Time-series in healthcare: challenges and solutions

AAAI 2022 Tutorial

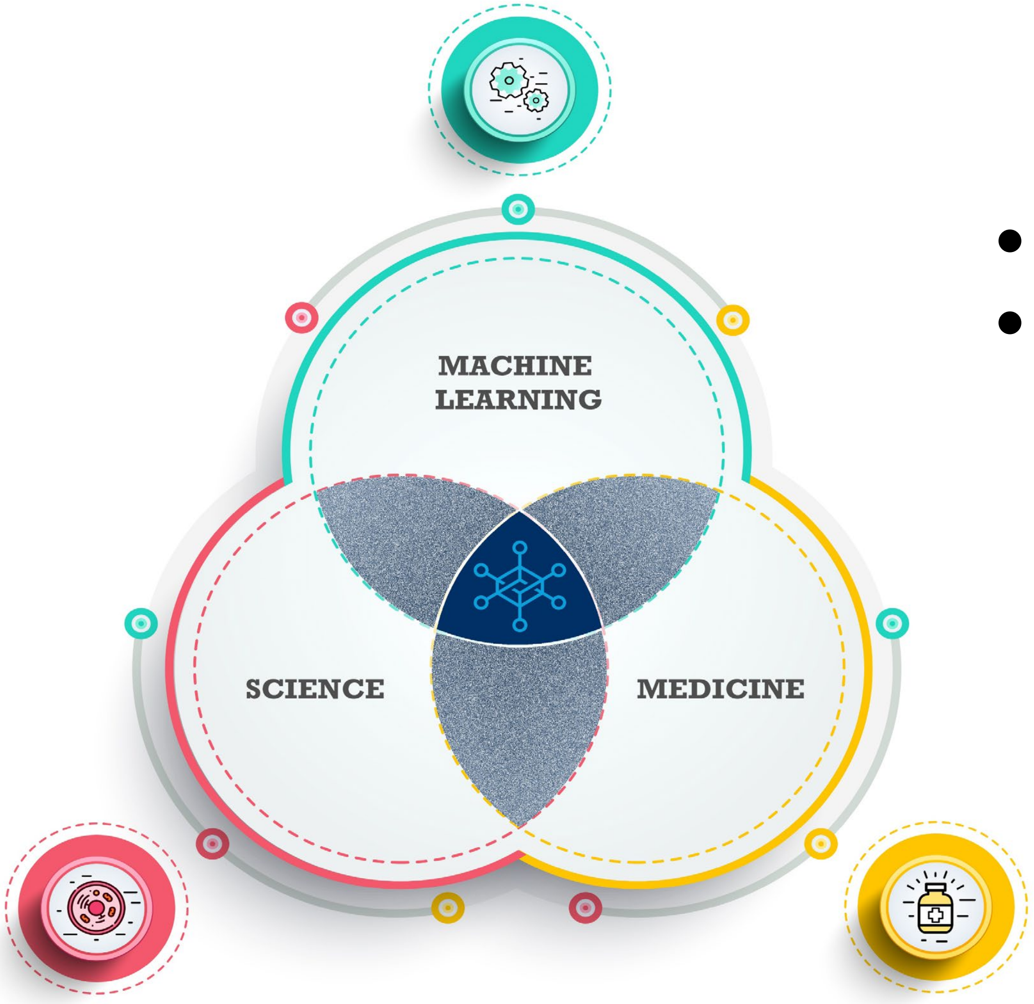
Mihaela van der Schaar & Fergus Imrie

University of Cambridge and University of California, Los Angeles

van_der_Schaar
\ LAB

vanderschaar-lab.com

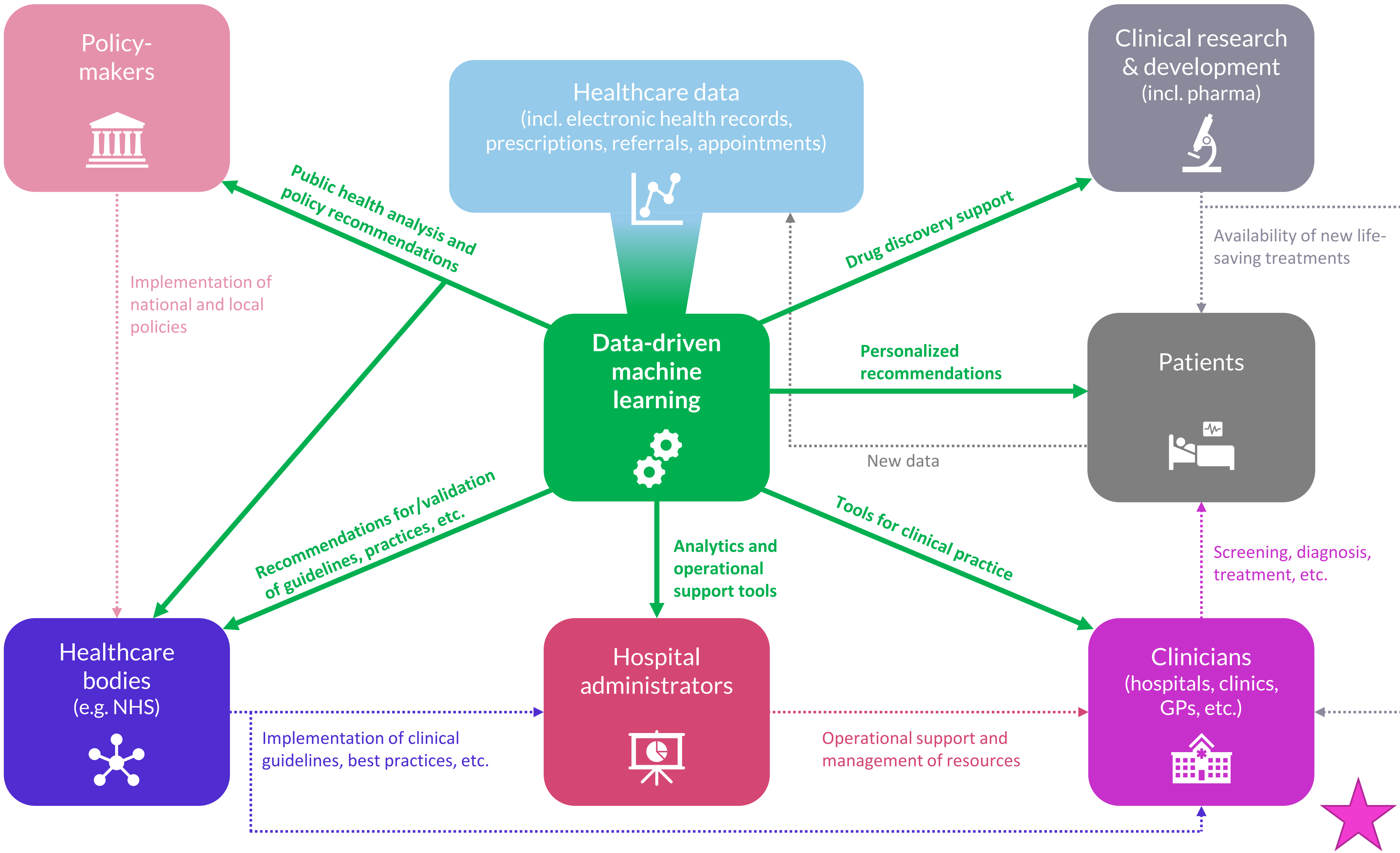
Machine learning & medicine/healthcare/bio-science



- ML/AI drives a revolution in medicine
- Medicine drives innovations in ML/AI

Machine learning can transform medicine & healthcare

1) deliver precision medicine at the patient level
2) understand the basis and trajectories of health and disease
3) empower healthcare professionals and patients
4) inform and improve clinical pathways, better utilize resources & reduce costs
5) transform population health and public health policy
6) enable new discoveries – clinical, therapeutics


Policy-makers
Healthcare data
(incl. electronic health records, prescriptions, referrals, appointments)
Clinical research & development
(incl. pharma)
Public health analysis and policy recommendations
Drug discovery support
Availability of new life-saving treatments
Implementation of national and local policies
Data-driven machine learning
Personalized recommendations
Patients
New data
Recommendations for/validation of guidelines, practices, etc.
Analytics and operational support tools
Tools for clinical practice
Screening, diagnosis, treatment, etc.
Healthcare bodies
(e.g. NHS)
Hospital administrators
Clinicians
(hospitals, clinics, GPs, etc.)
Implementation of clinical guidelines, best practices, etc.
Operational support and management of resources

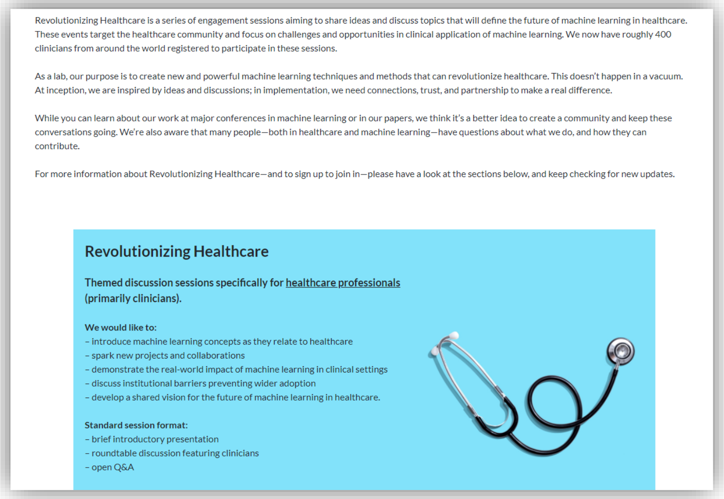
The "augmented" clinician, researcher, patient

Machine learning

…***can't*** do medicine!

…***can*** provide interpretable, trustworthy, <u>actionable information</u>!

Engagement sessions: Revolutionizing Healthcare


Revolutionizing Healthcare is a series of engagement sessions aiming to share ideas and discuss topics that will define the future of machine learning in healthcare. These events target the healthcare community and focus on challenges and opportunities in clinical application of machine learning. We now have roughly 400 clinicians from around the world registered to participate in these sessions.

As a lab, our purpose is to create new and powerful machine learning techniques and methods that can revolutionize healthcare. This doesn't happen in a vacuum. At inception, we are inspired by ideas and discussions; in implementation, we need connections, trust, and partnership to make a real difference.

While you can learn about our work at major conferences in machine learning or in our papers, we think it's a better idea to create a community and keep these conversations going. We're also aware that many people—both in healthcare and machine learning—have questions about what we do, and how they can contribute.

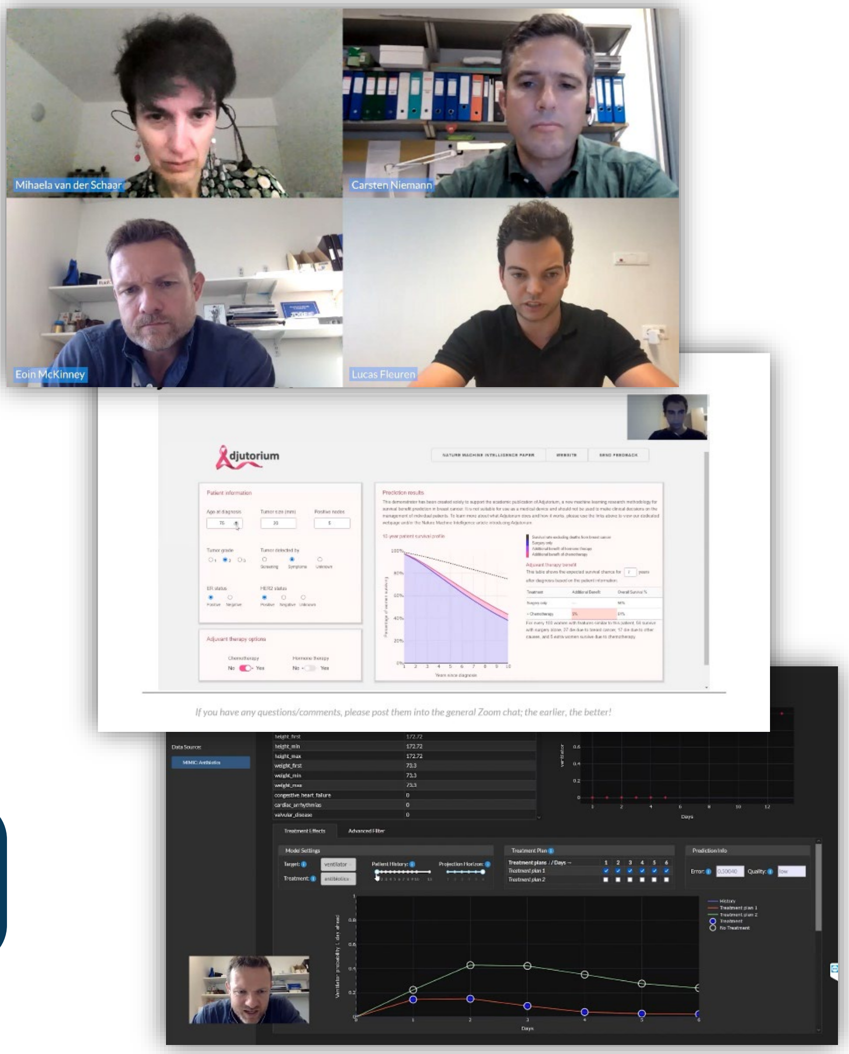
For more information about Revolutionizing Healthcare—and to sign up to join in—please have a look at the sections below, and keep checking for new updates.

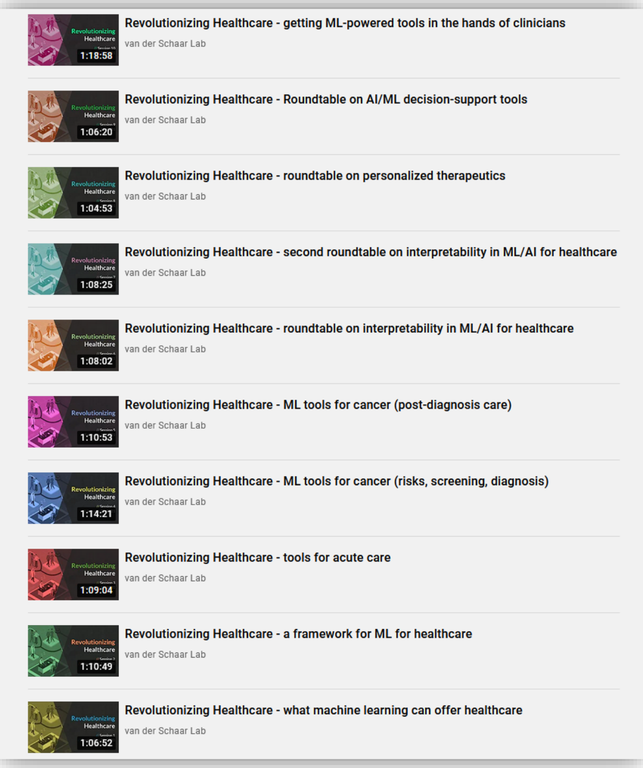
Revolutionizing Healthcare

Themed discussion sessions specifically for healthcare professionals (primarily clinicians).

We would like to:
– introduce machine learning concepts as they relate to healthcare
– spark new projects and collaborations
– demonstrate the real-world impact of machine learning in clinical settings
– discuss institutional barriers preventing wider adoption
– develop a shared vision for the future of machine learning in healthcare.

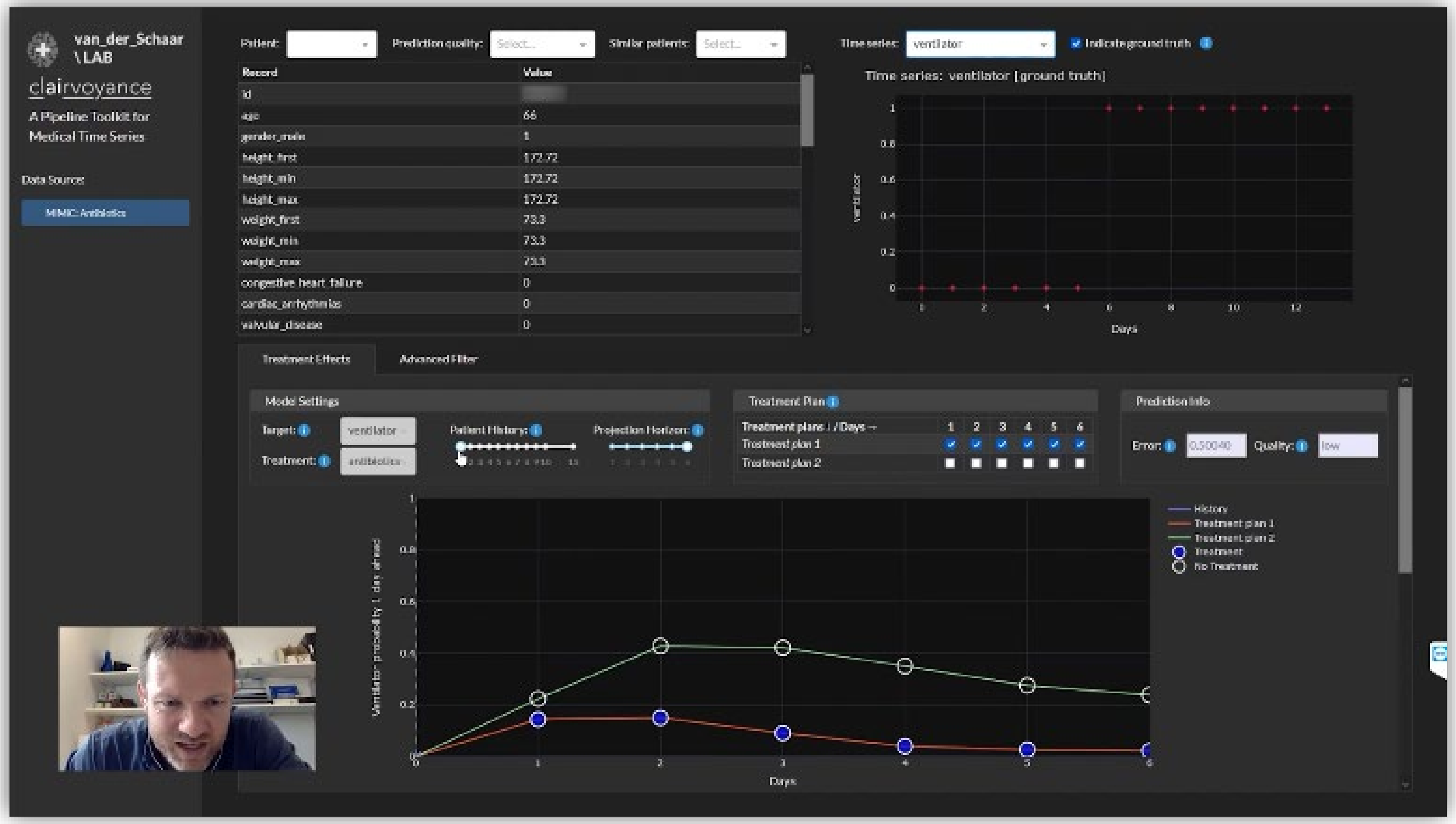
Standard session format:
– brief introductory presentation
– roundtable discussion featuring clinicians
– open Q&A





https://www.vanderschaar-lab.com/
→ Engagement sessions
→ Revolutionizing Healthcare

An integrated clinical decision ecosystem using ML



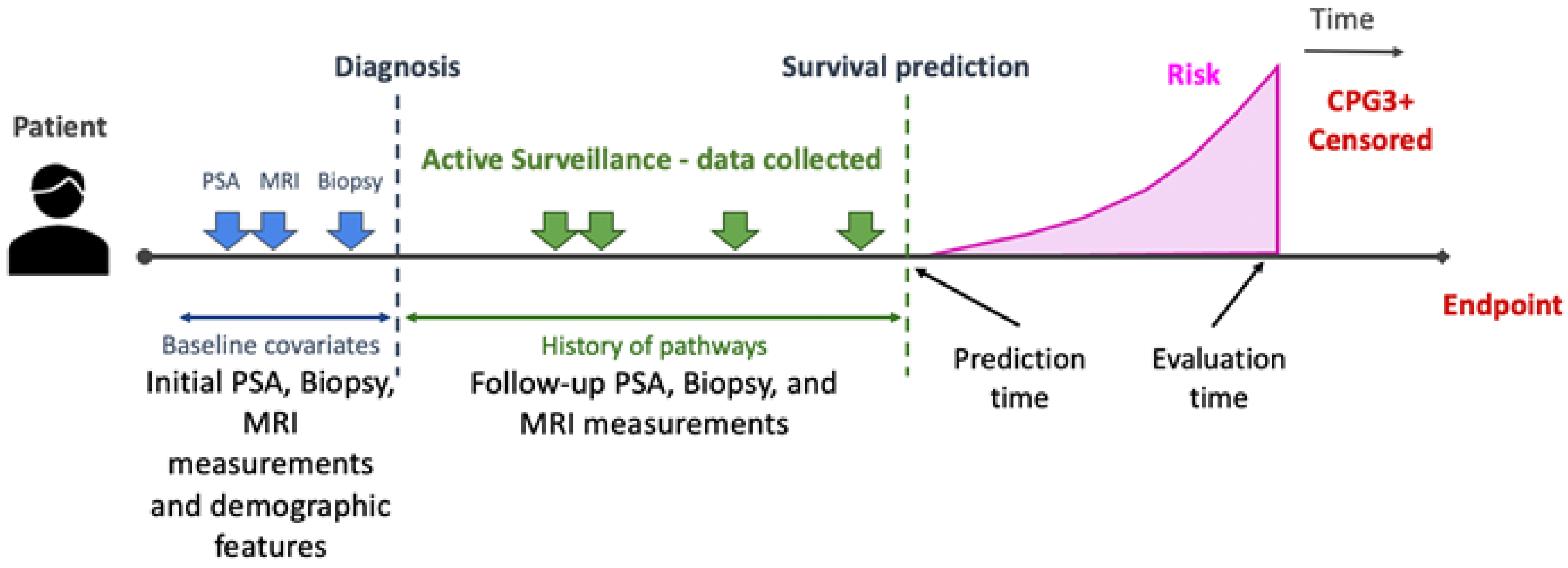
An integrated clinical decision ecosystem using ML

An integrated clinical decision support ecosystem using machine learning to provide **patient-level recommendations and support**

Integrated care:
- Prevention
- Screening
- (Early) Diagnosis
- Treatment
- Monitoring

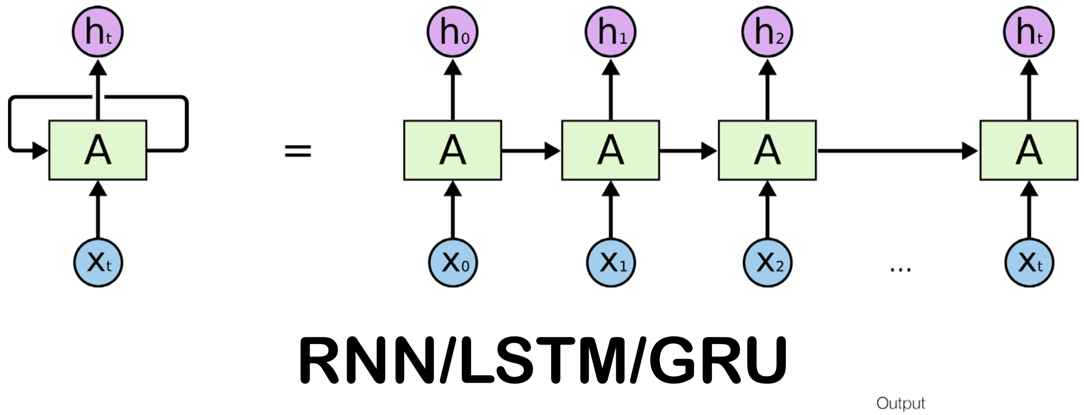
Multiple venues/areas:
- In-patient/out-patient
- At home

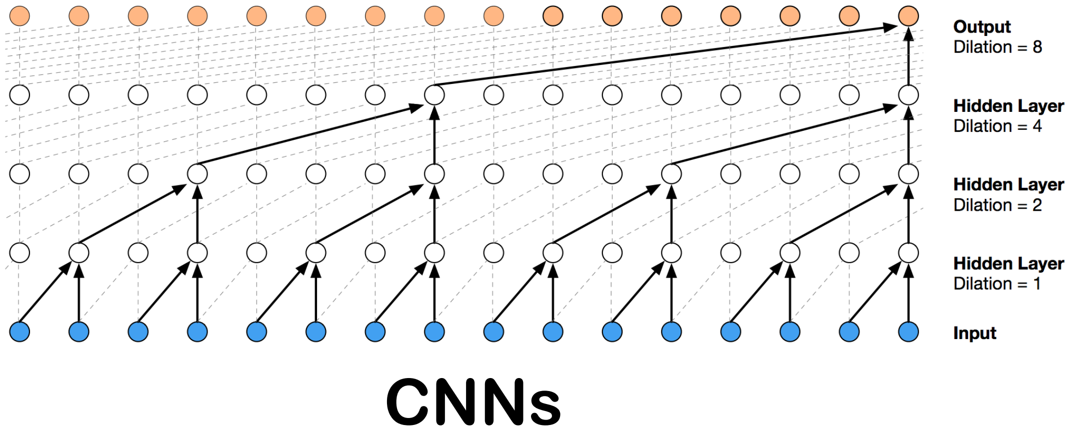
Many stakeholders in every stage of care
- Clinicians, nurses
- Healthcare planners
- Patients!



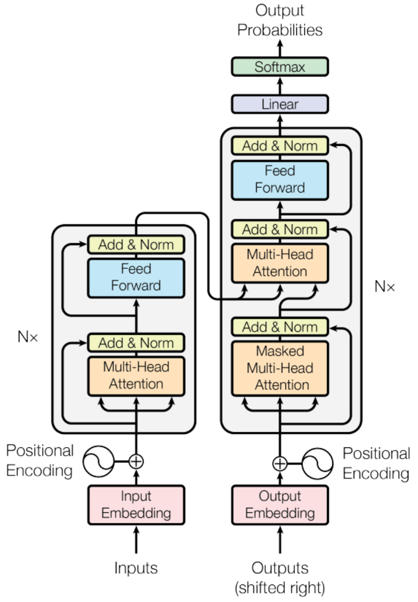
Today's tutorial

Time-series models

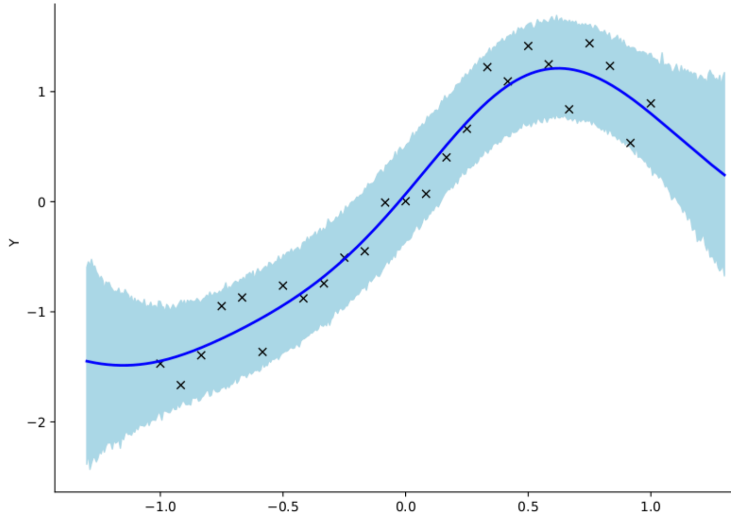




Transformers



Gaussian Process



Time-series models: Resources

Forecasting Big Time Series, Faloutsos et al., KDD Tutorial (2019) [Link]

Understanding LSTM Networks, Christopher Olah [Link]

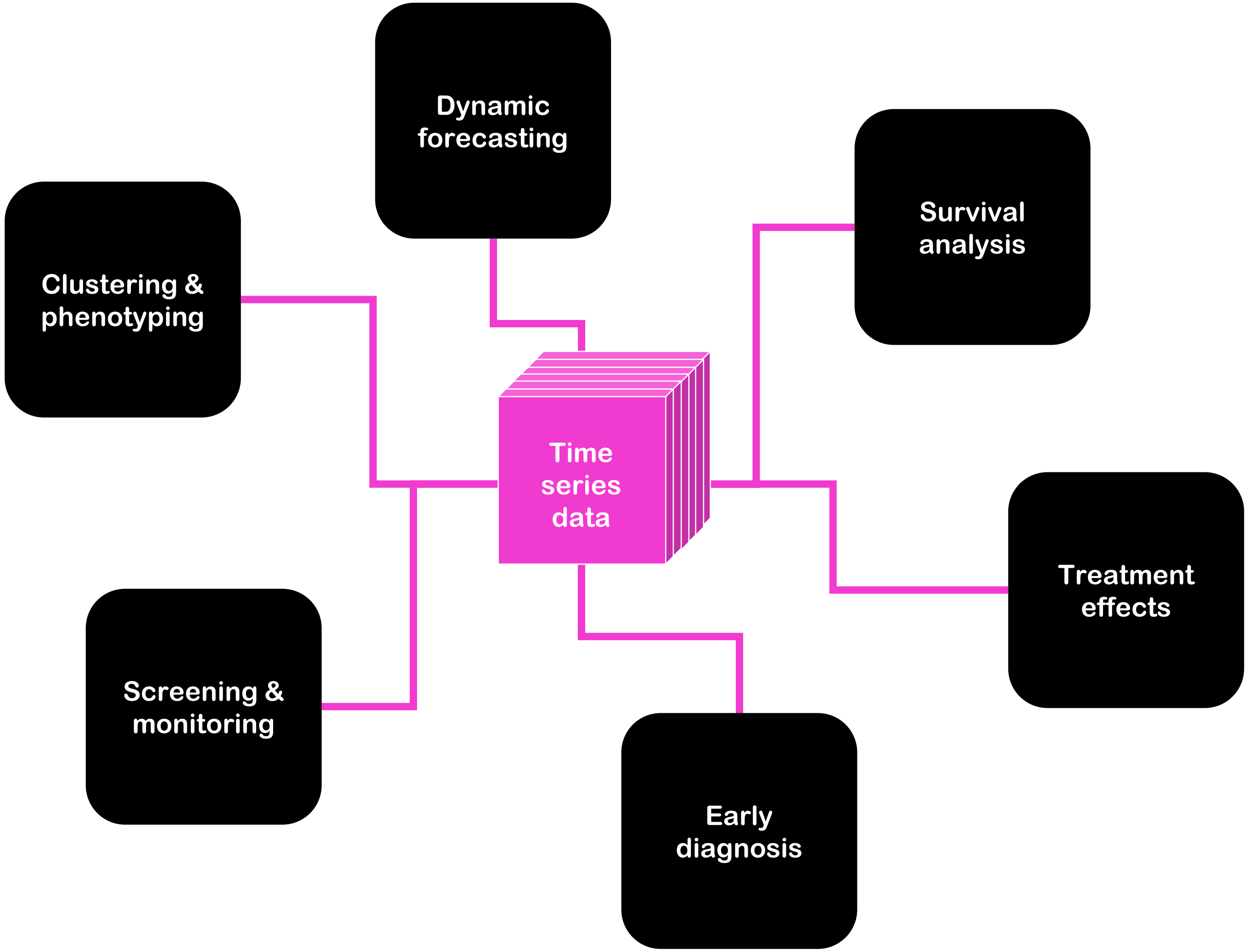
Gaussian processes for Machine Learning, Rasmussen & Williams [Link]

The Art of Gaussian Processes, Mattos & Tobar, NeurIPS Tutorial (2021) [Link]

Deep Implicit Layers - Neural ODEs, Deep Equilibrium Models, and Beyond, Kolter, Dubenaud & Johnson, NeurIPS Tutorial (2020) [Link]

…and many more!

Time-series in healthcare: a multi-faceted problem



Time-series in healthcare: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

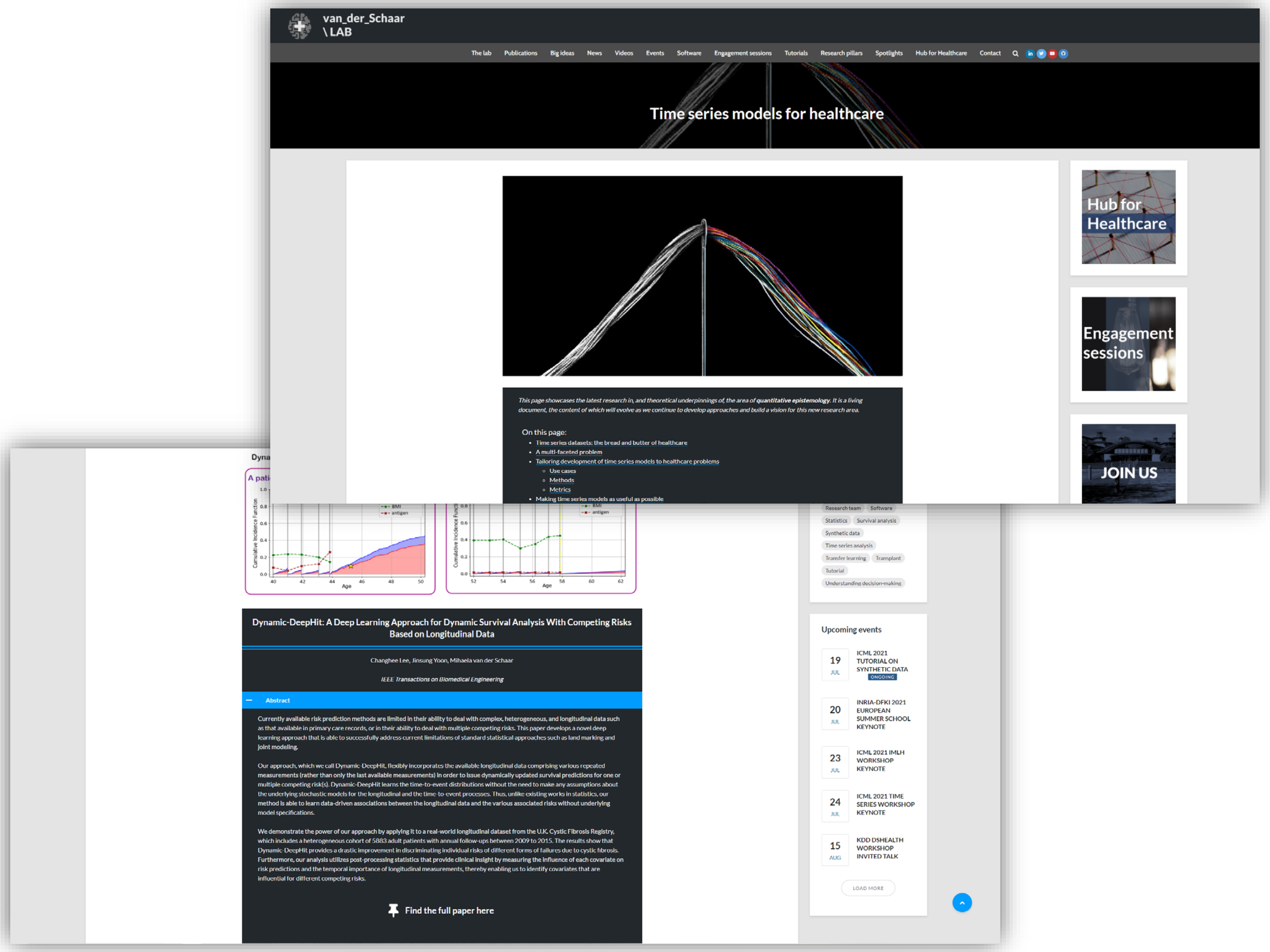
- Reproducibility and visualization

Part 1: tailoring development of time series models to healthcare challenges

Part 2: making time series models as useful as possible

More information and updates

vanderschaar-lab.com/
→ Research pillars
→ Time series



Engagement sessions: Inspiration Exchange

vanderschaar-lab.com/
→ Engagement sessions
→ Inspiration Exchange

Inspiration Exchange

Themed discussion sessions specifically for machine learning students (particularly masters, Ph.D., and post-docs).

We would like to:
– discuss machine learning models and techniques
– share ideas about how machine learning can revolutionize healthcare
– spark new projects and collaborations
– raise awareness about this unique and exciting area of machine learning.

Part 1:
tailoring development of time series models to healthcare challenges

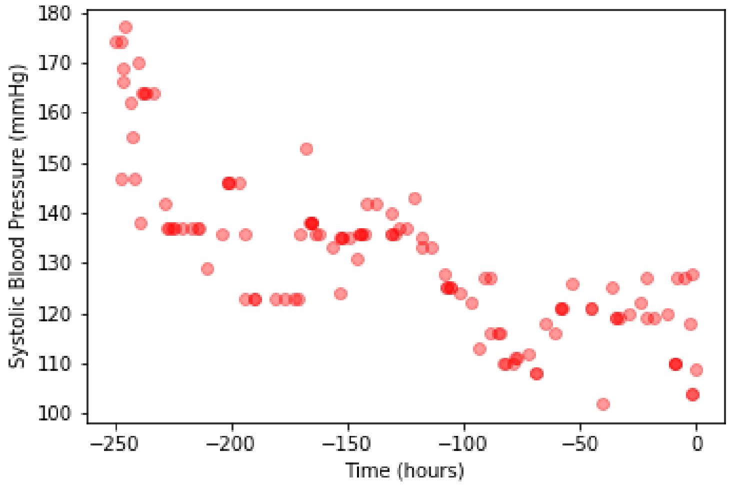
Time-series in healthcare: a multi-faceted problem

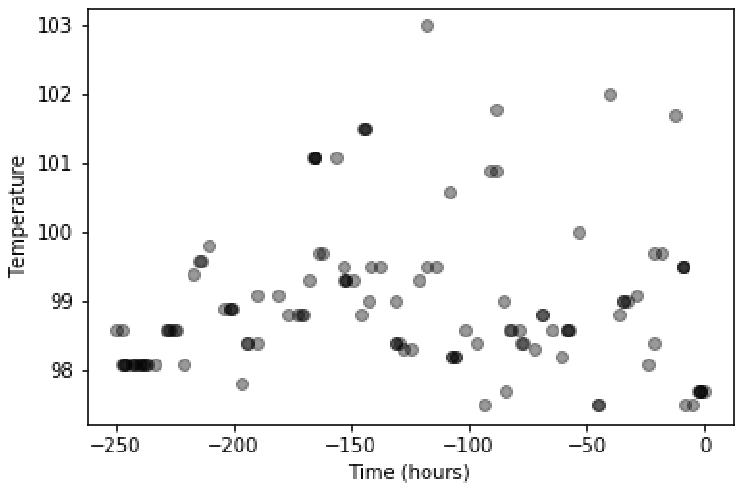
1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

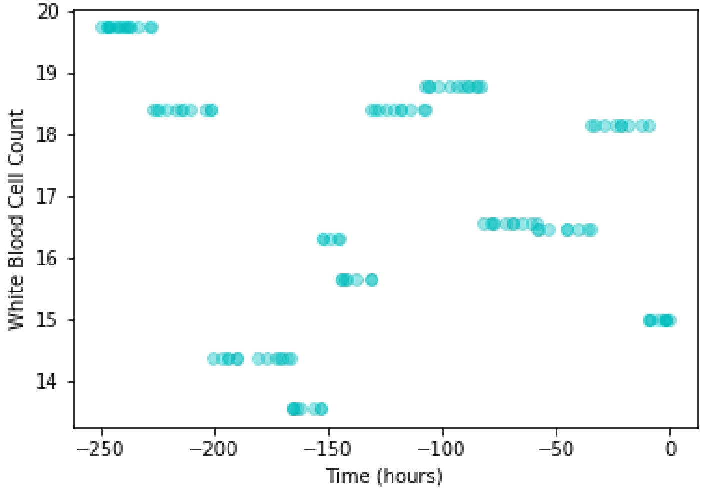
- Reproducibility and visualization

Healthcare data - Unique challenges

- Multiple streams of measurements
- Measurements are sparse, irregularly and informatively sampled
- Multiple outcomes of interest (various events of interest, various morbidities)
- True clinical states are unobserved (e.g., onset of diseases)
- Many possible patterns (heterogeneous phenotypes, comorbidities)







Time-series analysis and dynamic forecasting

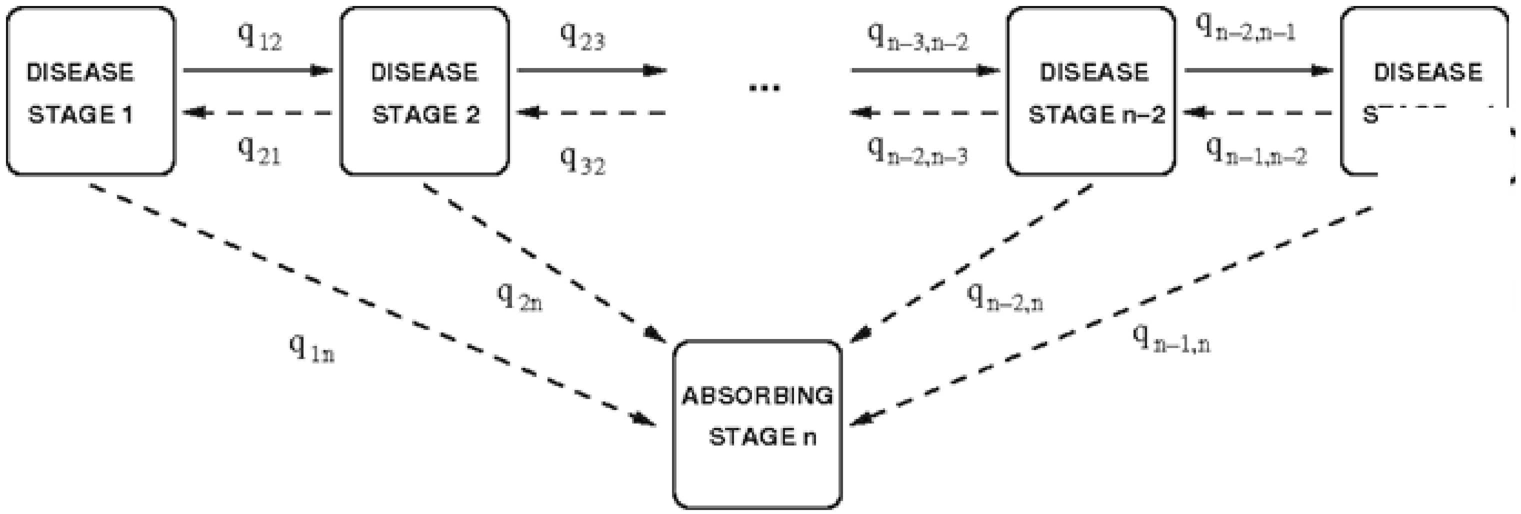
- **Build** disease progression **models**
  - Understand and model carefully the available data!
- Learn the model parameters from available EHR data (Training time)
- Issue dynamic forecasts for the patient at hand (Test time/Run-time)
- Unravel new understanding of disease progression
  - Population
  - Sub-groups of patients
  - Personalized

Input data (time-series) → **Disease progression model** $\mathcal{M}$ → Dynamic forecasts

Observational EHR data + clinical knowledge

Current disease progression models: formalisms

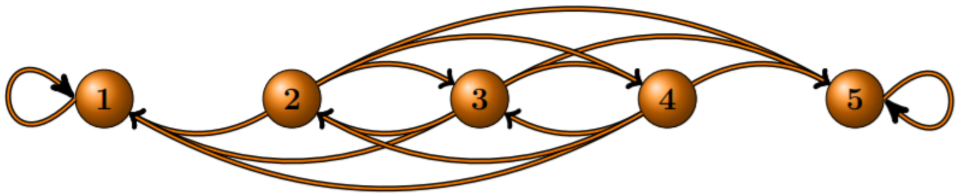
Markov Models $P(Z_{n+1} \mid \mathcal{H}_{t_n}) = P(Z_{n+1} \mid Z_n)$



Disadvantages

- Observable models
- One disease at a time
- "Average" patient

Population-level representation of disease states

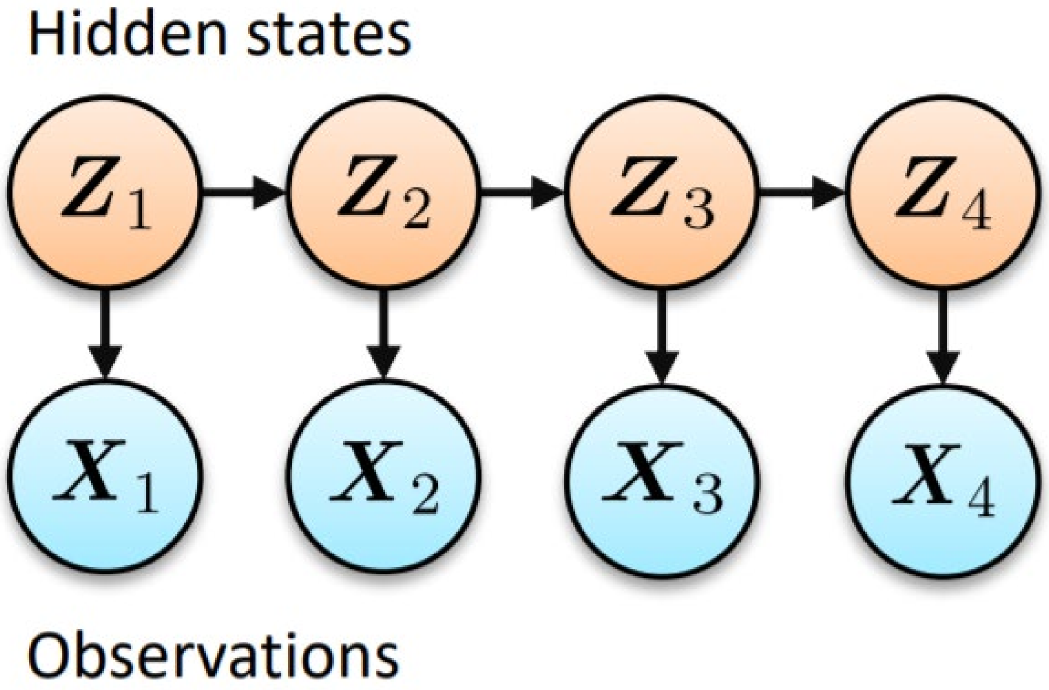


Current disease progression models: formalisms

Hidden Markov Models (HMMs)

Introducing latent (hidden/unobservable) disease states

Disease Stages



Clinical findings
Lab measurements
Vital signs
Treatments
Events of interest
Observation times

Markov models?

History matters!

Ignore history
- Previous states
- Order of states
- Duration in a state

One size fits all!

Only capture population-level transitions across progression stages
Ignores individual clinical trajectories

Recurrent Neural Nets?

Two central goals

Goal A: Accurately forecast individual-level disease trajectories

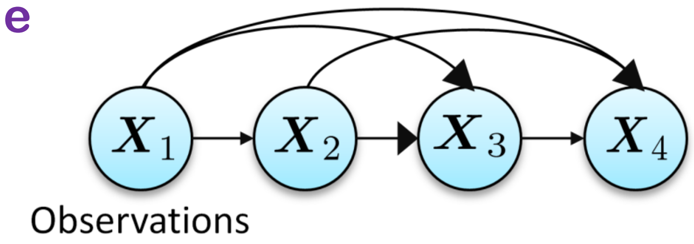
- What are the risks of mortality, relapse, comorbidities, complications, etc. in the future?

Goal B: Understand disease progression mechanisms.

- Underlying latent structure of disease evolution
- Patients' subgroup analysis
- Refined phenotypes

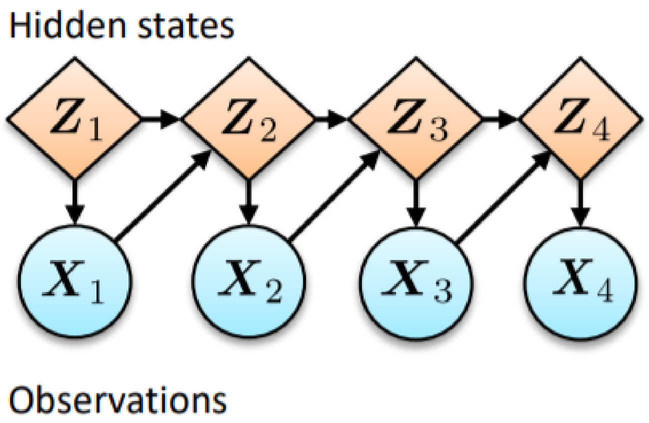
Deep learning models?

Observable models



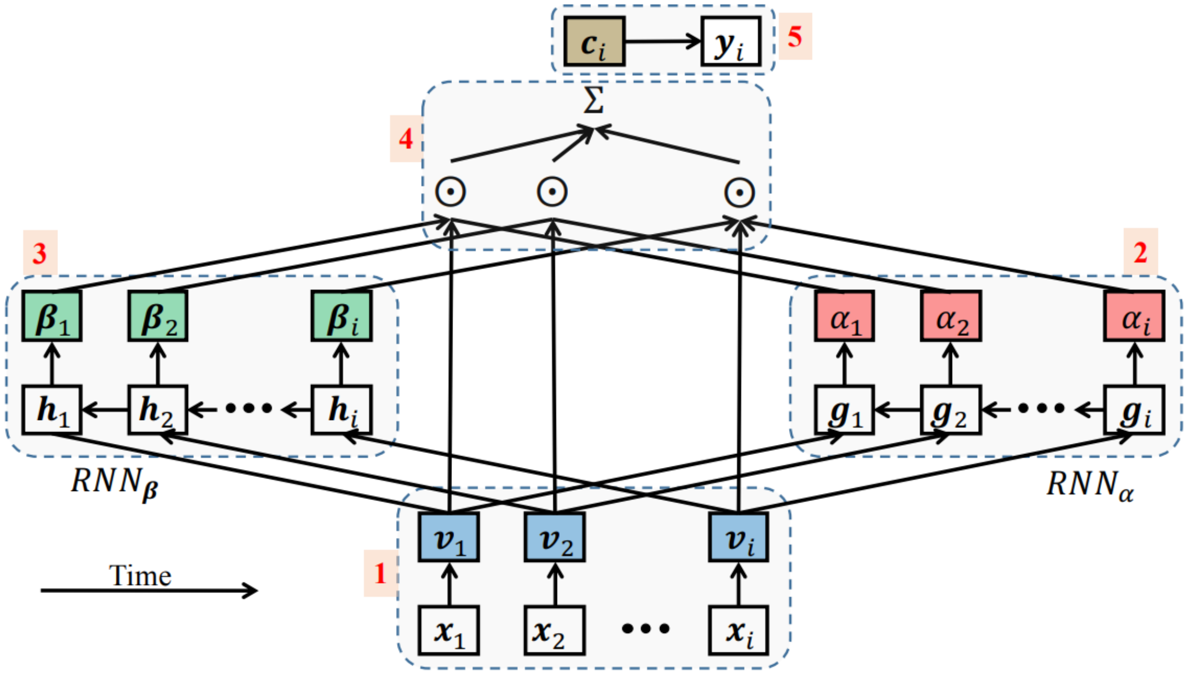
No latent structure

RNN



Uninterpretable predictions, Uninterpretable latent structure

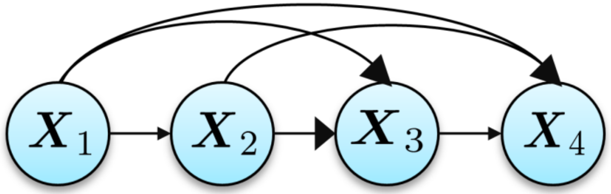
Retain [Choi et al., NeurIPS 2016]



1. Embed observations $\{x_1,\ldots,x_i\}$
2. Generate α using RNN_α
3. Generate ß using $RNN_ß$
4. Generate context vector using attention α, ß and representations ν
5. Make prediction

Deep learning models?

Observable models

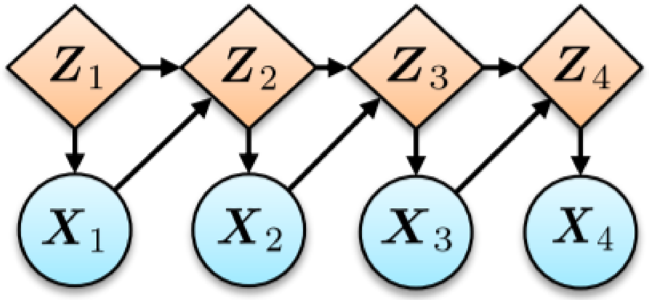


Observations

No latent structure

RNN

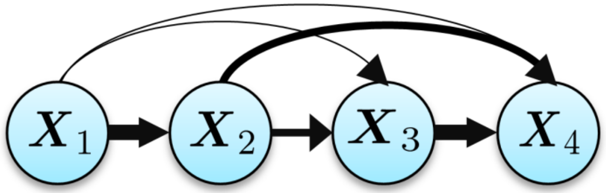
Hidden states



Observations

Uninterpretable predictions, Uninterpretable latent structure

RETAIN



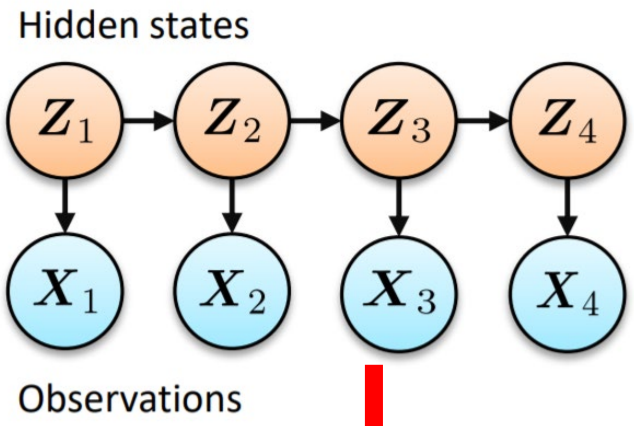
Observations

Interpretable predictions, Uninterpretable latent structure

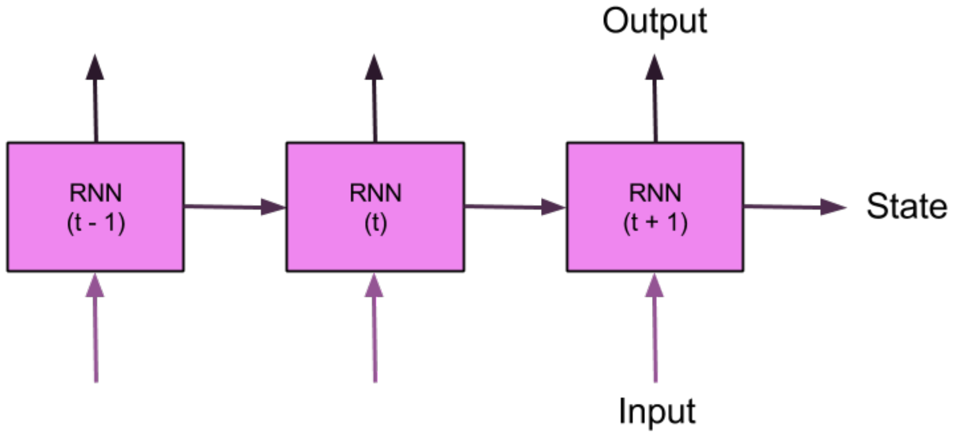
Attentive state space models [Alaa & vdS, 2018, NeurIPS 2019]

Main idea: a general and versatile deep probabilistic model capturing complex, non-stationary representations for patient-level trajectories

Maintain probabilistic structure of HMMs



But use RNNs to model state dynamics



$$P(\{\boldsymbol{Z}_m\}_m, \{\boldsymbol{X}_m\}_m \,|\, \boldsymbol{Y}, \{t_m\}_m) = \prod_{m'=1}^{m} P(\boldsymbol{X}_{m'} \,|\, \boldsymbol{Z}_{m'}) \cdot P(\boldsymbol{Z}_{m'} \,|\, \mathcal{F}_{t_{m'-1}})$$

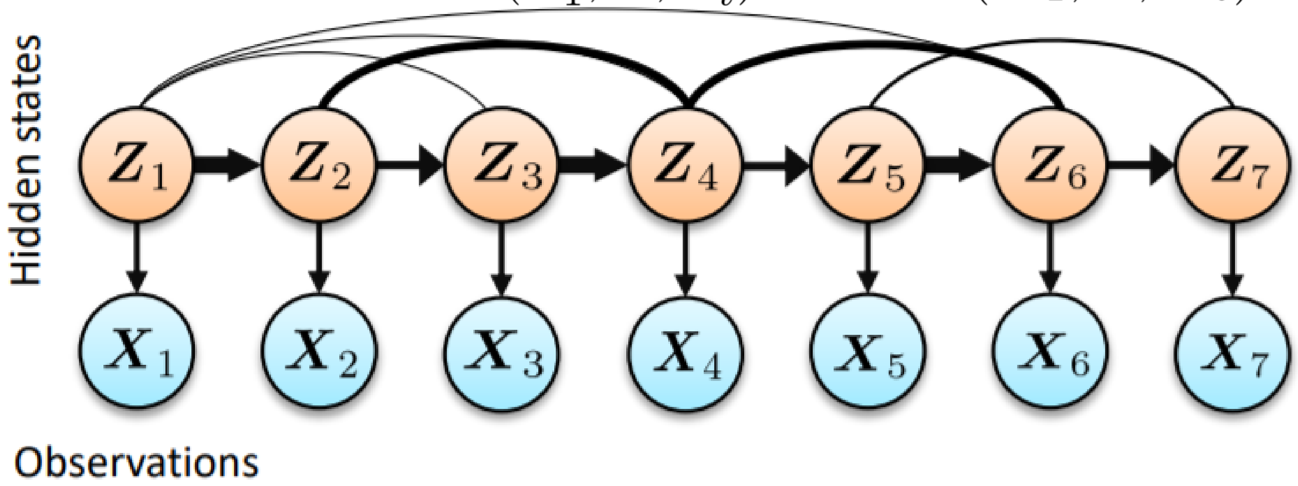
Emission

Transition

Going beyond Markov

- Attention weights determine the influences of past state realizations on future state transitions

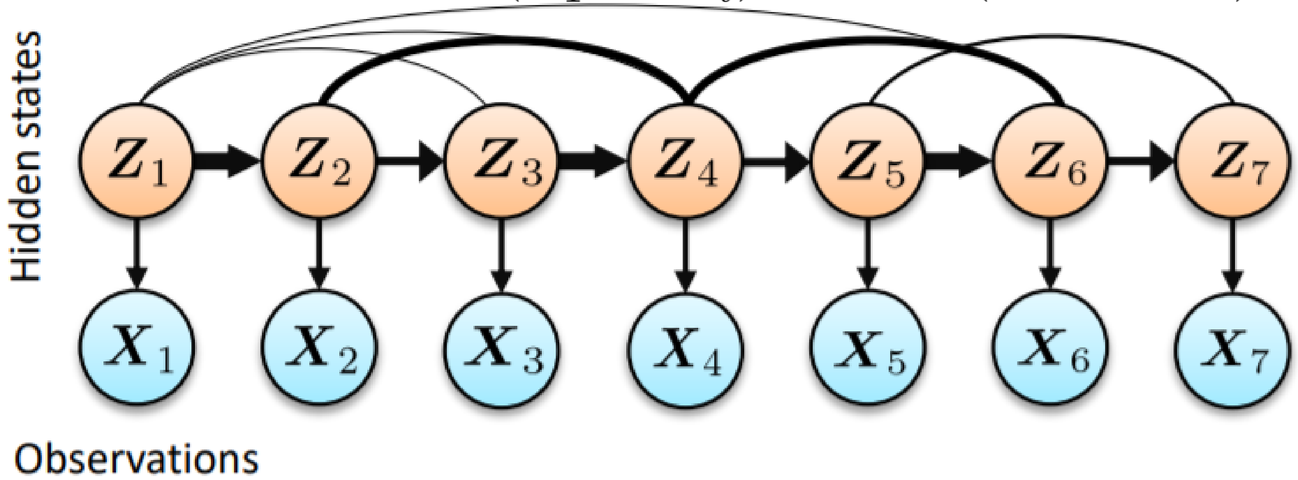
$$\underbrace{(\boldsymbol{\alpha}_1^t, \ldots, \boldsymbol{\alpha}_t^t)}^{\text{Attention weights}} = \mathrm{RNN}(\overbrace{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_t}^{\text{Patient context}})$$



Overcomes shortcoming of Markov Models

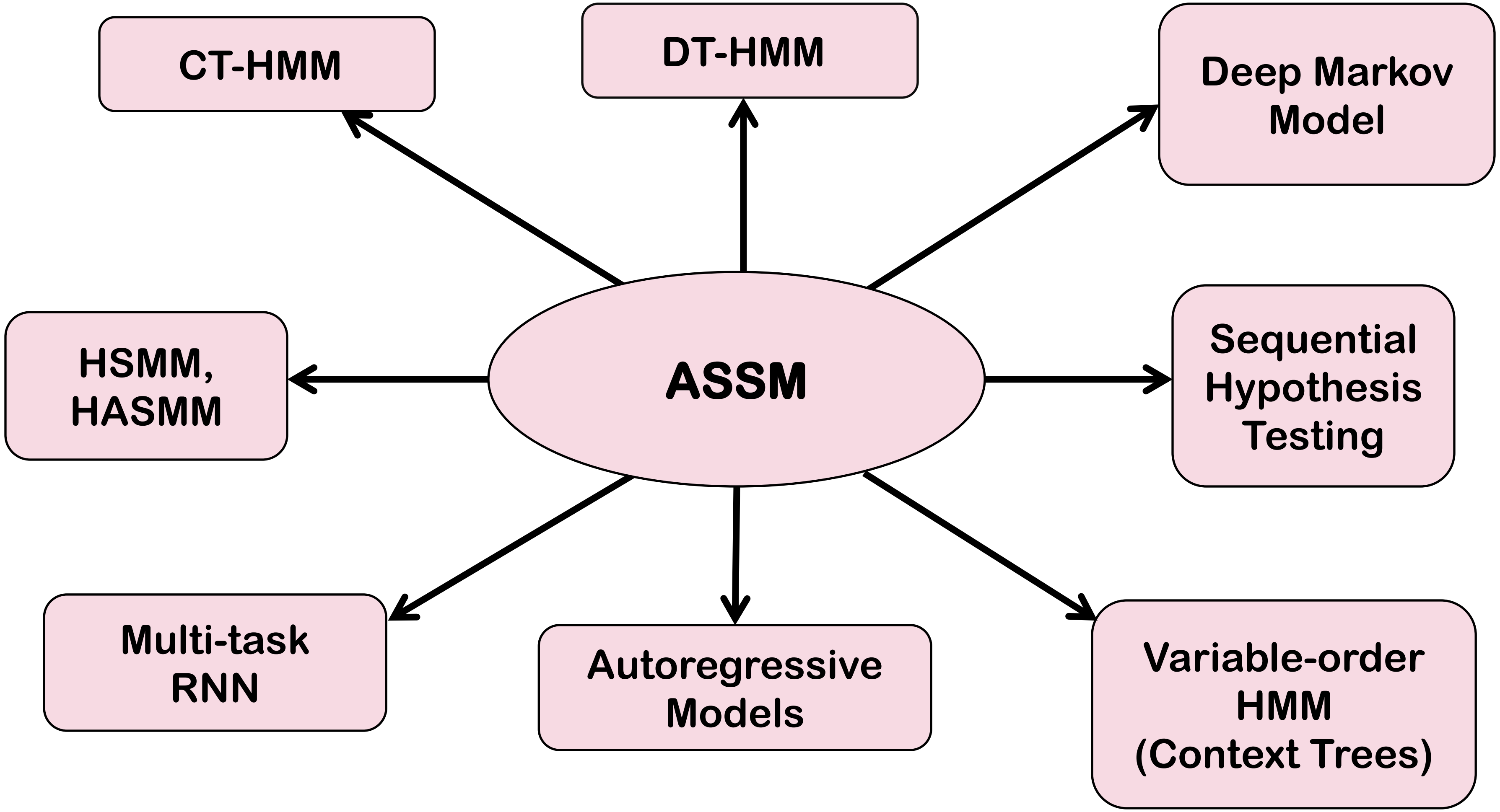
Attention weights create a "soft" version of a non-stationary, variable-order Markov model where underlying dynamics of a patient change over time based on an individual's clinical context!

$$\underbrace{(\boldsymbol{\alpha}_1^t, \ldots, \boldsymbol{\alpha}_t^t)}_{\text{Attention weights}} = \mathbf{RNN}\underbrace{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_t)}_{\text{Patient context}}$$



ASSM - "memory" is shaped by patient's current context (clinical events, treatments, etc.)

ASSM: A General, Versatile and Clinically Actionable Model



Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

Dynamic-DeepHit [Lee & vdS, TBME 2019]

Longitudinal survival data: $\mathcal{D} = \left\{\left(\mathcal{X}^{(i)}, \tau^{(i)}, k^{(i)}\right)\right\}_{i=1}^{N}$

- $\mathcal{X}^i$: History of longitudinal measurements until time the last measurement
  - $\mathcal{X}^i(t) = \{x^i(t_j^i): 0 \leq t_j^i \leq t \;\; for \; j = 1, \cdots, M^i\}$ where M^i is the number of measurements.
- τ : Time-to-event including right-censoring
- k : Event label

Dynamic-DeepHit [Lee & vdS, TBME 2019]

Estimation of the incidence of the occurrence of an event while taking competing risks into account!

New goal: Estimate "dynamic" Cumulative Incidence Function (CIF)

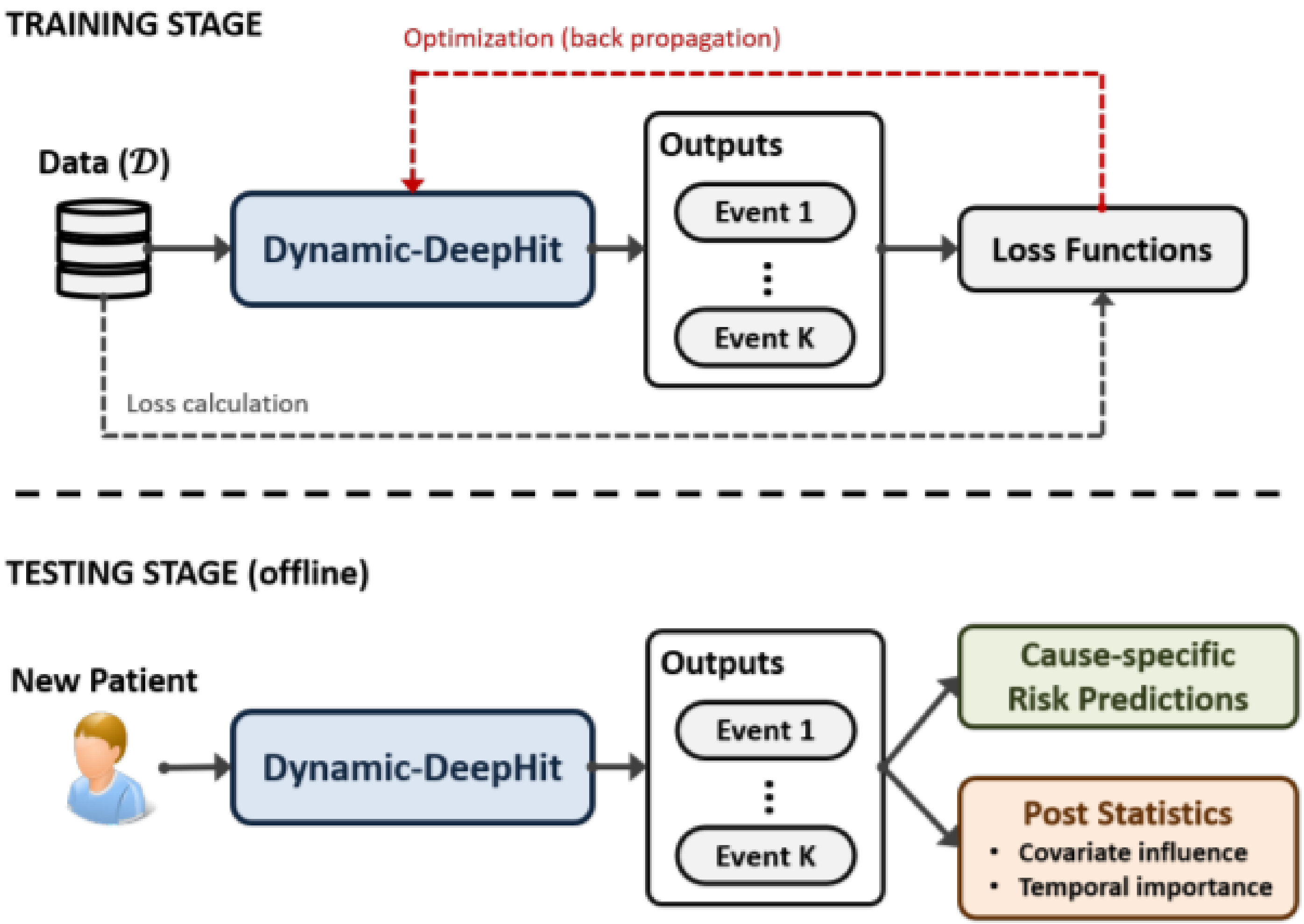
$$\hat{F}_k(\tau|\mathcal{X}^*) \stackrel{\text{def}}{=} P(T \leq \tau, E = k|\underline{\mathcal{X}^*}, \underline{T > t^*_{M^*}})$$

Longitudinal measurements accrued by the time of risk predictions

The patient was alive at the time of the last measurement!

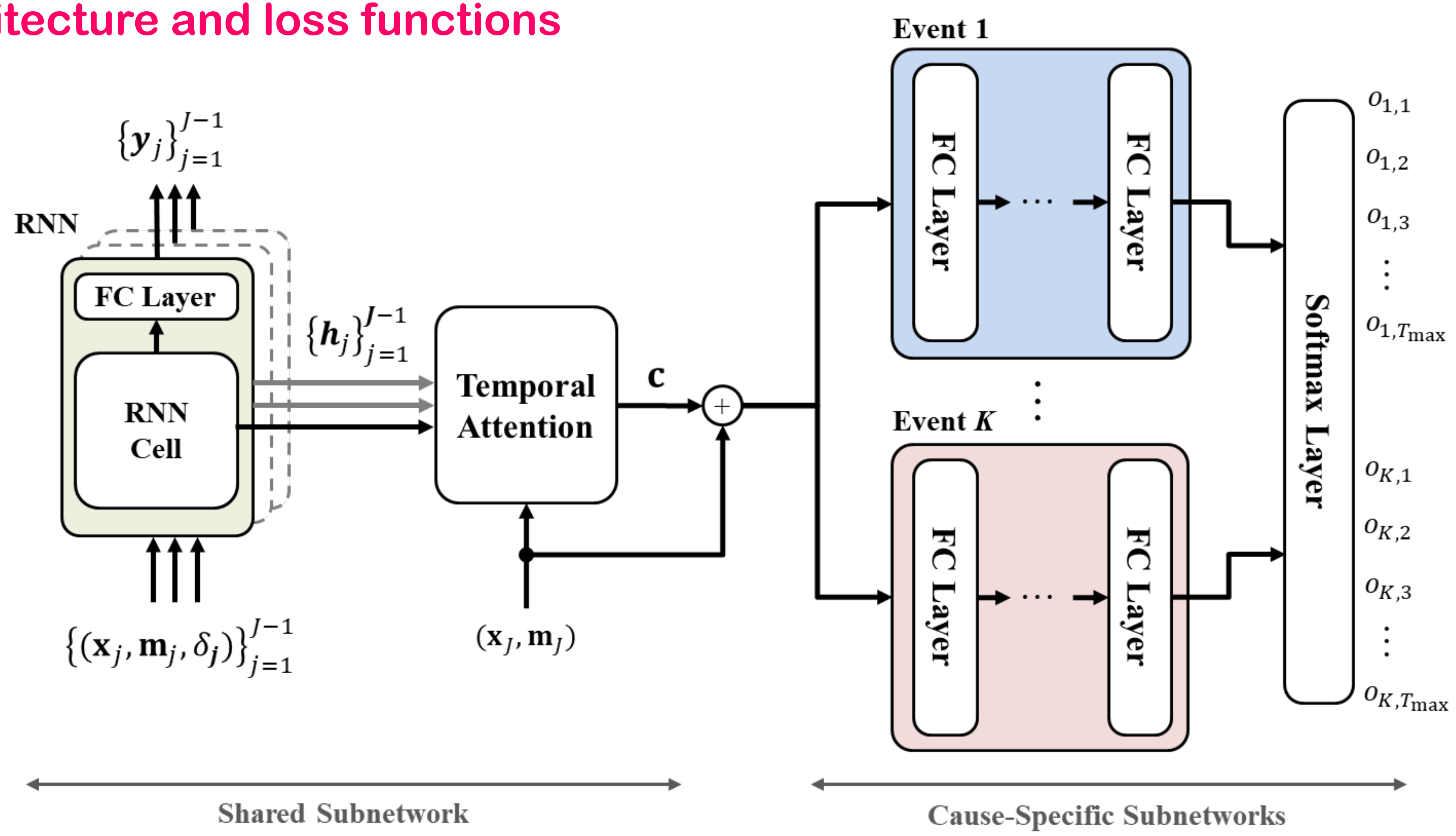
Dynamic-DeepHit [Lee & vdS, TBME 2019]

High-level schematic



Dynamic-DeepHit [Lee & vdS, TBME 2019]

Network architecture and loss functions



Dynamic-DeepHit [Lee & vdS, TBME 2019]

Network architecture and loss functions

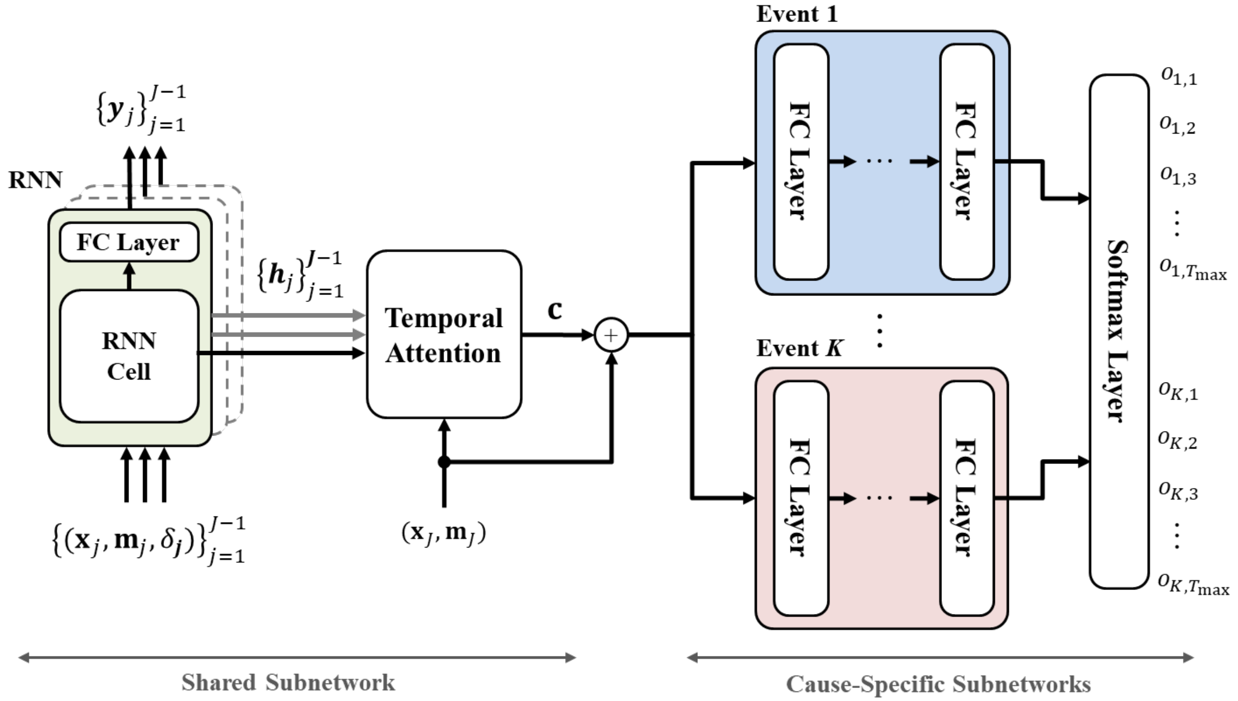
Loss functions:

$$\mathcal{L}_{Total} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$$

Log-likelihood of joint TTE distribution

Ranking loss

Step-ahead Prediction loss



Dynamic-DeepHit [Lee & vdS, TBME 2019]

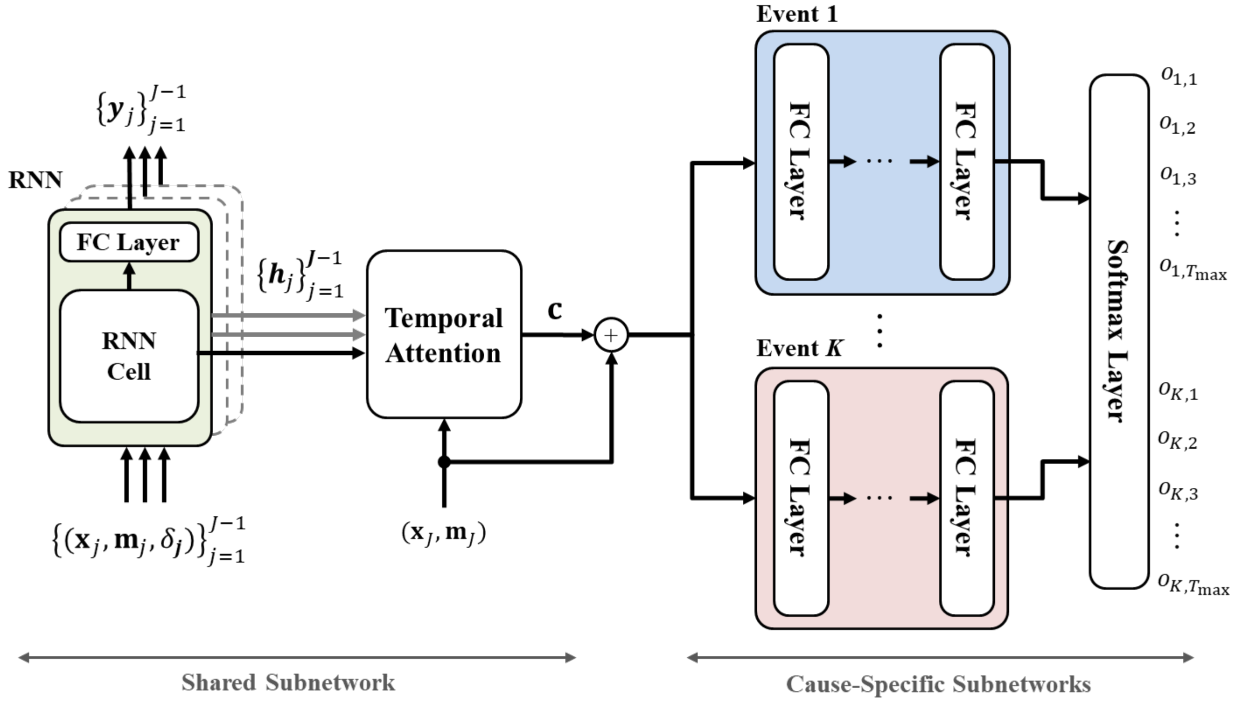
Network architecture and loss functions

Loss functions:

$$\mathcal{L}_{Total} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$$

Log-likelihood of joint TTE distribution

$$\mathcal{L}_1 = -\sum_{i=1}^{N} \left[\mathbb{1}(k^i \neq \varnothing) \cdot \log\left(\frac{o^i_{k^i,\tau^i}}{1 - \sum_{k\neq\varnothing}\sum_{n\leq t^i_{J^i}} o^i_{k,n}} \right) + \mathbb{1}(k^i = \varnothing) \cdot \log\left(1 - \sum_{k\neq\varnothing} \hat{F}_k(\tau^i|\mathcal{X}^i)\right) \right],$$



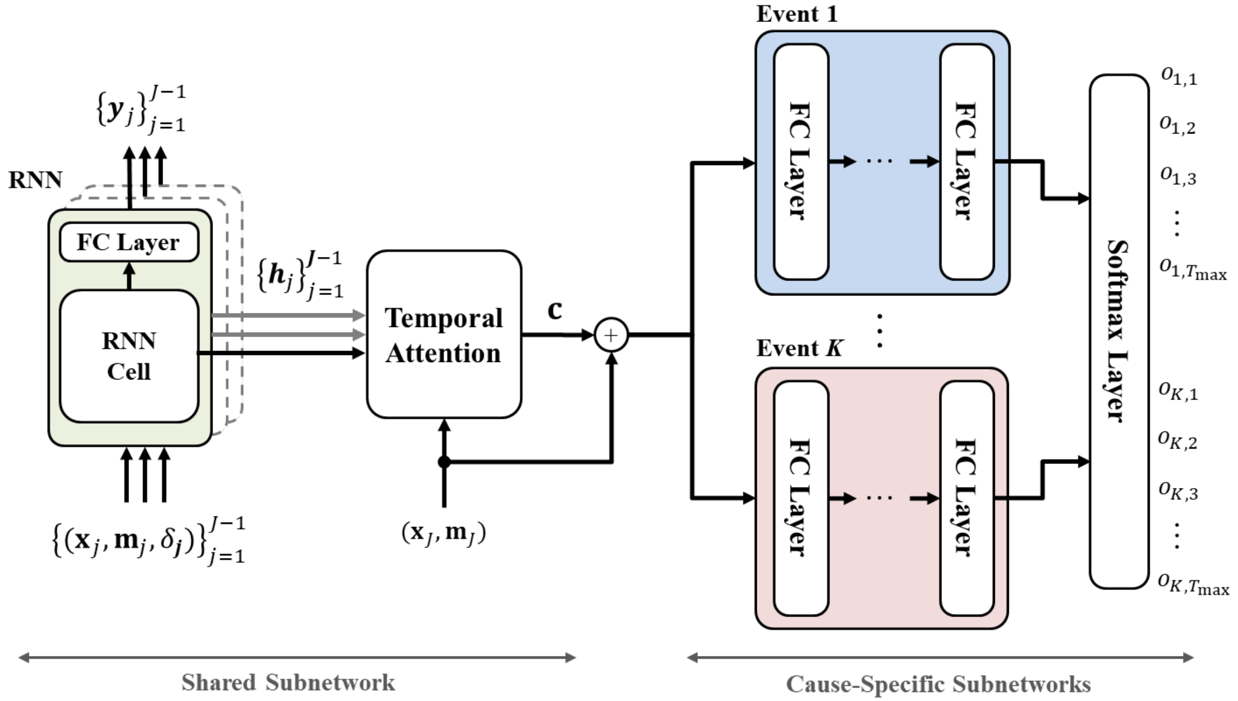
Dynamic-DeepHit [Lee & vdS, TBME 2019]

Network architecture and loss functions

Loss functions:

$$\mathcal{L}_{Total} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$$

Ranking loss



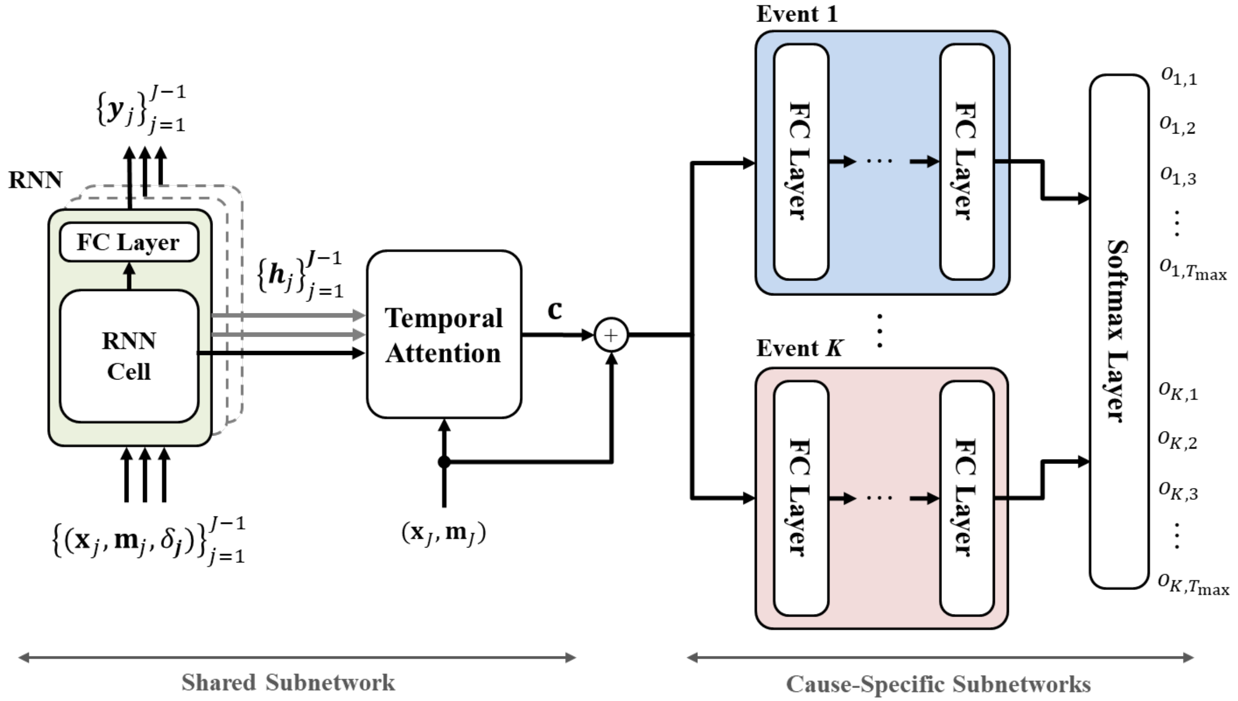
Dynamic-DeepHit [Lee & vdS, TBME 2019]

Network architecture and loss functions

Loss functions:

Step-ahead Prediction loss

$$\mathcal{L}_{Total} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$$



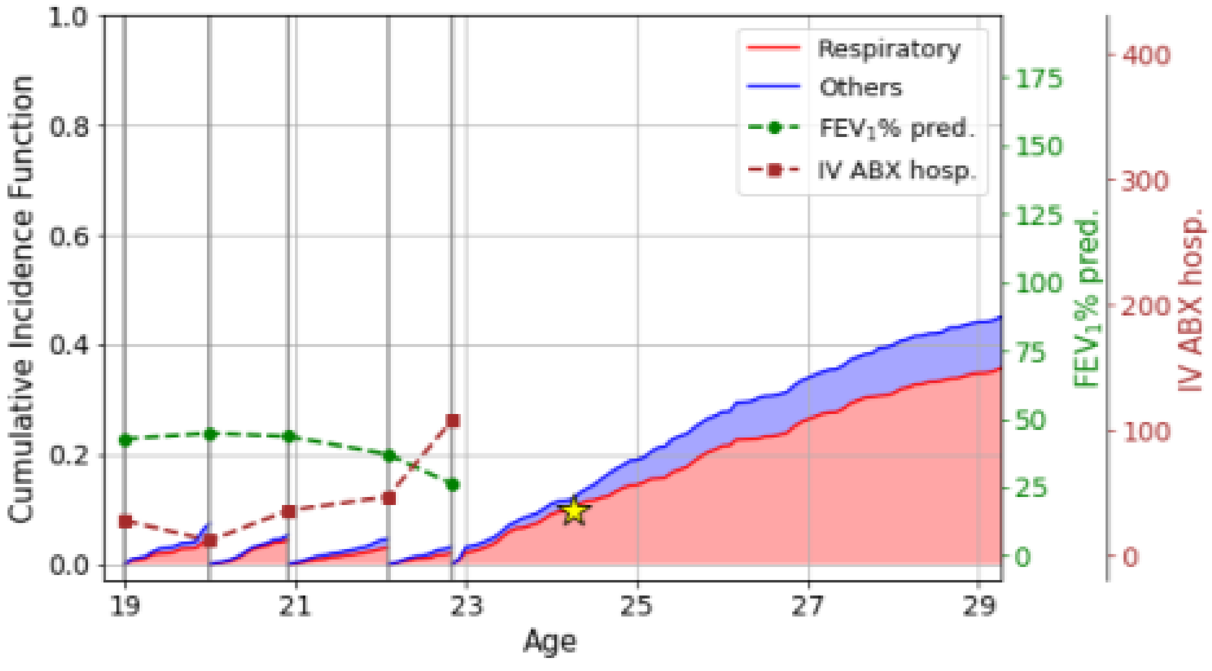
- **Prediction loss ($\mathcal{L}_3$):** penalizes error on the step-ahead predictions

$$\mathcal{L}_3 = \beta \cdot \sum_{i=1}^{N} \sum_{m=0}^{M_i - 1} \zeta(\mathbf{x}_{m+1}^i, \mathbf{y}_m^i),$$

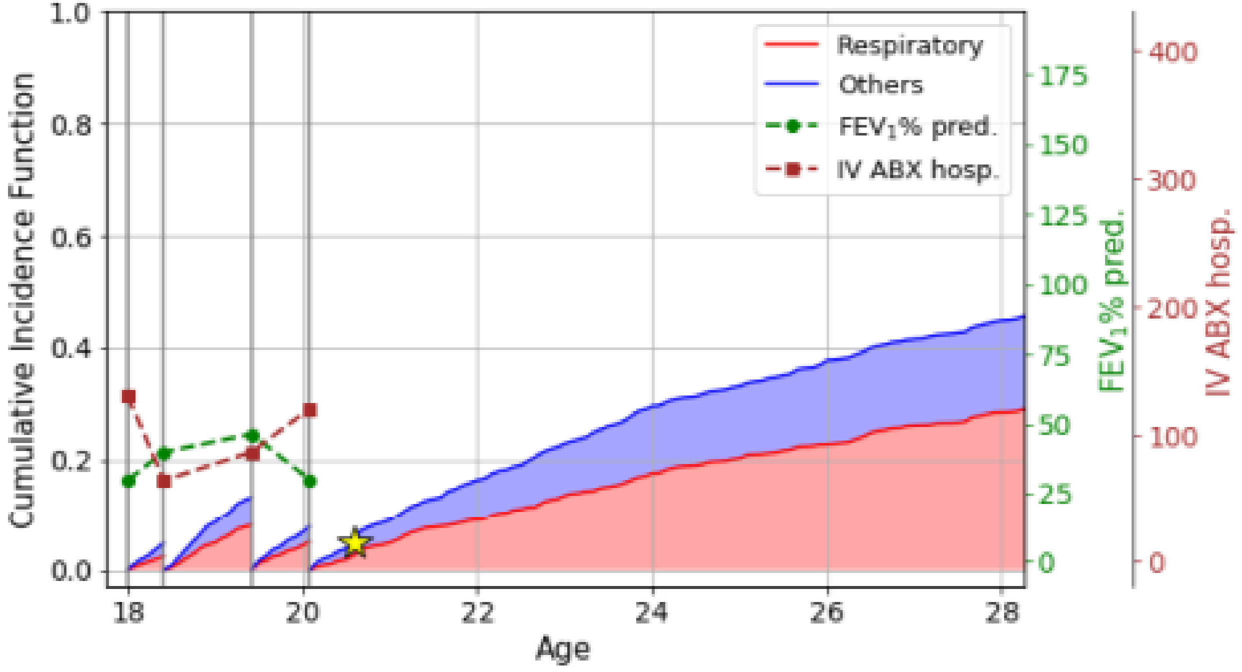
where $\zeta_d(a_d, b_d) = |a_d - b_d|^2$ or $\zeta_d(a_d, b_d) = a_d \log b_d + (1 - a_d) \log(1 - b_d)$

Dynamic-DeepHit [Lee & vdS, TBME 2019]

Dynamic-DeepHit updates the survival predictions as new observations are collected over time.

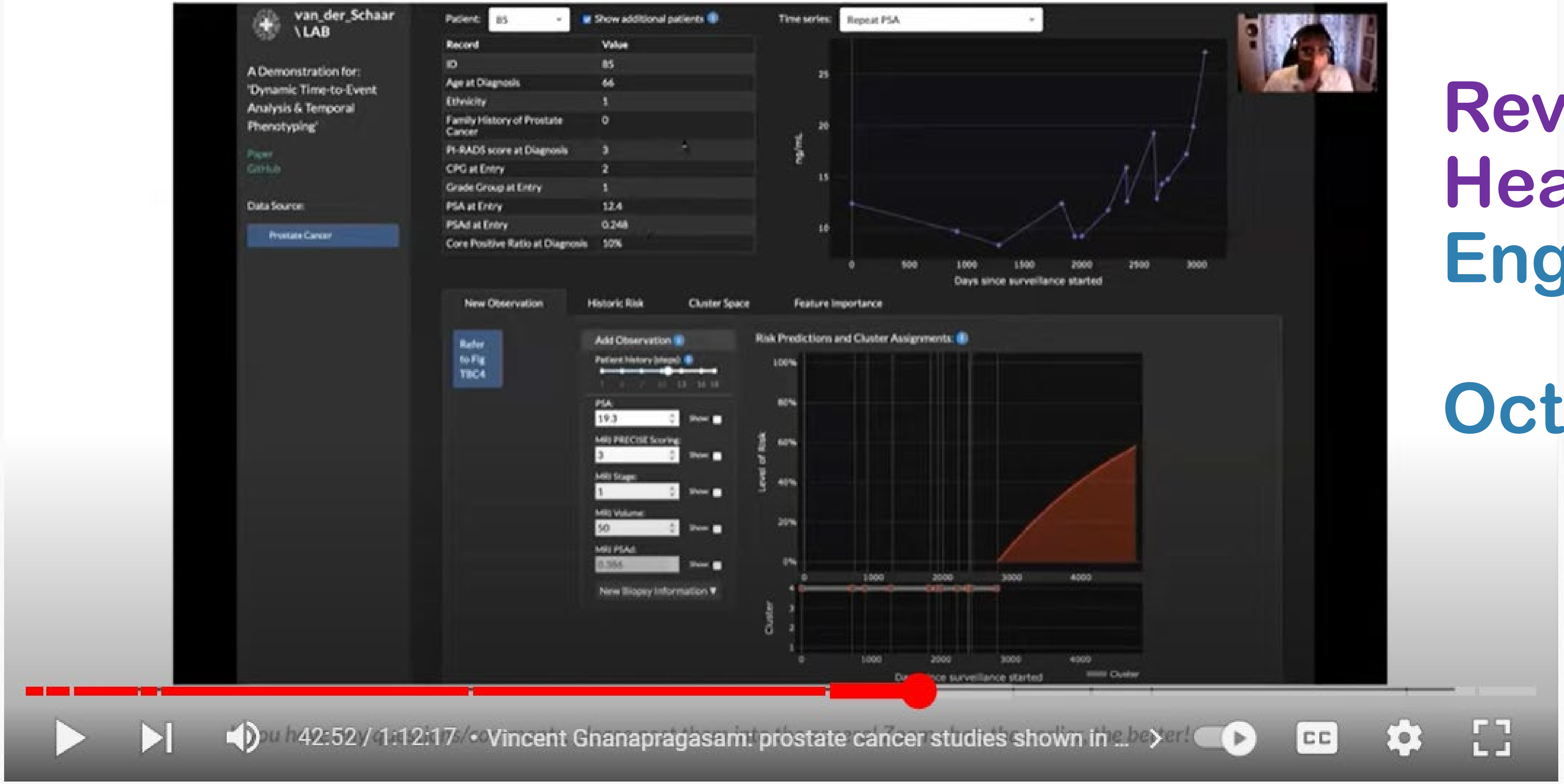


(a) A patient died of respiratory failure ($k = 1$)



(b) A patient died of other causes ($k = 2$)

Vincent Gnanapragasam



Revolutionizing Healthcare - getting ML-powered tools in the hands of clinicians (part 2)

Revolutionizing Healthcare Engagement

October 2021

Time-series: a multi-faceted problem

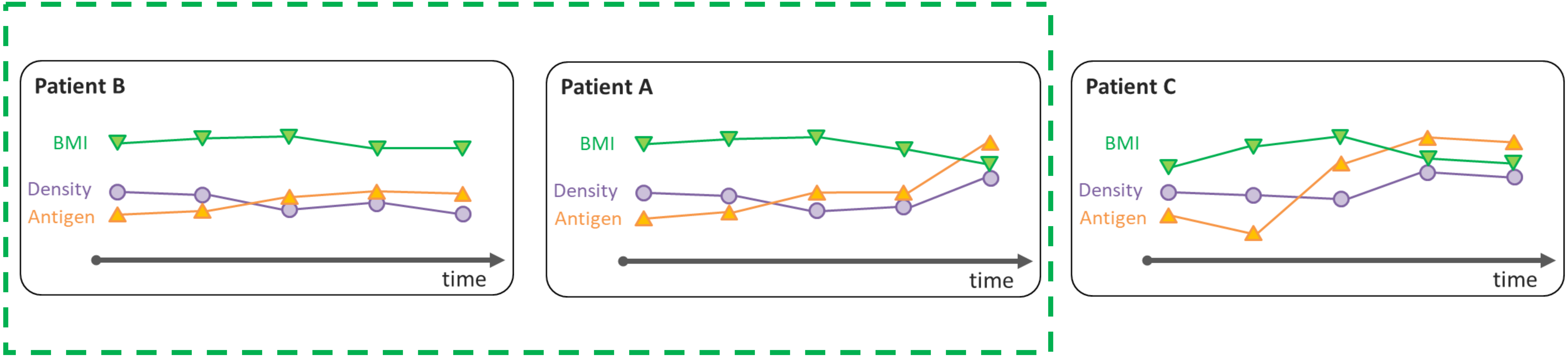
1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

Motivation: How should we group patients?

Example of 3 patients diagnosed with breast cancer (BC)

Should we group patients based on similarity in the time-series observations?



conventional notion of clustering

Key idea: similarity in time-series observations

(e.g. dynamic time warping, auto-encoders)

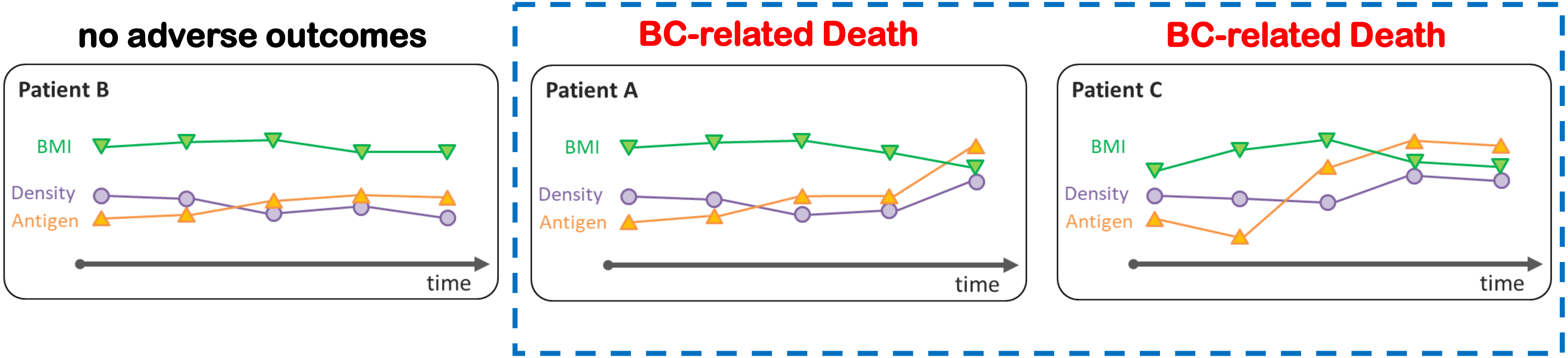
Autoencoder-based approaches

- N. S. Madiraju et al., 2018
- Q. Ma et al., 2019

Motivation: How should we group patients?

Example of 3 patients diagnosed with breast cancer (BC)

What if both Patient A and C will have an adverse event (e.g., death) that can be expected by increases in cancer antigen and mammographic density



New notion of clustering

Key idea: similarity in future outcomes

Temporal Phenotyping using Deep Predicting Clustering of Disease Progression [Lee, vdS, ICML 2020]

New notion of phenotype (clustering):
- Predictive of similar future outcomes
- Doctors and patients can actively plan

Learn discrete representations of past observations (time-series data) that best describe future events and outcomes of interest

Problem Formalism

Notation

- $\mathbf{x}_{1:t} = (\mathbf{x}_1, \cdots, \mathbf{x}_t)$ and y_t: input (sub)sequence and output label at time t
- s_t: cluster assignment at time t and $\mathcal{E} = \big(\mathbf{e}(1), \cdots, \mathbf{e}(K)\big)$: cluster centroids
- $\mathcal{C} = \{\mathcal{C}(1), \cdots, \mathcal{C}(K)\}$: a set of K predictive clusters where $\mathcal{C}(k) = \{\mathbf{x}^n_{1:t} | s^n_t = k\}$

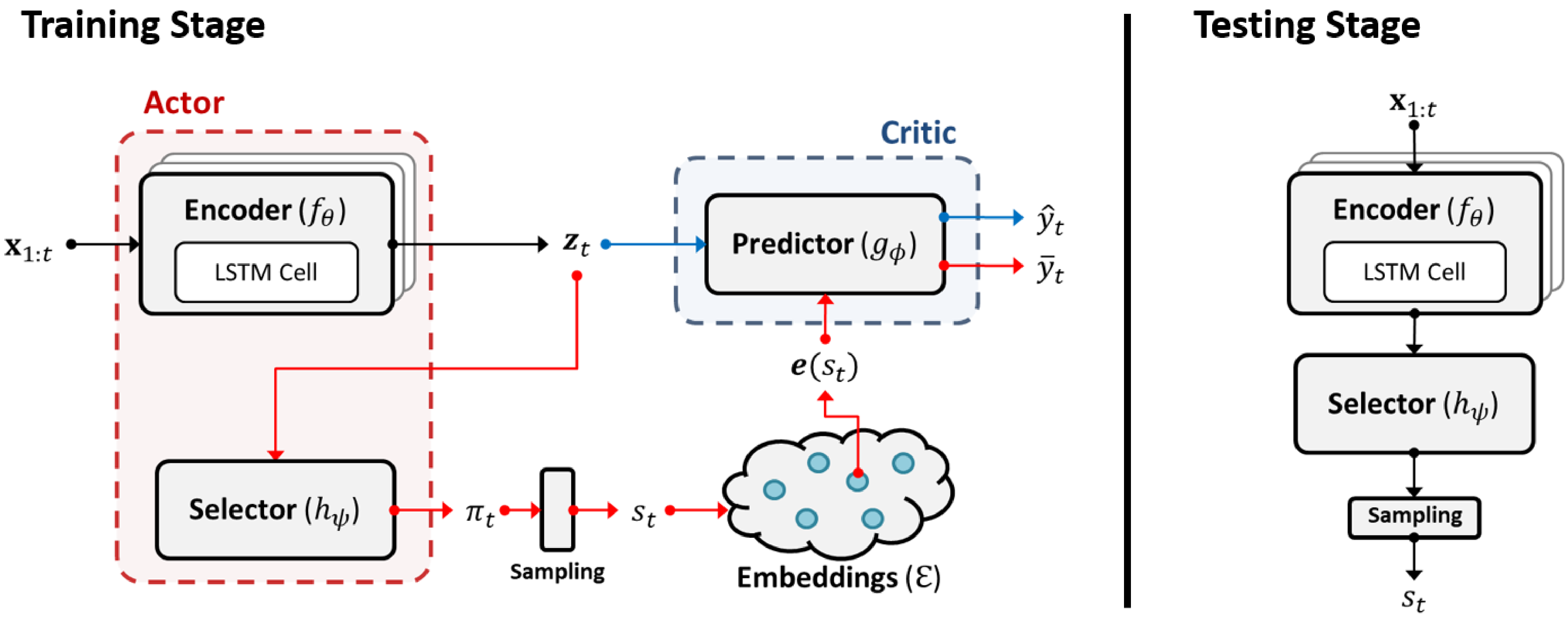
We establish identifying a set of predictive clusters, $\mathcal{C}$, as

$$\underset{\mathcal{C}}{\text{minimize}} \sum_{k \in \mathcal{K}} \sum_{\mathbf{x}_{1:t} \in \mathcal{C}(k)} KL\Big(\underbrace{Y_t | \mathbf{X}_{1:t} = \mathbf{x}_{1:t}}_{\text{label distribution given a sequence (continuous rep.)}} \Big\| \underbrace{Y_t | S_t = k}_{\text{label distribution given a cluster assignment (discrete rep.)}}\Big) \qquad (1)$$

Challenges

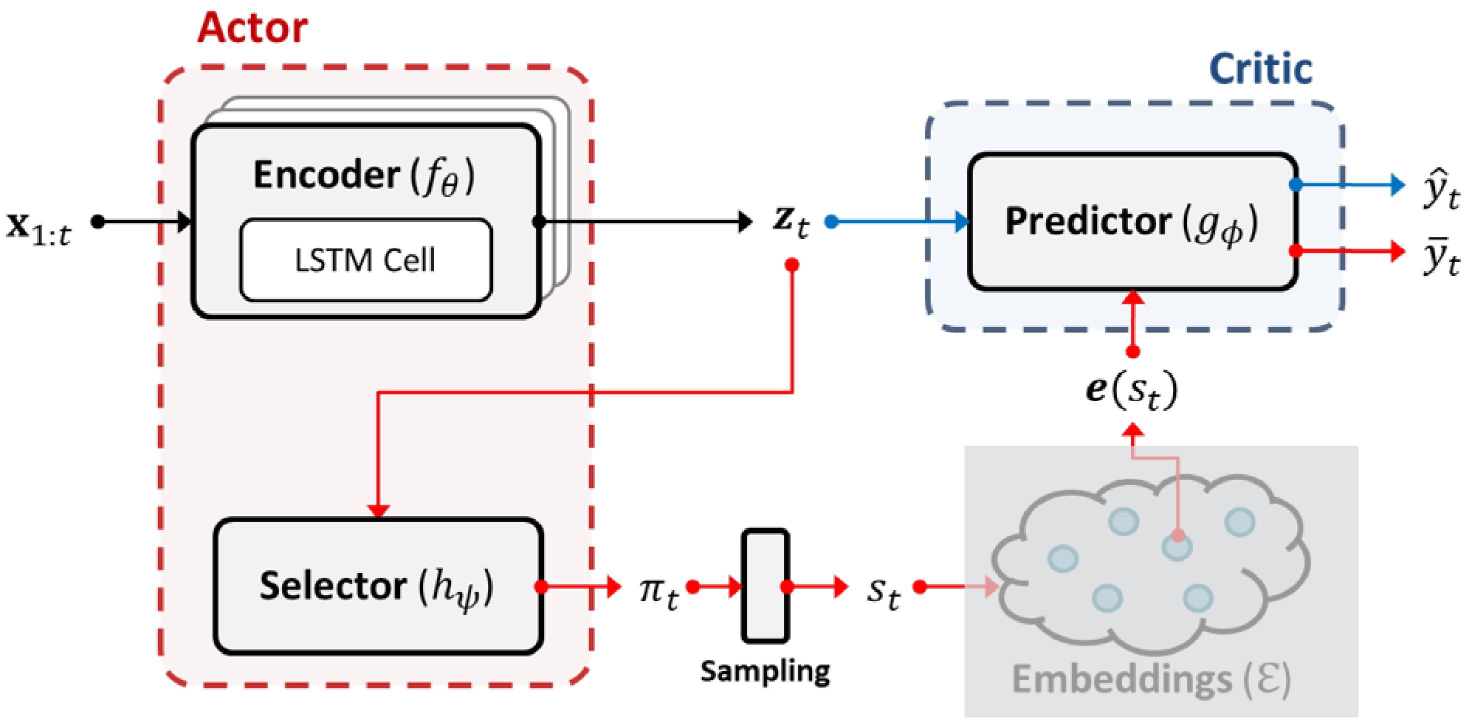
- NP-hard combinatorial problem → **iteratively solving two subproblems**
- Assigning clusters involves sampling process → **actor-critic training** [Konda & Tsitsiklis, 2000]

AC-TPC [Lee & vdS, ICML 2020]



AC-TPC [Lee & vdS, ICML 2020]

Subproblem 1 - Optimize network parameters (θ,ϕ,ψ)



Given $\mathcal{E} = (\mathbf{e}(1), \cdots, \mathbf{e}(K))$ fixed, update (θ, ϕ, ψ) based on:

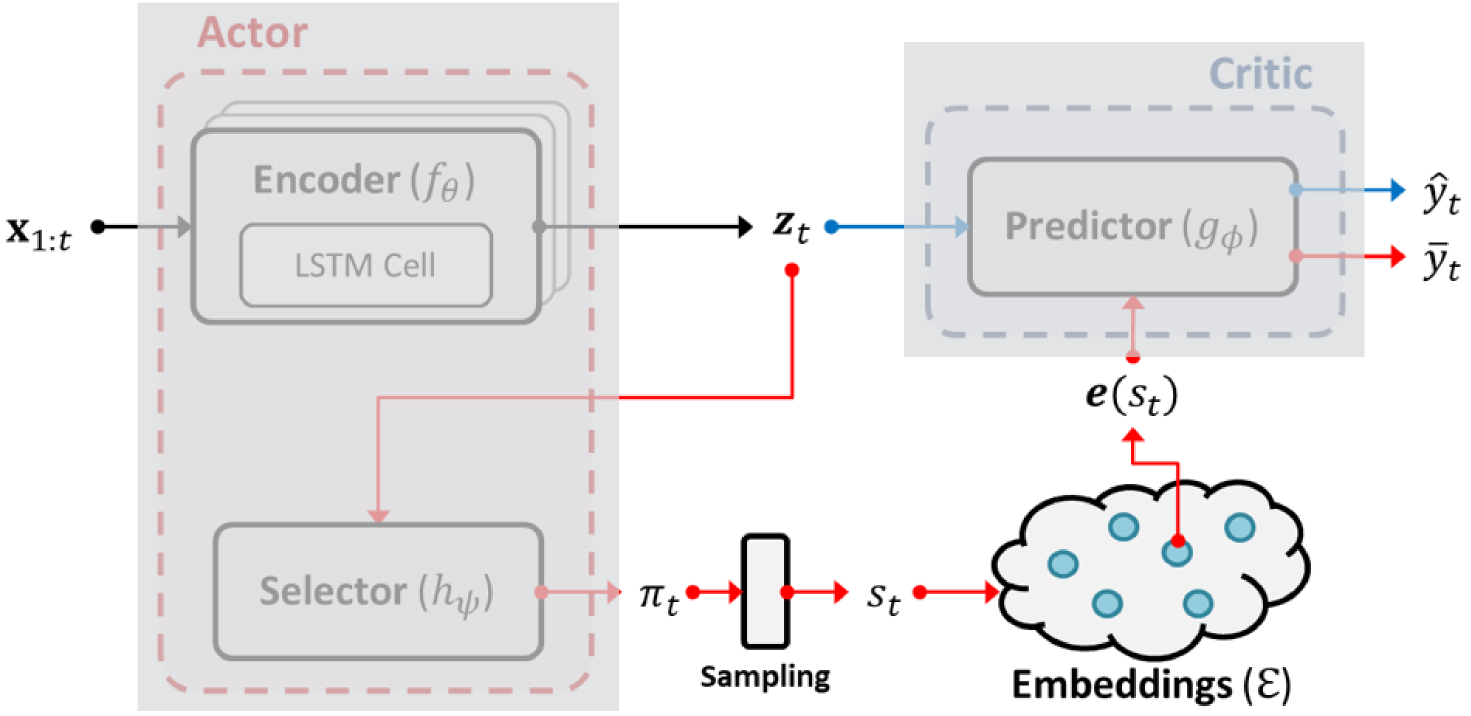
$$\mathcal{J}_1(\theta, \psi, \phi) =$$

$$\underbrace{\mathbb{E}_{\mathbf{x},y\sim p_{XY}}\left[\sum_{t\in\mathcal{T}} \mathbb{E}_{s_t\sim Cat(\pi_t)}\left[\ell_1(y_t, \bar{y}_t)\right]\right]}_{\text{predictive clustering loss } \mathcal{L}_1(\theta,\psi,\phi)}$$

$$+ \alpha \underbrace{\mathbb{E}_{\mathbf{x}\sim p_X}\left[-\sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}} \pi_t(k)\log \pi_t(k)\right]}_{\text{sample-wise entropy of cluster assignment } \mathcal{L}_2(\theta,\psi)}$$

AC-TPC [Lee & vdS, ICML 2020]

Subproblem 2 - Optimize embeddings ($\mathcal{E} = (\mathbf{e}(1), \cdots, \mathbf{e}(K))$)



Given (θ, ϕ, ψ) fixed, updated $\mathcal{E} = (\mathbf{e}(1), \cdots, \mathbf{e}(K))$ based on:

$$\mathcal{J}_2(\mathcal{E}) = \underbrace{\mathbb{E}_{\mathbf{x},y\sim p_{XY}}\left[\sum_{t\in\mathcal{T}} \mathbb{E}_{s_t\sim Cat(\pi_t)}\left[\ell_1(y_t, \bar{y}_t)\right]\right]}_{\text{predictive clustering loss } \mathcal{L}_1(\mathcal{E})}$$

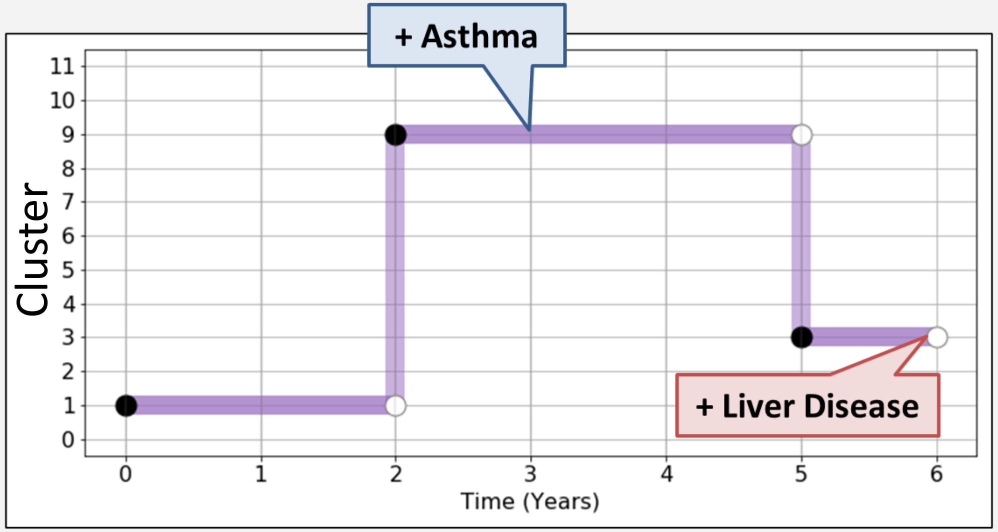
$$+ \beta \underbrace{\sum_{k\neq k'} \ell_1(g_\phi(\mathbf{e}(k)), g_\phi(\mathbf{e}(k')))}_{\text{embedding separation loss } \mathcal{L}_3(\mathcal{E})}$$

Example Trajectories [Lee & vdS, ICML 2020]

Patient A

Cluster 1 → Cluster 9 → Cluster 3

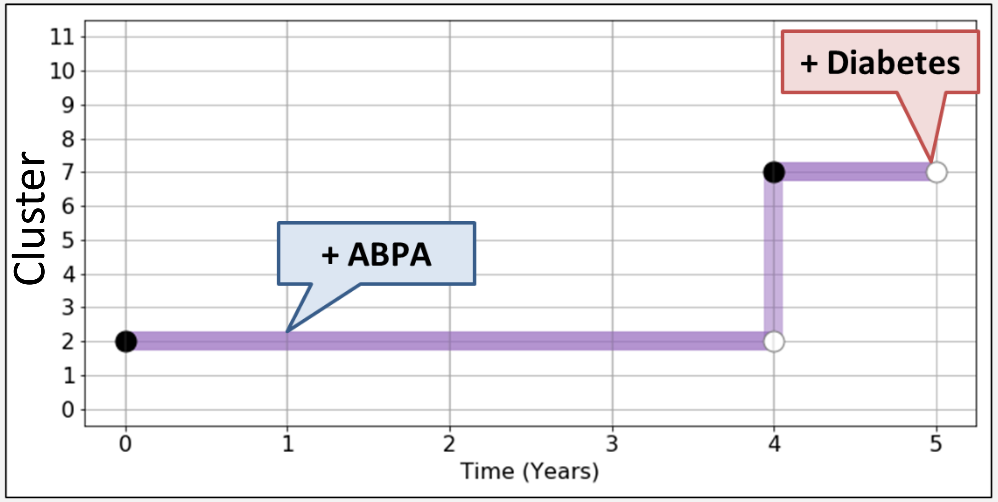
Cluster 1	Cluster 3	Cluster 9
Asthma (0.03) Liver Disease (0.04)	Asthma (0.89) Liver Disease (0.87)	Asthma (0.89) Liver Disease (0.02)



Patient B

Cluster 2 → Cluster 7

Cluster 2	Cluster 7
ABPA (0.77) Diabetes (0.06)	ABPA (0.83) Diabetes (0.78)

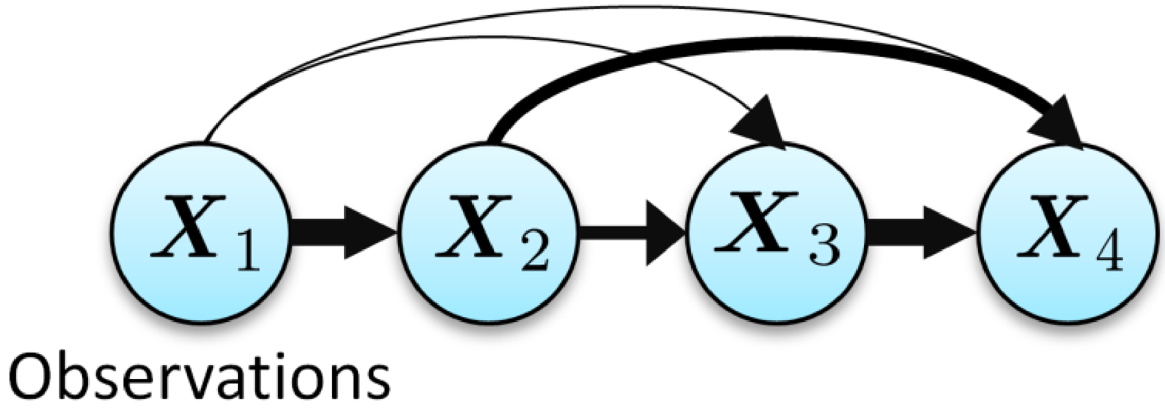


Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

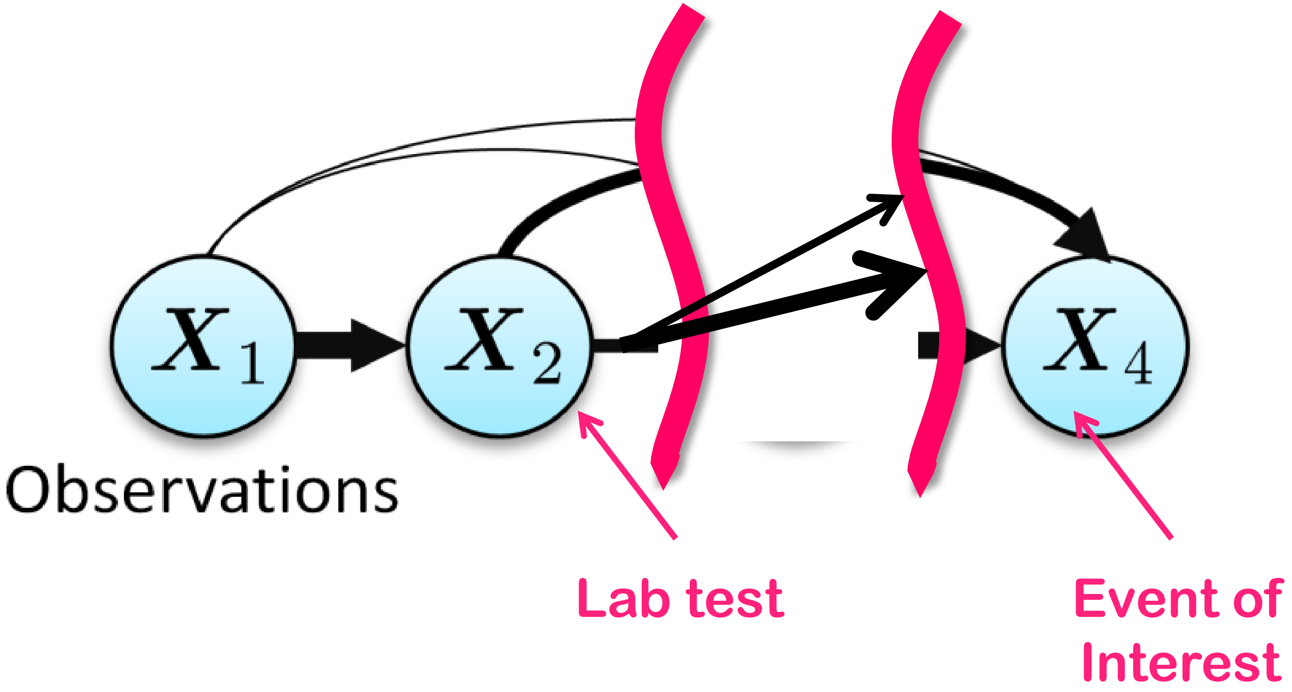
Personalized screening/monitoring



Who to Screen?
When to Screen?
What to Screen?

- **What is the *value of various information* over time for *this event* for *this individual*?**

How to formalize the personalized monitoring problem?



Who to Screen?
When to Screen?
What to Screen?

- Deep Sensing [Yoon, Jordon, vdS, 2018]
- Disease Atlas [Yoon, Jordon, vdS, 2019]
- Clairvoyance [Jarrett et al, 2021]

Deep Sensing: Active Sensing using multi-directional recurrent neural networks [Yoon, Zame, vdS, ICLR 2018]

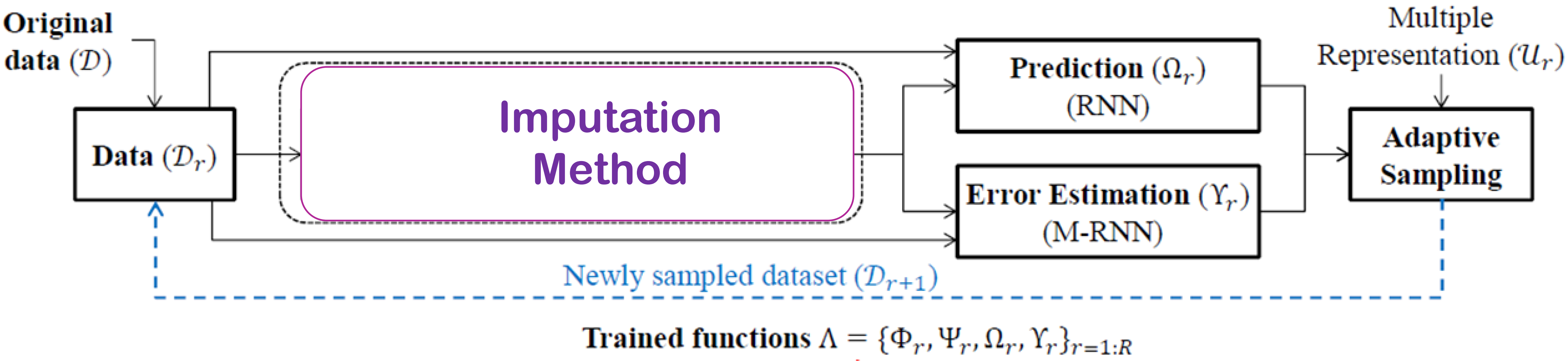
- **Motivation:**
  - Monitoring and screening (sensing) is costly
  - Trade-off between value of information and cost of sensing
  - Sensing should be an active choice

- **Challenges:**
  - Value of information is unknown & dynamically changing – needs to be learned!

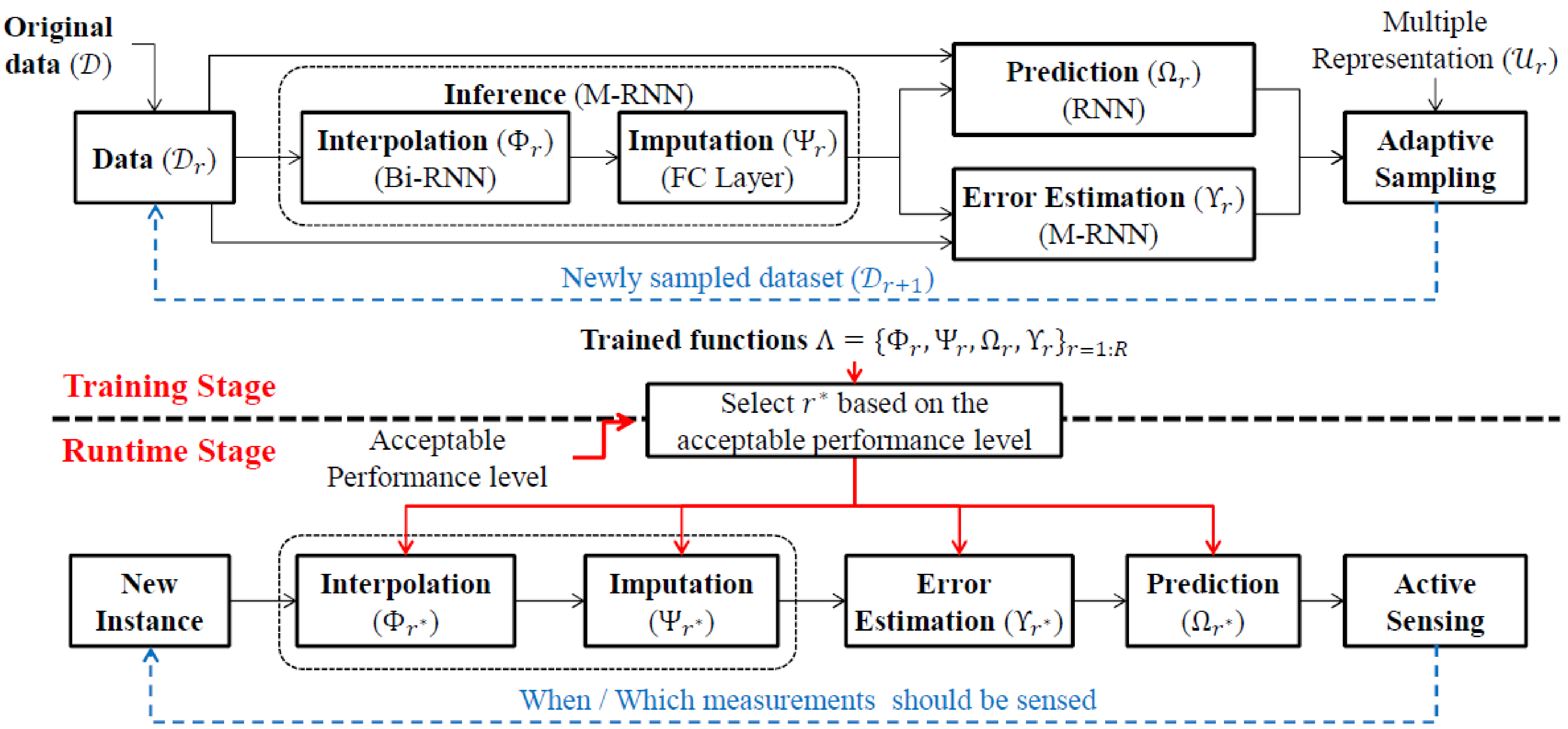
How to do this???

Deep sensing architecture

Training time



Deep sensing architecture

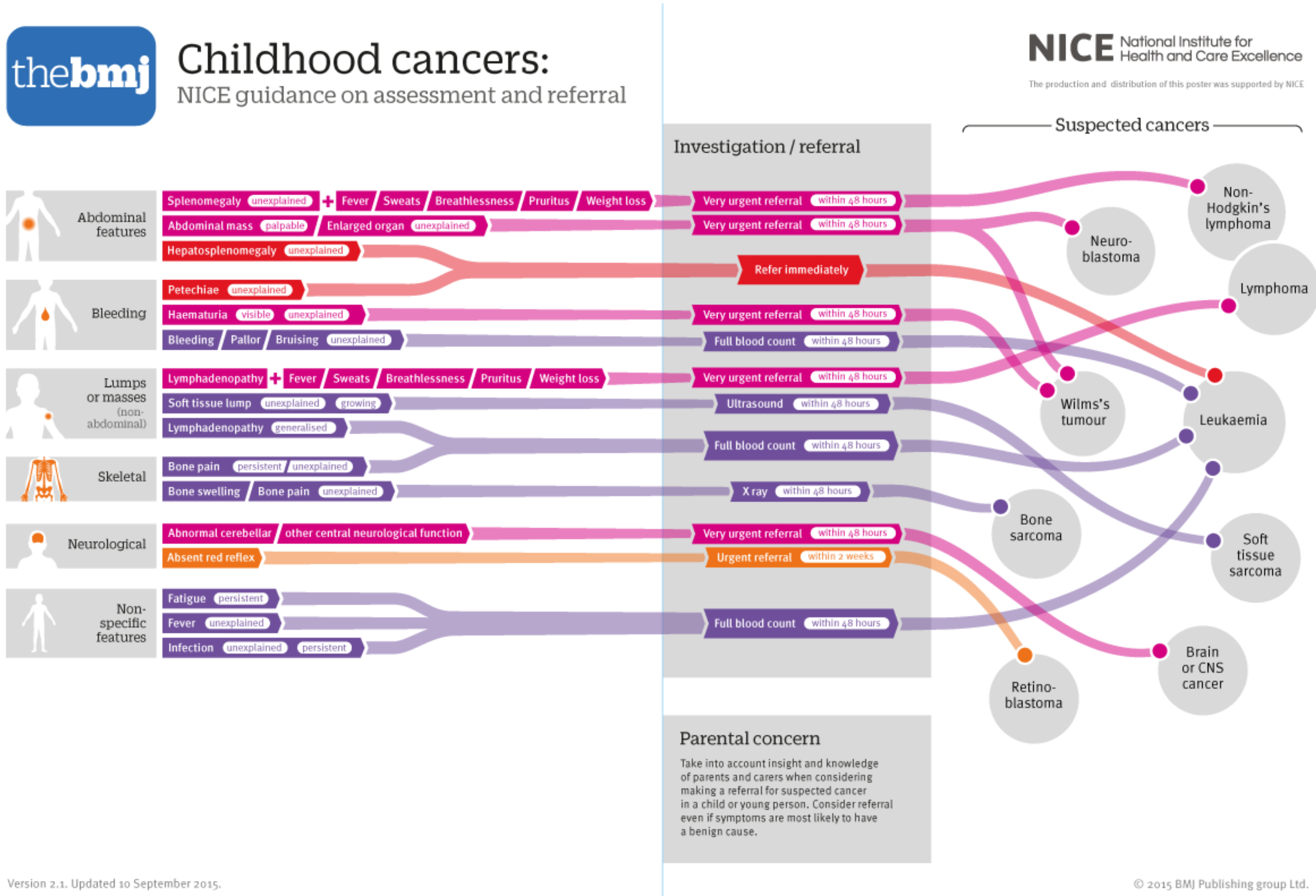


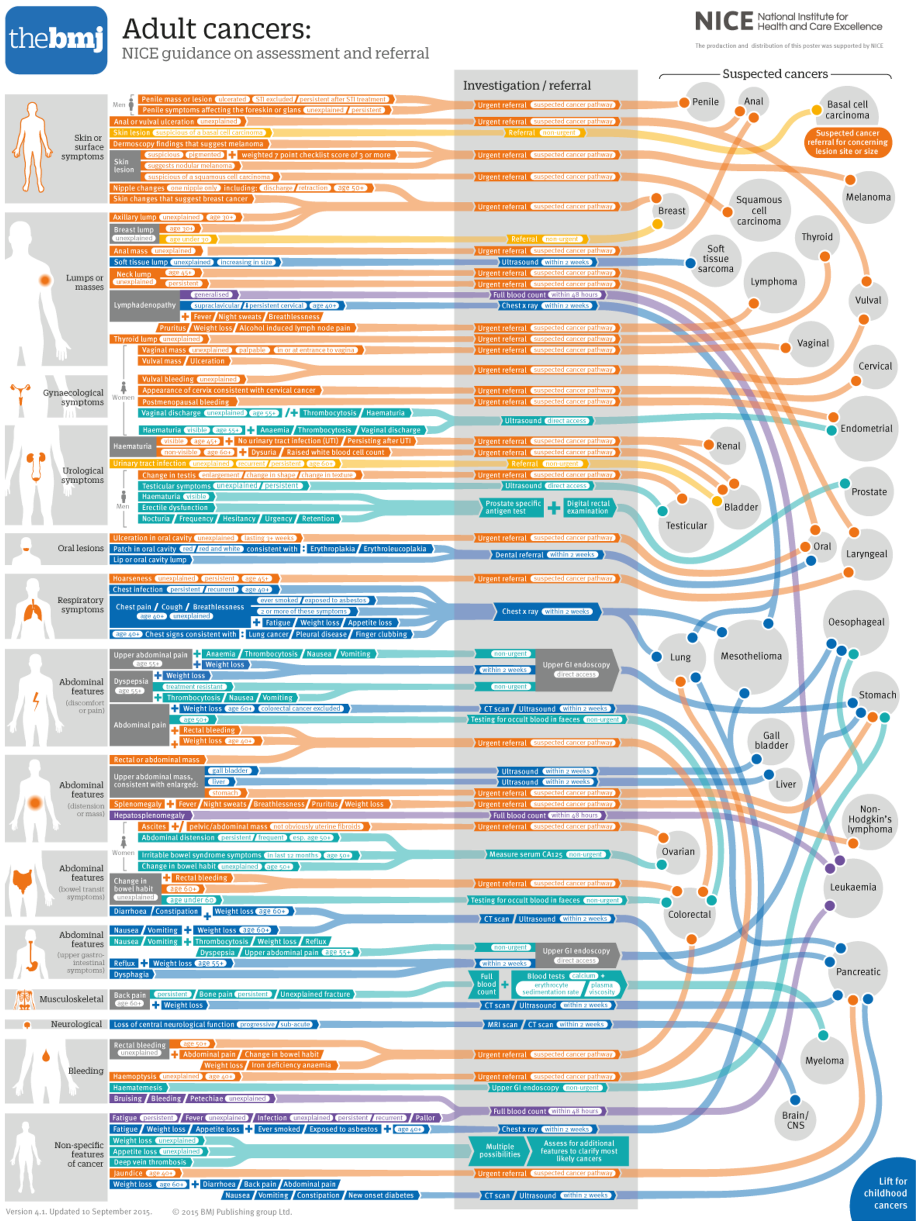
Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

ED&D – A complex problem





van_der_Schaar
\ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Current thinking in ED&D

Risk prediction

Segments individuals using population-based risks,

- usually based on few variables
- rarely uses longitudinal data
- usually only calculated once

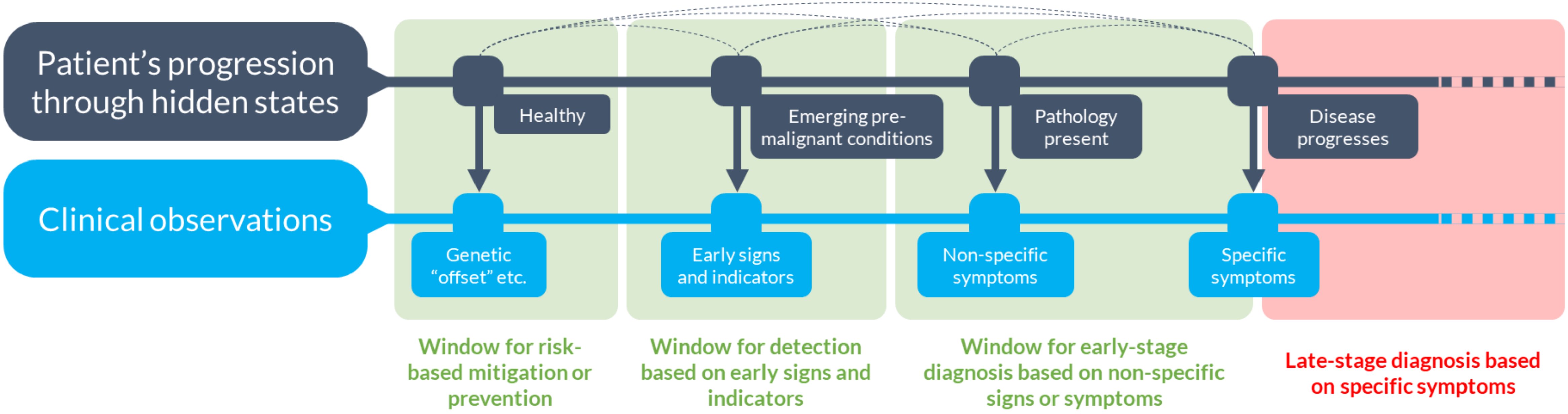
Risk scores then lead to guideline-driven management of patients

- often rigid
- many diseases lack guidelines and protocols

This is all predicated upon a quantitative understanding of disease progression

How can we detect disease early?

Early diagnosis is more than just event prediction/forecasting
- It involves unravelling and dissecting the underlying states of disease progression towards the event of interest

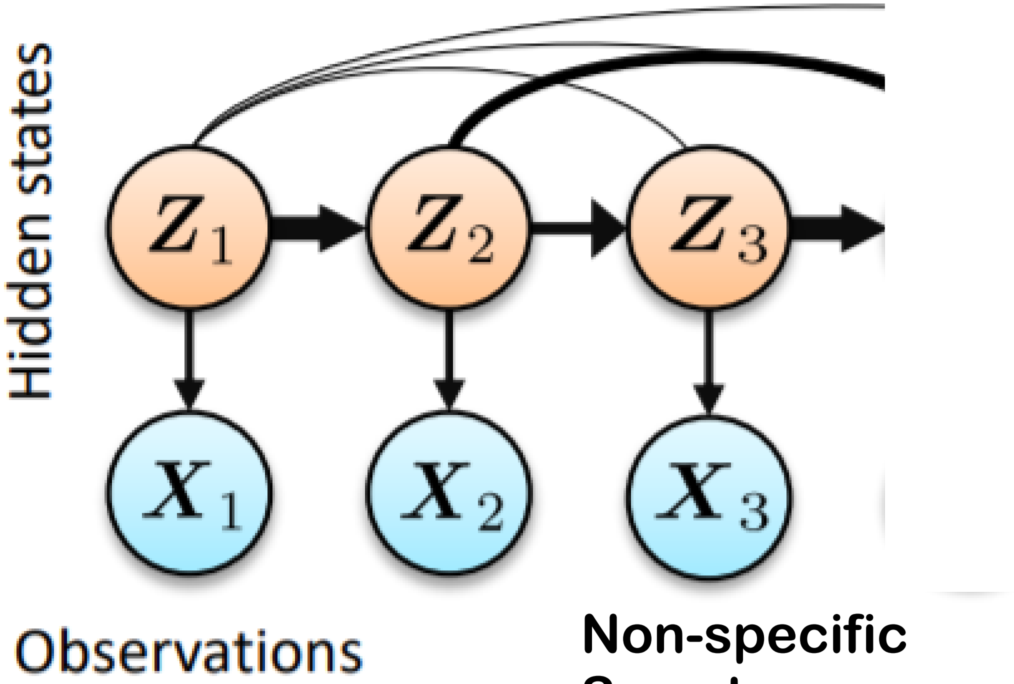


A quantitative understanding of disease progression is needed!

Early diagnosis: How?

Healthy

Emerging Pre-malignant Conditions (Dysplasia)



Non-specific Symptoms

Genetic "Offset"

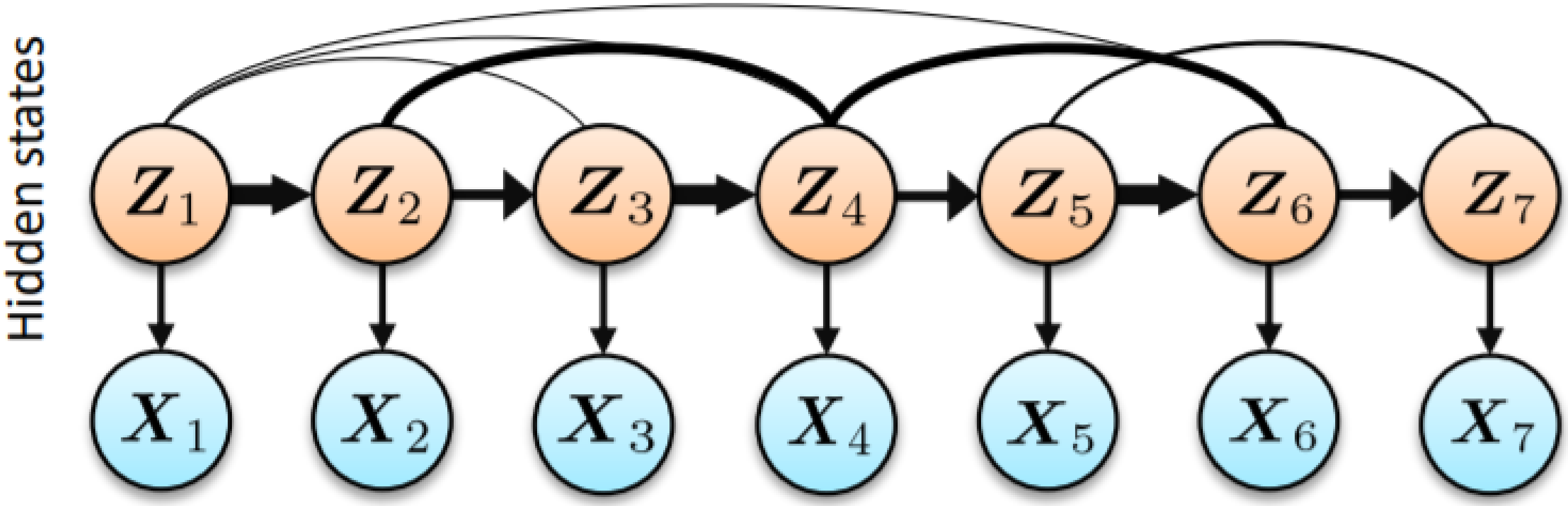
Early diagnosis: How?

Healthy

Emerging Pre-malignant Conditions (Dysplasia)

Pathology Present

Disease Progresses



Observations

Genetic "Offset"

Non-specific Symptoms

Early diagnosis Made (with ML support)

Diagnosis Made

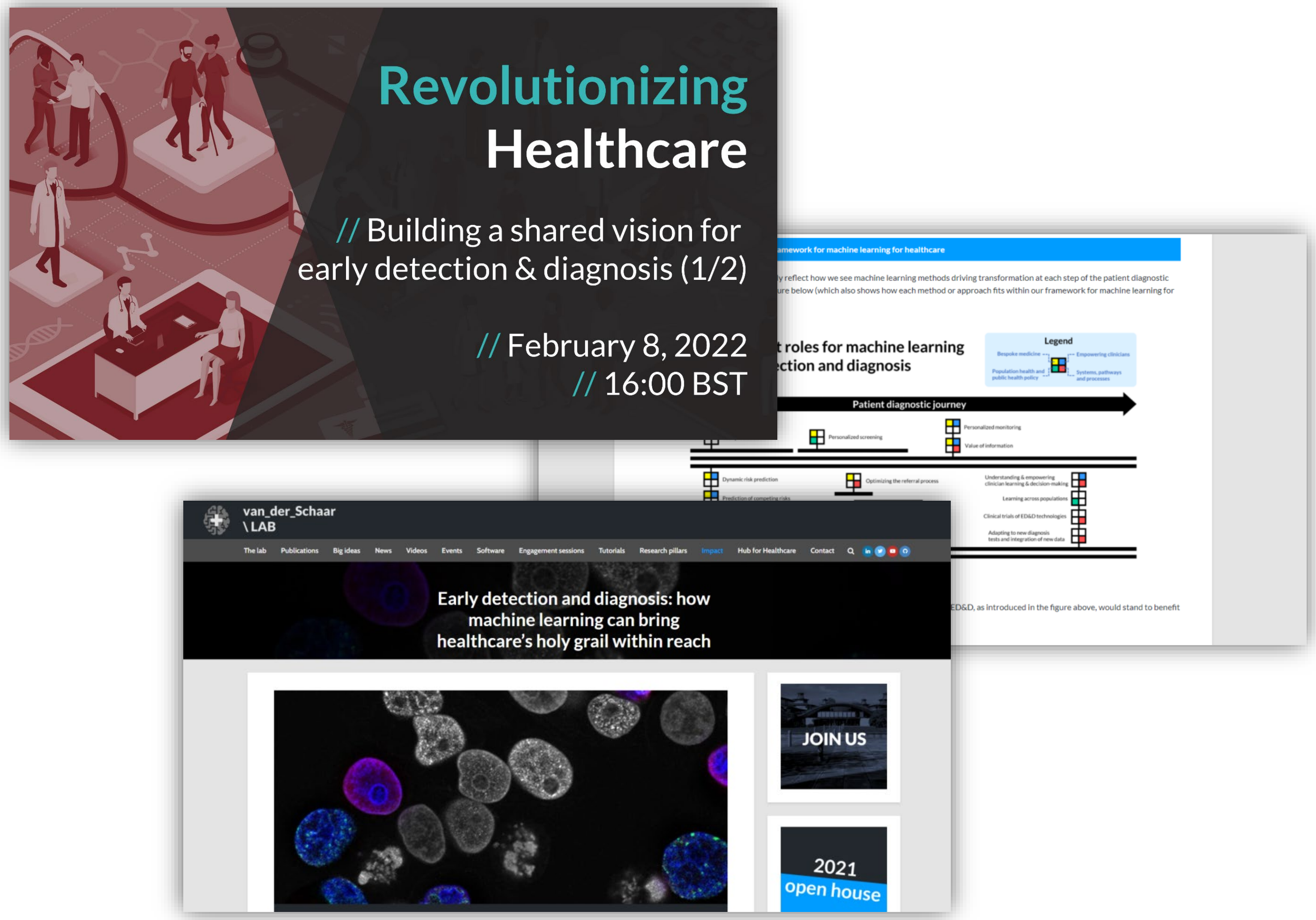
Revolutionizing Healthcare: roundtable on ED&D

Double-header (February 8 and March 10) on ED&D – one of healthcare's holy grails!

https://www.vanderschaar-lab.com/
→ Engagement sessions
→ Revolutionizing Healthcare

Visit our extensive new reference page on ML for ED&D!

https://www.vanderschaar-lab.com/
→ Impact
→ Early detection and diagnosis

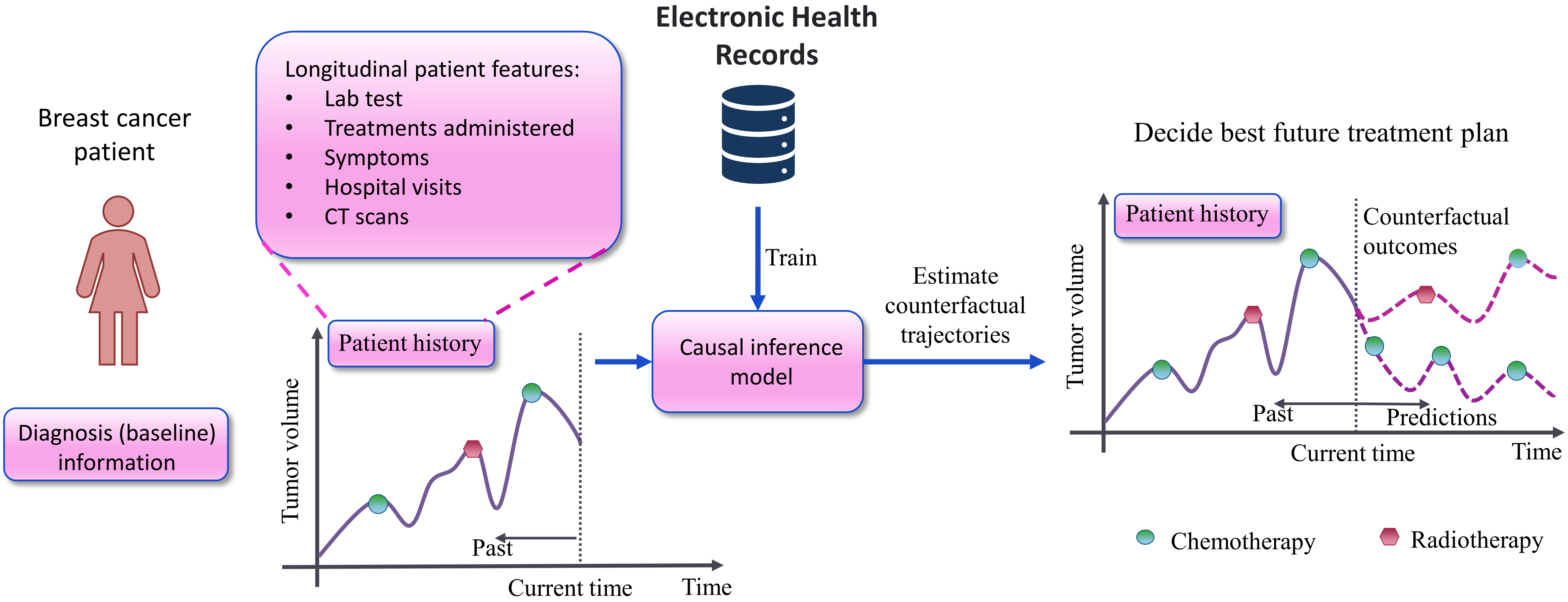


Time-series: a multi-faceted problem

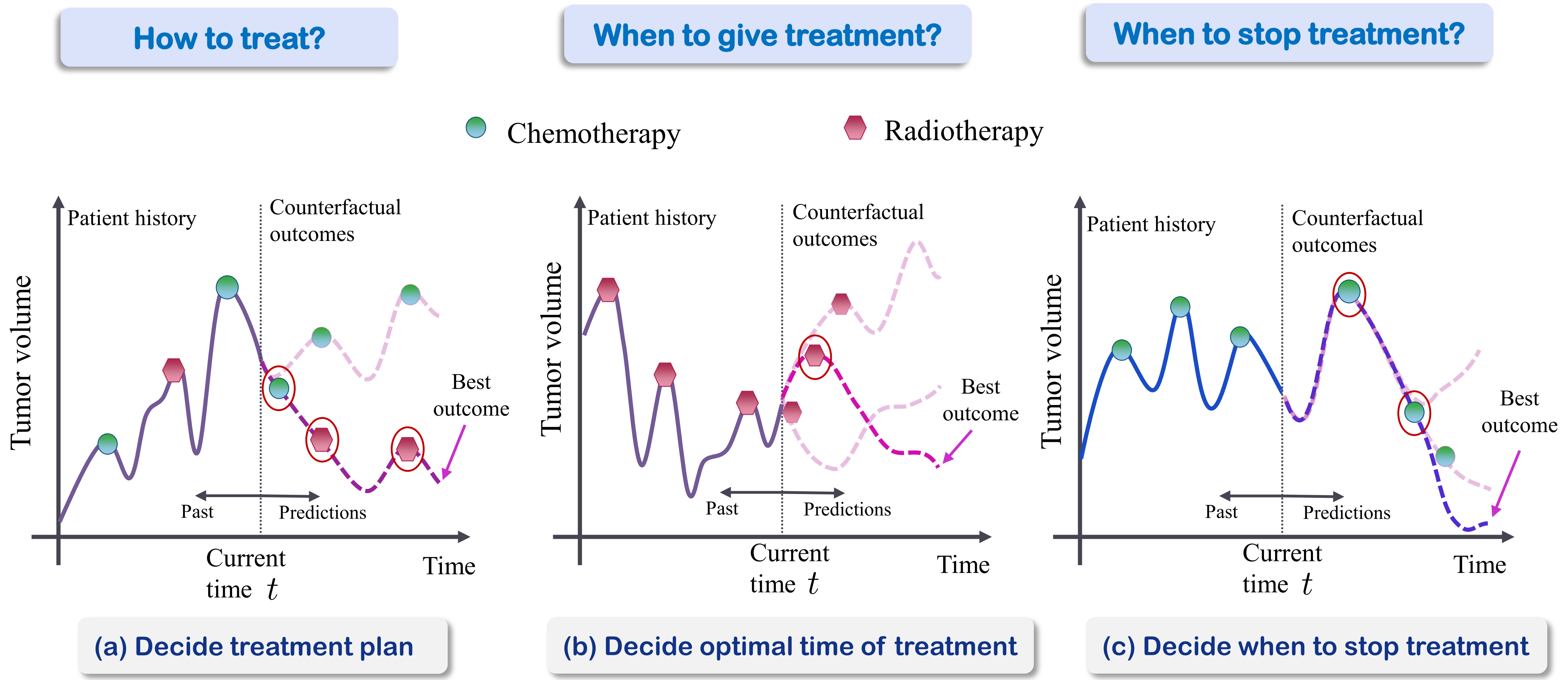
1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

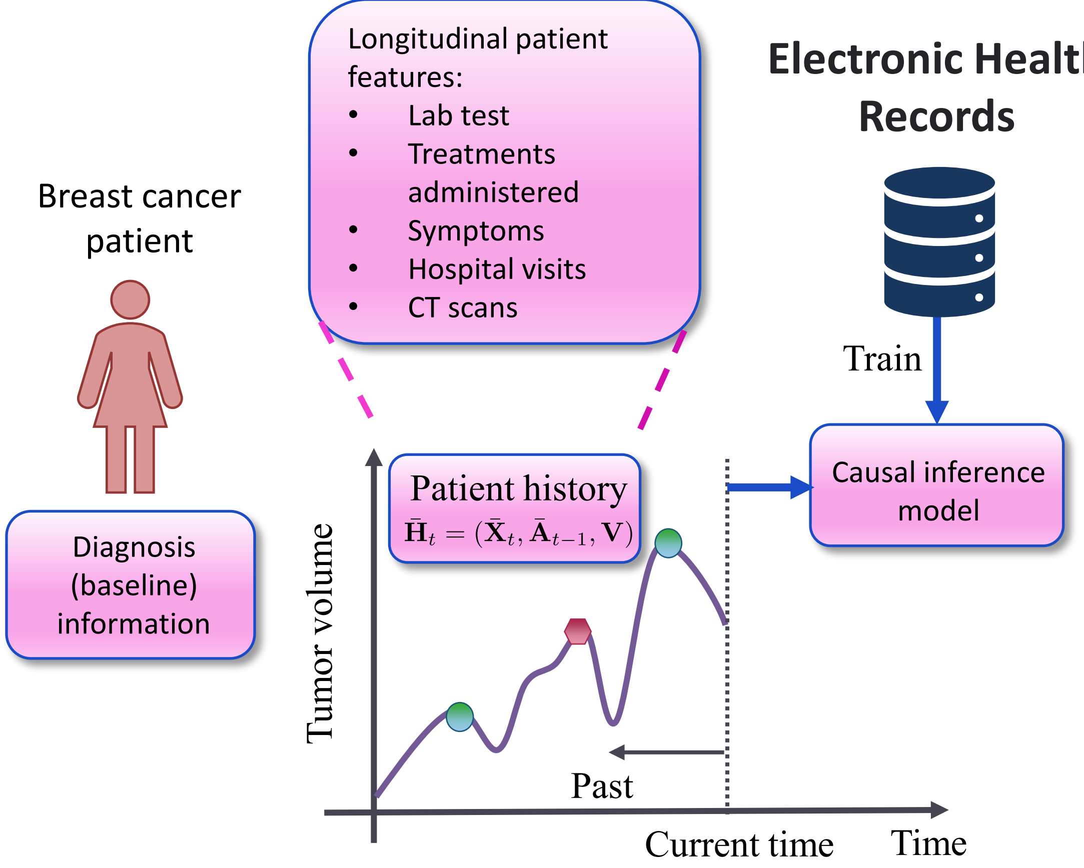
Individualized treatment effects over time



Individualized treatment effects over time



Causal effect inference based on longitudinal patient observational data



Longitudinal patient observational data

- Time-dependent patient features: $\bar{\mathbf{X}}_t = (\mathbf{X}_1, \ldots \mathbf{X}_t)$
- Time-dependent treatments: $\bar{\mathbf{A}}_t = (\mathbf{A}_1, \ldots \mathbf{A}_t)$ where

$$\mathbf{A}_t \in \{A_1, \ldots, A_K\}$$

- Static patient features: $\mathbf{V}$

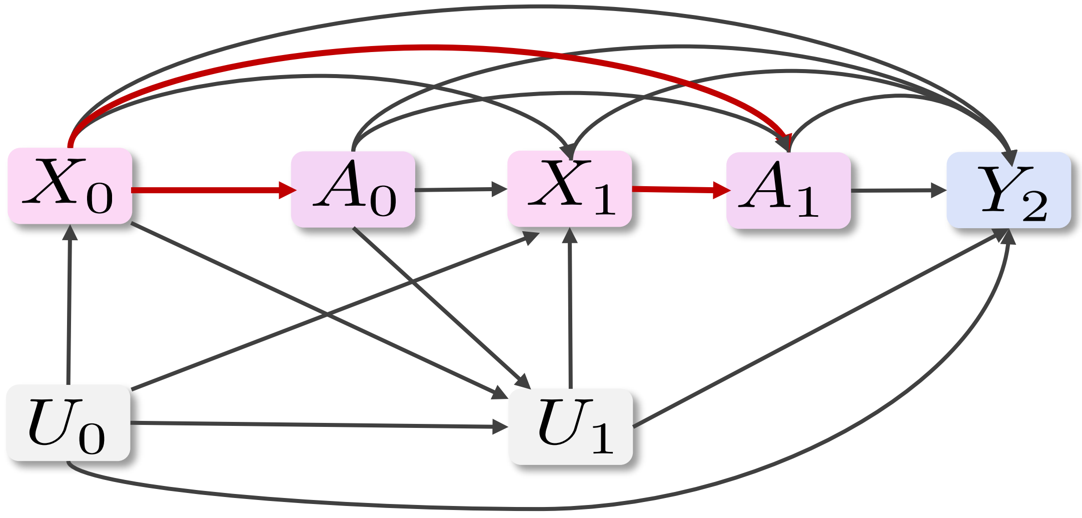
Patient history: $\bar{\mathbf{H}}_t = (\bar{\mathbf{X}}_t, \bar{\mathbf{A}}_{t-1}, \mathbf{V})$

Observed (factual) outcome for treatment $\mathbf{A}_t$ given patient history $\bar{\mathbf{H}}_t$: $\mathbf{Y}_{t+1}$

Challenges in using longitudinal observational data for estimating individualized outcomes

The patient history $\bar{\mathbf{H}}_t = (\bar{\mathbf{X}}_t, \bar{\mathbf{A}}_{t-1}, \mathbf{V})$ contains time-dependent confounders which bias the treatment assignment $\mathbf{A}_t$ in the observational dataset.

Patient covariates - affected by past treatments which then influence future treatments and outcomes

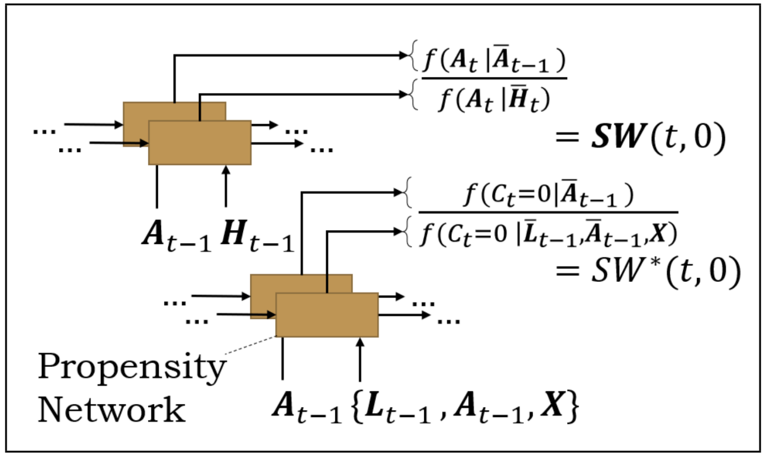


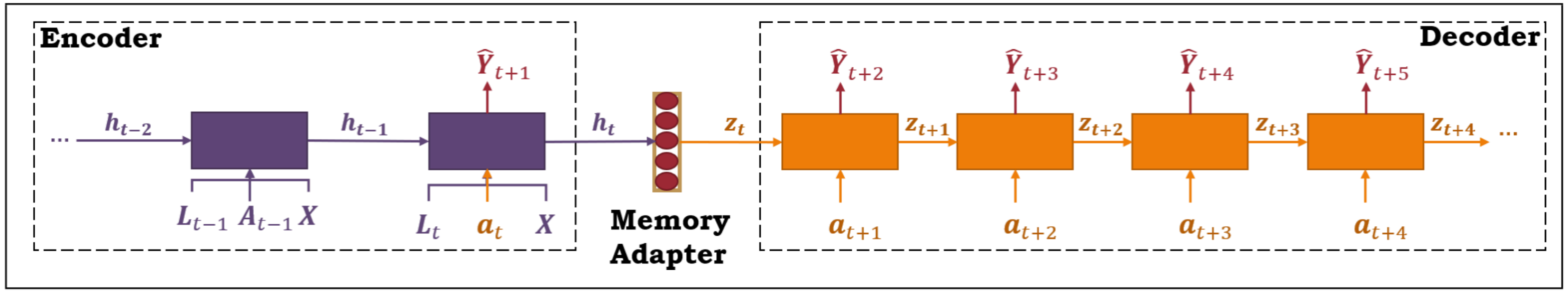
Bias from time-dependent confounders.

Handling time-dependent confounding bias

Inverse probability of treatment weighting

- → Marginal structural models [Robins, Hernan, Brumback, Epidemiology 2000]
- → Recurrent marginal structural networks [Lim, Alaa, van der Schaar, NeurIPS 2018]





$$e(i,t,\tau) = \mathbf{S\tilde{W}}_i(t,\tau-1) \times S\tilde{W}_i^*(t,\tau-1) \times \|\mathbf{Y}_{t+\tau,i} - g(\tau, a(t,\tau-1), \bar{\mathbf{H}}_t)\|^2$$

$$\mathbf{SW}(t,\tau) = \prod_{n=t}^{t+\tau} \frac{f(\mathbf{A}_n|\bar{\mathbf{A}}_{n-1})}{f(\mathbf{A}_n|\bar{\mathbf{H}}_n)} = \prod_{n=t}^{t+\tau} \frac{\prod_{k=1}^{\Omega_a} f(A_n(k)|\bar{\mathbf{A}}_{n-1})}{\prod_{k=1}^{\Omega_a} f(A_n(k)|\bar{\mathbf{H}}_n)}$$

Handling time-dependent confounding bias

Inverse probability of treatment weighting

→ Marginal structural models [Robins, Hernan, Brumback, Epidemiology 2000]

→ Recurrent marginal structural networks [Lim, Alaa, van der Schaar, NeurIPS 2018]

Numerically unstable

High variance

Representation Learning

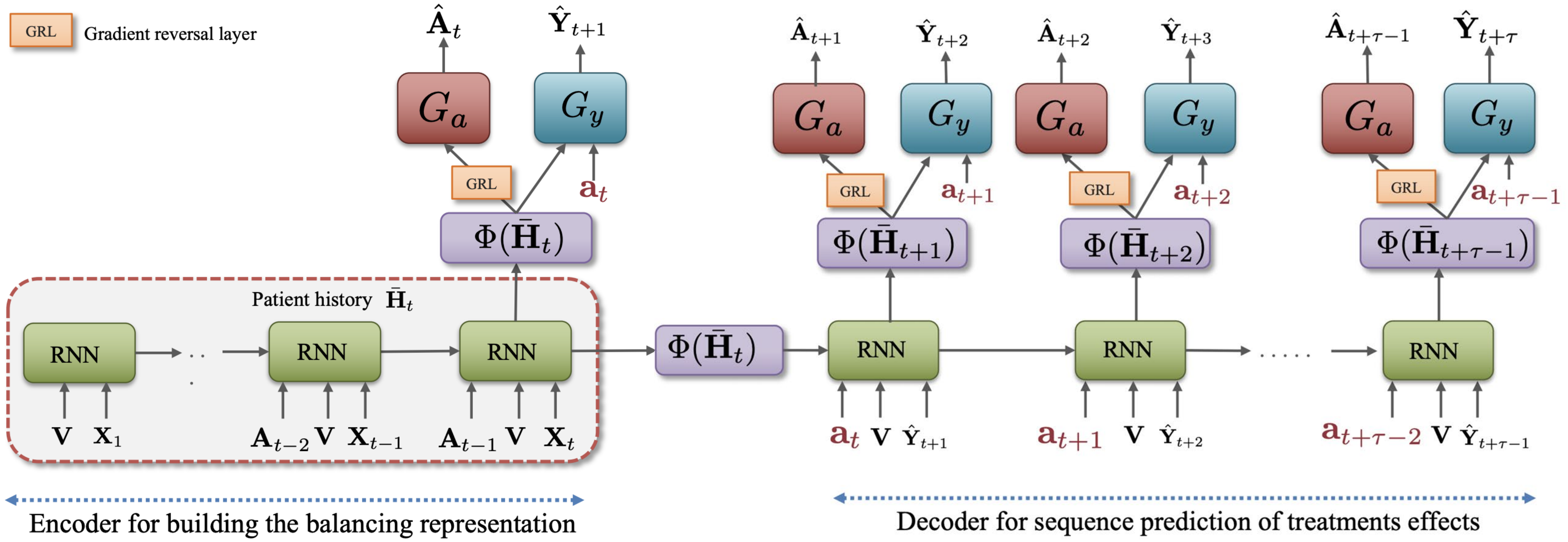
→ Counterfactual recurrent network [Bica, Alaa, Jordon, van der Schaar, ICLR 2020]

$$P(\Phi(\bar{\mathbf{H}}_t) \mid \mathbf{A}_t = A_1) = \cdots = P(\Phi(\bar{\mathbf{H}}_t) \mid \mathbf{A}_t = A_K)$$

Balanced representations/ Treatment invariant representations

Counterfactual Recurrent Network [Bica, Alaa, Jordon & van der Schaar, ICLR 2020]

- Builds **treatment invariant representations** using **domain adversarial training [Ganin et al., 2016]**.
- Estimates **counterfactual trajectories** using **sequence-to-sequence architecture**.



Part 2:
making time series models as useful as possible

Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

Which time-series method to select?

What is the challenge?

- RNN cells (e.g. LSTM, GRU)
- Architectures (e.g. Bidirectional, Encoder-decoder)
- Attention or not?
  Long or short memory?

Temporal distribution shifts, risk factors are changing! → Best model for each time step is different! Can't manually select the best model for each time step → Stepwise Model Selection

Stepwise Model Selection for Sequence Prediction via Deep Kernel Learning [Zhang, Jarrett, vdS, AISTATS 2020]

Solution: novel BO algorithm to tackle model selection challenge

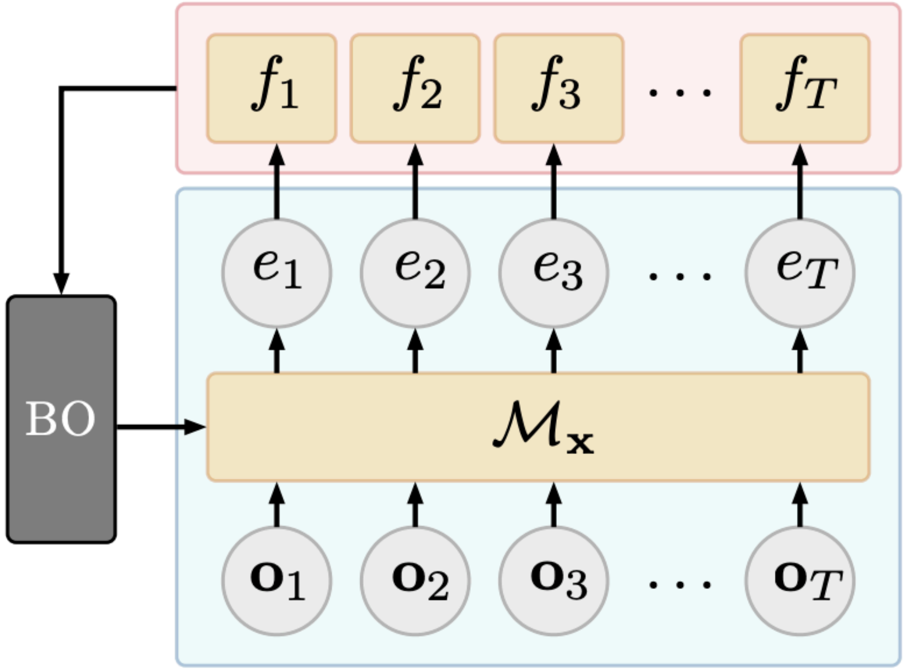
Select one optimal sequence model for all time steps? No!

Treat performance at each time step as its own black-box function

- Objective: Model performance at each time step
- Multi-Objective Bayesian Optimization finds *one* model with best trade-off across all objectives
- Expensive to compute volume gain w.r.t all the objectives ☹

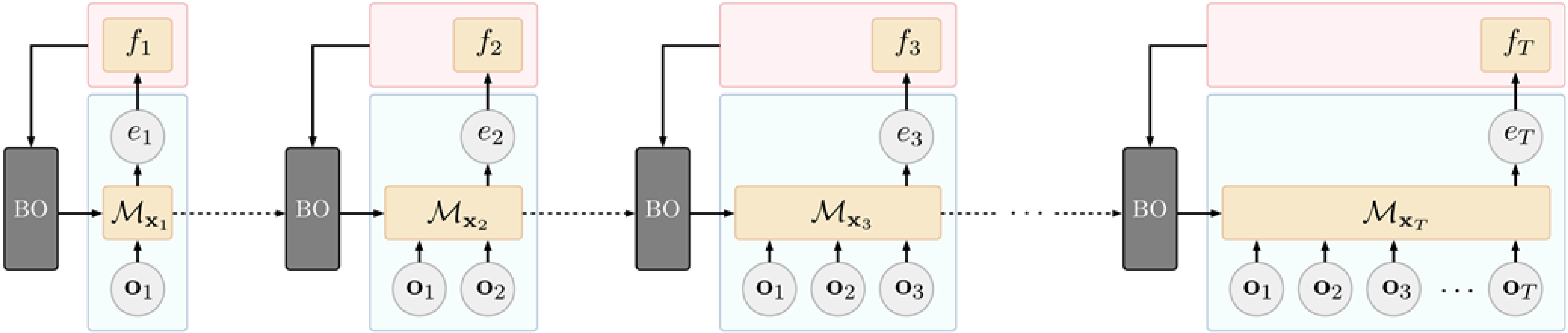
Other solutions?

Black box functions



Multi-Objective Bayesian Optimization (MOBO)

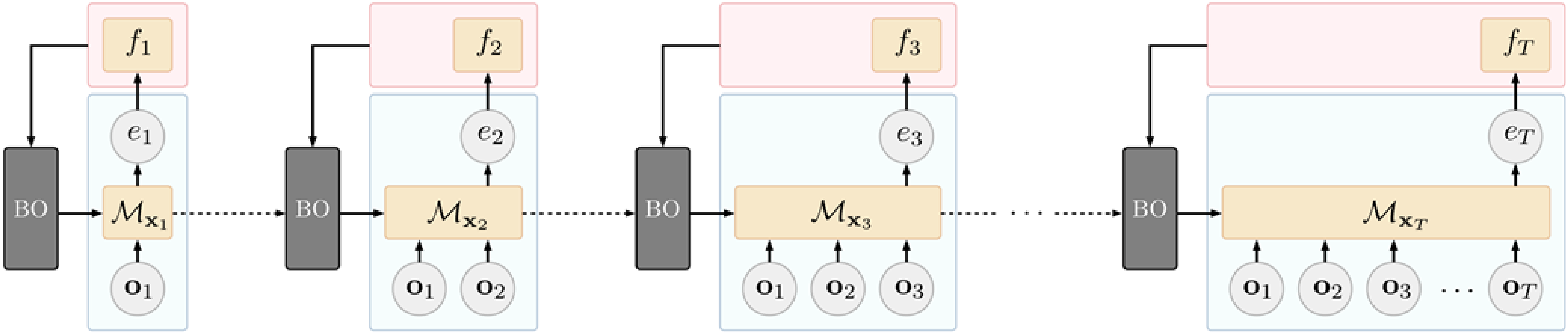
Apply BO sequentially across time-steps as multi-task? No!



Multi-Task Bayesian Optimization (MTBO)

☺ Warm-start: Transfer knowledge gained from previous optimizations to new tasks, such that subsequent optimizations are more efficient

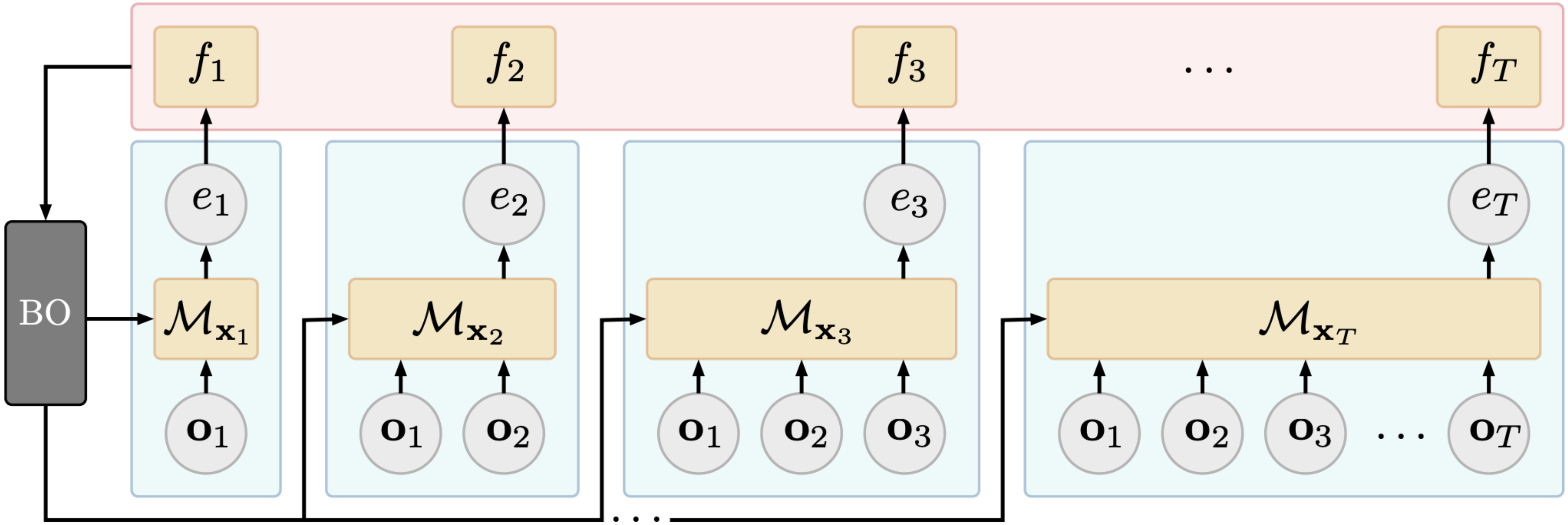
Apply BO sequentially across time-steps as multi-task? No!



- ☹ **MTBO requires evaluating deep learning models on large datasets which is prohibitively expensive**
- ☹ **MTBO requires solving T separate BO procedures in a sequence - unclear how to allocate evaluations among these subproblems**
- ☹ **MTBO does not take full advantage of information from all acquisition functions**

SMS-DKL [Zhang, Jarrett, vdS, AISTATS 2020]

A hyperparameter optimization tool for sequence model

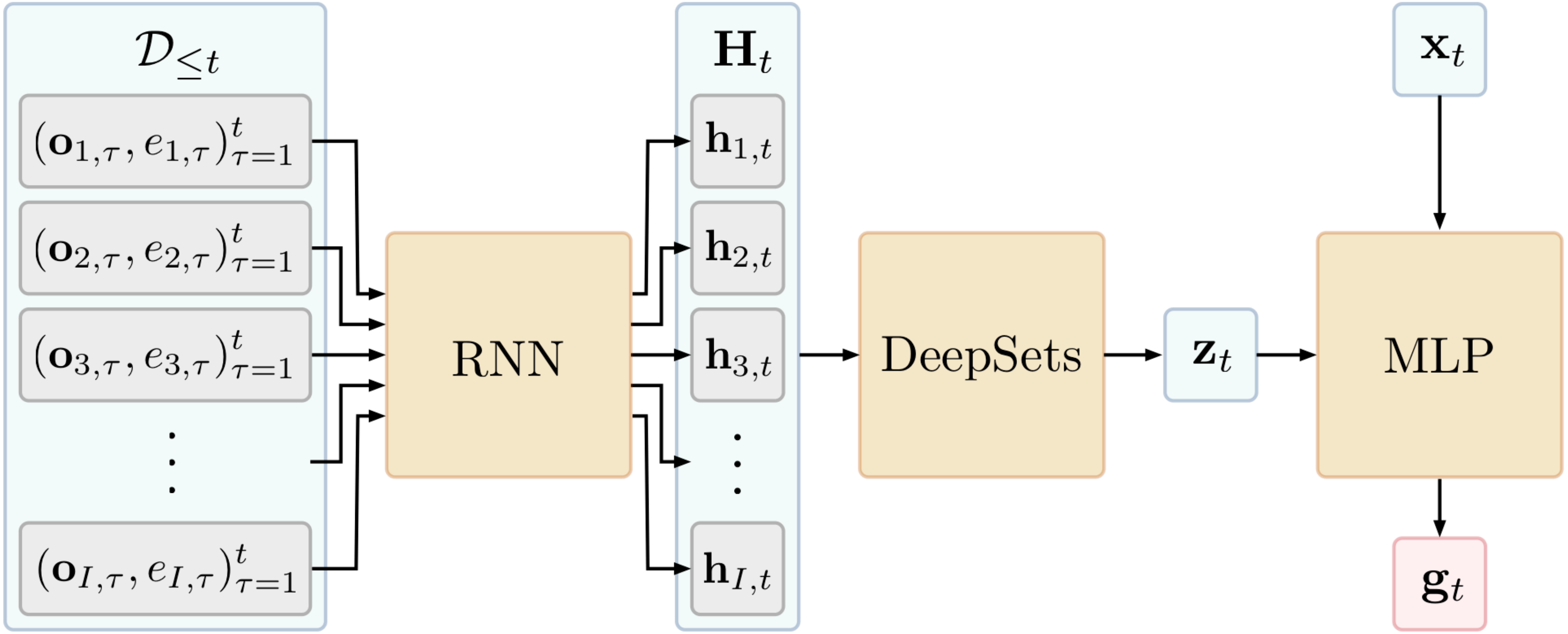


Solve the multiple black-box function optimization problem jointly and efficiently by learning and exploiting correlations among black-box functions using deep kernel learning

Stepwise Model Selection via Deep Kernel Learning – SMS-DKL

SMS-DKL [Zhang, Jarrett, vdS, AISTATS 2020]

How do we jointly and efficiently learn and exploit correlations among black-box functions?



Idea: Using deep kernel learning

Create feature maps to measure similarities between data tuples

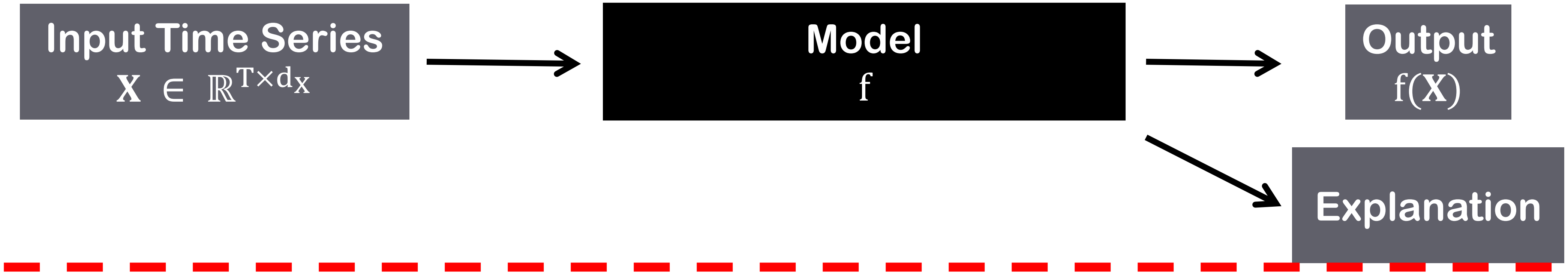
Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

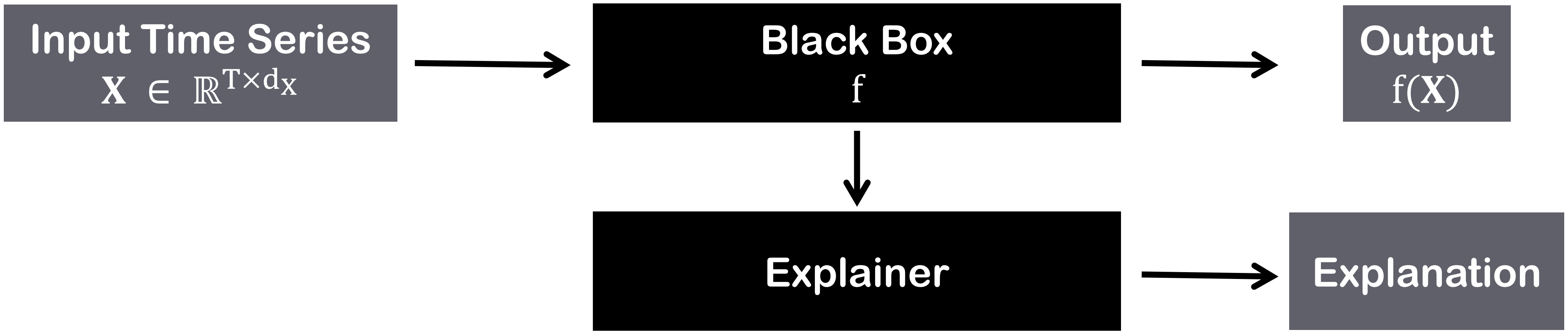
- Reproducibility and visualization

Intrinsic vs. Post-Hoc Interpretability

Intrinsic (e.g. linear models, trees, attention)



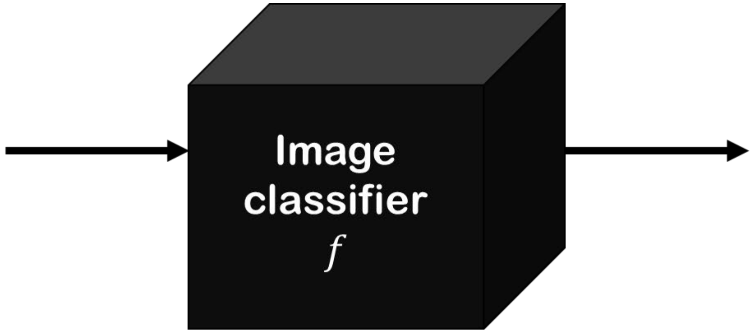
Post-Hoc (e.g. LIME, SHAP)



Feature Importance

Highlight most important features for the model

$\boldsymbol{x}$ = Image



$f(\boldsymbol{x}) = Dog$

- **Integrated Gradient [Sundararajan et al. 2017]**

$$a_i(f,\boldsymbol{x}) = \left(x_i - x_i^0\right) \times \int_0^1 \frac{\partial f[\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)]}{\partial x_i} dt$$

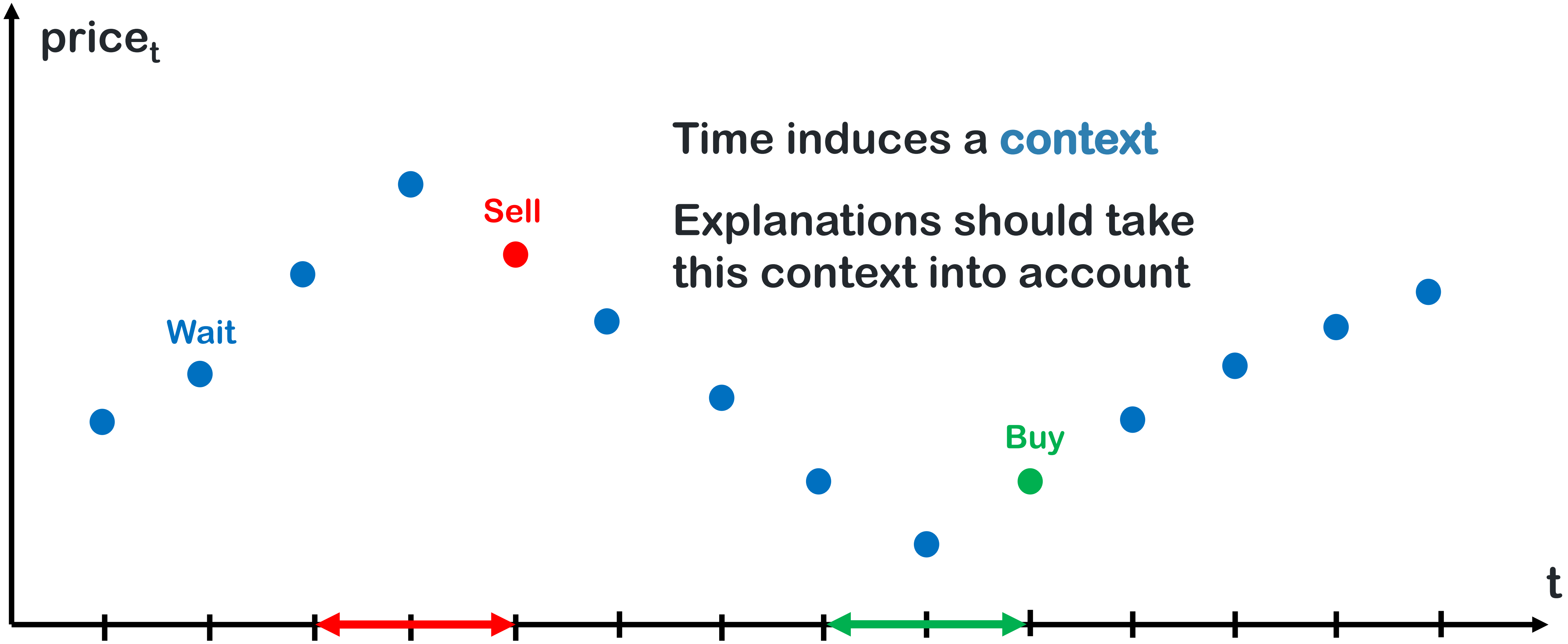
- **SHAP [Lundberg et al. 2017]**

$$a_i(f,\boldsymbol{x}) = \sum_{S \subset [\dim \mathcal{X}] \setminus \{i\}} \frac{|S|!\,(\dim \mathcal{X} - |S| - 1)}{(\dim \mathcal{X})!} \left[f\left(\boldsymbol{x}_{S \cup \{i\}}\right) - f(\boldsymbol{x}_S)\right]$$

"Standard" feature importance methods perform poorly for time-series

[Ismail et al., NeurIPS 2020]

What makes Time Series special?



Time induces a **context**

Explanations should take this context into account

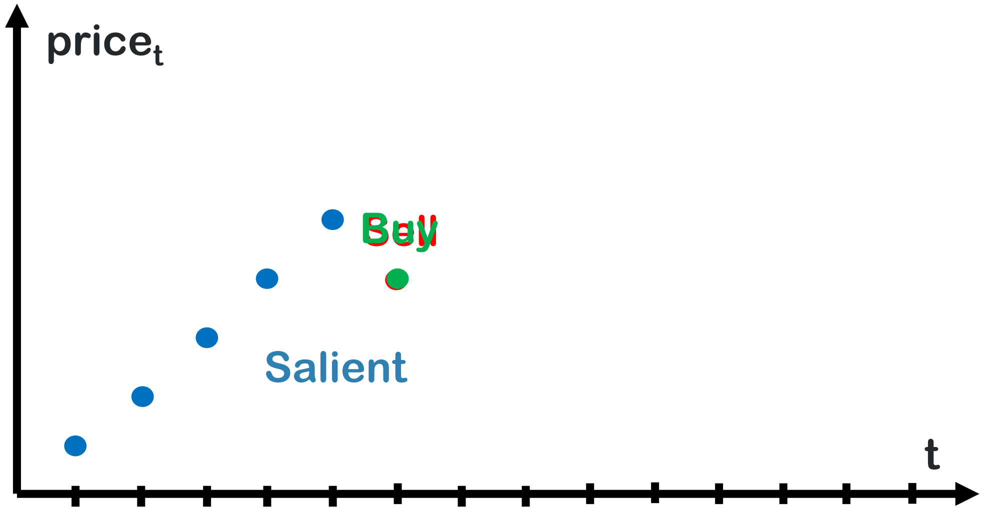
How to detect salient features?

Perturbation based detection

Premise: salient features **affect** the model's **prediction**

Detect salient features by **feature perturbations**

Feature perturbation affects prediction → **Salient** feature



How to take the time context into account? [Crabbé, vdS, ICML 2021]

Time context matters

Typical saliency methods treat each input $x_{t,i}$ as a feature

$\Longrightarrow$ Time dependency is **ignored** by the saliency method

Dynamic Perturbation Operator

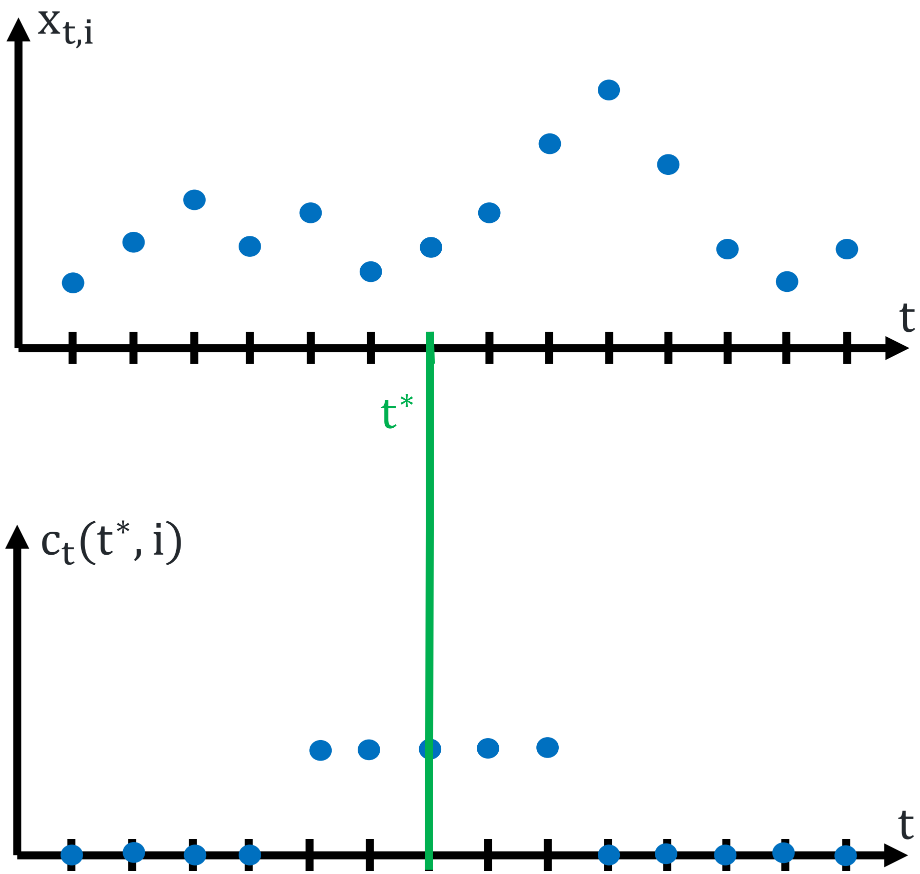
Idea: perturb each $x_{t^*,i}$ by using **neighbouring times:**

Perturbed input / Linear combination

$$\pi\left(x_{t^*,i}\,;t^*,i\right) = \sum_{t=t^*-W_1}^{t^*+W_2} c_t(t^*,i) \times x_{t,i}$$

$\Longrightarrow$ Time dependency is **integrated** in perturbation

Window perturbation:



How to take the time context into account? [Crabbé, vdS, ICML 2021]

Time context matters

Typical saliency methods treat each input $x_{t,i}$ as a feature

$\Longrightarrow$ Time dependency is ignored by the saliency method

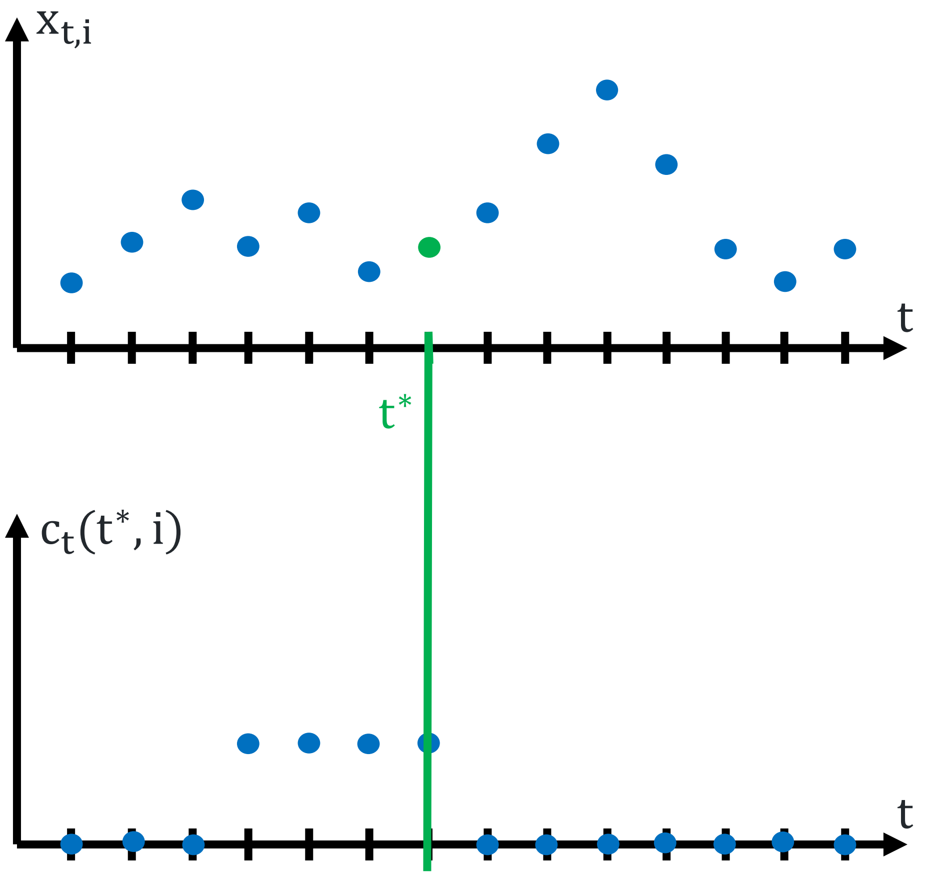
Dynamic Perturbation Operator

Idea: perturb each $x_{t^*,i}$ by using **neighbouring times:**

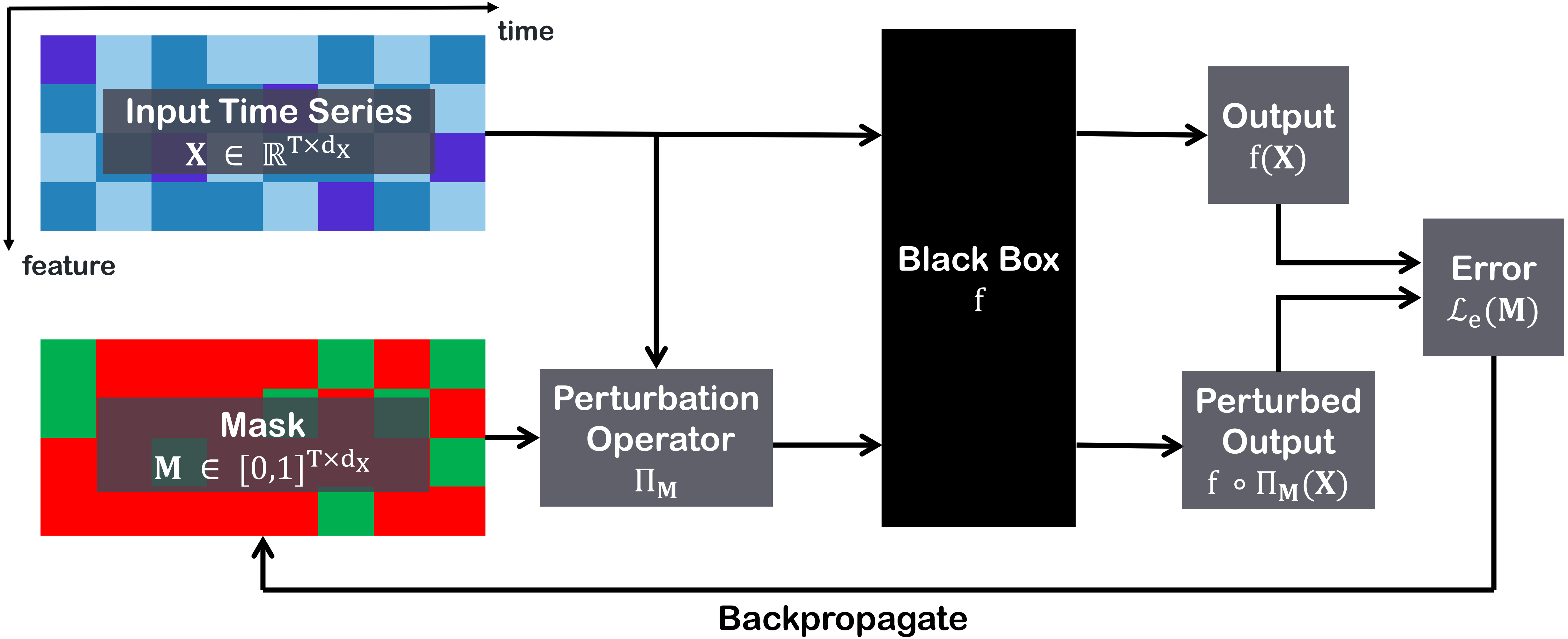
$$\pi\left(x_{t^*,i}\,; t^*, i\right) = \sum_{t=t^*-W_1}^{t^*+W_2} c_t(t^*, i) \times x_{t,i}$$

$\Longrightarrow$ Time dependency is integrated in perturbation

Past window perturbation:



Dynamask [Crabbé, vdS, ICML 2021]



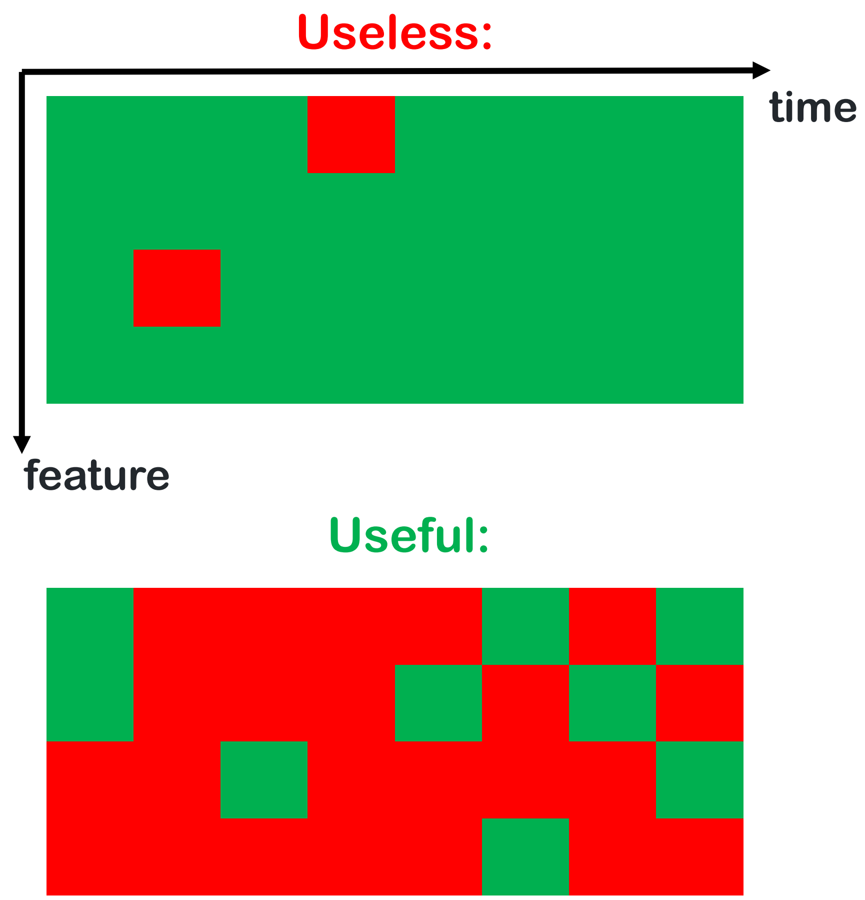
How to make the masks parsimonious?

[Crabbé, vdS, ICML 2021]

What do we mean by parsimonious?

Masks should **not** highlight more features than necessary

$\Longrightarrow$ We need to enforce feature **selection**



How to enforce parsimony?

The user selects the fraction a of most important features

We add a regularization to enforce sparsity:

$$\mathcal{L}_a(\mathbf{M}) = \|\text{vecsort}(\mathbf{M}) - \mathbf{r}_a\|^2$$

Sets the $(1-a) \times T \times d_X$ smallest mask coefficients to zero

How to avoid quick variations of saliency? [Crabbé, vdS, ICML 2021]

Quick time variations of the saliency

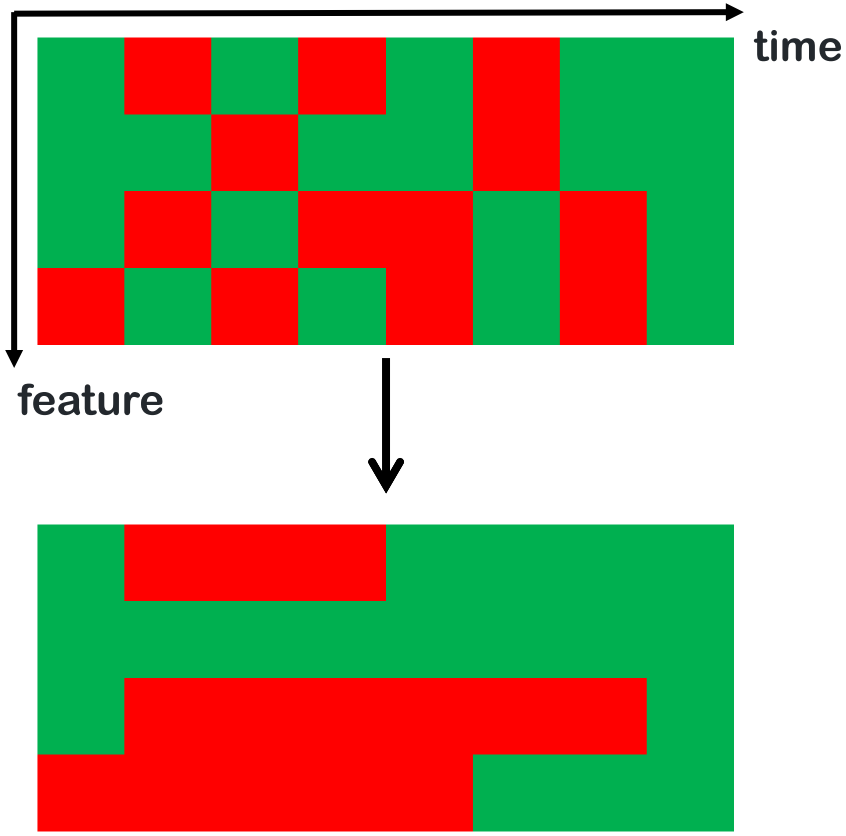
Might want to avoid **quick time variations** of the saliency

This can be a **prior belief** or a **preference** of the user

How to avoid this?

We add a regularization to penalize saliency jumps over time:

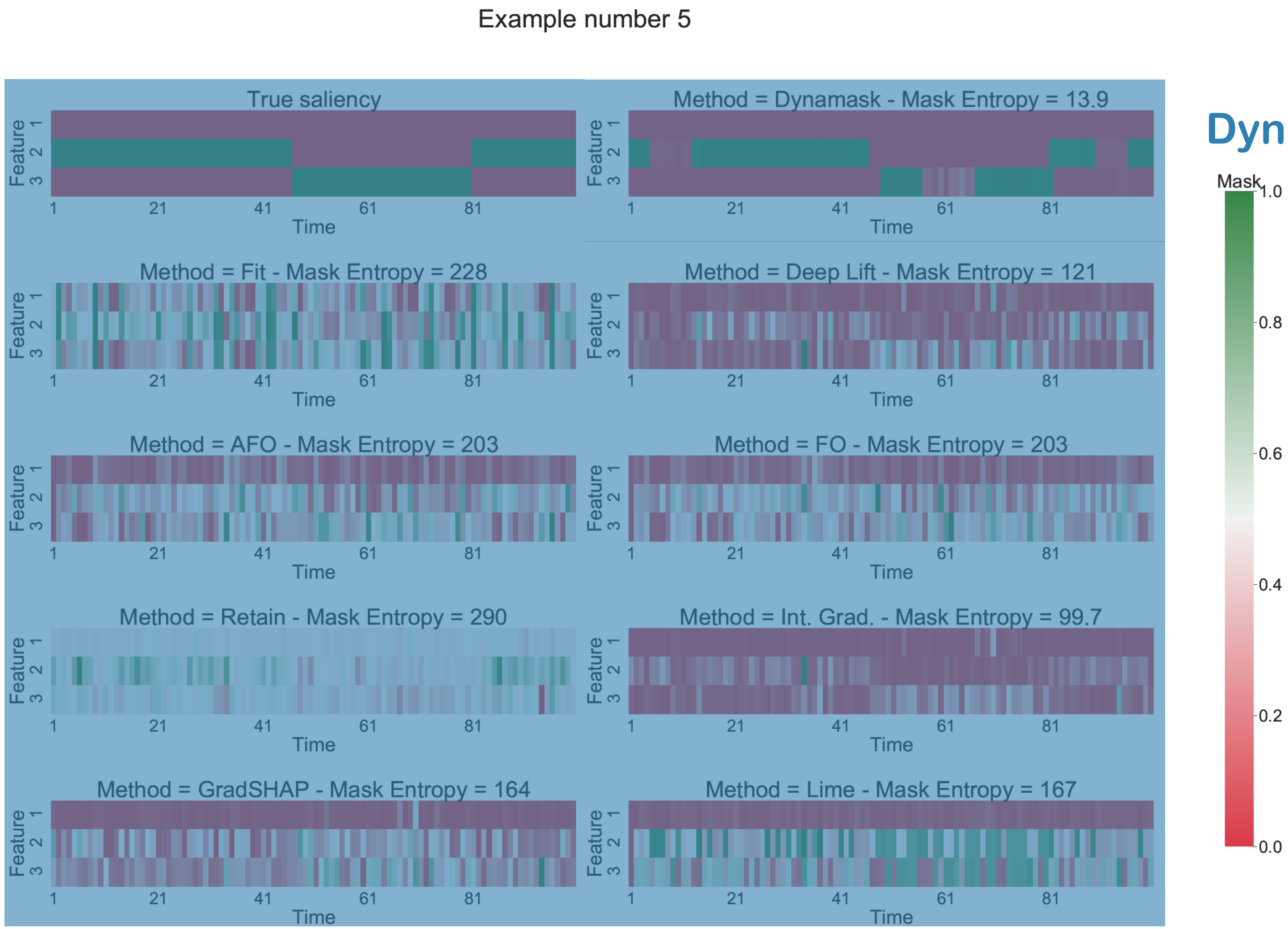
$$\mathcal{L}_c(\mathbf{M}) = \sum_{t=1}^{T-1} \sum_{i=1}^{d_X} \left| m_{t+1,i} - m_{t,i} \right|$$



Dynamask - Example

[Crabbé, vdS, ICML 2021]

True saliency



Dynamask saliency

Baseline saliency

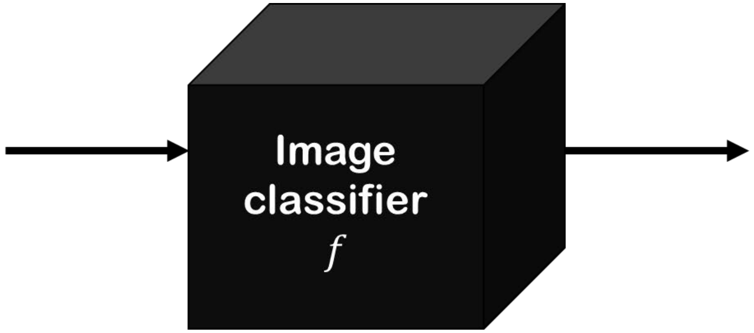
Feature Importance

Highlight **most important features** for the model

- Integrated Gradient [Sundararajan et al. 2017]

$$a_i(f,\boldsymbol{x}) = \left(x_i - x_i^0\right) \times \int_0^1 \frac{\partial f[\boldsymbol{x}^0 + t(\boldsymbol{x} - \boldsymbol{x}^0)]}{\partial x_i} dt$$

$\boldsymbol{x}$ = Image



$f(\boldsymbol{x}) = Dog$

- SHAP [Lundberg et al. 2017]

$$a_i(f,\boldsymbol{x}) = \sum_{S \subset [\dim \mathcal{X}] \setminus \{i\}} \frac{|S|!\,(\dim \mathcal{X} - |S| - 1)}{(\dim \mathcal{X})!} \left[f\left(\boldsymbol{x}_{S\cup\{i\}}\right) - f(\boldsymbol{x}_S)\right]$$

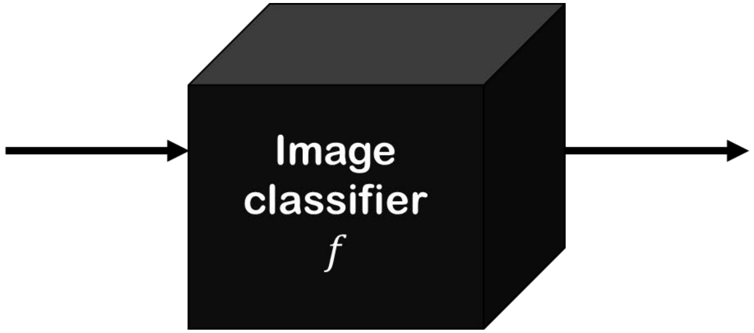
Example Based Explanations

Highlight **relevant examples** seen by the model

- Influence Functions [Koh & Liang 2017]

$$a_{\boldsymbol{z}^i}(f_\theta, \boldsymbol{z}) = -\left\langle \nabla_\theta L(\boldsymbol{z}),\ \boldsymbol{H}_\theta^{-1} \nabla_\theta L\left(\boldsymbol{z}^i\right) \right\rangle$$

$\boldsymbol{x}$ = Image

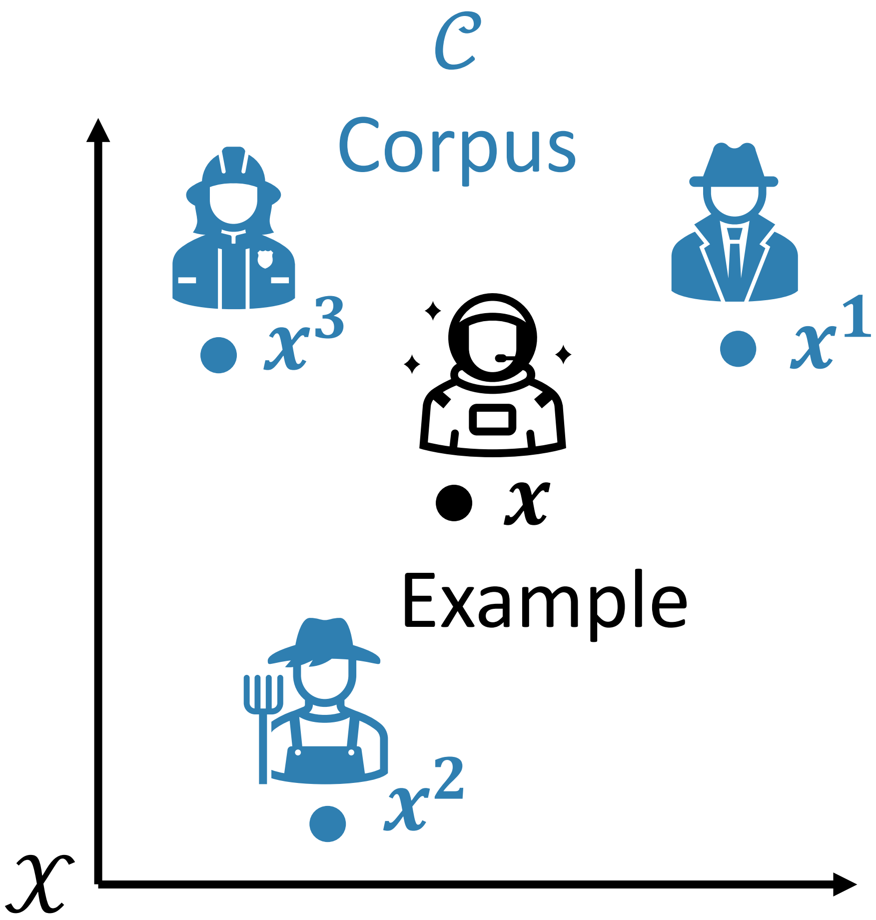


$f(\boldsymbol{x}) = Dog$

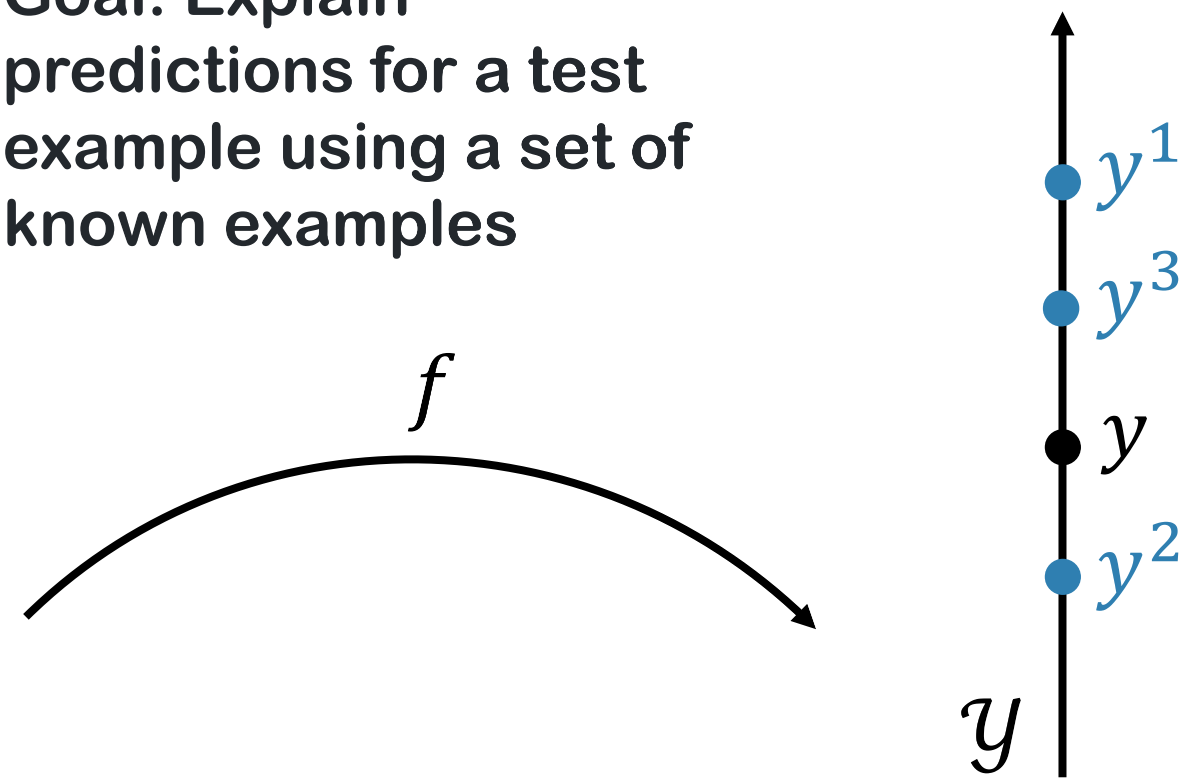
- Representer Theorem [Yeh et al. 2018]

$$a_{\boldsymbol{z}^i}(f_\theta, \boldsymbol{z}) = k(\boldsymbol{x}, \boldsymbol{x}^i, \alpha^i)$$

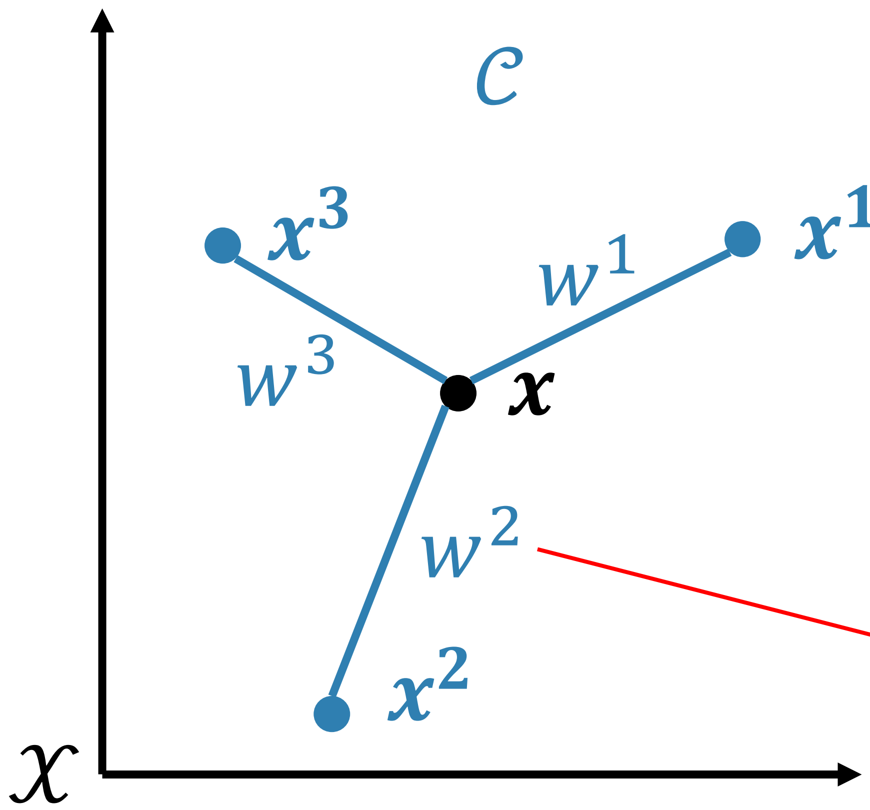
Problem Setup



Goal: Explain predictions for a test example using a set of known examples



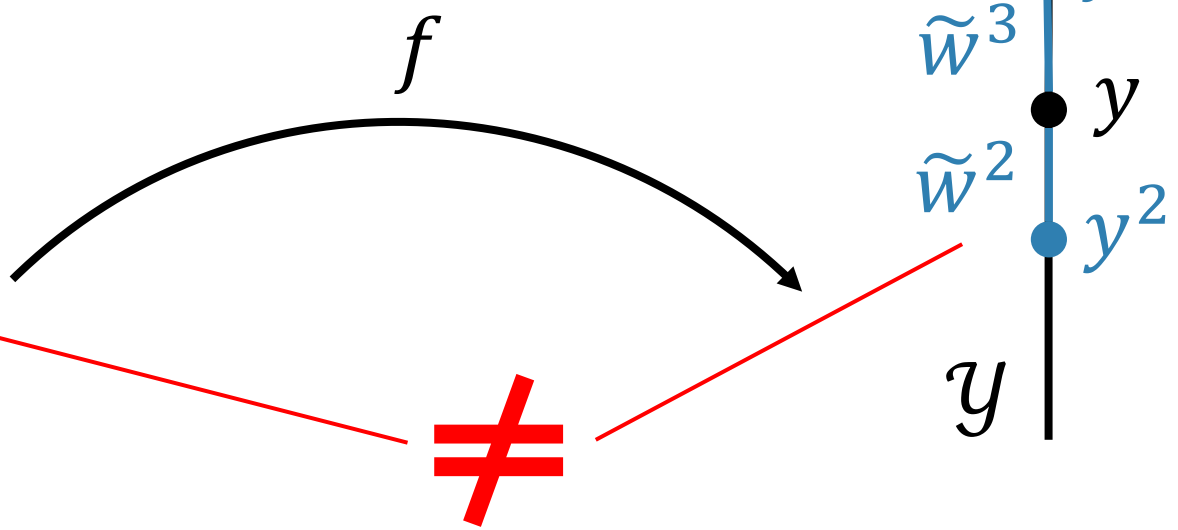
A First Attempt – Input Similarity



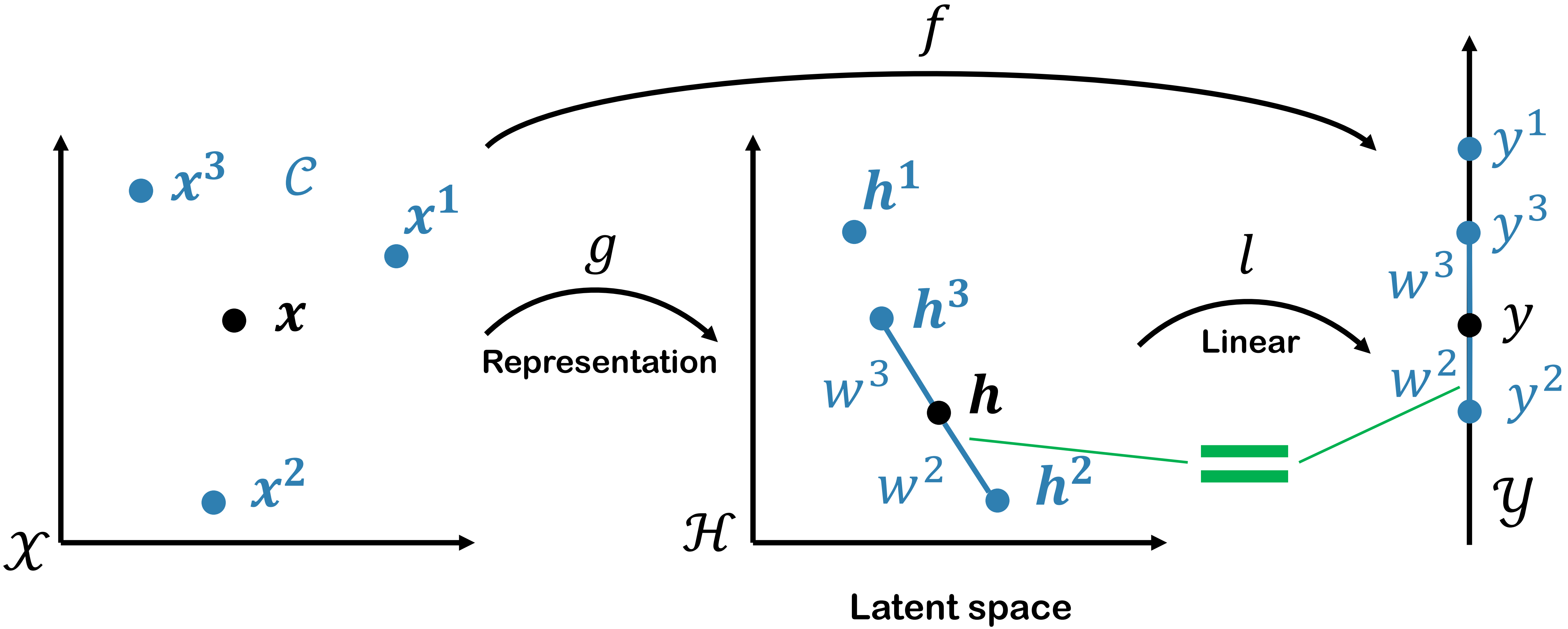
Weights not conserved if f nonlinear

Convex sets not conserved if f nonconvex

Input space $\mathcal{X}$ does not reflect the black-box f



Leveraging Learned Features



Corpus Decomposition

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

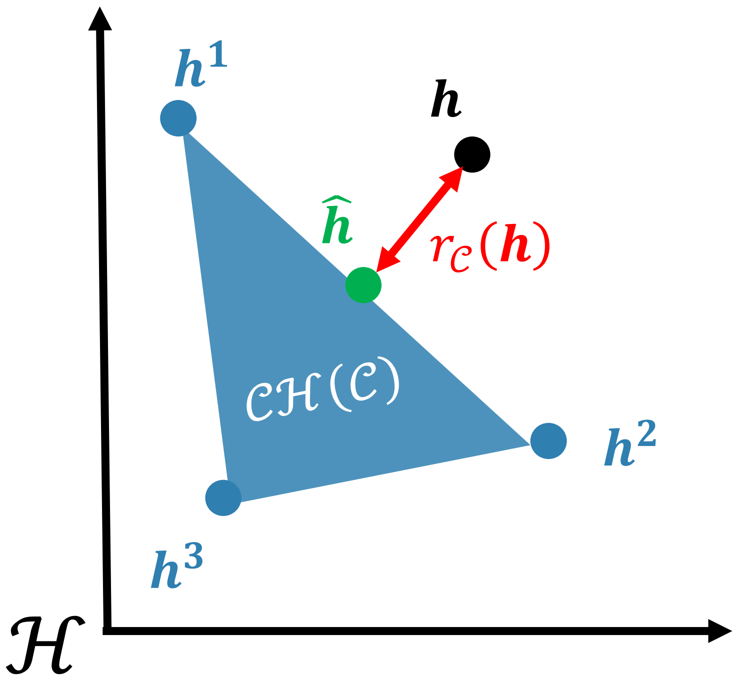
- **Corpus hull in latent space**

$$\mathcal{CH}(\mathcal{C}) \equiv \left\{\sum_{c=1}^{C} w^c \boldsymbol{h}^c \mid w^c \in [0,1] \forall c \in [C] \wedge \sum_{c=1}^{C} w^c = 1\right\}$$

- **Find the best corpus decomposition of the example**

$$\widehat{\boldsymbol{h}} = \arg\min \left\| \boldsymbol{h} - \widetilde{\boldsymbol{h}} \right\|_{\mathcal{H}} \quad s.t. \ \widetilde{\boldsymbol{h}} \in \mathcal{CH}(\mathcal{C})$$

- **Might have a residual** $r_{\mathcal{C}}(\boldsymbol{h}) = \left\| \boldsymbol{h} - \widehat{\boldsymbol{h}} \right\|_{\mathcal{H}}$

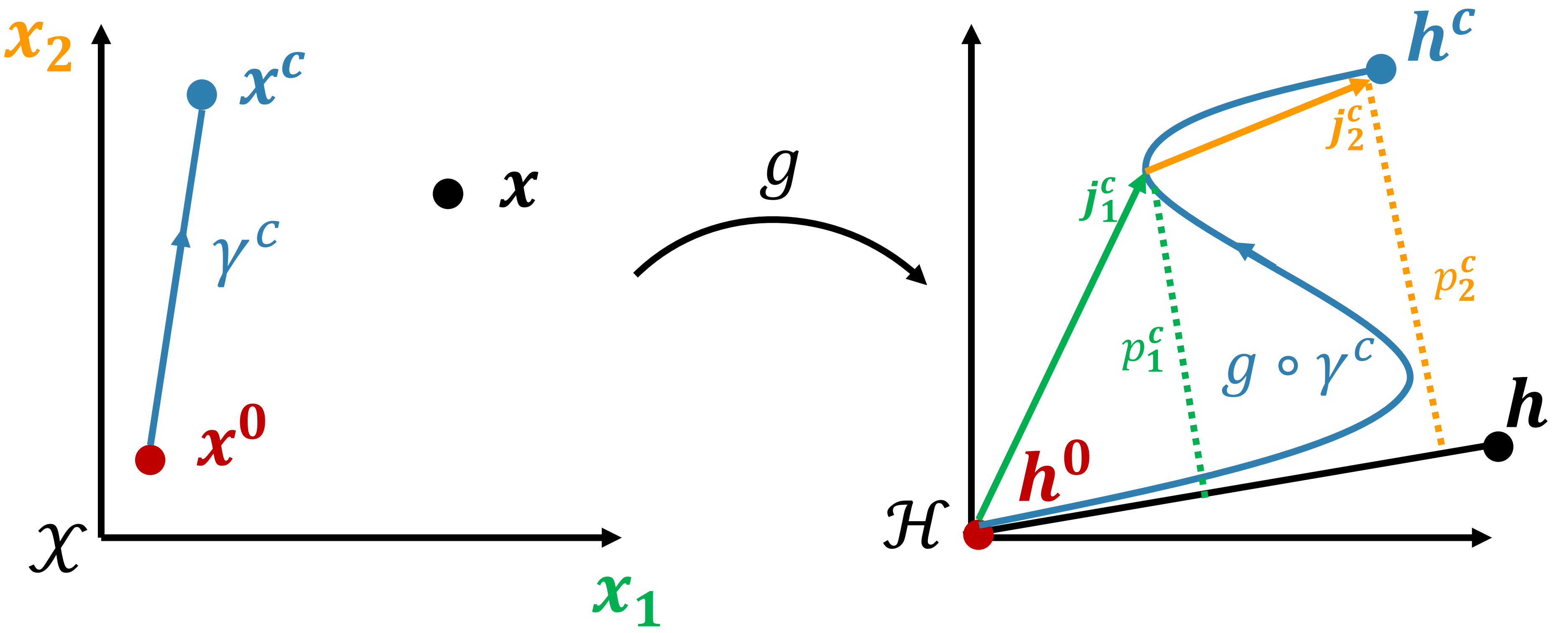


Explaining At The Feature Level

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

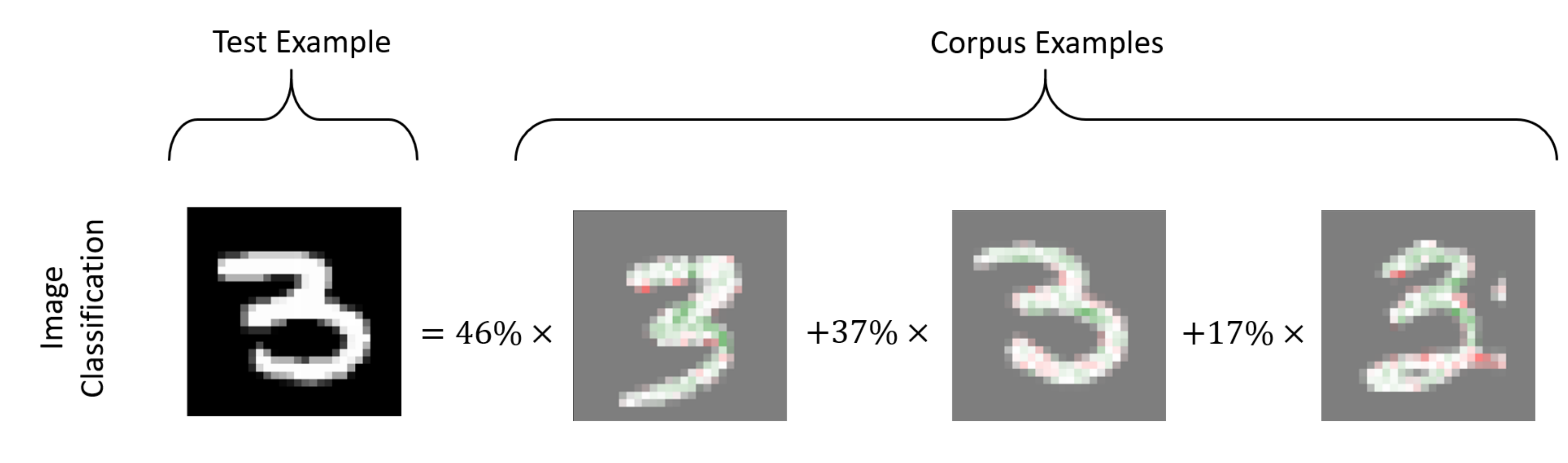
$$j_i^c = \int_0^1 \frac{\partial g \circ \gamma^c}{\partial x_i}(t)\, dt$$

$$p_i^c = \frac{\langle \boldsymbol{h} - \boldsymbol{h}^0, \boldsymbol{j}_i^c \rangle}{\langle \boldsymbol{h} - \boldsymbol{h}^0, \boldsymbol{h} - \boldsymbol{h}^0 \rangle}$$



SimplEx

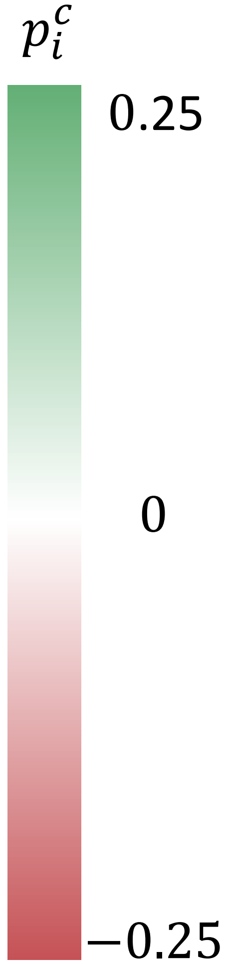
[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]



Clinical Risk Estimation

= 58% × + 25% × + 17% ×

Feature	Test Example	Corpus Example 1 (58%)	Corpus Example 2 (25%)	Corpus Example 3 (17%)
Age	78.0	73.0	73.0	79.0
PSA	10.2	8.0	7.1	10.5
Comorbidities	0.0	0.0	0.0	0.0
Treatments	['RT-RDx']	['RT-RDx']	['RT-RDx']	['RT-RDx']
Grade	1	1	1	2
Stage	2	2	2	2
Gleason1	2	2	2	2
Gleason2	3	4	2	3



What Makes Simplex Special?

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

- SimplEx gives the user freedom to choose the corpus of examples to explain the model predictions
- **Advantage:**
  - No need for this corpus to be from the model's training set

    (a) The training set of a model is not always accessible

    (b) The user might want explanations in terms of examples that make sense for them

What Makes Simplex Special?

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

- Keep humans in the loop

  Leverage user's knowledge: SimplEx explains with a corpus **chosen by the user**

- Increase the scientific content of the models

  Expand the picture: SimplEx **unifies** example and feature-based explanations

  Enhance the picture: SimplEx **captures insights** from the model's **latent space**

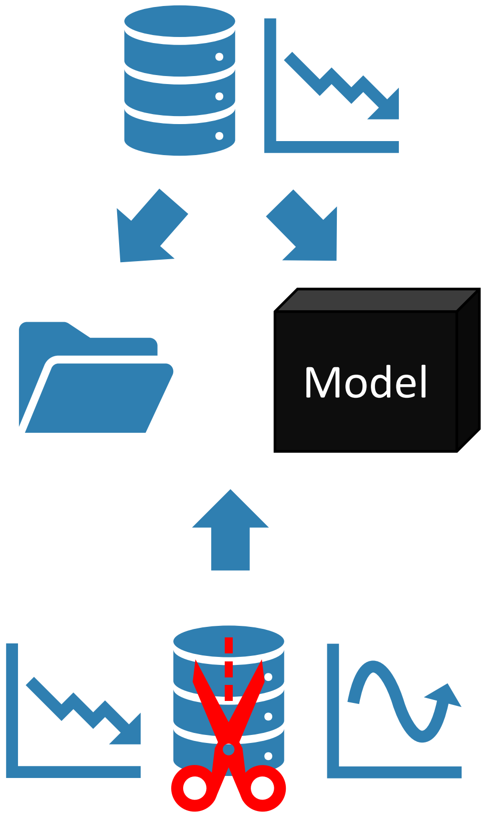
- Debug the models

  Trigger user's scepticism: residual $r_{\mathcal{C}}$ detects examples for which the model **extrapolates**

Detecting Model's Limitations

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

- Start with non-oscillating time AR time series dataset $\mathcal{D}$
- Split it into a training and testing set $\mathcal{D} = \mathcal{D}_{\text{train}} \sqcup \mathcal{D}_{\text{test}}$
- Train a forecasting RNN on $\mathcal{D}_{\text{train}}$
- Sample a corpus from training set $\mathcal{C} \subset \mathcal{D}_{\text{train}}$
- Corrupt the testing set with oscillating AR time series $\mathcal{T} = \mathcal{D}_{\text{test}} \sqcup \mathcal{D}_{\text{oscil}}$
- Make a corpus decomposition of each example in $\mathcal{T}$, compute the residual $r_{\mathcal{C}}$
- Can we detect oscillating time series with corpus residual?



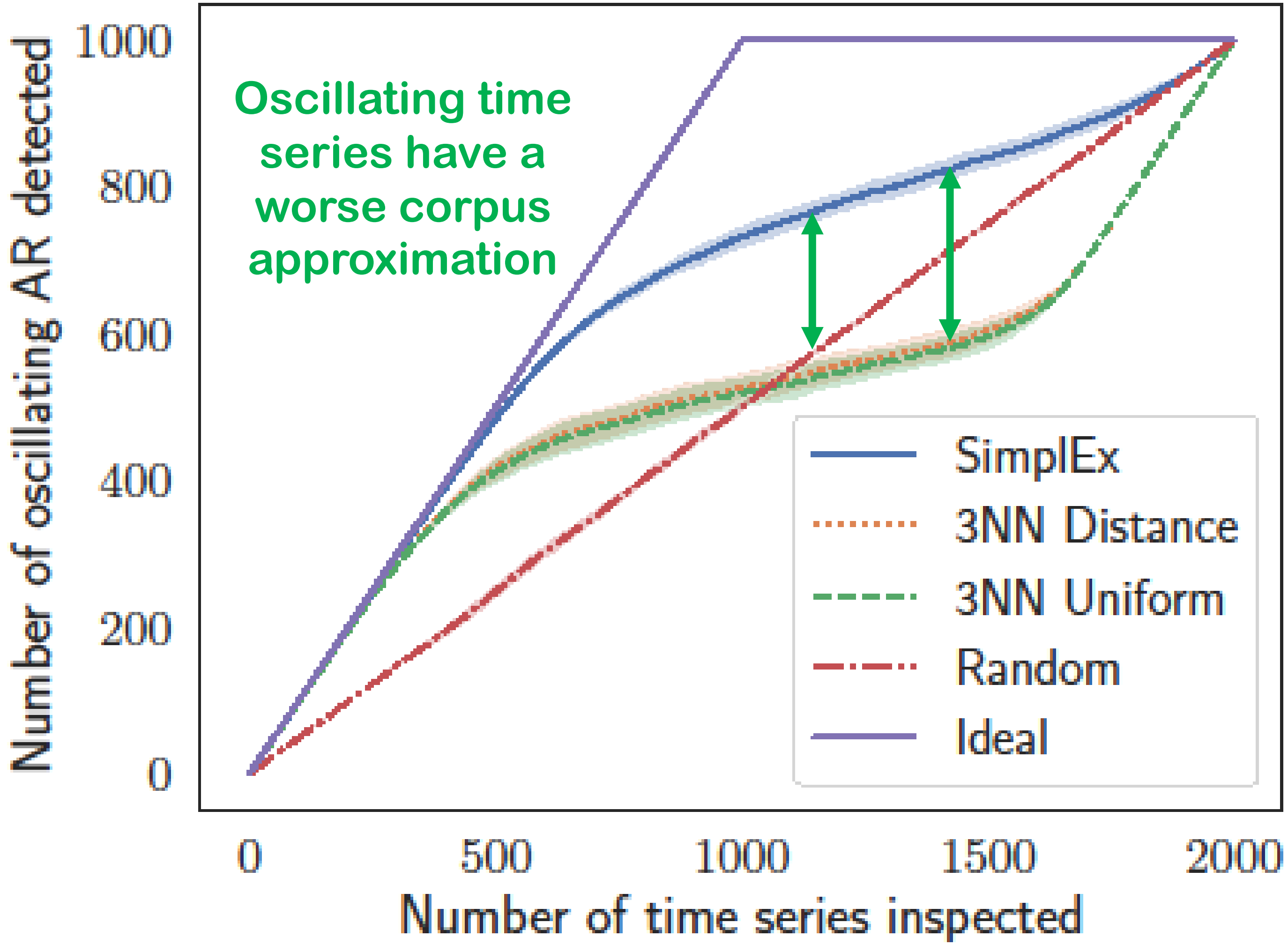
Detecting Model's Limitations

[Crabbé, Qian, Imrie, vdS, NeurIPS 2021]

Sort the time series by decreasing order of residual $r_{\mathcal{C}}$

Inspect the time series in this order

Increase the counter each time an oscillating time series is detected



SimplEx offers a better detection than 3NN baselines

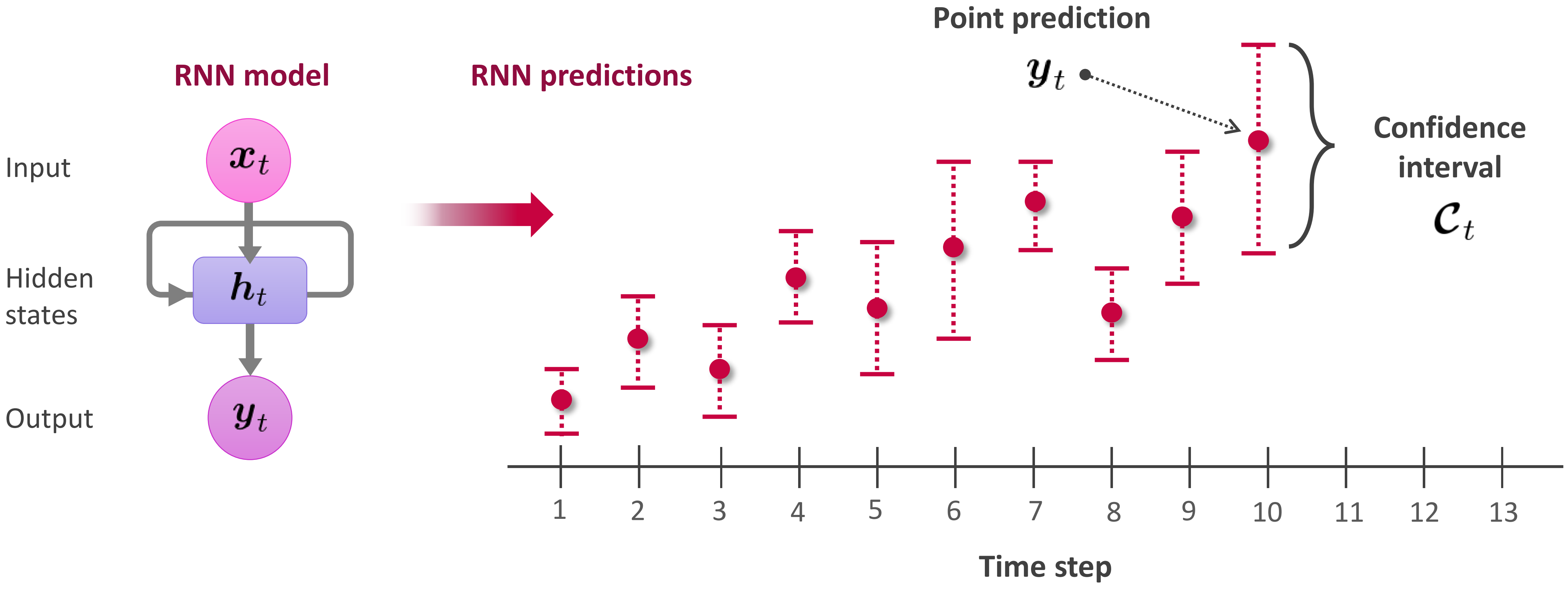
Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

Objective: sequential confidence intervals for RNNs

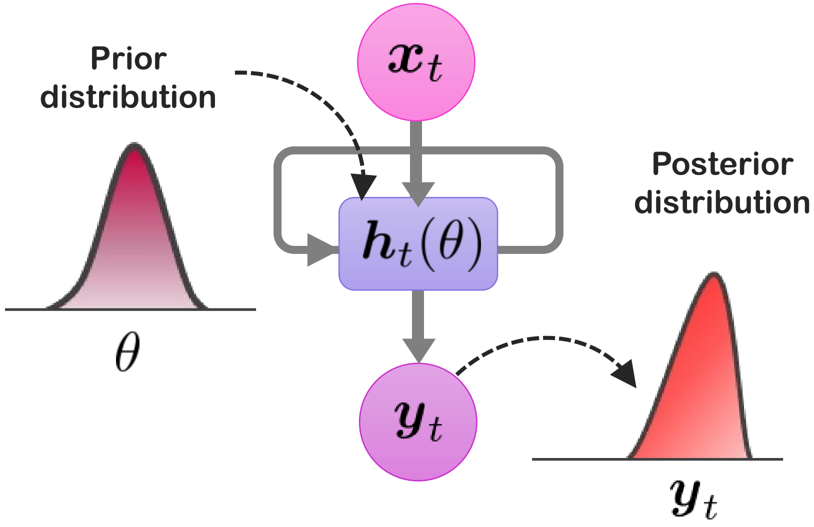
Predictive intervals for Recurrent Neural Networks (RNNs).



Some solutions

Bayesian RNNs

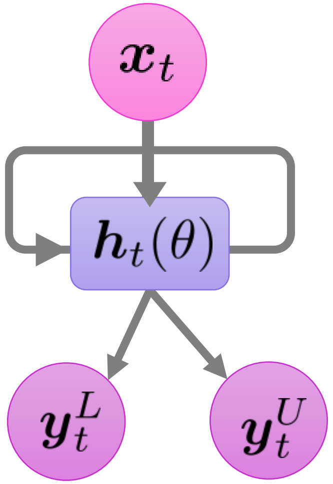
Prior over RNN parameters
Uncertainty = credible intervals



Posterior is intractable = Monte Carlo dropout
(Gal & Ghahramani, 2016)

Quantile RNNs

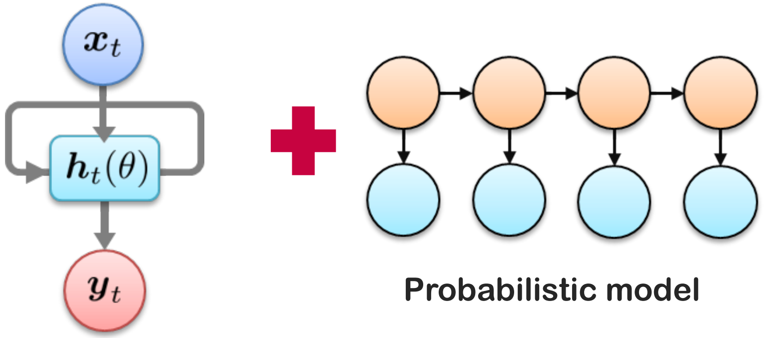
Explicitly train a multi-output RNN to predict intervals



Quantile loss for RNN training
(Gasthaus et al., 2019)

Probabilistic RNNs

Combine RNNs with variants of state-space models



Attentive state-space model (Alaa & van der Schaar, 2019)

Deep state-space model
(Rangapuram et al., 2018)

Why are these solutions not enough in healthcare?

Post-hoc application

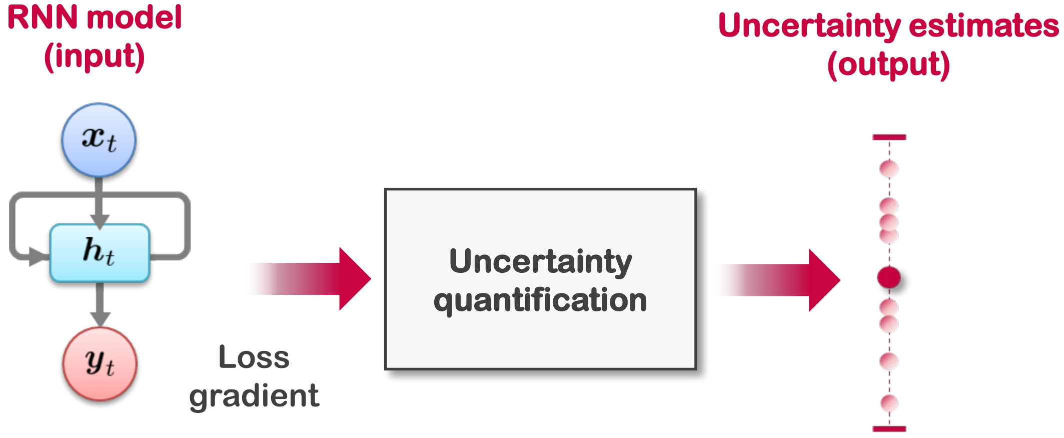
- Does not affect model accuracy
- Does not interfere with model training

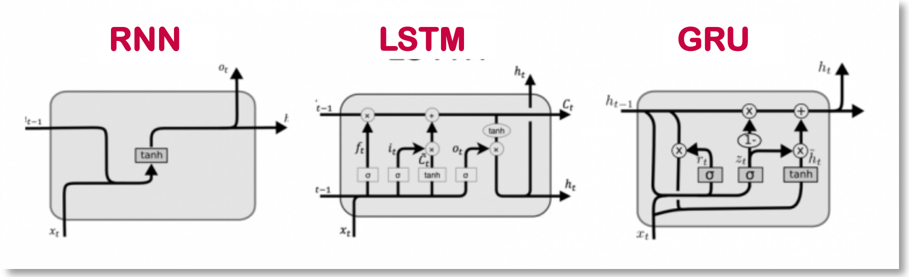
Generality and versatility

- Does not require changes to model architecture
- Applies to a wide range of sequence prediction settings

Frequentist coverage guarantees

- Formal frequentist procedure

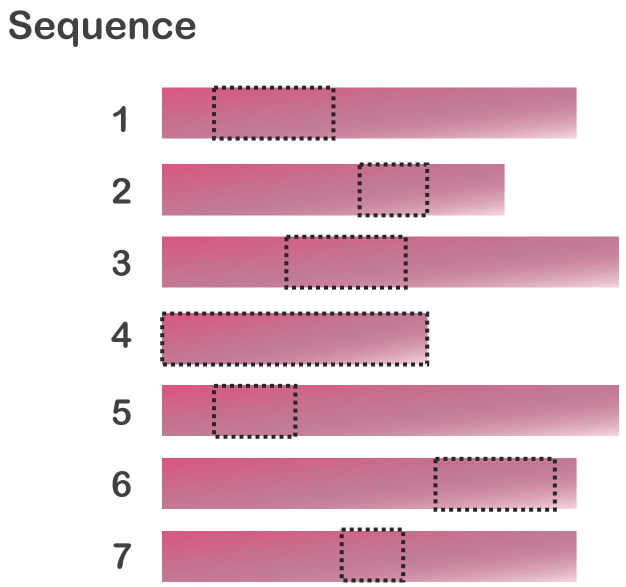




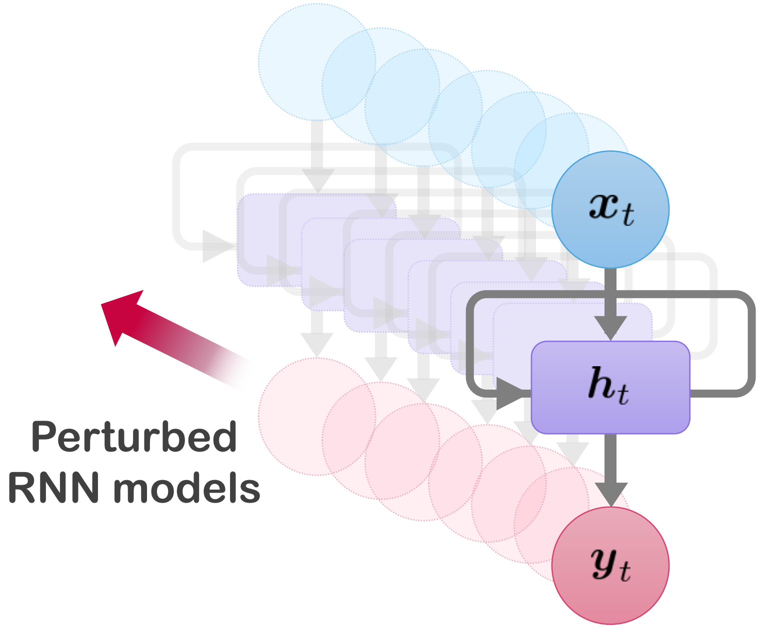
Frequentist Uncertainty in Recurrent Neural Networks via Blockwise Influence Functions [Alaa & vdS, ICML 2020]

Uncertainty intervals = variability in re-sampled RNN outputs.
RNN outputs are re-sampled by perturbing the model parameters through iterative deletion of blocks of data and re-training the model on the remaining data

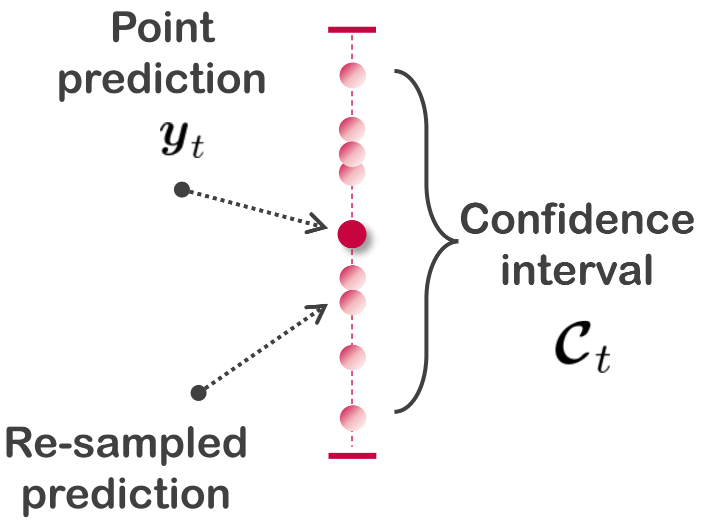
Block deletion



RNN model re-training



RNN prediction re-sampling



van_der_Schaar \LAB

vanderschaar-lab.com

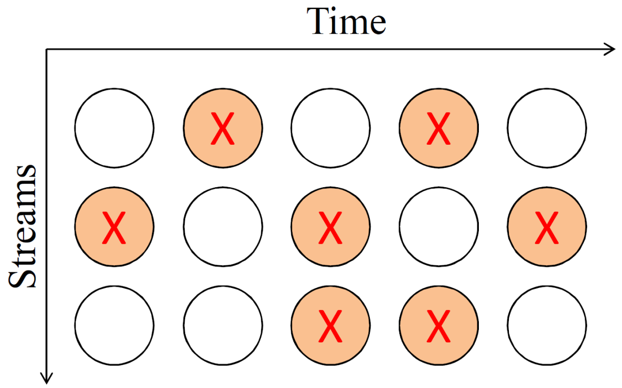
UNIVERSITY OF CAMBRIDGE

Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

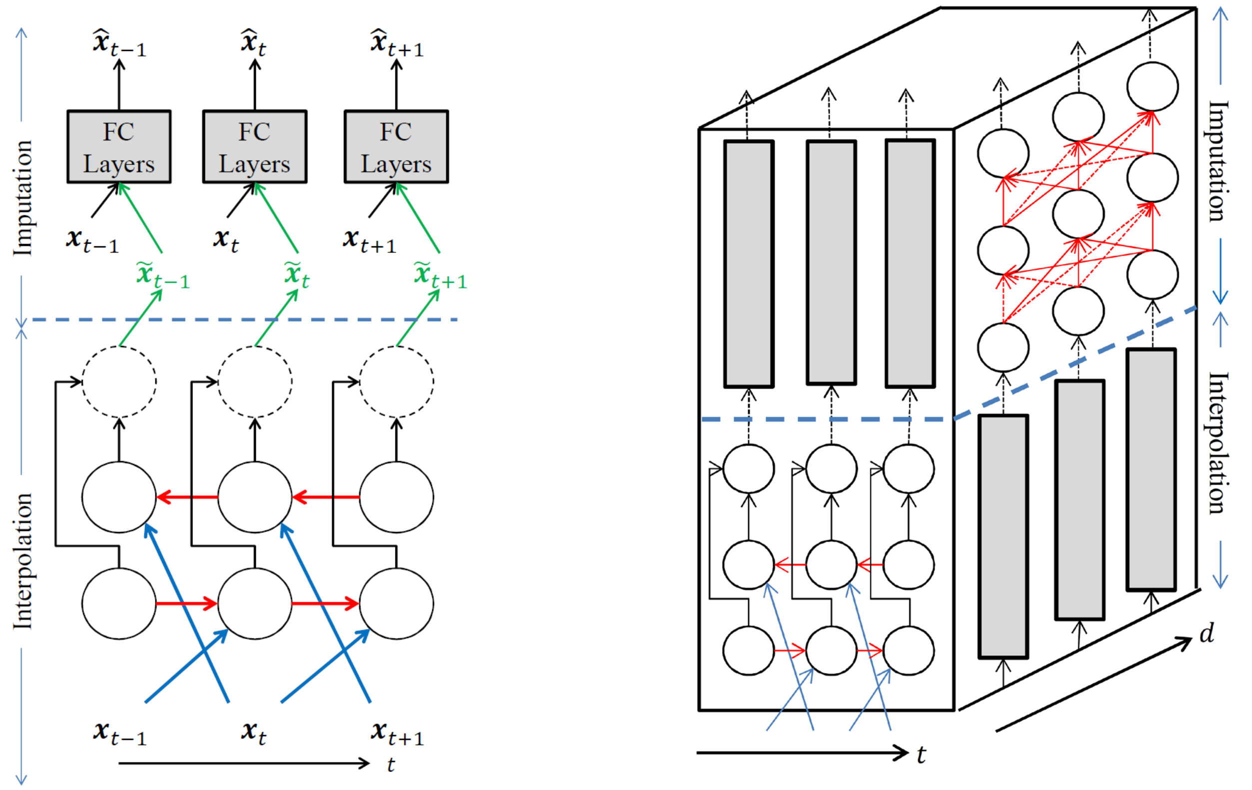
Multi-directional RNN (M-RNN) [Yoon, Zame, vdS, TBME 2018]



Temporal data streams

- **Interpolation – temporal correlations**
- **Imputation –cross-features correlations**
- Both correlations must be **simultaneously** learned

Multi-directional RNN (M-RNN) [Yoon, Zame, vdS, TBME 2018]

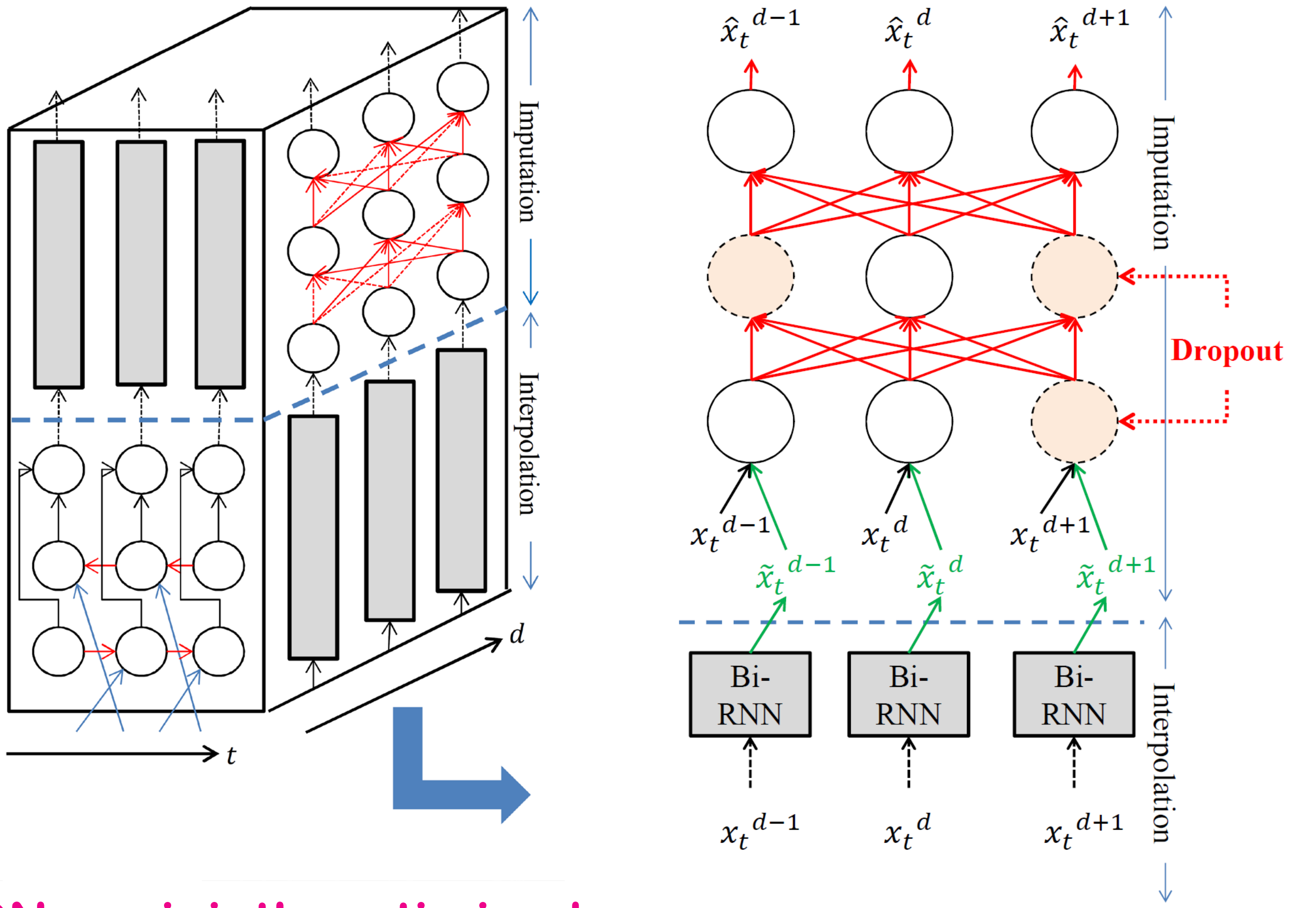


Sequentially operate across streams and within streams

Timing of inputs into hidden layers is both lagged in forward direction and advanced in the backward direction

Multi-directional RNN (M-RNN) [Yoon, Zame, vdS, TBME 2018]

- **Correlations across features:**
  FC network
- **Multiple imputations:**
  Dropout



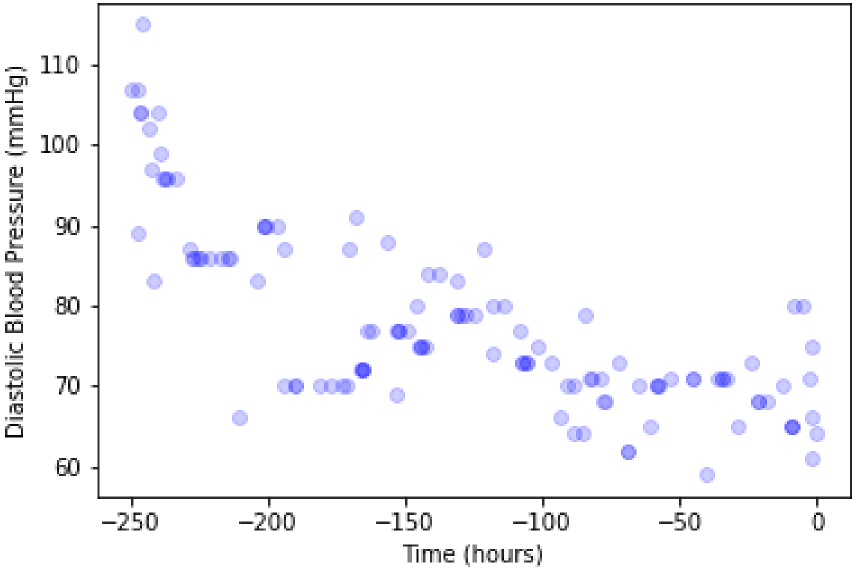
Bi-RNN and FCN are jointly optimized

Can we do better? Learn from clinical judgements!

[Alaa, Hu, vdS, ICML 2017]

Data - shaped by **clinical judgments!**

Probabilistic model for learning from observational data

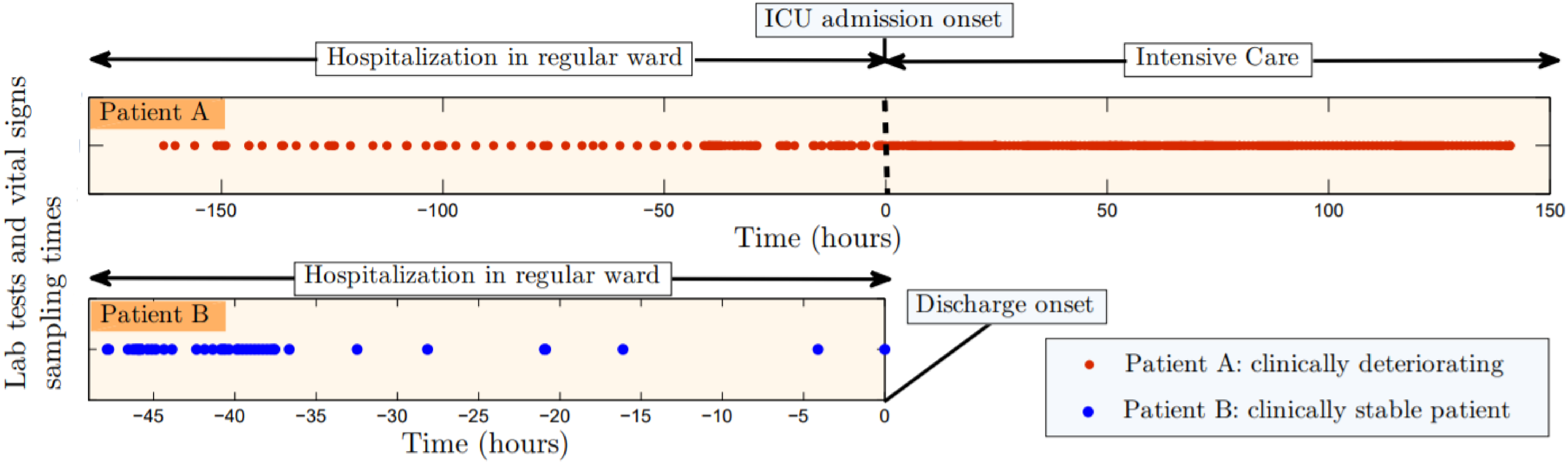


Informative sampling:
Time-varying sampling frequency

Model a patient's trajectory as a marked point process modulated by their health state

Elements of the probabilistic model (I): the observation process

- Nature of Informative Sampling is **Problem-dependent**
- **E.g. Cancer patient in regular hospital wards:** evidence that sampling rate increases when patient is in a bad health state



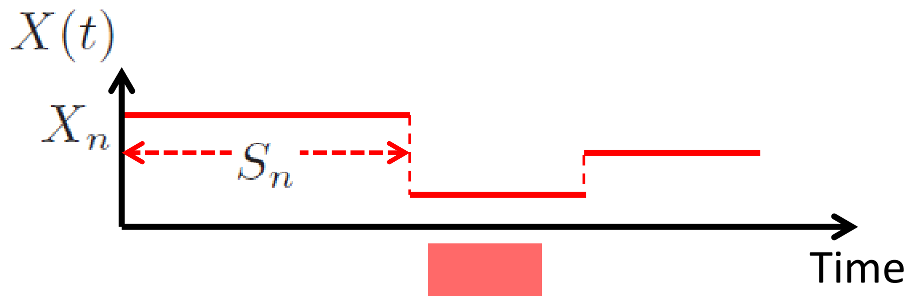
Elements of the probabilistic model (II): the observation process

Clinicians observe the patient's vital signs and lab tests according to a **Hawkes process**

$\{t_m\}_{m \in \mathbb{N}_+}$

...doubly stochastic point process

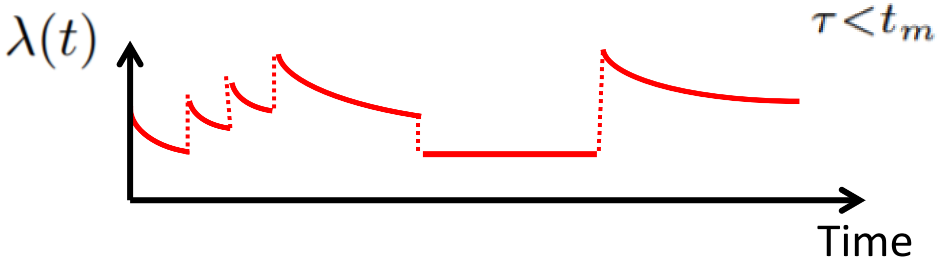
Captures impact of patient's health state on clinicians' sampling behavior



...with a self-exciting Triggering kernel

Captures dependence between observation events

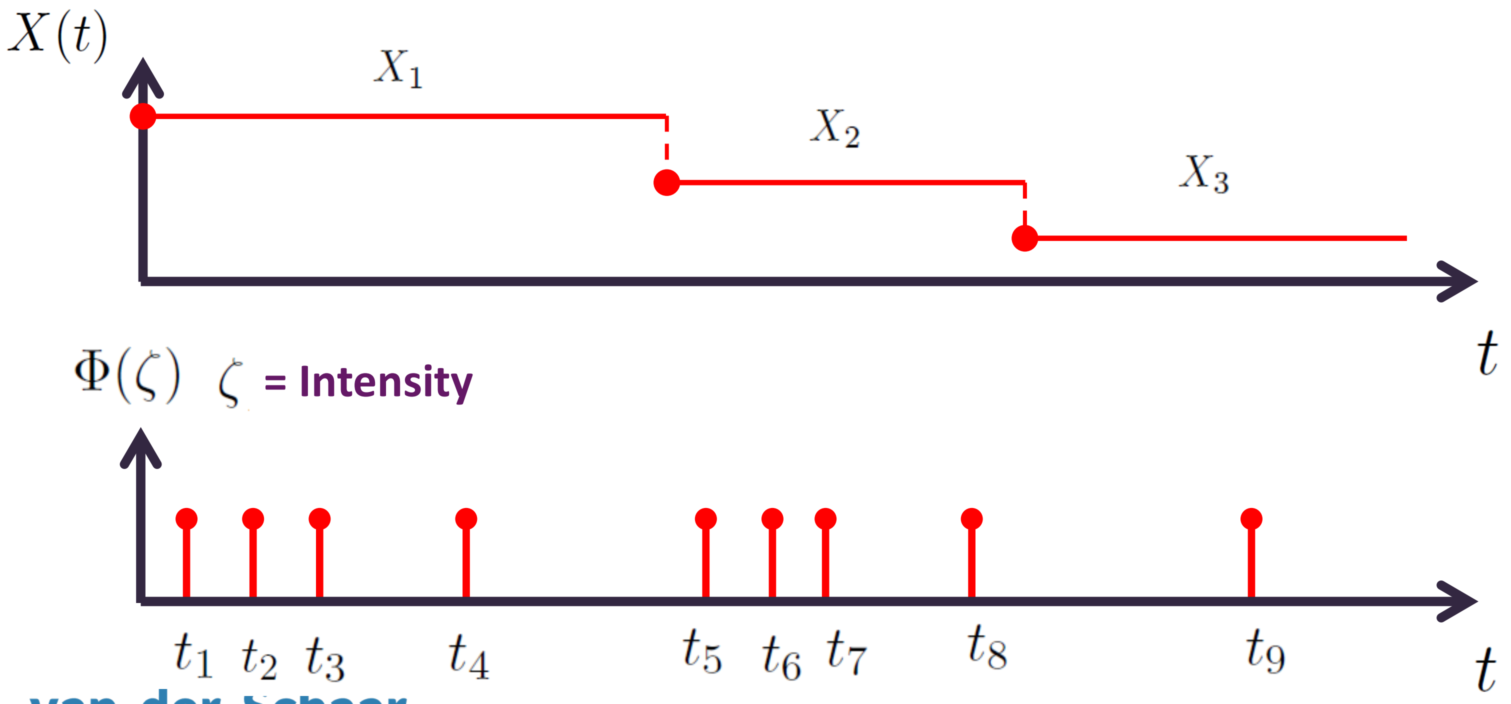
$$\lambda(t, X(t) = i) = \lambda_i^o + \alpha_i \sum_{\tau < t_m < t} e^{-\beta_i (t - t_m)}$$



Better inference using informative observations

Observation times are modeled as a **Hawkes process**

- **Continuous-time jump process (like Poisson)**
- **Jump intensities depend on state (unlike Poisson)**



Time-series: a multi-faceted problem

1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Uncertainty estimation
10) Missing data and informatively missing data
11) Synthetic data generation

- Reproducibility and visualization

Healthcare data: not easy to access

Strict regulations for data access

…the result of perfectly valid concerns regarding privacy

Lack of high-quality healthcare data: impedes ML research in healthcare!

De-identified data vs synthetic data

De-identified/anonymized data: real data with all personal identifiers removed/data fields scrambled

Synthetic data: data created from scratch, cannot be synced back to any individual (if modeled properly)

Requires ML/statistical modelling!

ICML 2021 Tutorial

Generating synthetic data to be used for machine learning modeling is itself a machine learning problem!

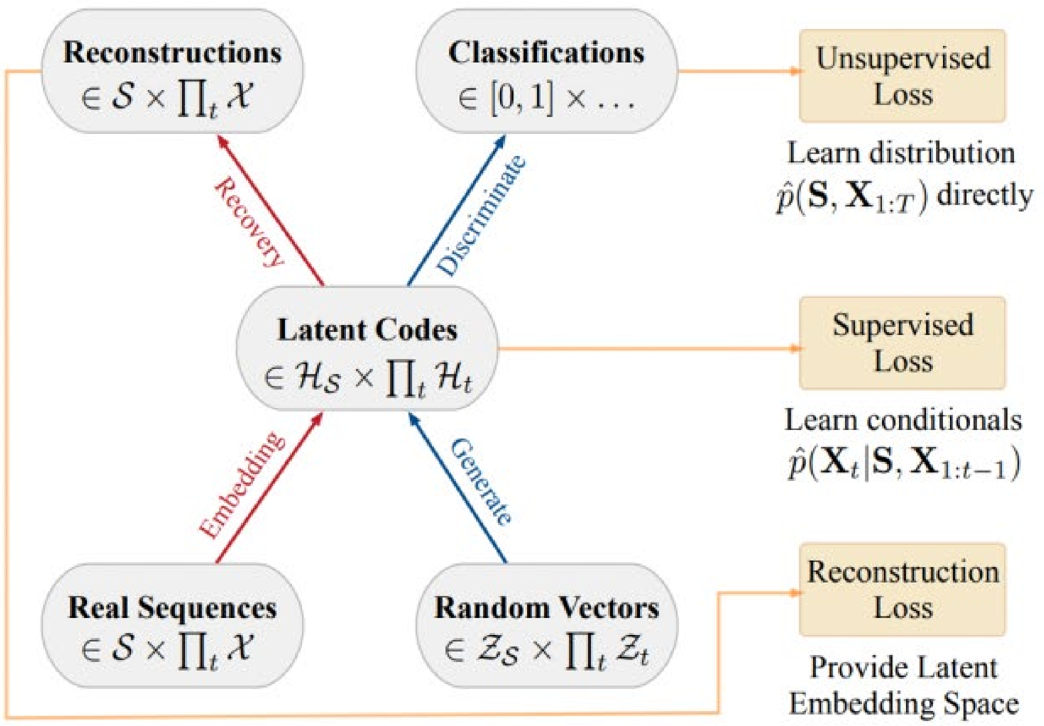
Time-series generation

Objective: To generate time-series data with preserving temporal dynamics

Key Example: Synthetic time-series healthcare data generation

Challenges: Capture the distributions of features within each time point as well as complex dynamics of those variables across time points

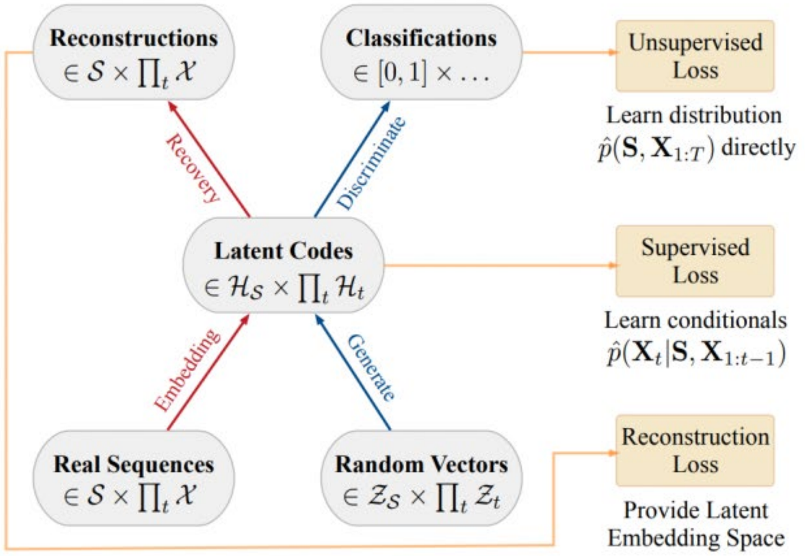
Time-series generative adversarial networks [Yoon, Jarrett, vdS, NeurIPS 2019]



Block diagram of component functions and objectives.

Generating time-series synthetic data

TimeGAN - intersection of multiple strands of research
- GAN-based methods for sequence generation
- autoregressive models for sequence prediction
- time-series representation learning.



Block diagram of component functions and objectives.

Important: Time-GAN handles mixed-data setting, where both static and time-series data can be generated at the same time

TimeGAN: some limitations

GAN-based models - powerful way to synthesize time-series data, but....

- difficult to train (especially for time-series data) – [Srivastava et al. (2017)]
- hard to evaluate quantitatively due to the absence of an explicitly computable likelihood function (only implicit likelihood modeling)
- vulnerable to training data memorization [Nagarajan et al. (2018)]

— a problem that would be exacerbated in the temporal setting

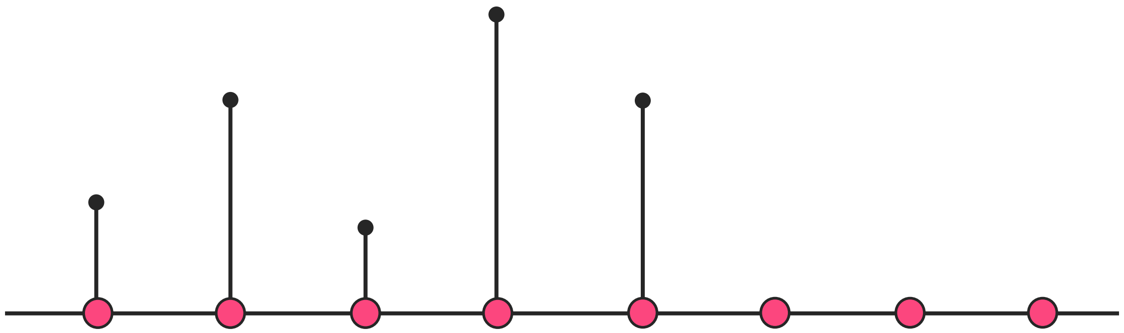
Generative Time-series Modeling with Fourier Flows

[Alaa, Chan, vdS, ICLR 2021]

Variable-length and variable-frequency sequences of vectors.

$$\boldsymbol{x} = [\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{T-1}],\ \boldsymbol{x}_t \in \mathcal{X},\ \forall\, 0 \leq t \leq T-1$$

$$x_{t,d}[r_d] \triangleq \begin{cases} x_{t,d}, & t \bmod r_d = 0, \\ *, & t \bmod r_d \neq 0. \end{cases}$$



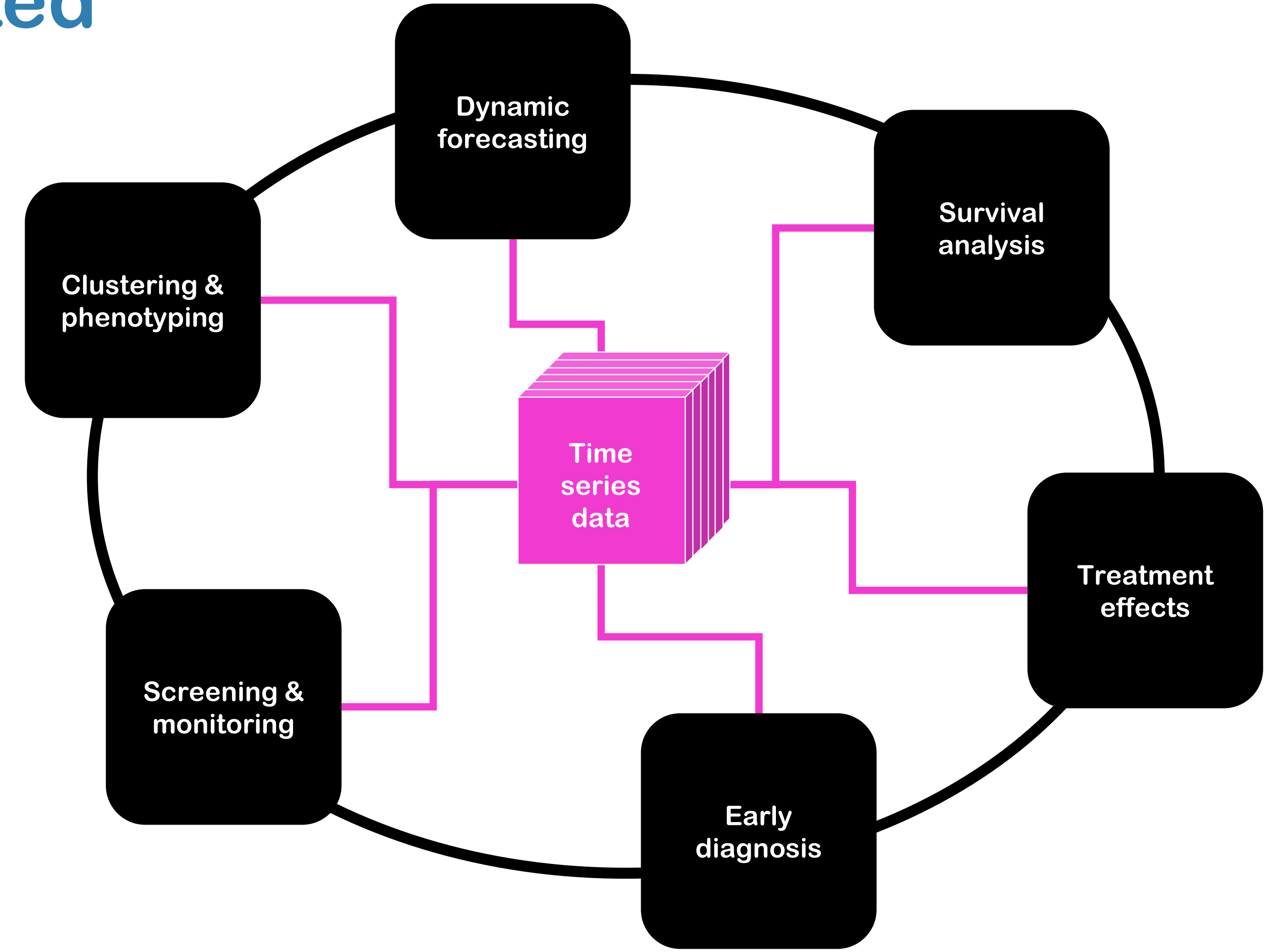
Goals:

Generative model: to enable sampling synthetic time series

Explicit likelihood model: easy to optimize the model, easy to evaluate the model

Conclusions

New Frontiers: Healthcare problems (and models) are interconnected



Patient-oriented | Profession-oriented

Individual

Bespoke medicine

- Risk scores
- Competing risks
- Screening and monitoring
- Diagnostic support
- Longitudinal disease trajectories
- Treatment effects

Empowering healthcare professionals

- Personalised ML assistants to support clinicians
- Interpretable, explainable, trustworthy
- Multi-disciplinary clinical contributions

At scale

Population health and public health policy

- Discover & disentangle public risks and risk factors
- Population risk assessment → personalized risk
- Data-driven guidelines, protocols, standards
- Cross-country learning and interventions

Systems, pathways and processes

- Improving healthcare pathways
- Integrating and curating data sources
- Integrating a multitude of analytics into delivery systems
- Cooperation, interaction and learning

Catalyze a revolution in healthcare

Want to learn more?

vanderschaar-lab.com/
→ Engagement sessions
→ Inspiration Exchange

Inspiration Exchange

Themed discussion sessions specifically for machine learning students (particularly masters, Ph.D., and post-docs).

We would like to:

– discuss machine learning models and techniques
– share ideas about how machine learning can revolutionize healthcare
– spark new projects and collaborations
– raise awareness about this unique and exciting area of machine learning.