Context c

N samples (positive + negatives)

X = [x1, x2,...,xN]

f is the density ratio used to predict the future x observations using the context

$$\mathcal{L} = -\log \frac{f(\mathbf{x},\mathbf{c})}{\sum_{\mathbf{x}' \in \mathbf{X}} f(\mathbf{x}',\mathbf{c})}$$

NT-Xent 2020

InfoNCE with f as cosine similarity

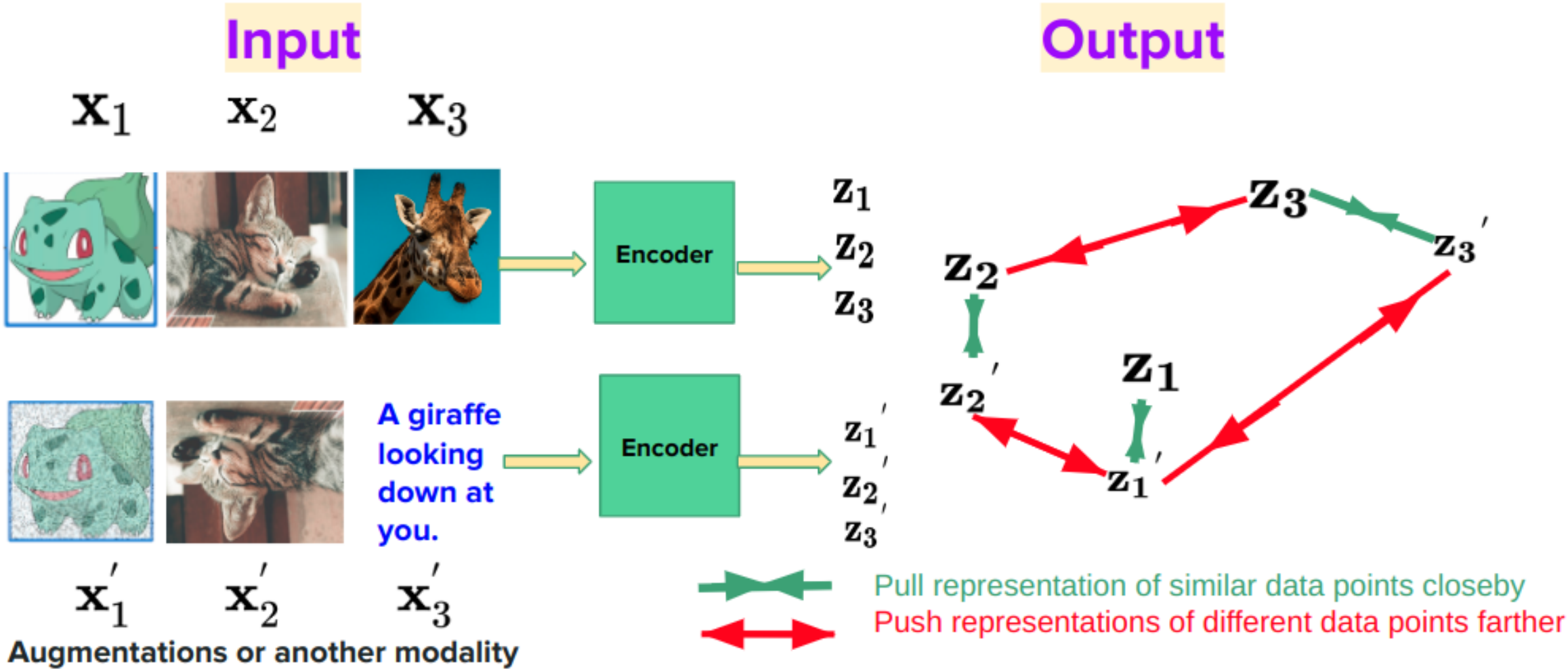
Positive and negatives are normalized embeddings

Addition of temperature parameter

$$\mathcal{L} = -\log \frac{exp(sim(\mathbf{x}_i,\mathbf{x}^+)/\tau)}{exp(sim(\mathbf{x}_i,\mathbf{x}^+)/\tau) + \sum_{j=1}^{N-1} exp(sim(\mathbf{x}_i,\mathbf{x}_j^-)/\tau)}$$

Takeaways

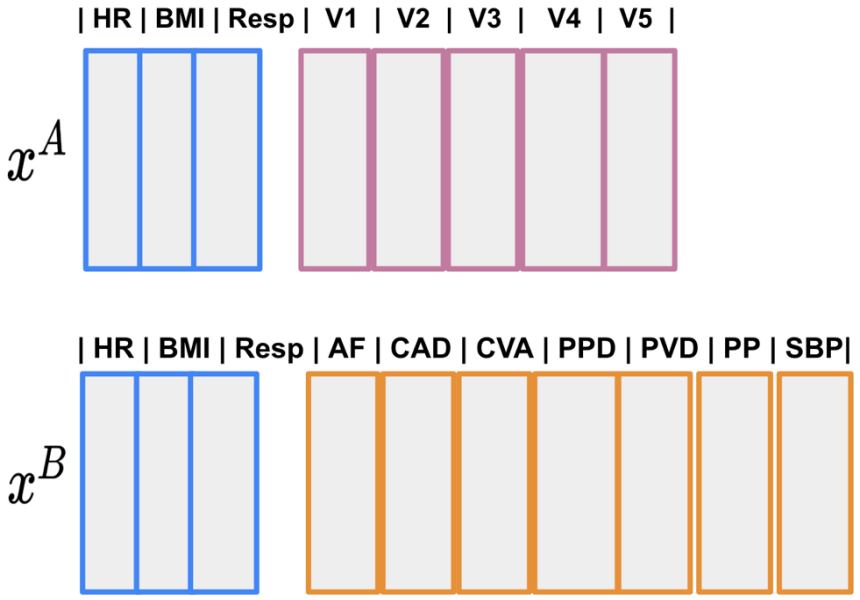
Contrastive Learning: Similar closeby, Different far apart



Database matching from multiple sources

What to do with so many EHR systems?

Problem



Epic feature name	Contextually derived name
BW#608	Hypertension
EPIC#55845	Diagnosis year
EPIC#31000009721	Typical SBP
EPIC#31000040796	Typical DBP
EPIC#5266	Hyperlipidemia
EPIC#HPI0118	CAD
EPIC#20764	CCS angina class
BW#164	No revasc but >50% stenotic coronary
EPIC#10072	MI
BW#165	Number of MIs
EPIC#62854	Date of last MI
EPIC#3710	History of CABG
EPIC#62856	Prior CABG date

Ur Creatinine ≠ POC creatinine

- Learning from a single healthcare organization's dataset isn't very informative.
- Data from each hospital needs to be reconciled to an agreed upon schema
- Multiple Electronic Health Record (EHR) systems across time (Change from Metavision to Epic) at Barnes Jewish Hospital, St Louis, MO
- Inconsistent documentation or surrogate features (same concept but different name)

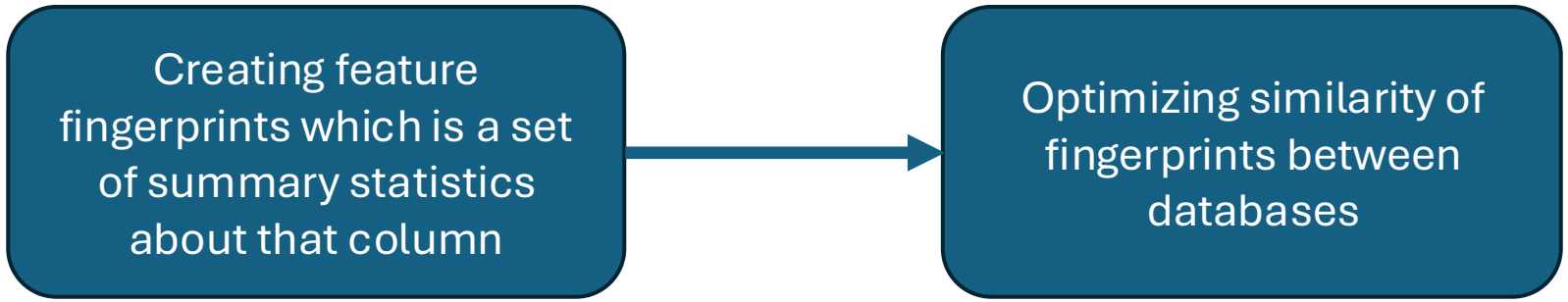
- Limited by binary variables and distribution shifts
- Absent but reparameterization of other datasets

Our proposed solution

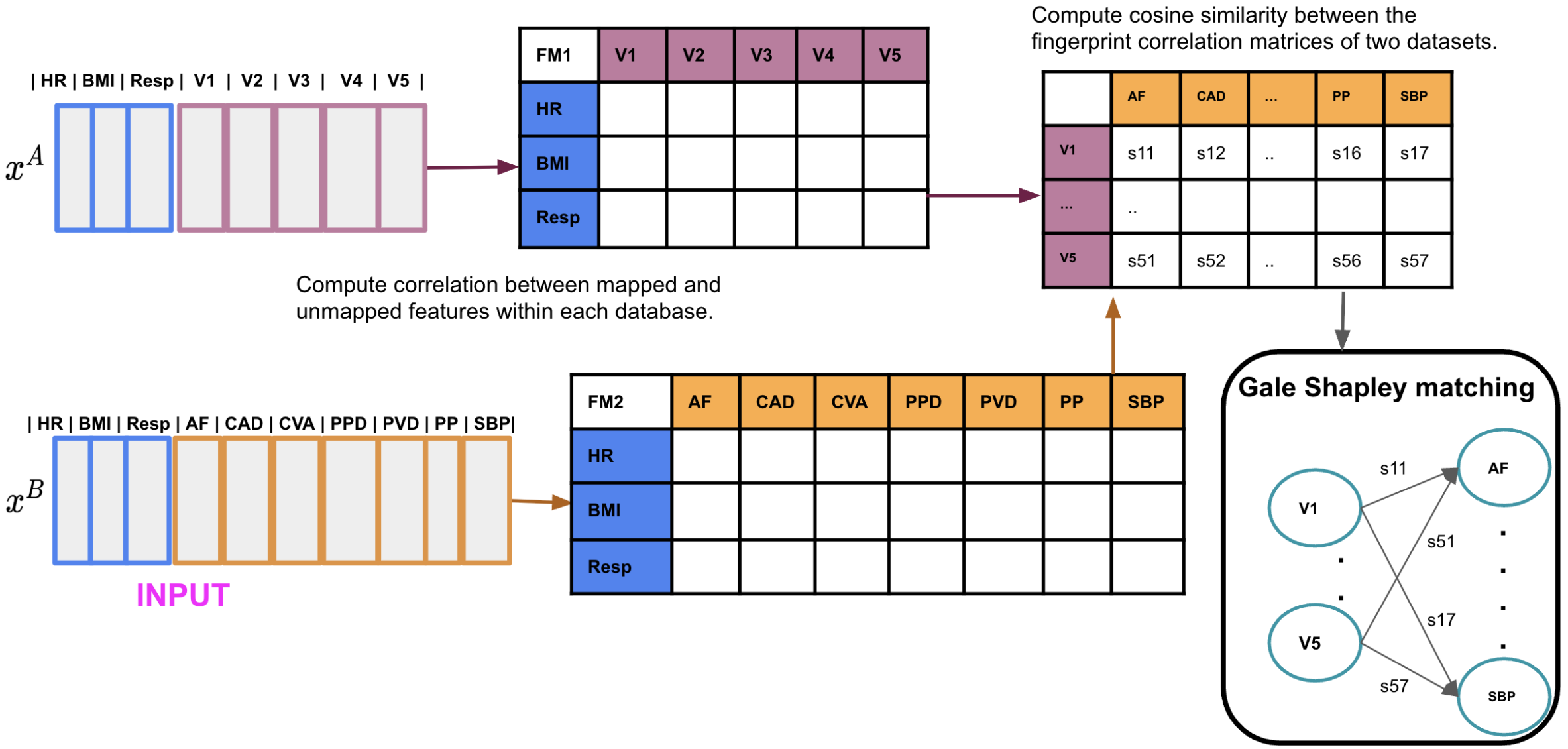
- *Underlying concept:* Relationships between variables are plausibly unchanged across databases even with substantial variable shift.

- *Example:* Although the fraction of men at a given hospital can vary, the relationship between sex and height is likely very similar.

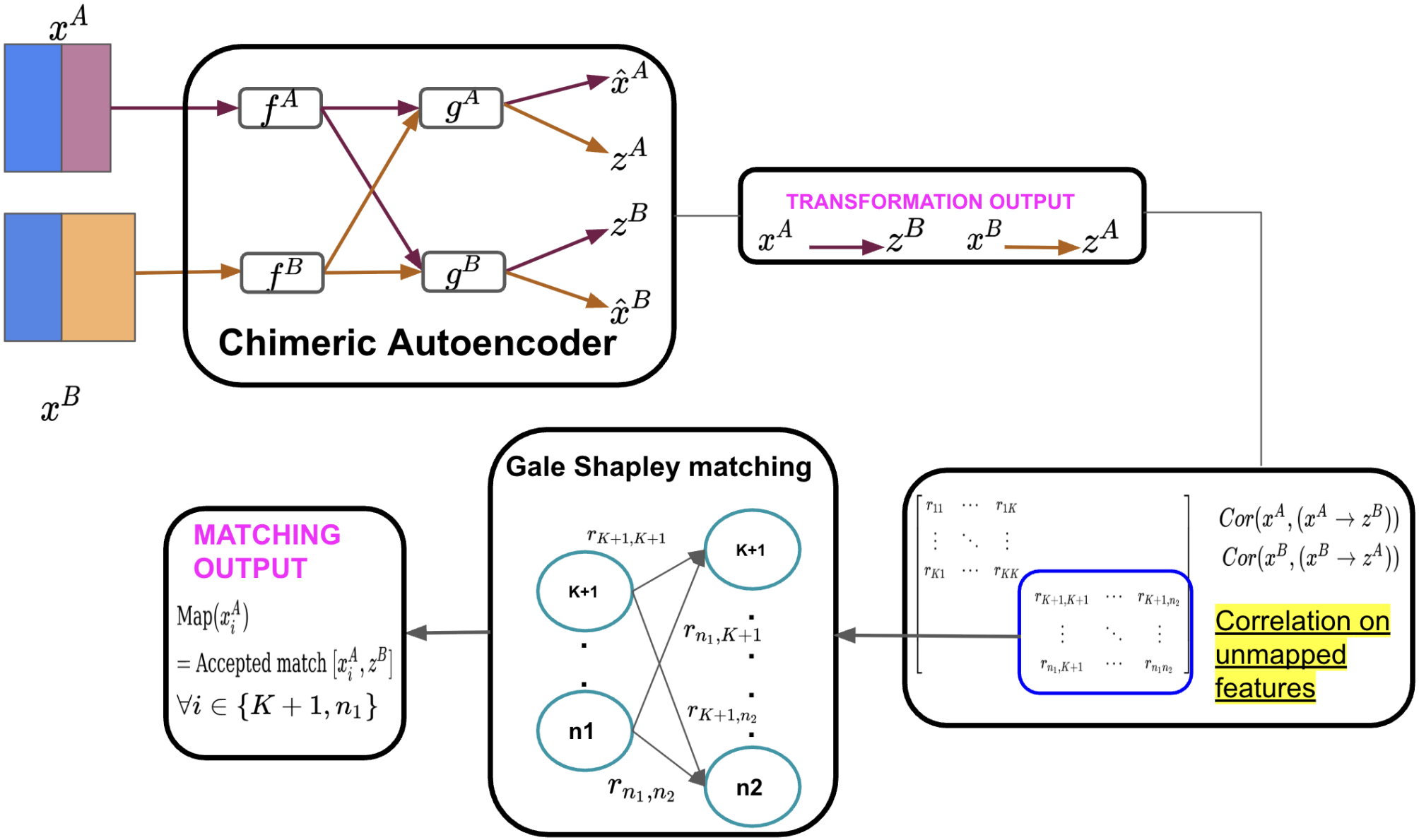
- *Assumption:* Some features are mapped by auxiliary knowledge or manual verification (known-mapped columns), e.g. categorical features.



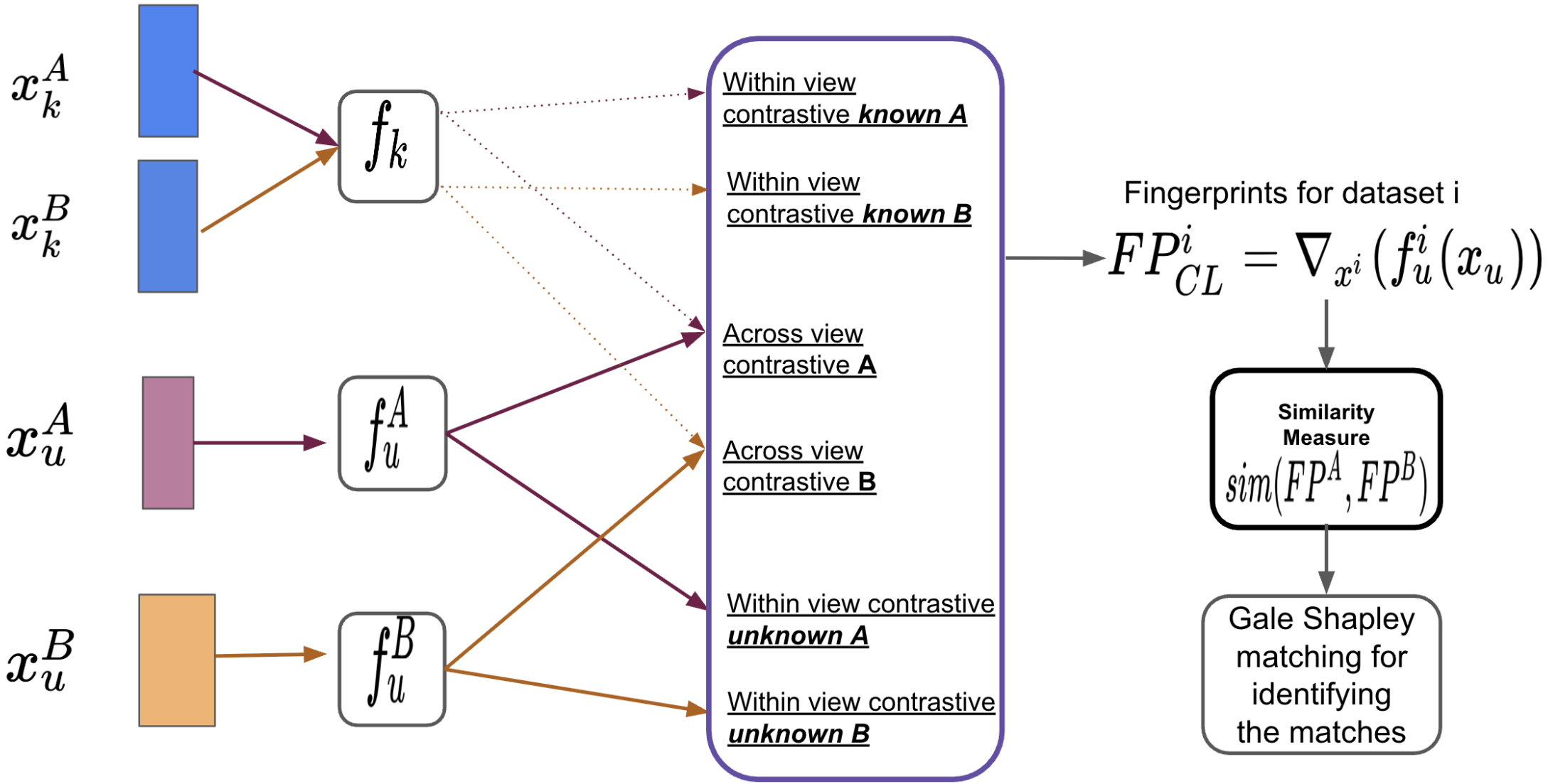
Simple correlation method



Chimeric encoders



Contrastive learning-based encoders

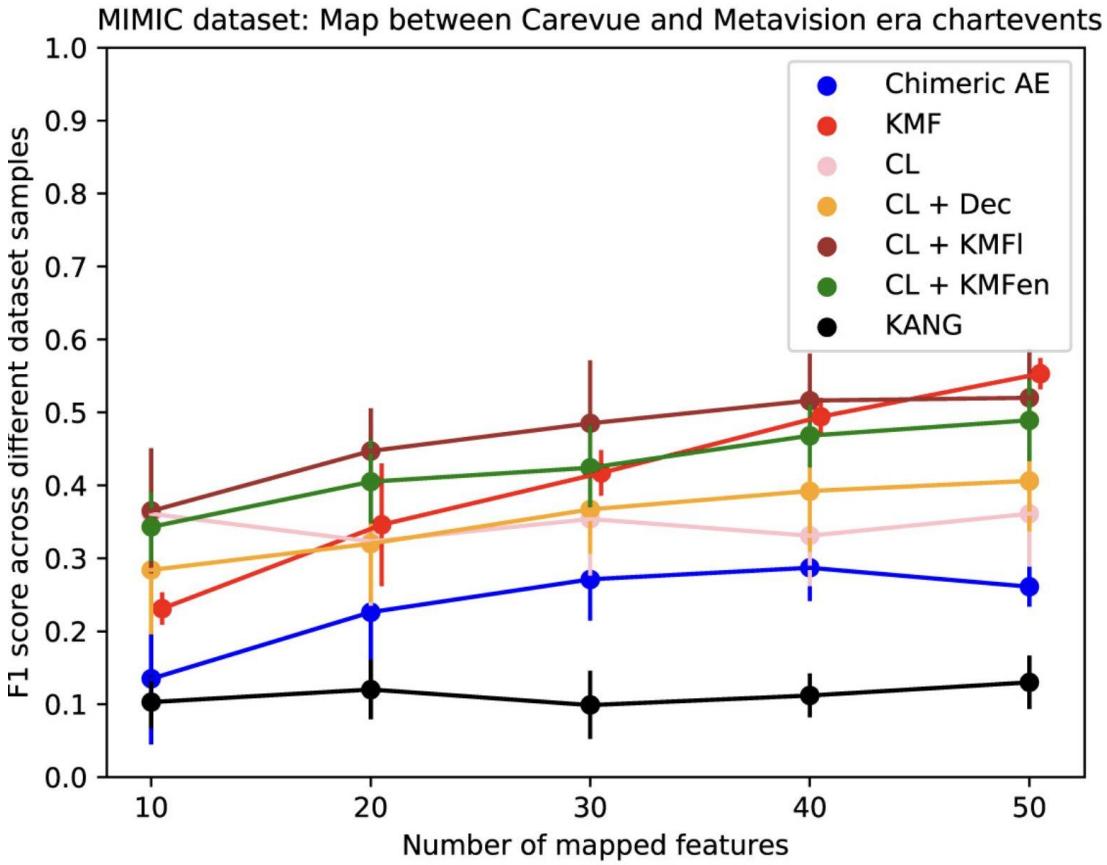


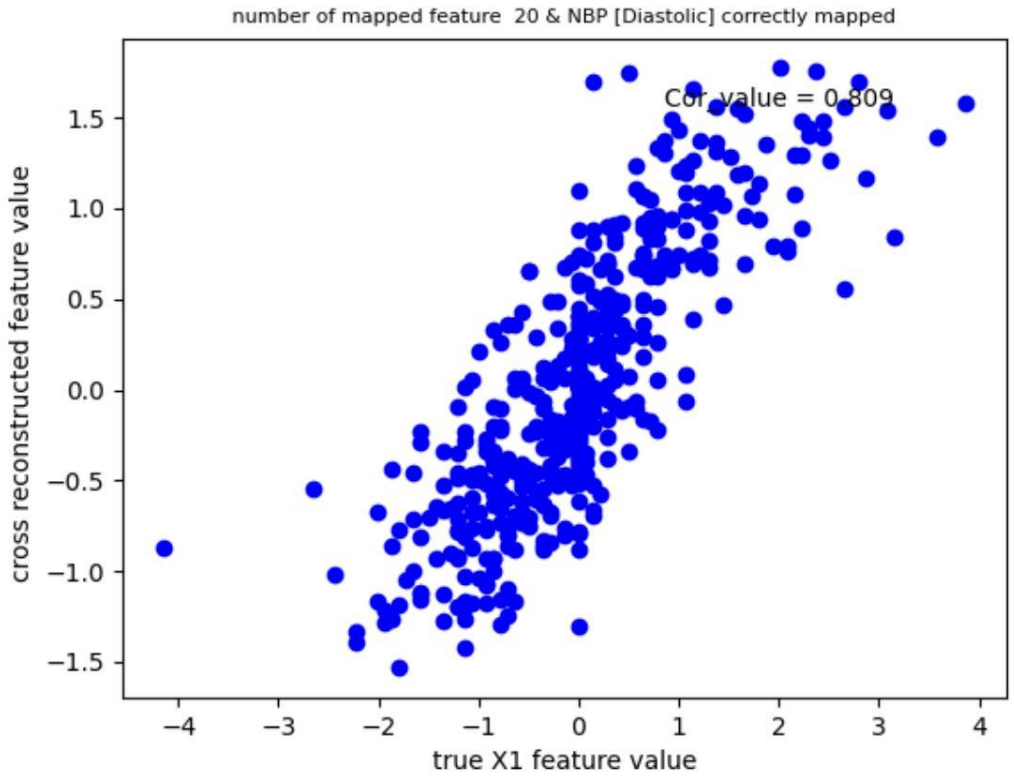
Experimental setup

- **MIMIC III** (Two EHRs across time, *Carevue (CV)* and *Metavision (MV))*
  - Labs are mapped (51), chart (vitals) need mapping (56 in MV, 55 in CV)

- **ACTFAST** (*Metavision (MV) (June 2012–June 2018)* and *Epic (November 2018–March 2022)*)
  - Preoperative data is mapped (64),
  - Lab values from two timelines need to be mapped (75 in Epic, 167 in MV)

MIMIC III: Matching and Transformation

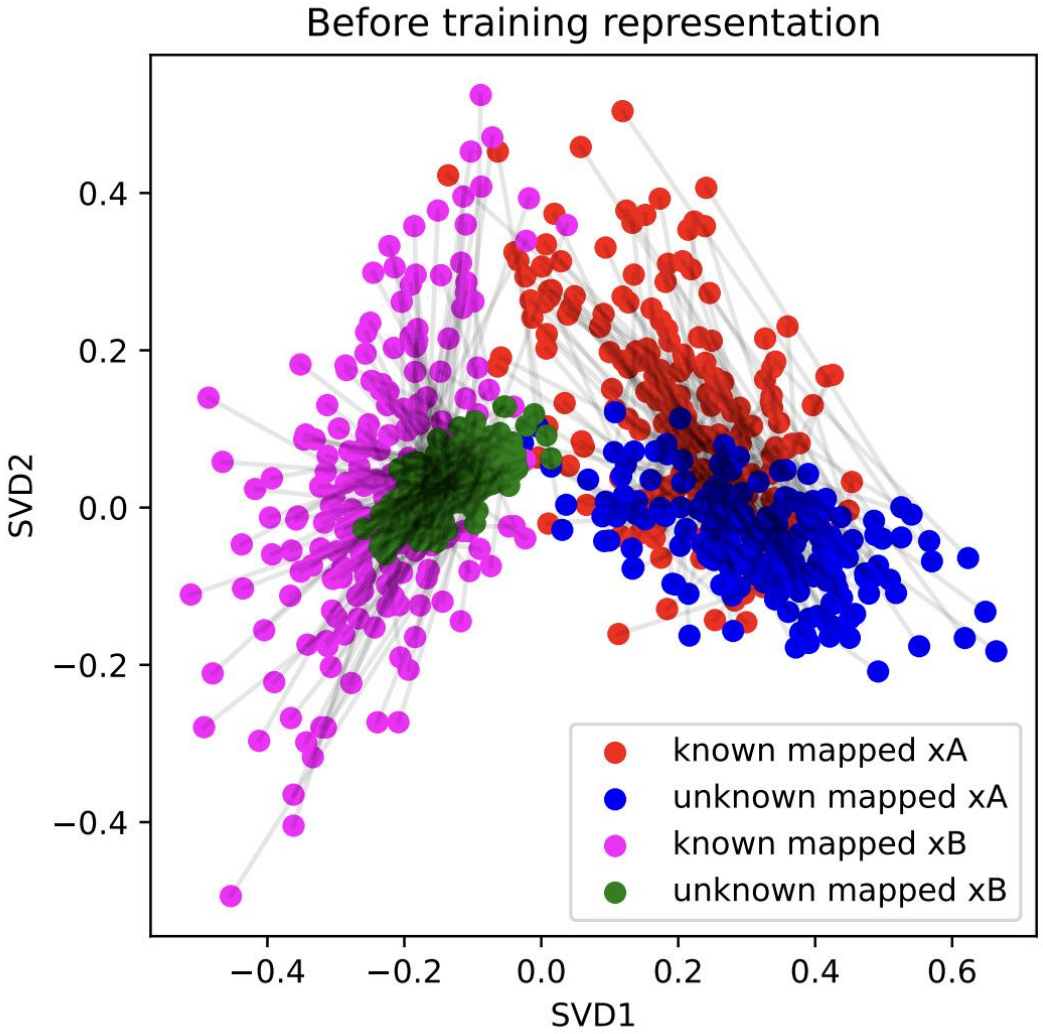


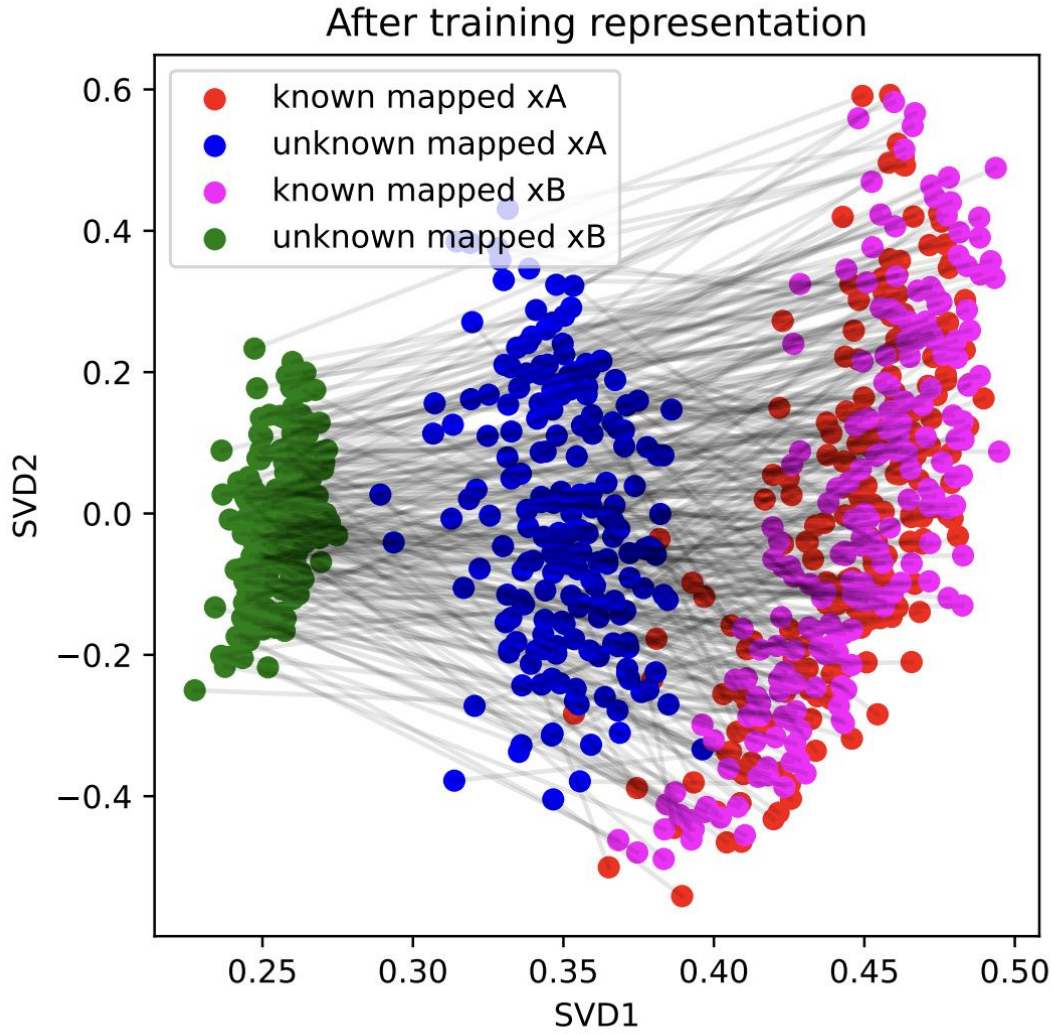


ACTFAST: so, what if the match is incorrect?

ump_feature_in_X1	match_byGS	Method	MatchFreq	Related	Agreement
BILIRUBIN, DIRECT	ALKALINE_PHOSPHATASE_MEASUREMENT_UNITS/L_P3-G8644Y	KMF	36	Strong	Yes
WHITE BLOOD CELLS, URINE	NITRITE_URINE_P3-90141Y	KMF	32	Strong	Yes
SPECIFIC GRAVITY, URINE	HYALINE_CASTS_URINE_P3-90018Y	KMF	24	Strong	Yes
HEMATOCRIT	HEMOGLOBIN_[HGB]_ESTIMATION_TOTAL_VENOUS_BLOOD_P3-G0053Y	KMF	24	Strong	Yes
FIBRINOGEN, CLAUSS	MAXIMUM_CLOT_FIRMNESS_P3-G5C7ZZ	KMF	24	Strong	Yes
HEMOGLOBIN	HEMATOCRIT_[HCT]_DETERMINATION_P3-34080	CL	21	Strong	Yes
URINE BLOOD	ERYTHROCYTE_[RBC]_COUNT_URINE_P3-3051BY	KMF	20	Strong	Yes
RDW	ERYTHROCYTE_[RBC]_DISTRIBUTION_WIDTH_[RDW]_DETERMINATION_COEFFICIENT_OF_VARIATION_[RDW-%CV]_P3-3108AY	KMF	20	Strong	Yes
TRANSFERRIN SATURATION	IRON_MEASUREMENT_NOS_P3-73330	KMF	48	Medium	Yes
IRON, TOTAL	TRANSFERRIN_SATURATION_P3-73343Y	KMF	44	Medium	Yes
SPECIFIC GRAVITY, URINE	UROBILINOGEN_MEASUREMENT_URINE_P3-74211	CL	31	Medium	–
IMMATURE GRANULOCYTE ABSOLUTE	MONOCYTE_COUNT_[ABSOLUTE]_P3-3056DY	CL	24	Medium	Yes
C REACTIVE PROTEIN	IMMATURE_GRANULOCYTE_COUNT_ABSOLUTE_K/CUMM_P3-G725ZZ	KMF	24	Medium	Yes
GLUCOSE	GLUCOSE_URINE_QUALITATIVE_P3-90100	chimeric	22	Medium	Yes
CALCIUM	ALBUMIN_MEASUREMENT_P3-71260	KMF	20	Medium	Yes

Sanity check: Modality gap





Takeaways for dataset integration

- Relationships between variables are unchanged across databases and hence we were able to tap on those similarities for matching.
- Simple correlation is great when small number of features and bijective mapping.
- Validation is easier than search so providing surrogate matches as in ACTFAST is still useful.

- Unanswered question: What happens to the timeseries variables?

Contrastive learning for making clinical predictions

Extracting maximum information from EHR

EHR has plethora of different data sources but..

- **<u>What are the challenges with current ML methods?</u>**
  - Simultaneously learning from highly complex multi modal data including time series, unstructured notes, etc.
  - Reduced effective sample size due to highly skewed and inconsistent quality labels.

- **<u>What could be a possible solution?</u>**
  - Self supervised multimodal learning particularly contrastive learning (CL) that bypasses the need for explicit supervision [1].
  - **Use CL techniques to exploit similarities or complementary relationships between different modalities, such as home medications and problem lists for learning time series representations which can be used for diverse prediction tasks.**

[1] Muse, E.D. and Topol, E.J., 2024. Transforming the cardiometabolic disease landscape: Multimodal AI-powered approaches in prevention and management. Cell Metabolism.

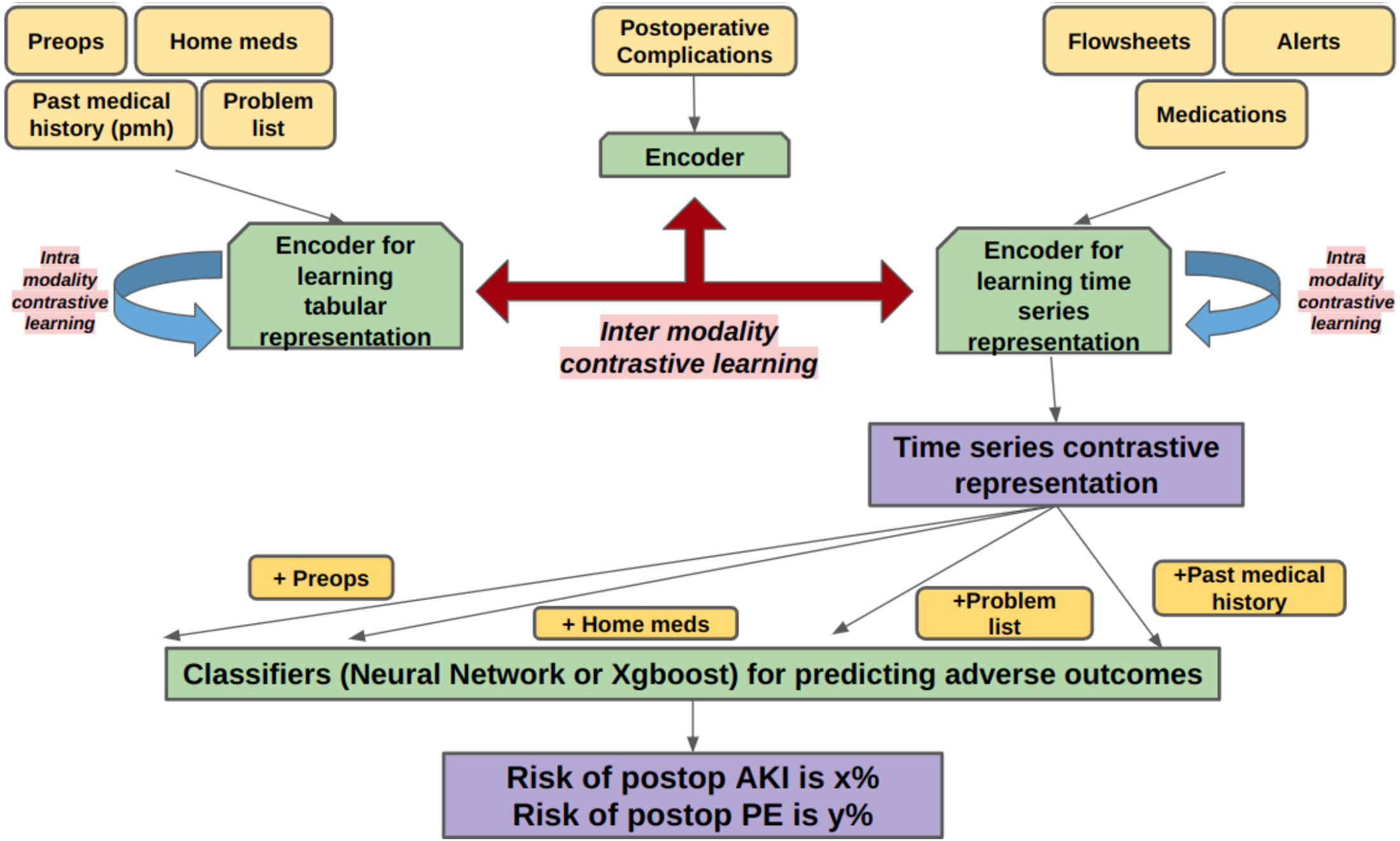
Experimental setup

- **Modalities of interest:** preoperative data, home medications, problem lists, past medical history, intraoperative data (flowsheets, medications and alerts), and postoperative complications.

- **Datasets**: ACTFAST-Epic with 69K train & 15K test data points[1].

- **Models**: TS2Vec for time series CL [2], SCARF for tabular CL [3], XGBoost for classifier training (with isotonic calibration post facto).

1.Xue, B., Li, D., Lu, C., King, C.R., Wildes, T., Avidan, M.S., Kannampallil, T. and Abraham, J., 2021. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. JAMA network open, 4(3), pp.e212240-e212240.
2.Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y. and Xu, B., 2022, June. Ts2vec: Towards universal representation of time series. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 8, pp. 8980-8987).
3.Bahri, D., Jiang, H., Tay, Y. and Metzler, D., 2021. Scarf: Self-supervised contrastive learning using random feature corruption. arXiv preprint arXiv:2106.15147.


Preops
Home meds
Past medical history (pmh)
Problem list
Postoperative Complications
Encoder
Flowsheets
Alerts
Medications
Intra modality contrastive learning
Encoder for learning tabular representation
Inter modality contrastive learning
Encoder for learning time series representation
Intra modality contrastive learning
Time series contrastive representation
+ Preops
+ Home meds
+Problem list
+Past medical history
Classifiers (Neural Network or Xgboost) for predicting adverse outcomes
Risk of postop AKI is x%
Risk of postop PE is y%

Results

Modalities								Outcomes to predict																	
Case #	Preops	Home meds	PMH &PL	Intraop Flowsheets	Intraop Medications	Intraop Alerts	Postoperative complications	30-day mortality (incidence rate 2.1%)			Acute Kidney Injury KDIGO grade 2+ (incidence rate 5.6%)			Cardiac (new CHF or MI incidence rate 1.1%)			Pulmonary Embolism (PE, incidence rate 0.3%)			Pneumonia (incidence rate 0.3%)			Unplanned ICU (incidence rate 8%)		
								AUROC	AUPRC	Imp	AUROC	AUPRC	Imp	AUROC	AUPRC	Imp	AUROC	AUPRC	Imp	AUROC	AUPRC	Imp	AUROC	AUPRC	Imp
1	✓							0.918	0.336		0.939	0.705		0.866	0.082		0.713	0.018		0.724	0.011		0.889	0.545	
2	✓	✓						0.924	0.367	++	0.941	0.702	+	0.880	0.102	++	0.668	0.023		0.871	0.050	++	0.905	0.601	++
3	✓	✓	✓					0.920	0.435	+	0.940	0.696		0.887	0.104		0.859	0.102	++	0.891	0.067	++	0.904	0.591	
4				✓				0.839	0.158		0.799	0.196		0.750	0.031		0.602	0.005		0.788	0.015		0.921	0.634	
5				✓	✓			0.881	0.227	++	0.880	0.379	++	0.776	0.035	+	0.688	0.007	+	0.821	0.018	+	0.929	0.680	+
6				✓	✓	✓		0.886	0.234	+	0.885	0.381	++	0.779	0.039		0.682	0.007		0.828	0.019		0.928	0.672	
7	✓	✓	✓	✓				0.837	0.164		0.798	0.194		0.756	0.032		0.641	0.006		0.801	0.015		0.920	0.636	
8	✓	✓	✓	✓	✓			0.936	0.484	++	0.954	0.735	++	0.869	0.098	++	0.828	0.039	++	0.872	0.027	++	0.957	0.782	++
9	✓	✓	✓	✓	✓	✓		0.940	0.471	+	0.952	0.729		0.874	0.106	++	0.843	0.045	++	0.874	0.026		0.956	0.781	
10	✓	✓	✓	✓	✓	✓	✓	0.938	0.478		0.952	0.730		0.880	0.108	+	0.849	0.050	+	0.888	0.031	++	0.956	0.780	

Table reading notes: + and ++ in **Imp** column denotes at least second decimal improvement without last data modality (previous row) in only AUROC/AUPRC and both AUROC and AUPRC respectively. Intraop alerts and postoperative complications are used only during representation learning and not in classifier training. PMH: Past medical history, PL: Problem list

Takeaways from CL for clinical risk prediction

- Adding time series modality improves performance for 30-day mortality, AKI and Unplanned ICU.
- Both home meds and intraoperative medications modality lead to better representations and further improved classifier performance.
- Our intention was to improve the time series representation but using CL also to learn tabular data representation is an overkill!

Societal implications during and after deployment

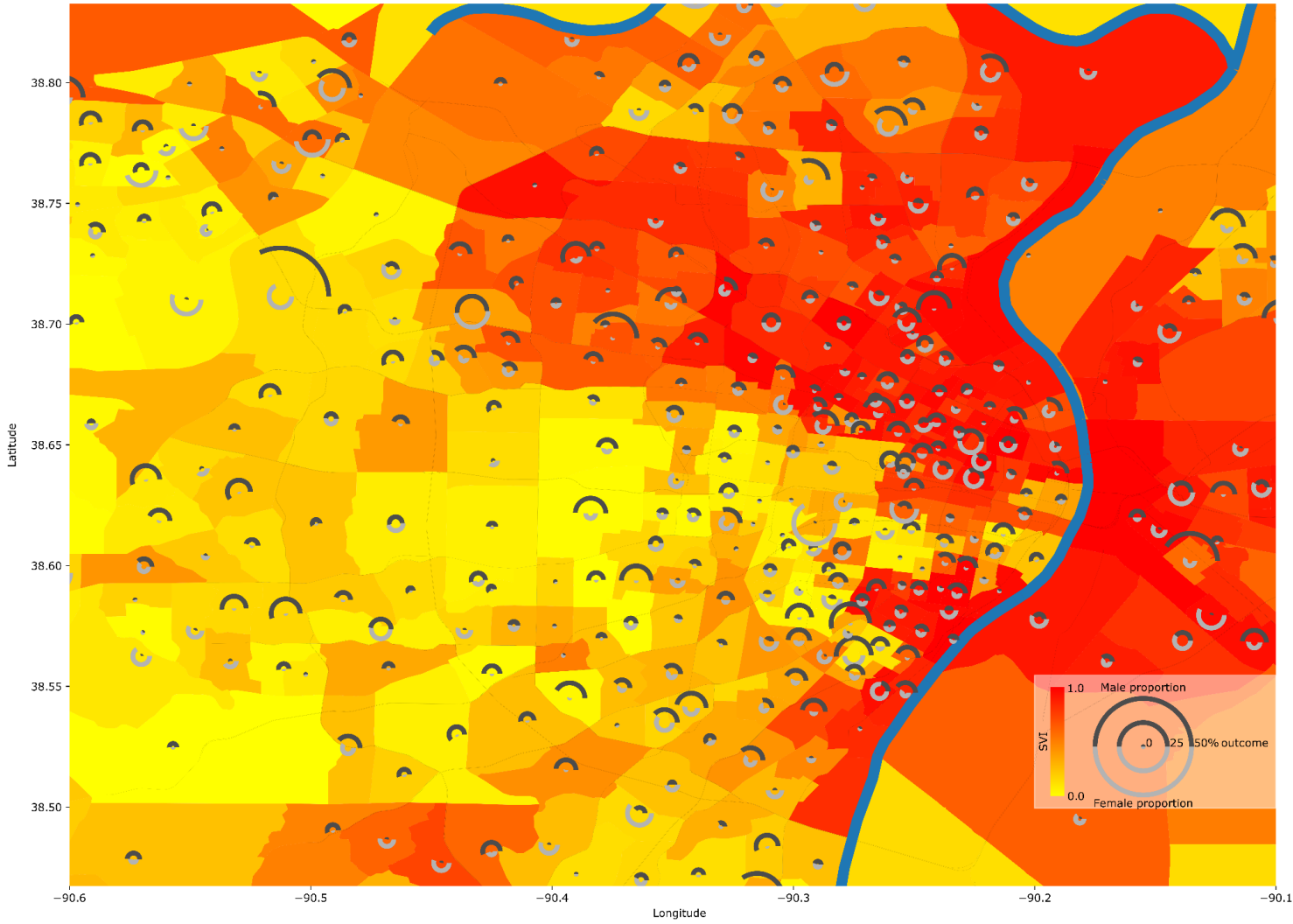
What to be cautious of ?

Postoperative outcomes and social vulnerability

Census tract level social vulnerability index (SVI) as a combination of economic, personal, social, and ethnic factors into a single metric.

Subthemes: socioeconomic status , household composition & disability, minority status & language, and housing type & transportation.

Abnormal heart rhythm associated with SVI with **socioeconomic status** and **household composition** as the main association factors.



Algorithmic bias in delirium evaluation

Delirium treatments/interventions are not scalable and could negatively impact the patient. ML can help in accurate targeting.

Algorithmic bias: ML model reinforcing existing biases against certain population groups (categories of protected attributes: race, age, sex, language, sometime insurance status).

Found that currently used random forest based models suffer from **disparate mistreatment** (algorithm's performance differs for otherwise similar people).

What could help?

Summary Metric

Removing the sensitive variable could reduce discrepancy

	Max discrepancy in AUROC across subgroups for each protected attribute			
Protected attribute	Model with protected attribute		Model without protected attribute	
	Before matching	After Matching	Before Matching	After Matching
	MIMIC-III dataset			
Race	0.03	0.04	0.01	0.01
Sex	0.02	0.04	0.02	0.02
Age	0.02	0.02	0.02	0
	ACTFAST-Epic dataset			
Race	0.02	0.04	0.02	0.02
Sex	0.02	0.02	0.02	0.02
Age	0.03	0.01	0.04	0

Propensity score matching could reduce discrepancy

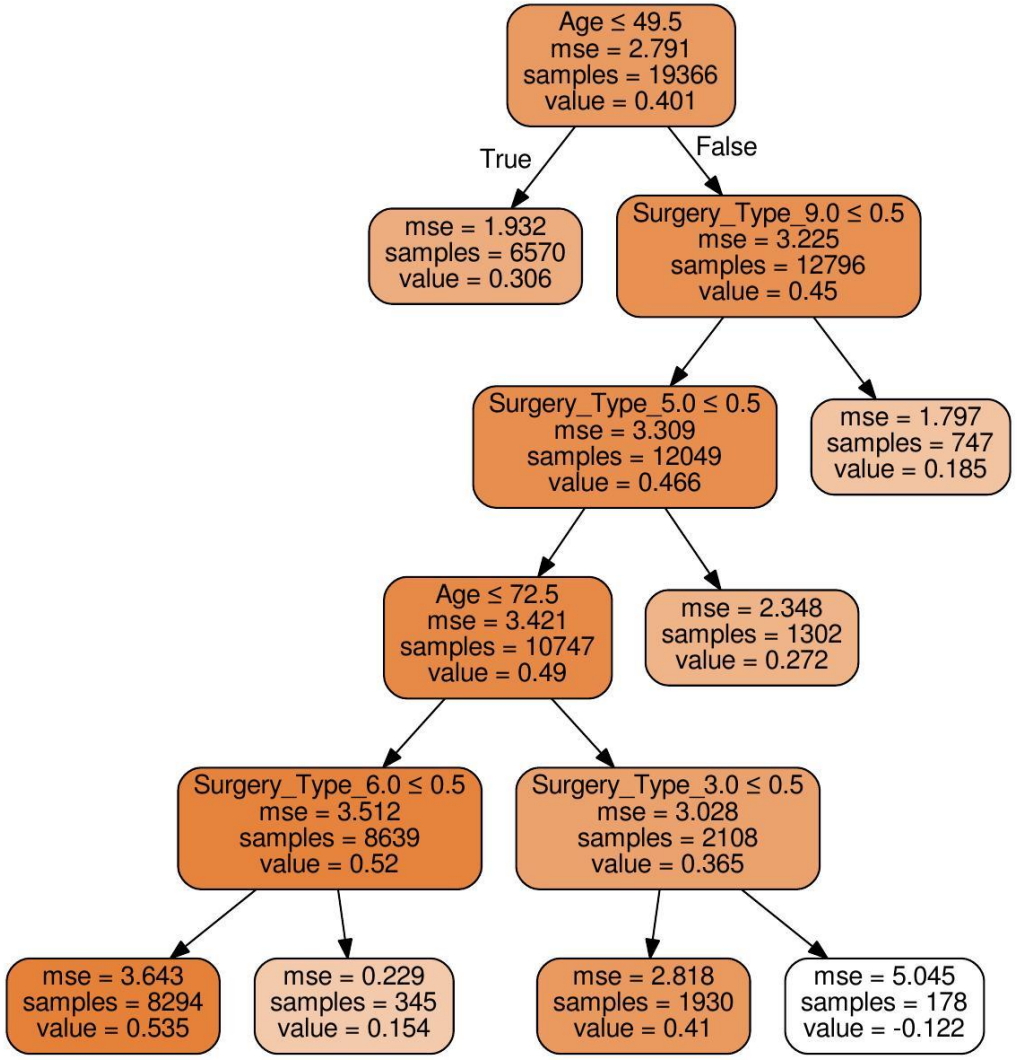
Aware decision making in the presence of algorithmic bias

- Check whether the patient benefits from the use of proposed model (**M_Full**) by comparing the transparent additional utility it provides over a basic model (**M_base**)
- Utility for a patient is defined as the weighted absolute probability of correct prediction

$$IU = w_1 y\hat{p} + w_2(1-y)(1-\hat{p})$$

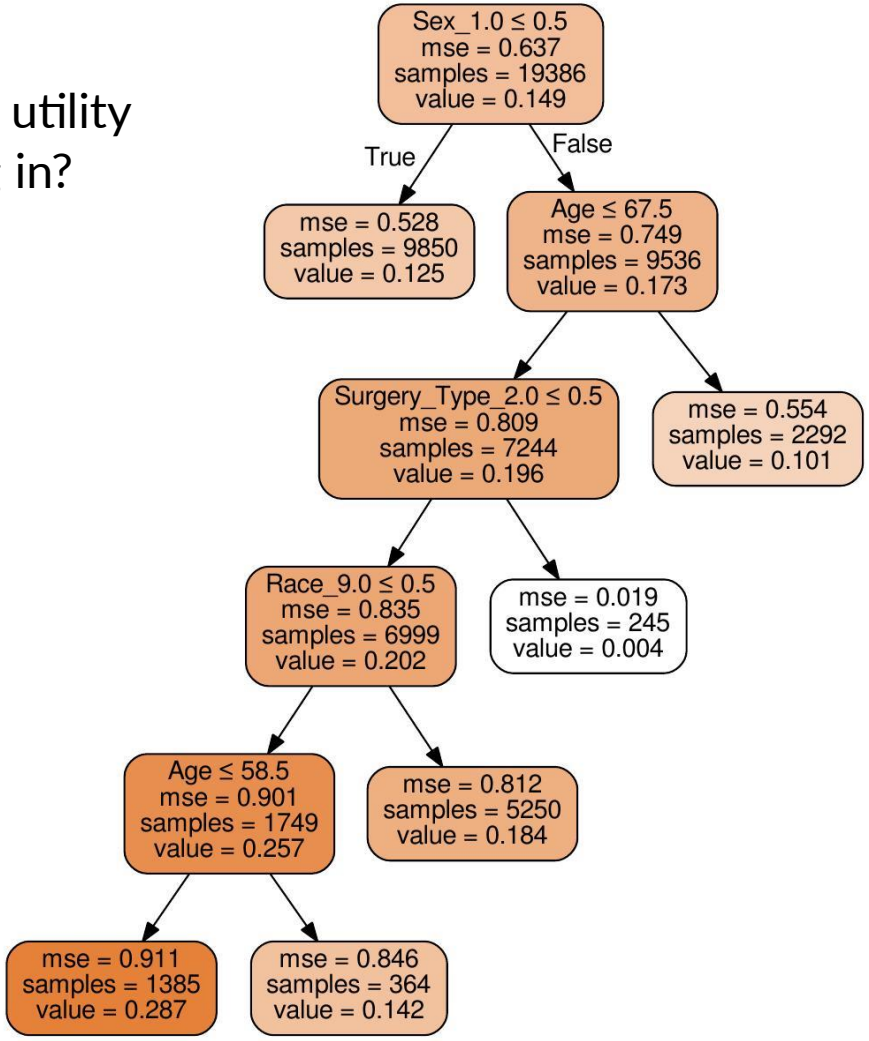
- Transparency is provided by regressing the additional utility (difference of that from **M_full** and **M_base**) on variables like age, sex, race and surgery type
- Decision trees are good candidates to identify the cause of increase or decrease in utility.

Utility explanation for 30 Day Mortality (L) and AKI (R)



Logistic regression as underlying classifier

How much additional utility does this model bring in?



What lies ahead?

- Demonstrated potential in structured data and longitudinal data.

- Tradeoff between the complexity of the model and perioperative stage where the model is being triggered.

- Already work started in use of LLM
  - Foundation models that are similar to the ConVirt model but on larger scale
  - Usage for discharge summaries by searching through the EHR
  - Radiology report summarization
  - Caveats: Trustworthiness of these models, user studies where the clinicians will decide if it correct or not.

- In Indian context, we have Kranium healthcare system, that seems to be dominating one here.
  - Ayushmaan Bharat Digital Mission' s AB Health account that takes care of integration part but more in a federated sense.
  - National Resource Centre for EHR Standards (NRCeS) supports in adoption, planning, procurements, design, development, deployment, and use of EHR in the healthcare system.

Acknowledgements

Collaborators

- Christopher R King
- Bradley A Fritz
- Mohamed Abdelhack
- Michael S Avidan
- Yixin Chen

Funding support

- NIH
- ICTS

WashU Medicine

Thank you!

Contrastive learning technical details in dataset integration

$$L_{within}(i,m,j) = -log(\frac{exp(sim(f_m(x^i_{m,j}) \cdot f_m(\tilde{x}^i_{m,j}))/\tau)}{\sum_{j'} exp(sim(f_m(x^i_{m,j}) \cdot f_m(\tilde{x}^i_{m,j'}))/\tau)}), \quad j, j' \in \{1, \ldots, N\}.$$

Here, $m \in \{k, u\}$ denotes the type of encoder, τ denotes a temperature parameter, $i \in \{A, B\}$ and *sim* denotes the similarity measure.

Next, the multiview contrastive loss for data point j from dataset i is defined as follows:

$$L_{across}(i,j) = -log(\frac{exp(sim(f_k([x^i_{k,j}, \tilde{x}^i_{k,j}]) \cdot f^i_u([x^i_{u,j}, \tilde{x}^i_{u,j}]))/\tau)}{\sum_{j'} exp(sim(f_k([x^i_{k,j}, \tilde{x}^i_{k,j}]) \cdot f^i_u([x^i_{u,j'}, \tilde{x}^i_{u,j'}]))/\tau)}), \quad j, j' \in \{1, \ldots, N\}.$$

The combined within and across-view contrastive loss can be written as follows:

$$L_{CL} = \sum_{i \in \{A,B\}} \{\sum_{j=1}^{N} L_{across}(i,j) + \sum_{m \in \{k,u\}} \sum_{j=1}^{N} L_{within}(i,m,j)\}$$

Contrastive learning decoder version technical details in dataset integration

1. Direct reconstruction loss

$$L_{dir} = \sum_{i \in \{A,B\}} MSE(g_u^i(f_u^i(x_u^i)), x_u^i) + MSE(g_k(f_k(x_k^i)), x_k^i)$$

2. Supervised reconstruction loss that uses the unknown encoder representation to reconstruct the known mapped features for the respective databases.

$$L_{sup\_k} = \sum_{i \in \{A,B\}} MSE(g_k(f_u^i(x_u^i)), x_k^i)$$

3. Combined supervised reconstruction loss that uses a combination of unknown and known mapped encoder representation to reconstruct the unknown and known mapped features respectively for each database. Suppose the combined cross-reconstruction for known mapped and unknown column subsets of dataset i is defined as $z_k^i = g_k(0.5 * f_k(x_k^i) + 0.5 * f_u^i(x_u^i))$ and $z_u^i = g_u^i(0.5 * f_k(x_k^i) + 0.5 * f_u^i(x_u^i))$. Then,

$$L_{comb\_sup\_k} = \sum_{i \in \{A,B\}} MSE(z_k^i, x_k^i) + MSE(z_u^i, x_u^i)$$

Finally, the joint CL autoencoder loss can be written as

$$L_{CL+Dec} = L_{CL} + \lambda * (w_d * L_{dir} + w_s * L_{sup\_k} + w_{sc} * L_{comb\_sup\_k}).$$

Here, MSE is the mean squared error, and w_d, w_s, w_{sc} are the hyperparameters to weigh the three reconstruction losses appropriately. As the CL encoder loss and the MSE are in very different scales, a λ multiplier is needed which needs to be decided during initial experiments.

Partners
HMIS
LMIS
PHR App
Health Locker
Insurance
Connectors
Health Tech
Pharmacy
Government Programs

HMIS

Filter By

Compare MVP 4.0

Website
BIOSOFT HEALTH TECH PVT LTD
Category: HMIS
Integration Completion Date
28/05/2025
Certificate Buy

Website
Karnavati Software Technologies Pvt. Ltd.
Category: HMIS
Integration Completion Date
06/05/2025
Certificate Buy

Website
Suvarna Arogya Suraksha Trust
Category: HMIS
Integration Completion Date
02/05/2025
Certificate Buy

Website
UNNATI DIGITAL SOLUTIONS PRIVATE LIMITED
Category: HMIS
Integration Completion Date
29/04/2025
Certificate Buy

Website
NEXTGEN ESOLUTIONS PVT LTD SUBSIDIARY COMPANY OF AIMIL LIMITED
Category: HMIS
Integration Completion Date
28/04/2025
Certificate

Website
NALAM HCIT SOLUTIONS PVT LTD.
Category: HMIS
Integration Completion Date
24/04/2025
Certificate Buy