How machine learning can drive the transformation of healthcare delivery

Mihaela van der Schaar

John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine, University of Cambridge // Director, Cambridge Center for AI in Medicine // Turing Faculty Fellow, The Alan Turing Institute

van_der_Schaar \ LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

mv472@cam.ac.uk

@MihaelaVDS

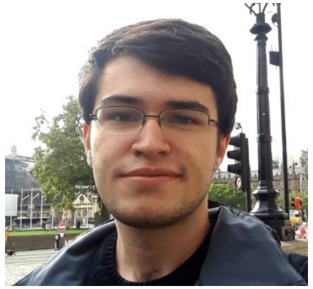
linkedin.com/in/mihaela-van-der-schaar/

Our research team

https://www.vanderschaar-lab.com/
→ Research Team


Alex Chan


Alicia Curth


Alihan Hüyük


Boris van Breugel


Dan Jarrett


Hao Sun


Jeroen Berrevoets


Jonathan Crabbé

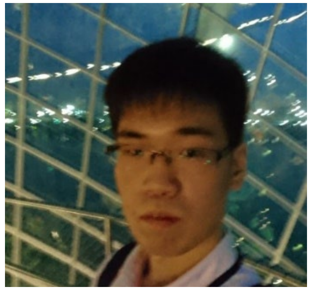

Kamile Stankeviciute


Krzysztof Kacprzyk


Nabeel Seedat


Sam Holt

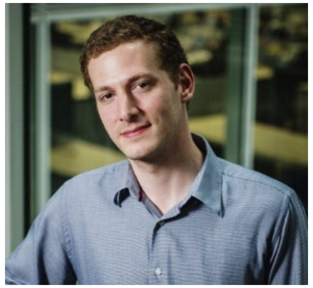

Ioana Bica


Yuchao Qin

Zhaozhi Qian


Bogdan Cebere


Evgeny Saveliev


Nick Maxfield

"If it were not for the great variability between individuals, medicine might as well be a science, not an art"

Sir William Osler (1892)

Variabilities between individuals	Lead to
- Different genetic background	- Different risks (and different risks over time)
- Different environmental exposures	- Variation in symptoms
- Different life-styles	- Different health and disease trajectories
- Different histories and interventions	- Different responses to treatment
etc.	etc.

The "art" of medicine has been to make judgements on the basis of this information to manage patients

What machine learning can do is turn this art into a science!

Machine learning can transform 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

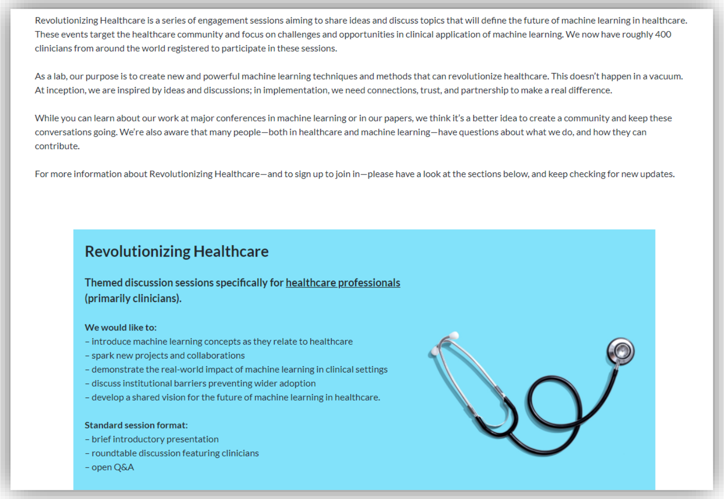
The "augmented" clinician, researcher, patient

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.

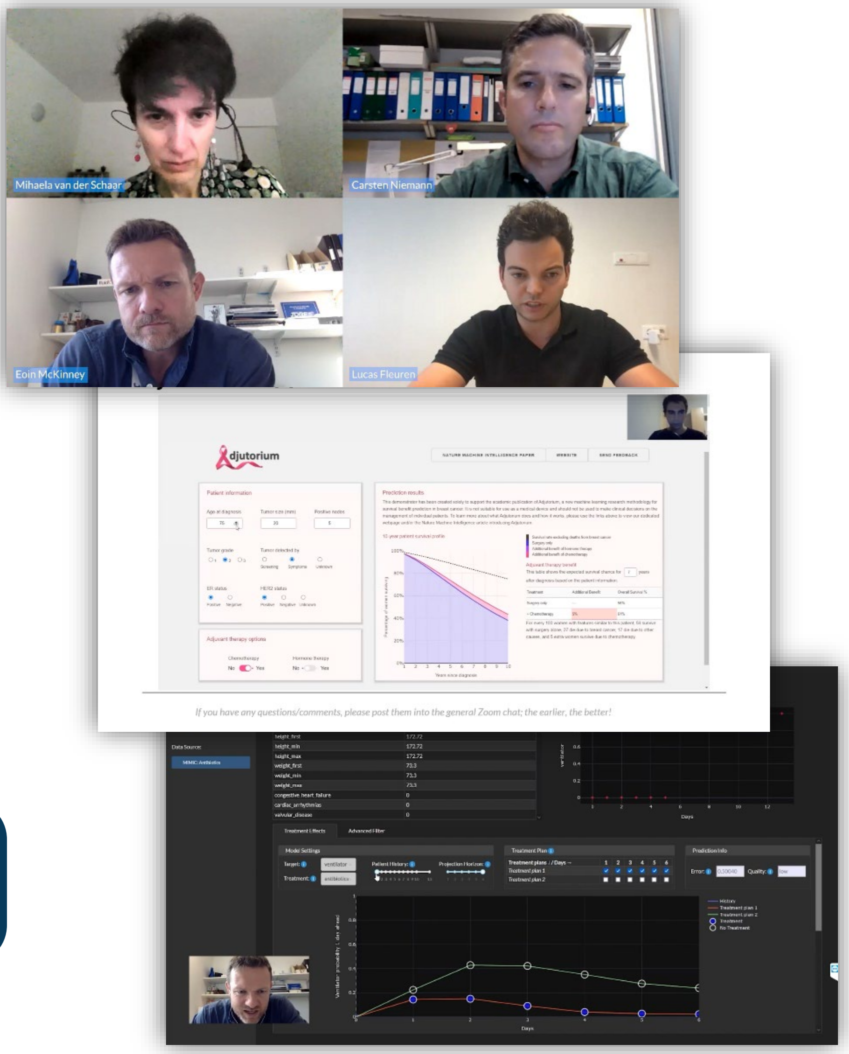
For more information about Revolutionizing Healthcare—and to sign up to join in—please have a look at the sections below, and keep checking for new updates.

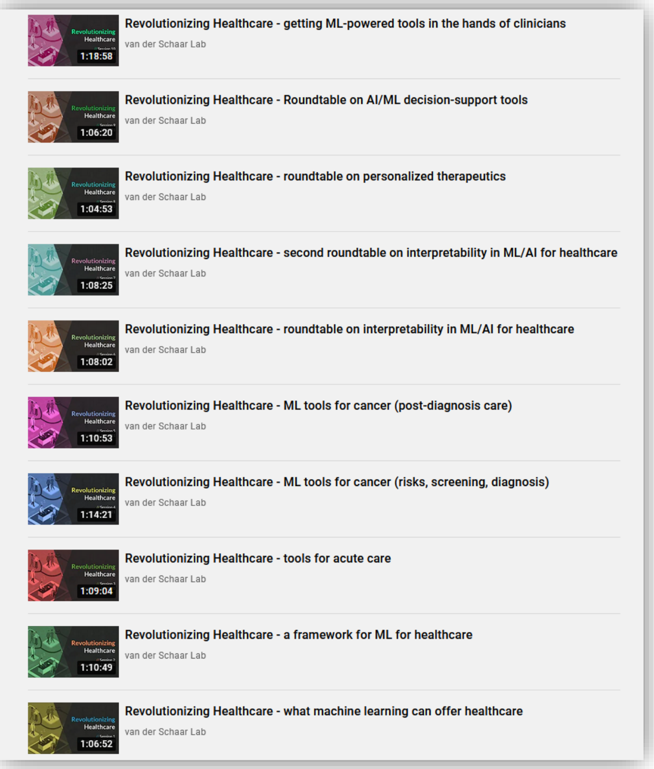
Revolutionizing Healthcare

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

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.

Overview

A healthcare ecosystem powered by Machine Learning (ML)

Time-series in healthcare

How can ML transform healthcare delivery?

Getting the right care to the right patients

Achieving the best outcomes at the lowest cost

This is possible through
- personalization (for patients, clinicians, and healthcare entities)
- learning at scale;
- identifying better ways of working and inefficiencies;
- providing concrete policies for improvement
- optimal allocation of resources (over time)

	Patient-oriented	Profession-oriented
Narrow	**Bespoke medicine**	**Empowering clinicians**
Broad	**Population health and public health policy**	**Systems, pathways and processes**

Why is now the right time?

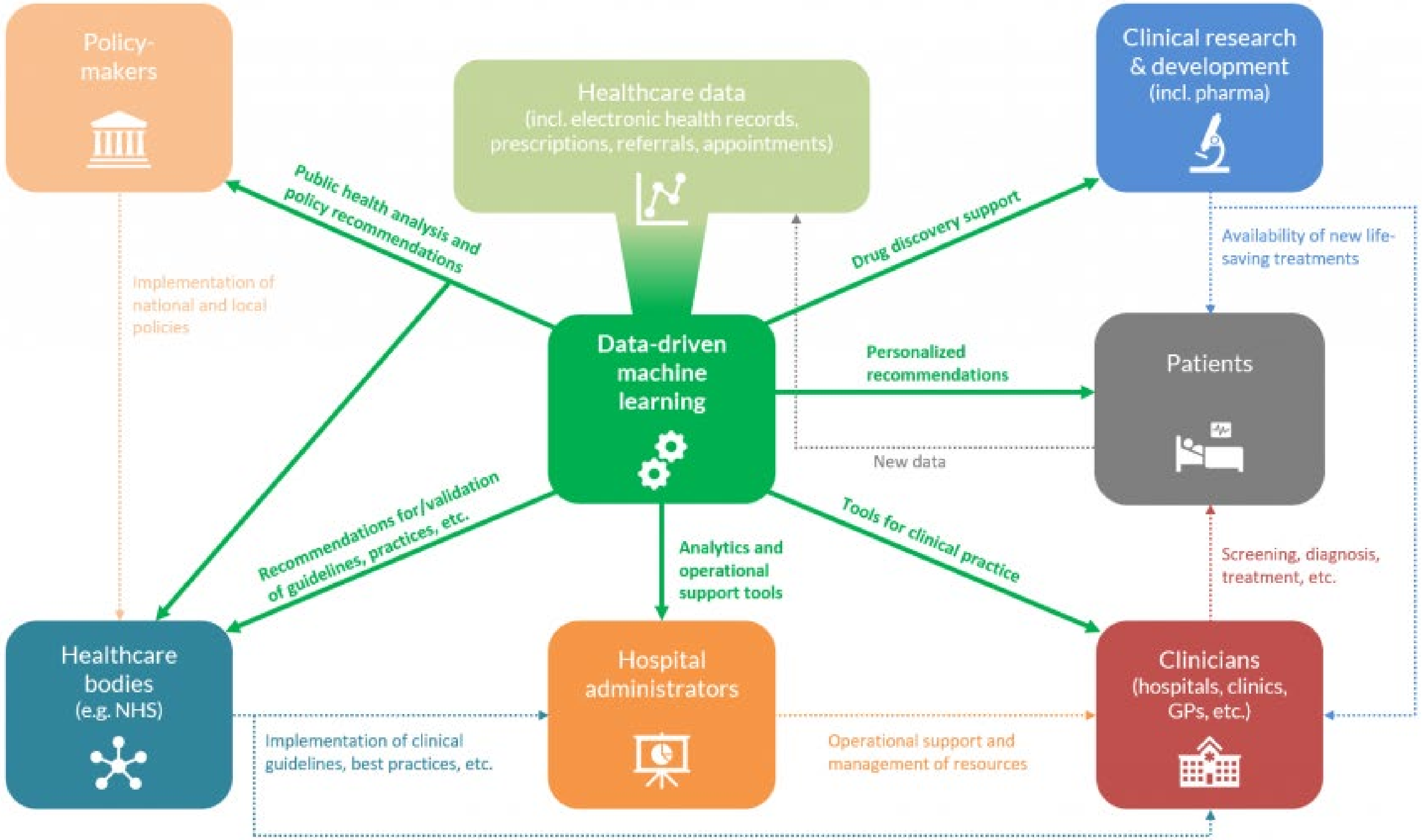
Machine learning methods have come of age and are ready to be used

Unprecedented access to diverse sources of valuable info, including:

- clinical notes
- electronic health records
- clinical registries
- prescription info
- appointment info
- wearables

The COVID-19 pandemic has accelerated digital info collection

A healthcare ecosystem powered by Machine Learning



van_der_Schaar \ LAB

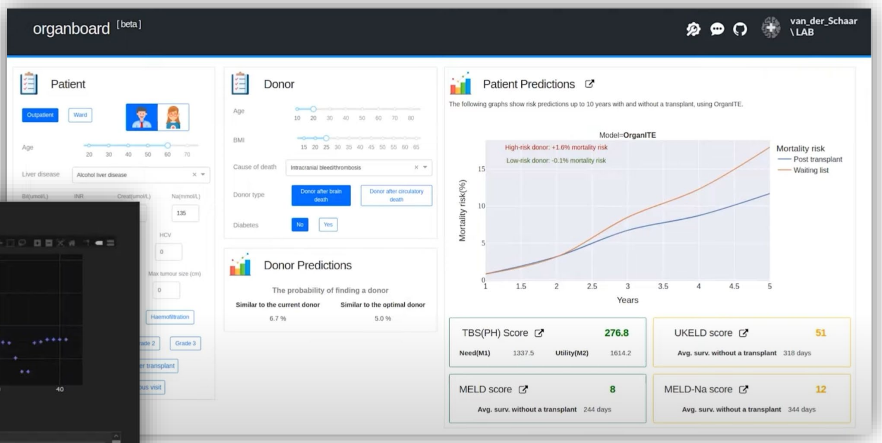
vanderschaar-lab.com

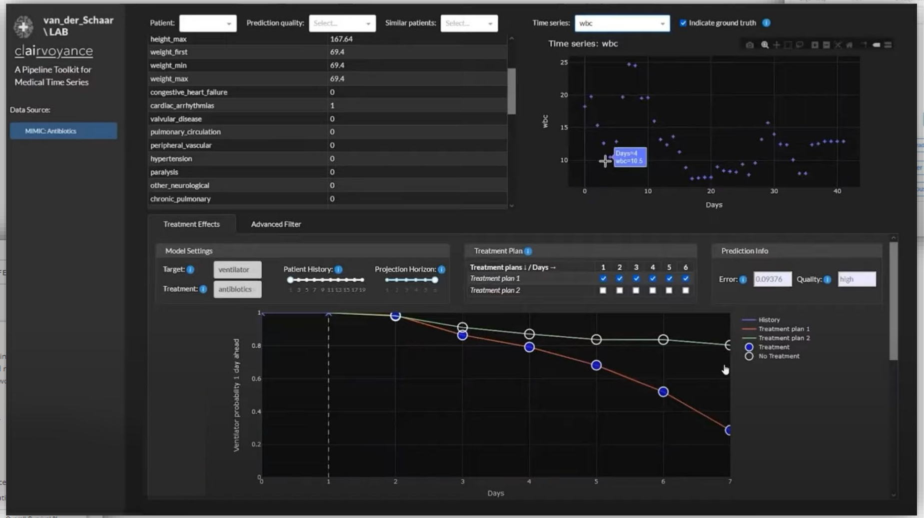
UNIVERSITY OF CAMBRIDGE

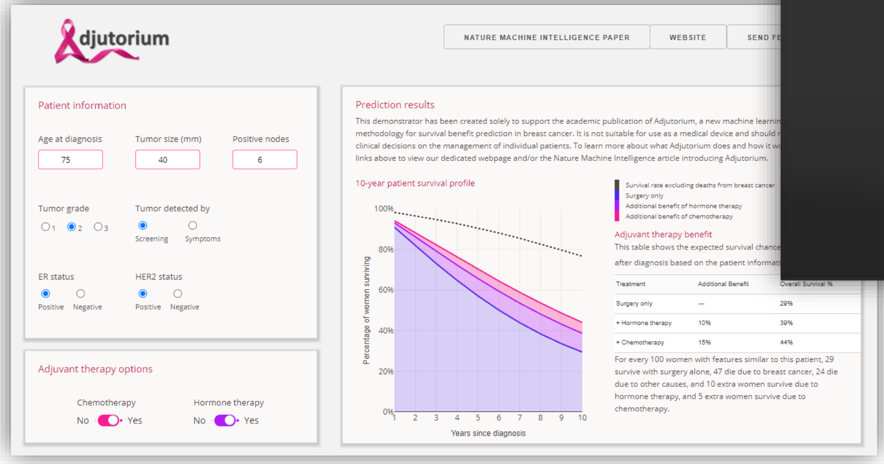
Personalized, comprehensive care with machine learning

Machine learning can offer personalized

- forecasts
- screening
- monitoring
- treatment plans
- recommendations







e.g.

- Adjutorium
- AutoPrognosis
- Clairvoyance

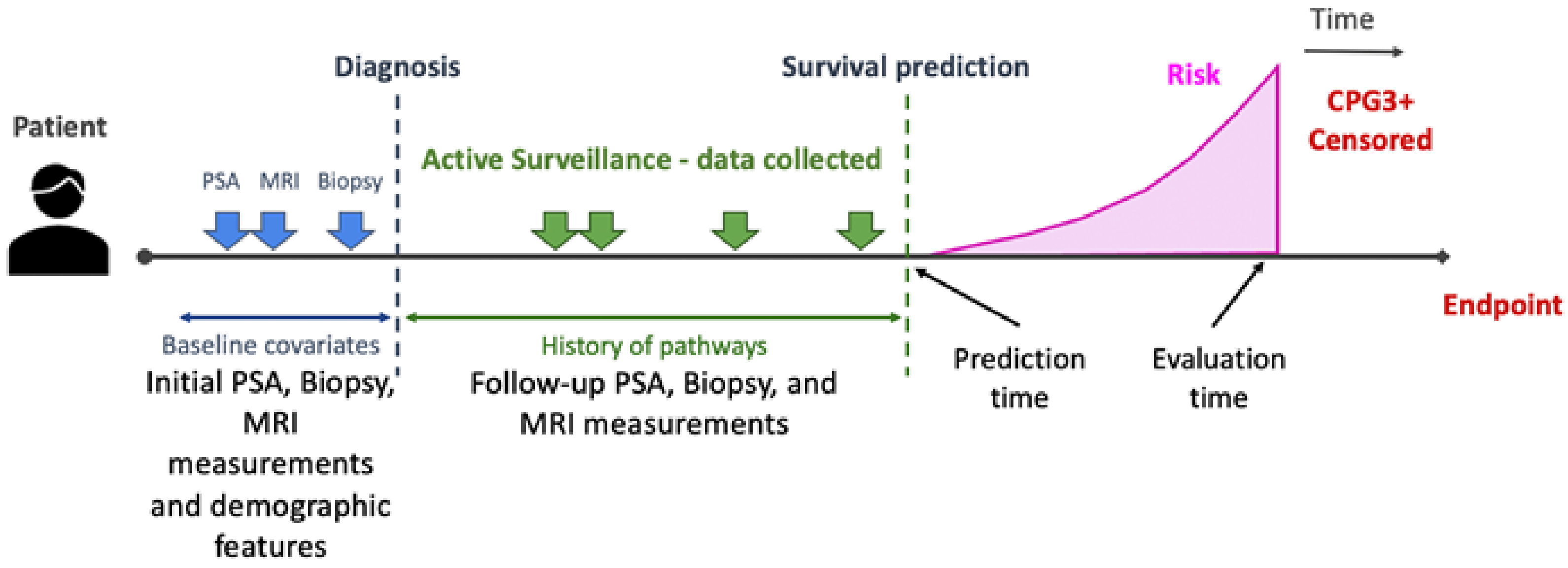
Actionable intelligence across the patient pathway

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

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

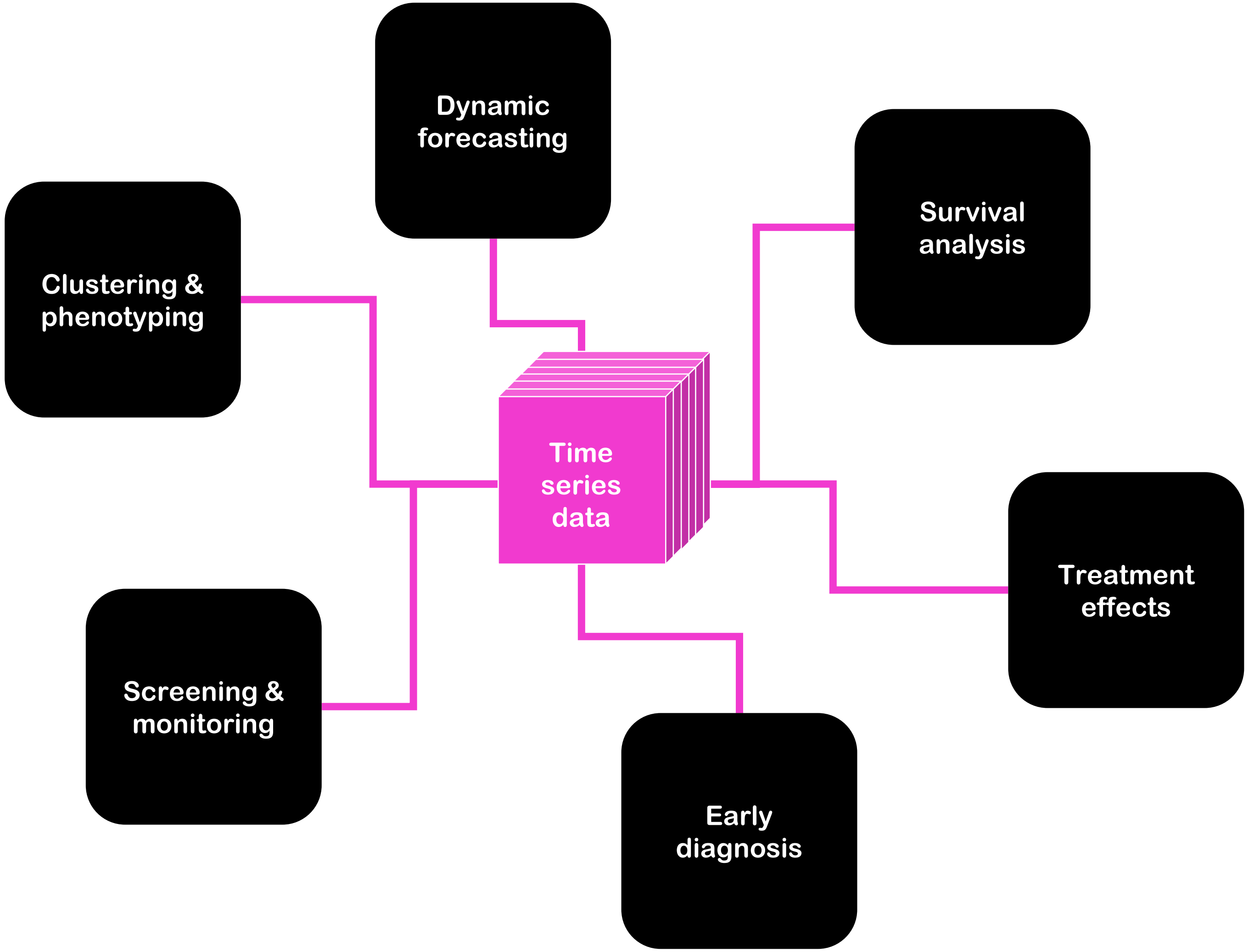
Multiple venues/areas:
- In-patient/out-patient
- At home

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



UNIVERSITY OF CAMBRIDGE

Time-series: a multi-faceted problem



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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) 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

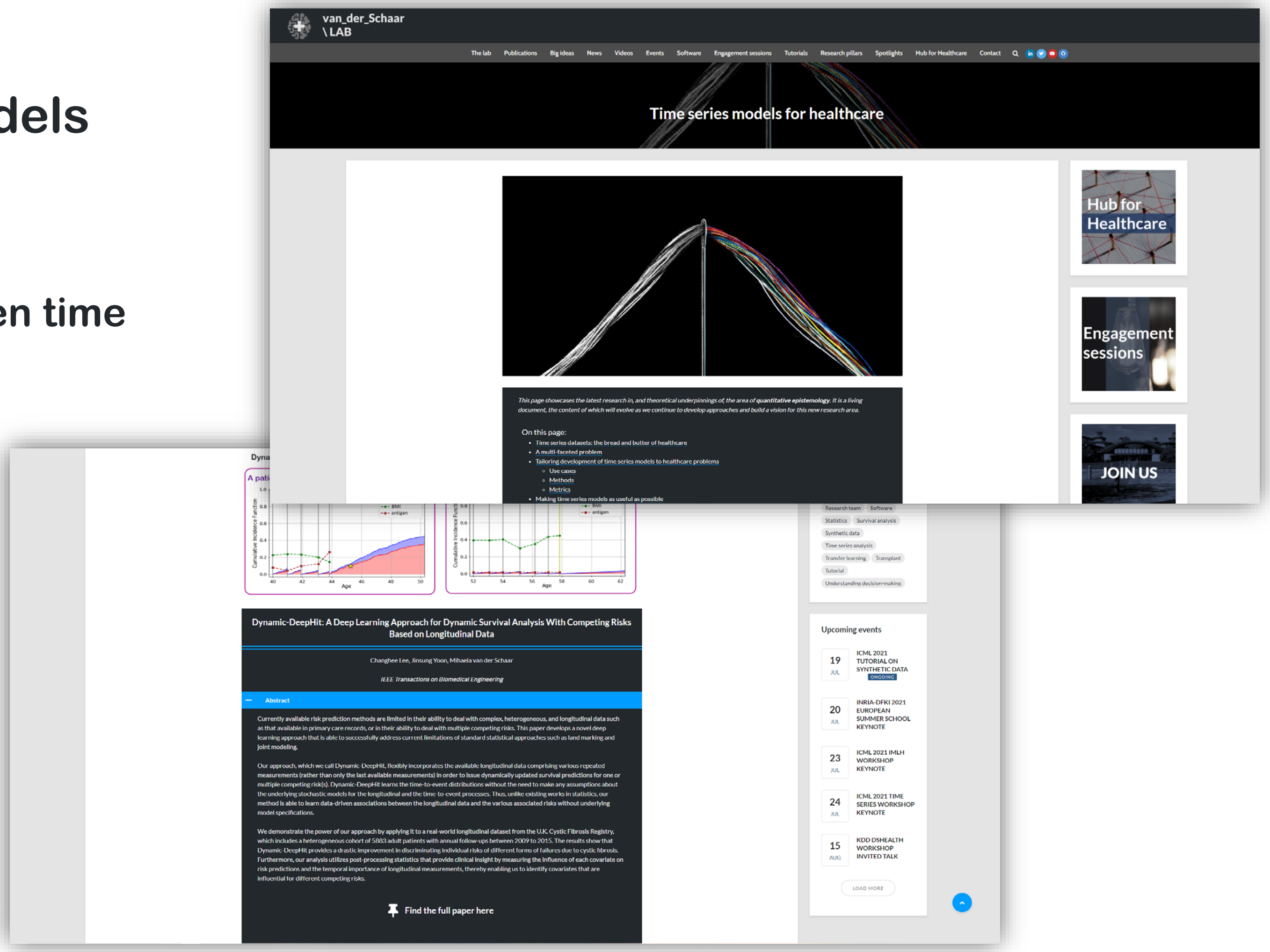
More information and updates

Overview of our work on time series models

Introduction for a variety of audiences

Explores the “cross-sectional” interactions between time series and many other areas of research

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



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

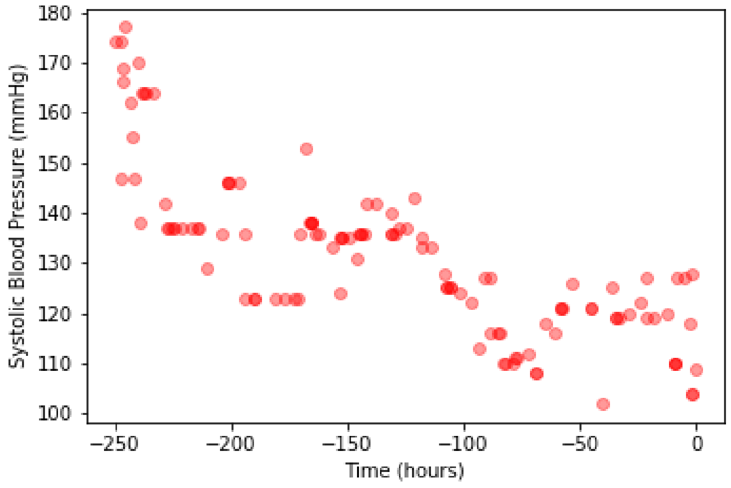
Time-series: a multi-faceted problem

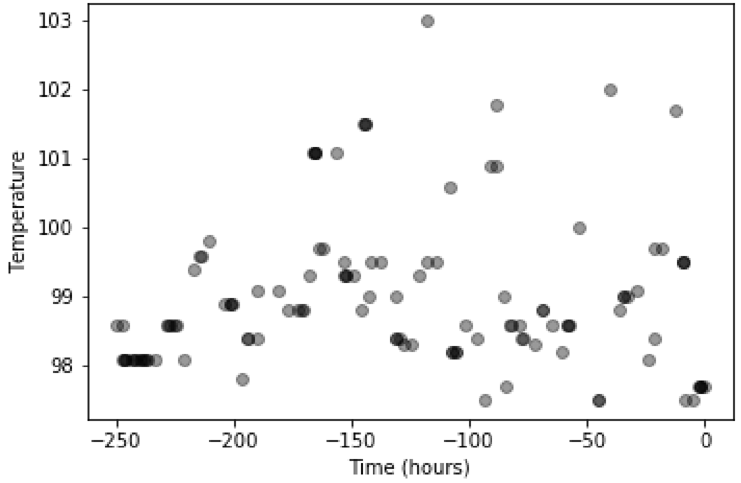
1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing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

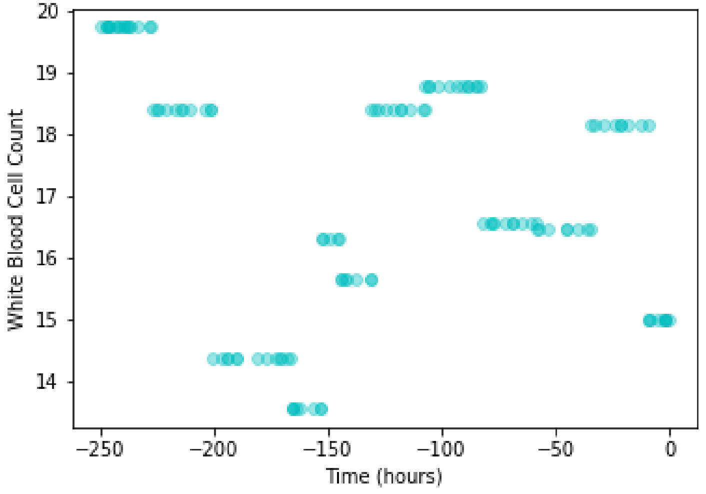
- Reproducibility and visualization

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)**







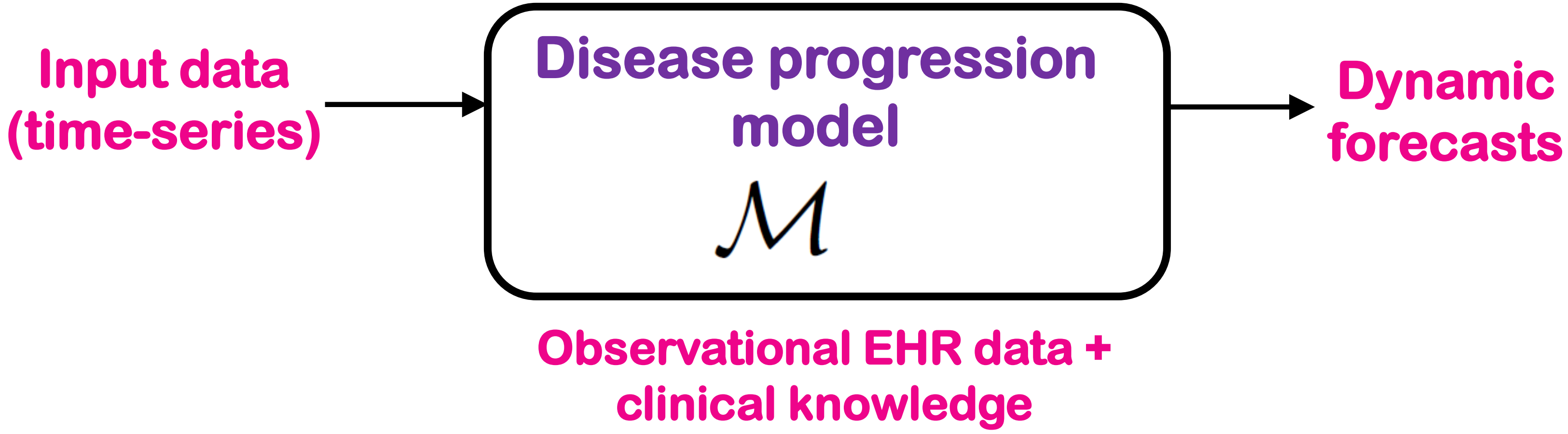
van_der_Schaar \LAB

vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Time-series analysis and dynamic forecasting

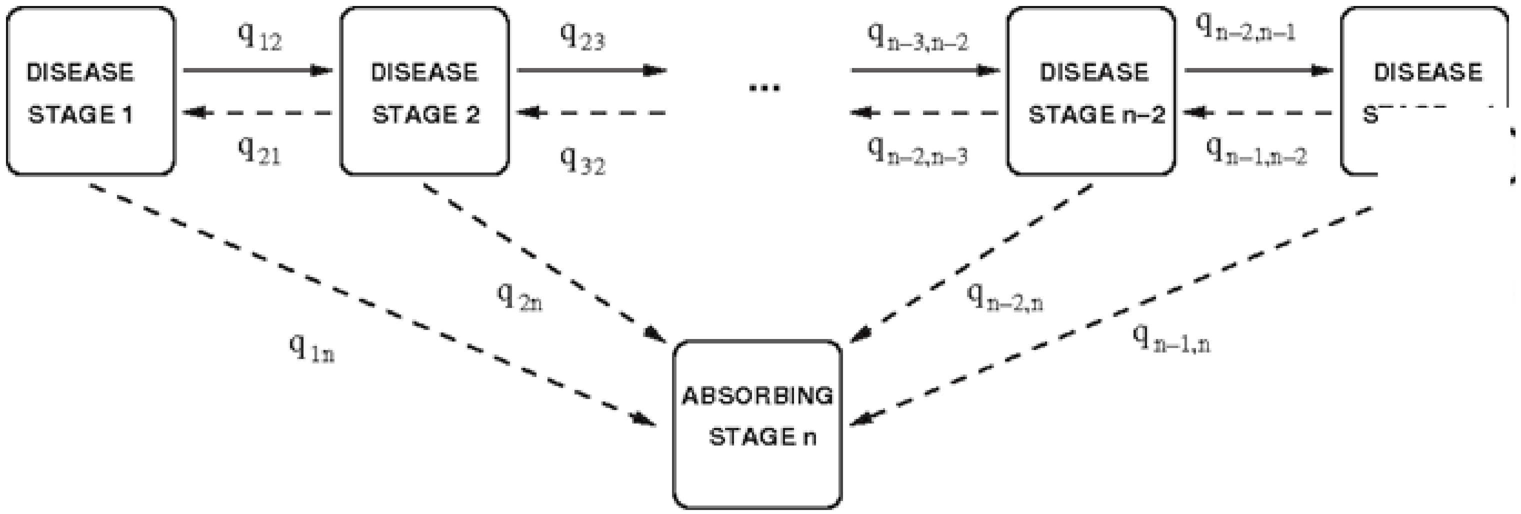
- **Build** disease progression **models**
  - **Understand** and **model** carefully the available data!
- **Learn** the **model parameters** from available EHR data (Training time)
- Issue **dynamic forecasts** for the patient at hand (Test time/Run-time)
- **Unravel new understanding** of disease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



Current disease progression models: formalisms

Markov Models

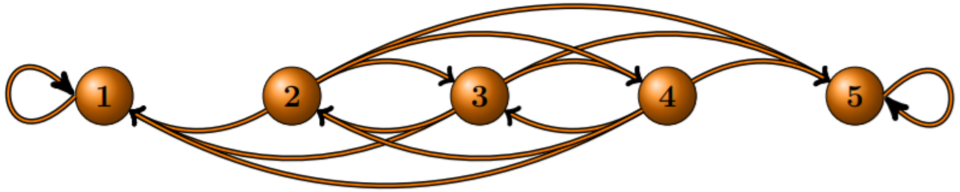
$$P(Z_{n+1} \mid \mathcal{H}_{t_n}) = P(Z_{n+1} \mid Z_n)$$



Disadvantages

- Observable models
- One disease at a time
- “Average” patient

Population-level representation of disease states

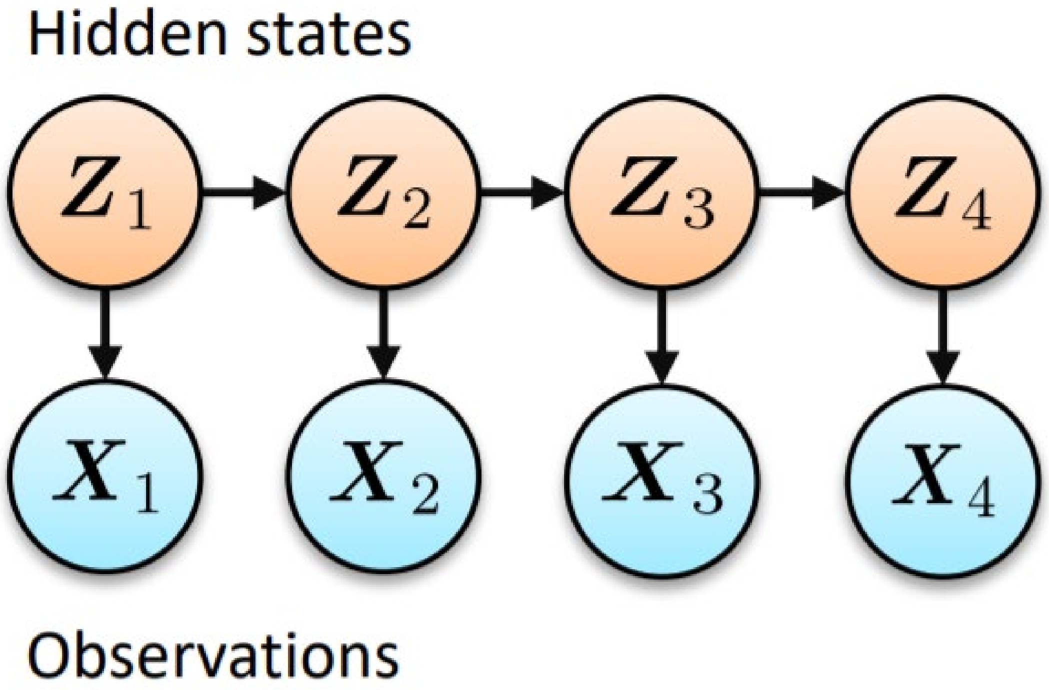


Current disease progression models: formalisms

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

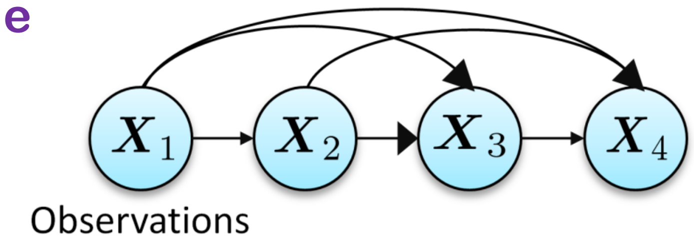
One size fits all!

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

Deep learning models?

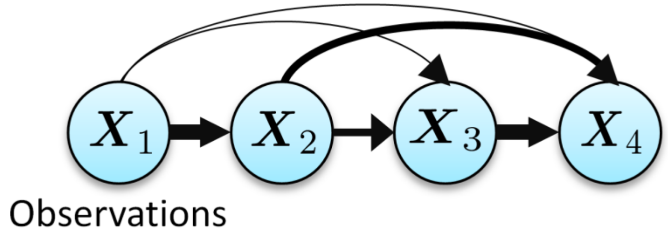
Deep learning models?

Observable models



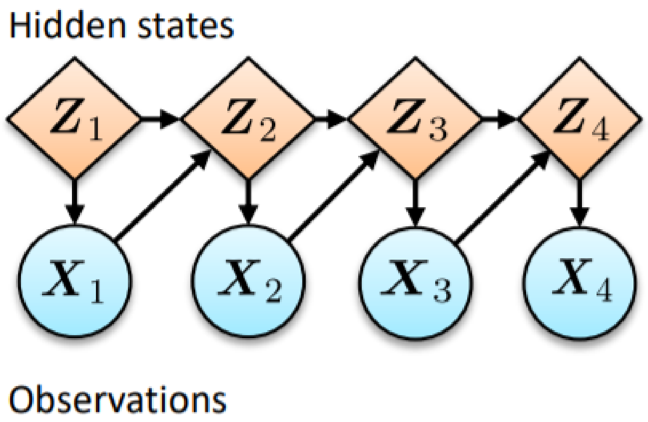
No latent structure

RETAIN



Interpretable predictions, Uninterpretable latent structure

RNN



Uninterpretable predictions, Uninterpretable latent structure

Two central goals of longitudinal models

Goal A: Accurately forecasting individual-level disease trajectories

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

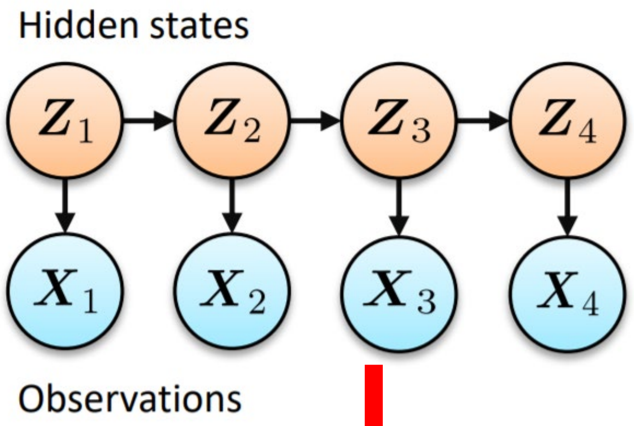
Goal B: Understanding disease progression mechanisms.

- Underlying latent structure of disease evolution
- Causal pathways and comorbidity networks
- Patients' subgroup analysis
- Refined phenotypes

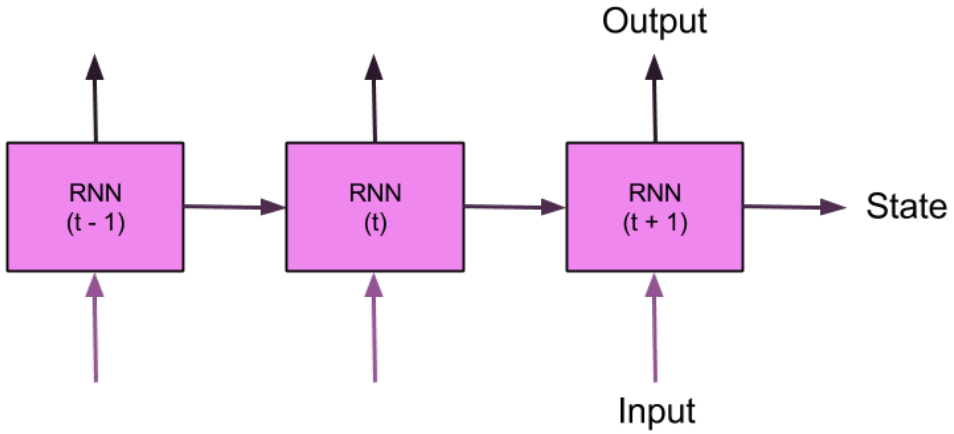
Attentive state space models [Alaa & vdS, 2018, NeurIPS 2019]

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



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

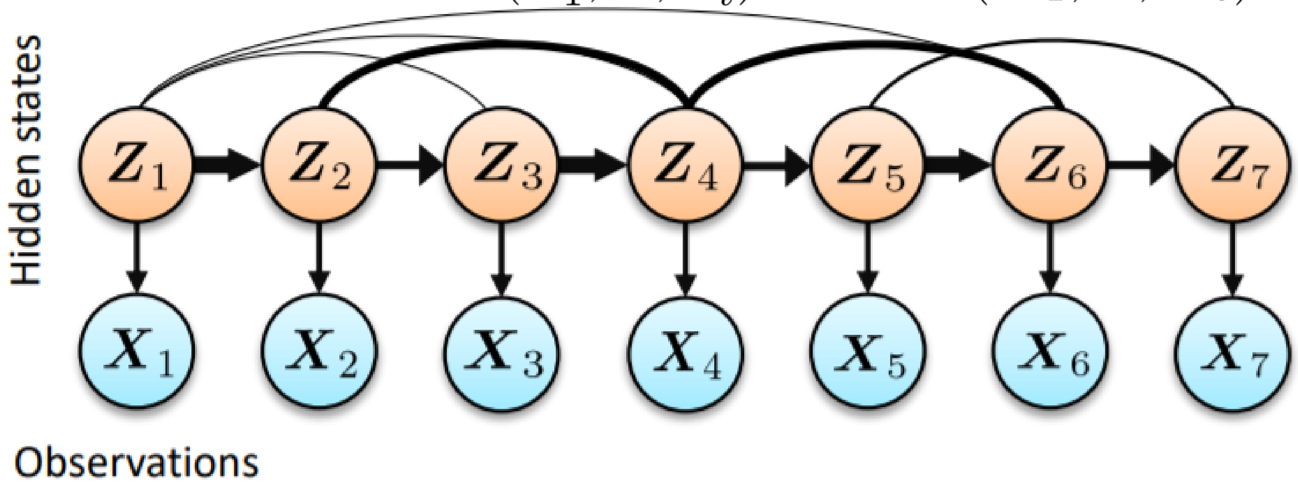
Emission

Transition

Going beyond Markov

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

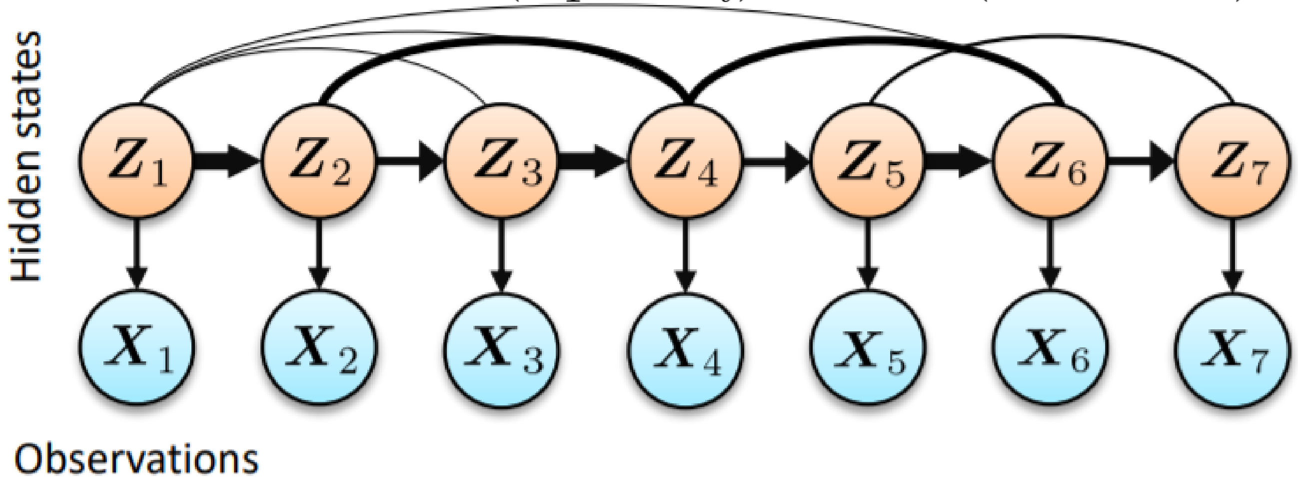
$$\underbrace{(\boldsymbol{\alpha}_1^t, \ldots, \boldsymbol{\alpha}_t^t)}_{\text{Attention weights}} = \mathbf{RNN}(\underbrace{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_t}_{\text{Patient context}})$$



Overcomes shortcoming of Markov Models

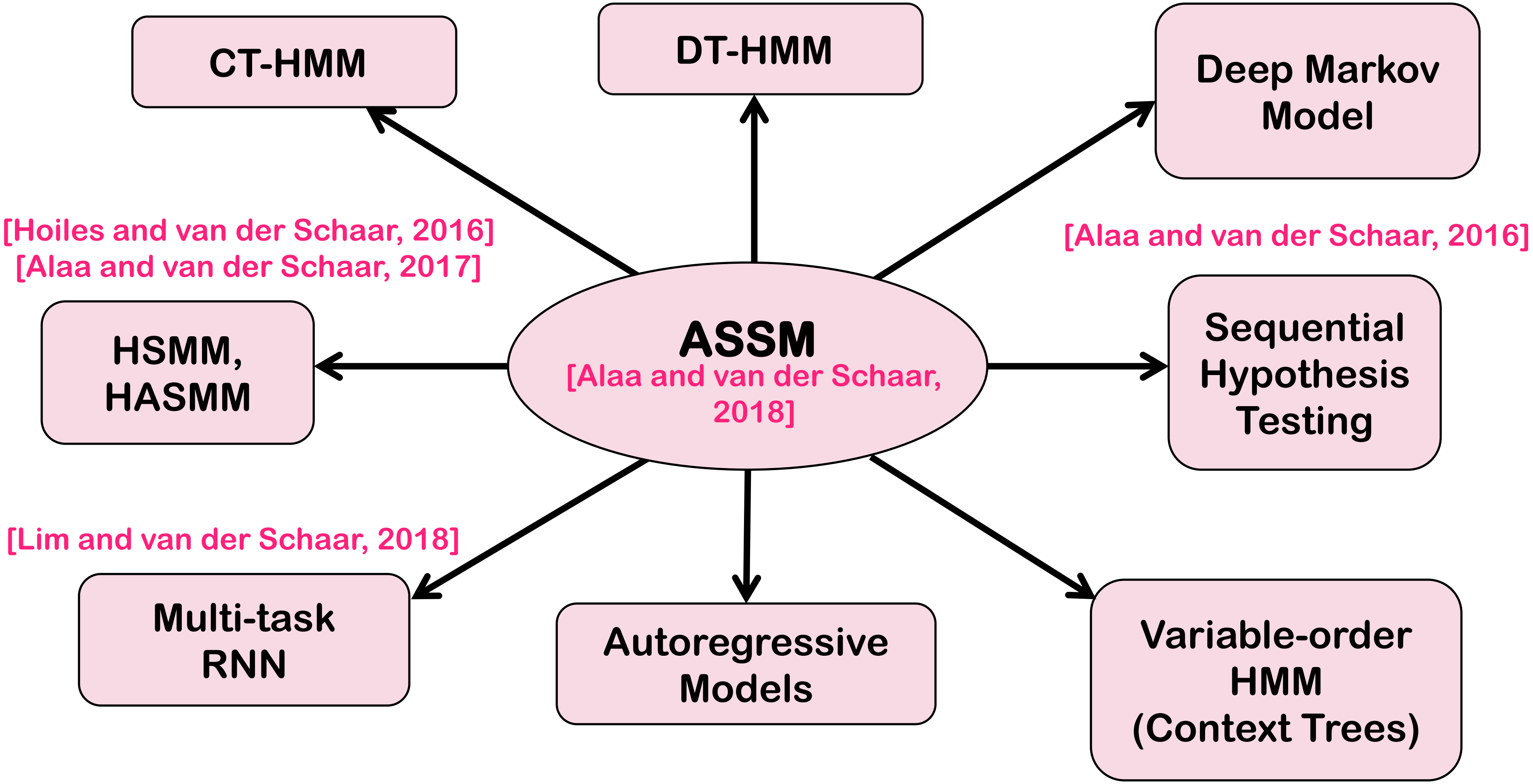
Attention weights create a "soft" version of a non-stationary, variable-order Markov model where underlying dynamics of a patient change over time based on an individual's clinical context!

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



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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) 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

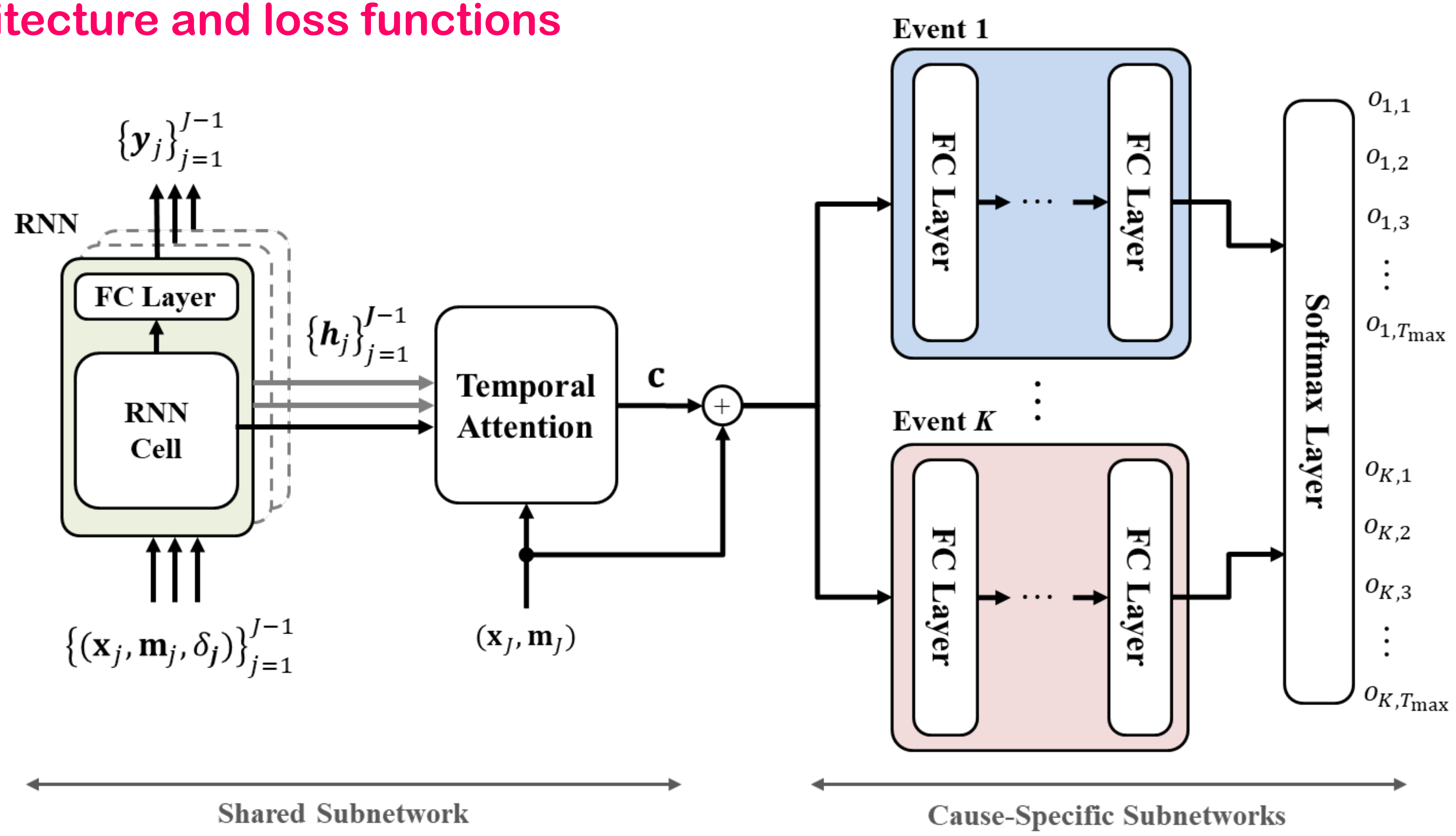
$$\hat{F}_k(\tau|\mathcal{X}^*) \stackrel{\text{def}}{=} P(T \leq \tau, E = k|\underline{\mathcal{X}^*}, \underline{T > t^*_{M^*}})$$

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]

Network architecture and loss functions



Dynamic-DeepHit [Lee & vdS, TBME 2019]

Network architecture and loss functions

Loss functions:

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

Step-ahead Prediction loss

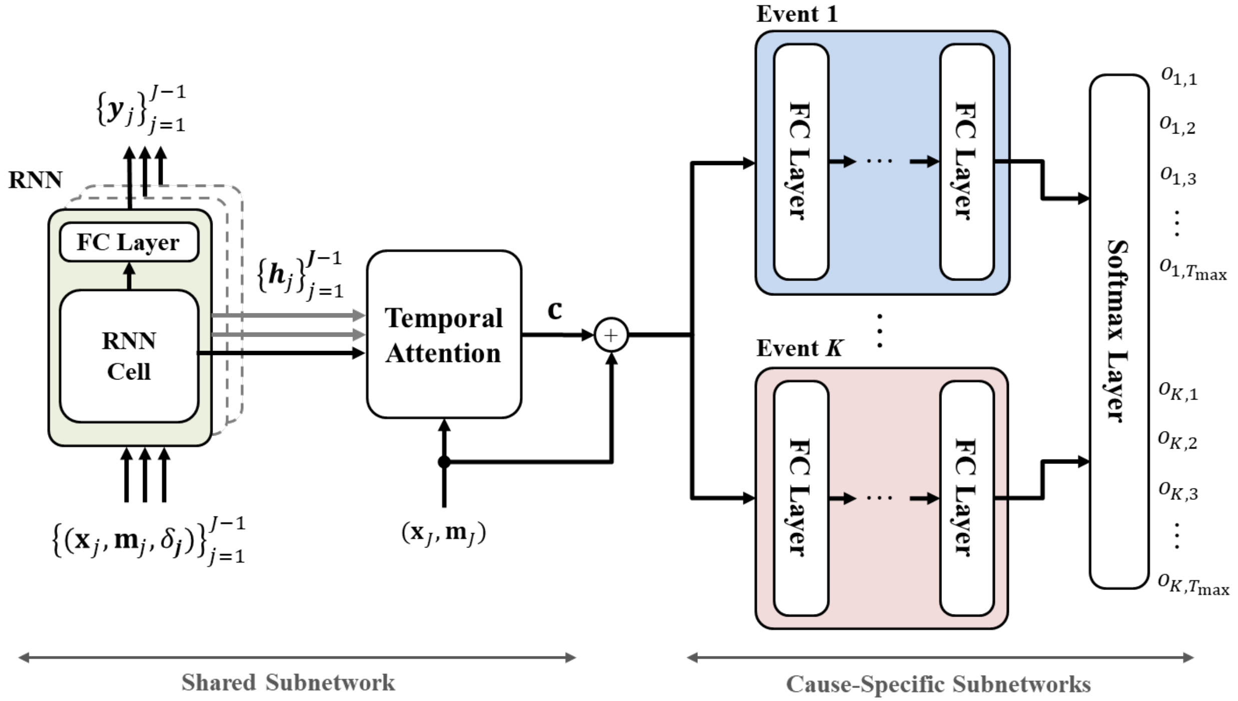
Log-likelihood of joint TTE distribution

Ranking loss

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

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

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



Time-series: a multi-faceted problem

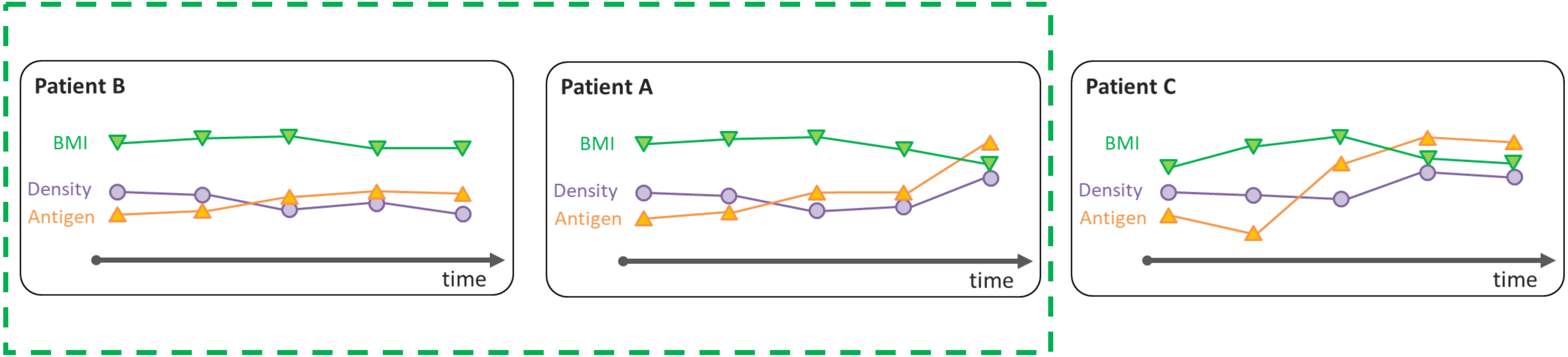
1) Dynamic forecasting
2) Time-to-event and survival analysis
3) Clustering and phenotyping
4) Screening and monitoring
5) Early diagnosis
6) Treatment effects
7) AutoML
8) Interpretability
9) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) 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

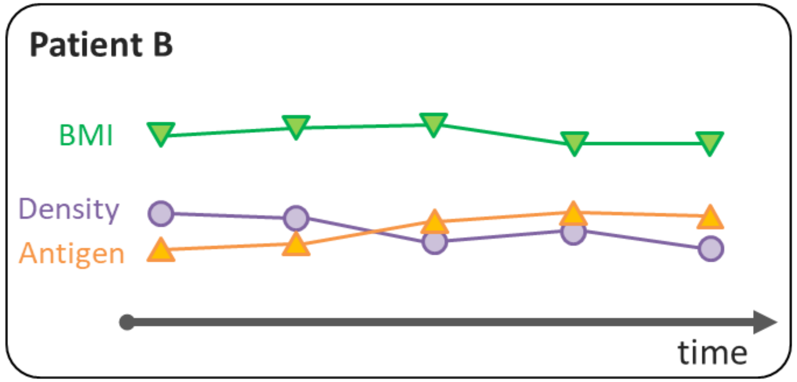
- N. S. Madiraju et al., 2018
- Q. Ma et al., 2019

Motivation: How should we group patients?

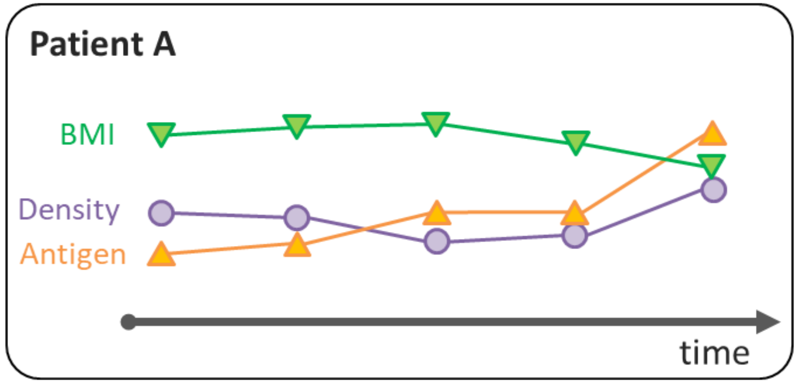
Example of 3 patients diagnosed with breast cancer (BC)

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

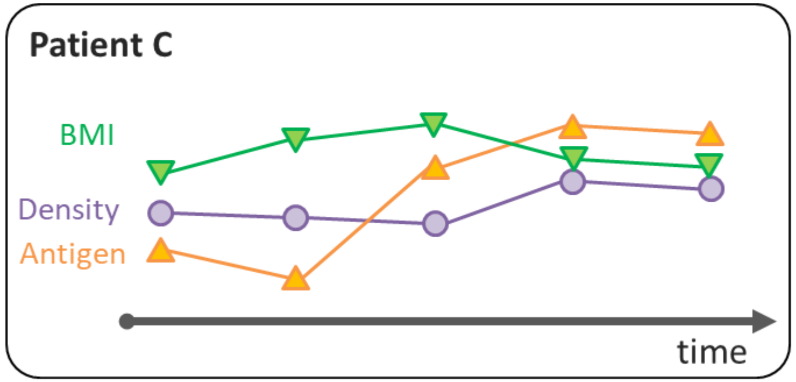
no adverse outcomes



BC-related Death



BC-related Death

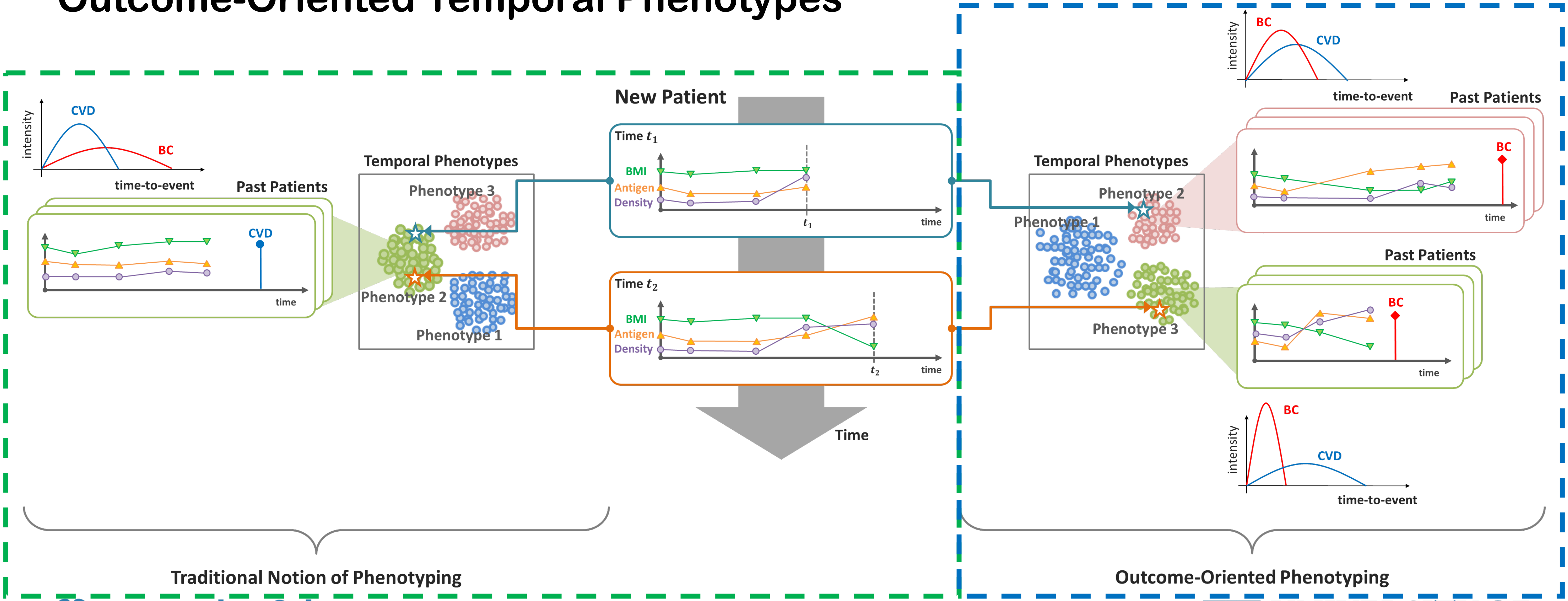


<u>our notion</u> of clustering

Key idea: similarity in future outcomes

Outcome-Oriented Temporal Phenotyping [Lee & vdS, ICML 2020] [Lee, Rashbass, vdS, TBME 2021]

Outcome-Oriented Temporal Phenotypes



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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

- Reproducibility and visualization

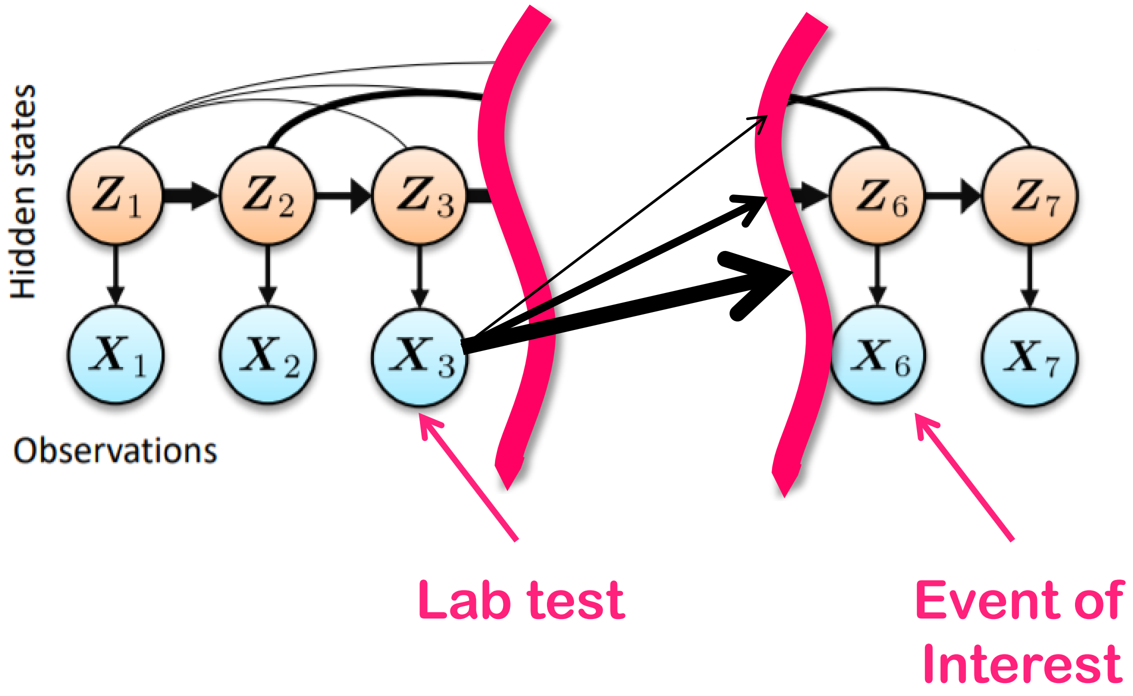
Personalized screening/monitoring

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

- Deep Sensing [Yoon, Zame, vdS, ICLR 2018]
- Disease Atlas [Lim, vdS, ML4HC 2018]

Which Modality of Screening?

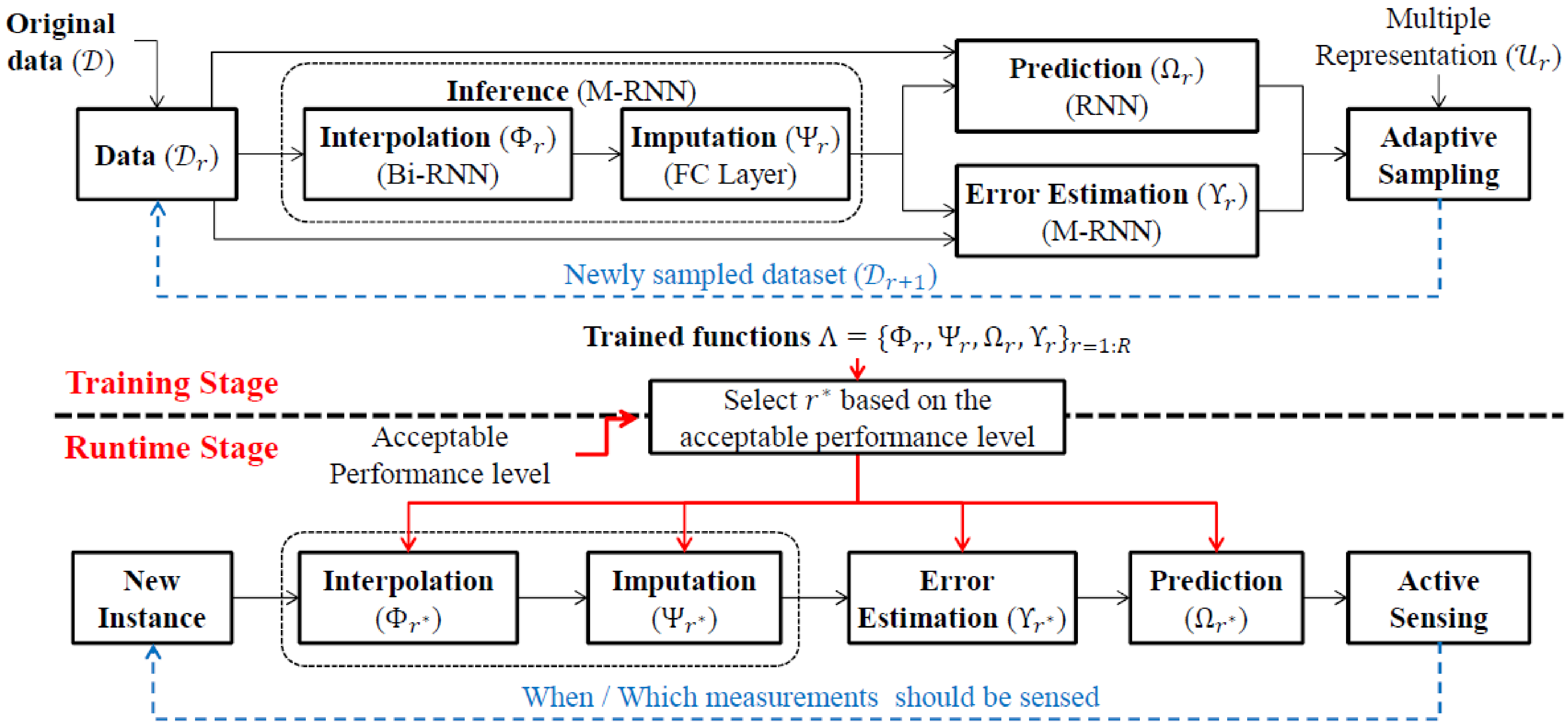
- [Alaa, Moon, Hsu, vdS, TMM 2016]



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

- Ideas:
  - A neural network must learn – at training time – how to issue predictions at various cost-performance points.
  - To do this, it creates multiple representations at various performance levels associated with different measurement rates (costs).
  - Each representation is learned and constructed recursively and adaptively learned by deliberately introducing missing data

Deep sensing architecture

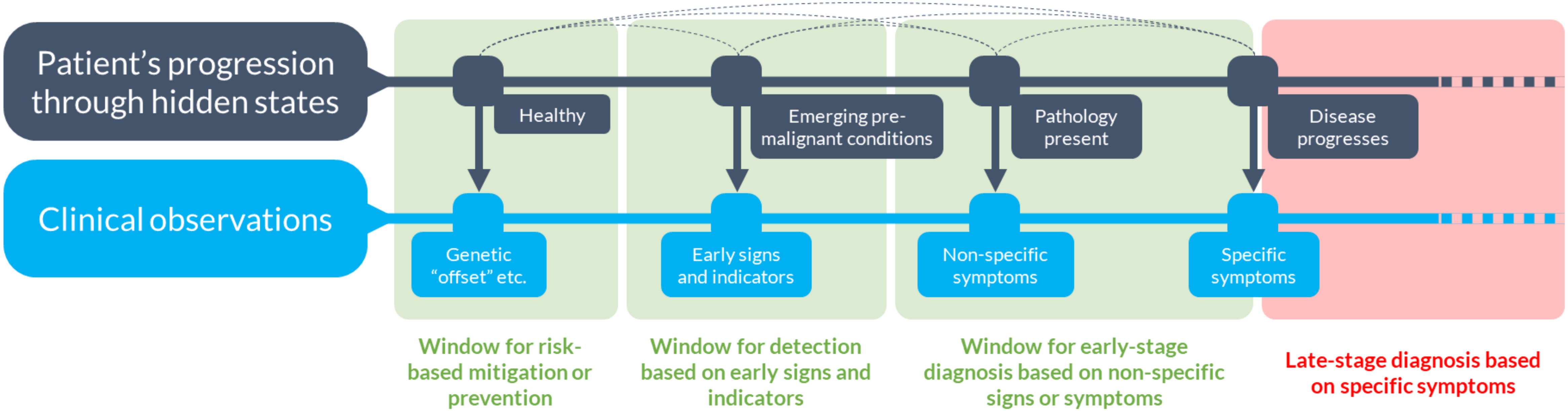


Time-series: a multi-faceted problem

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

- Reproducibility and visualization

Why is Early Diagnosis and Detection (ED&D) hard?



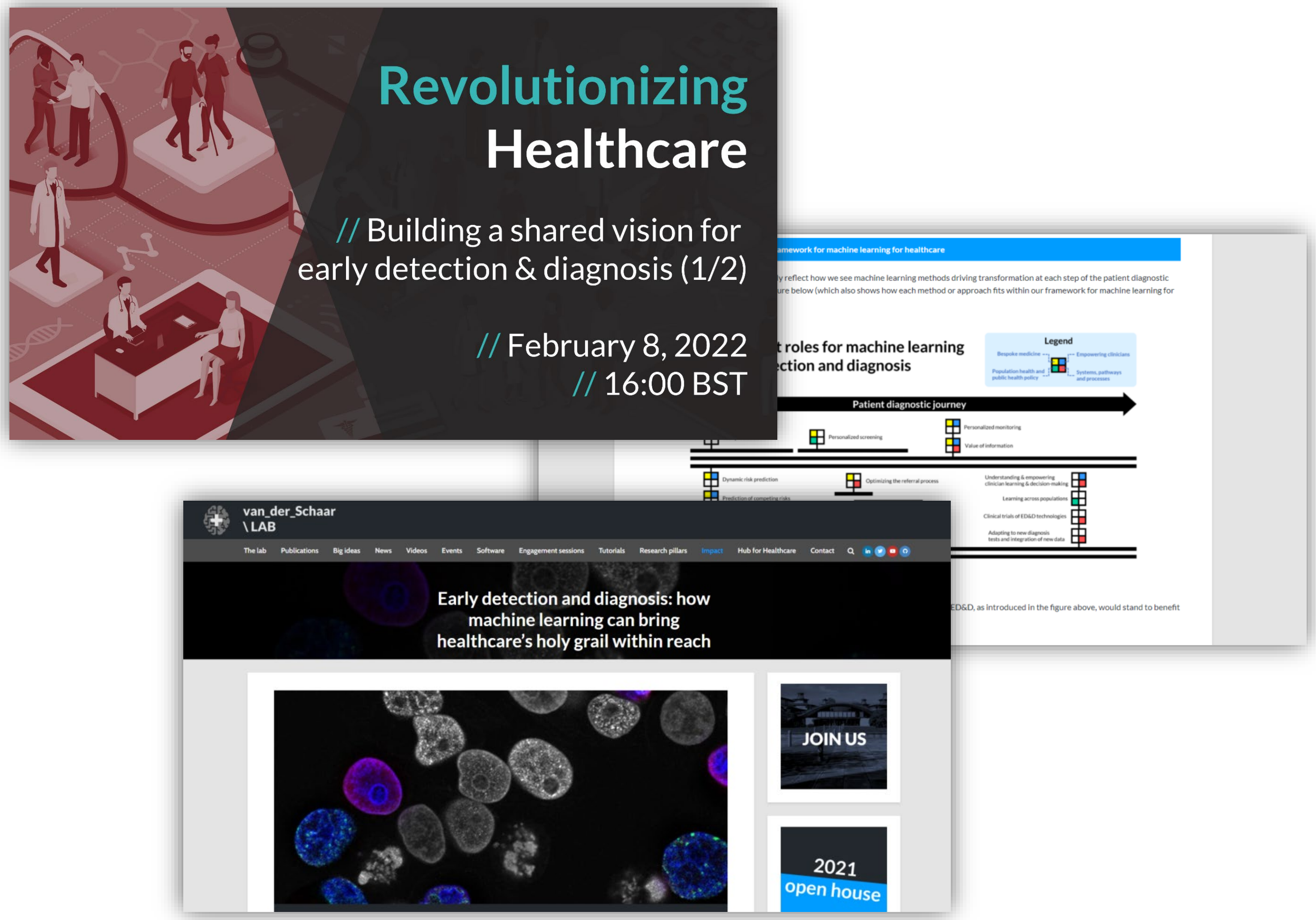
Revolutionizing Healthcare: roundtable on ED&D

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



van_der_Schaar \LAB

vanderschaar-lab.com

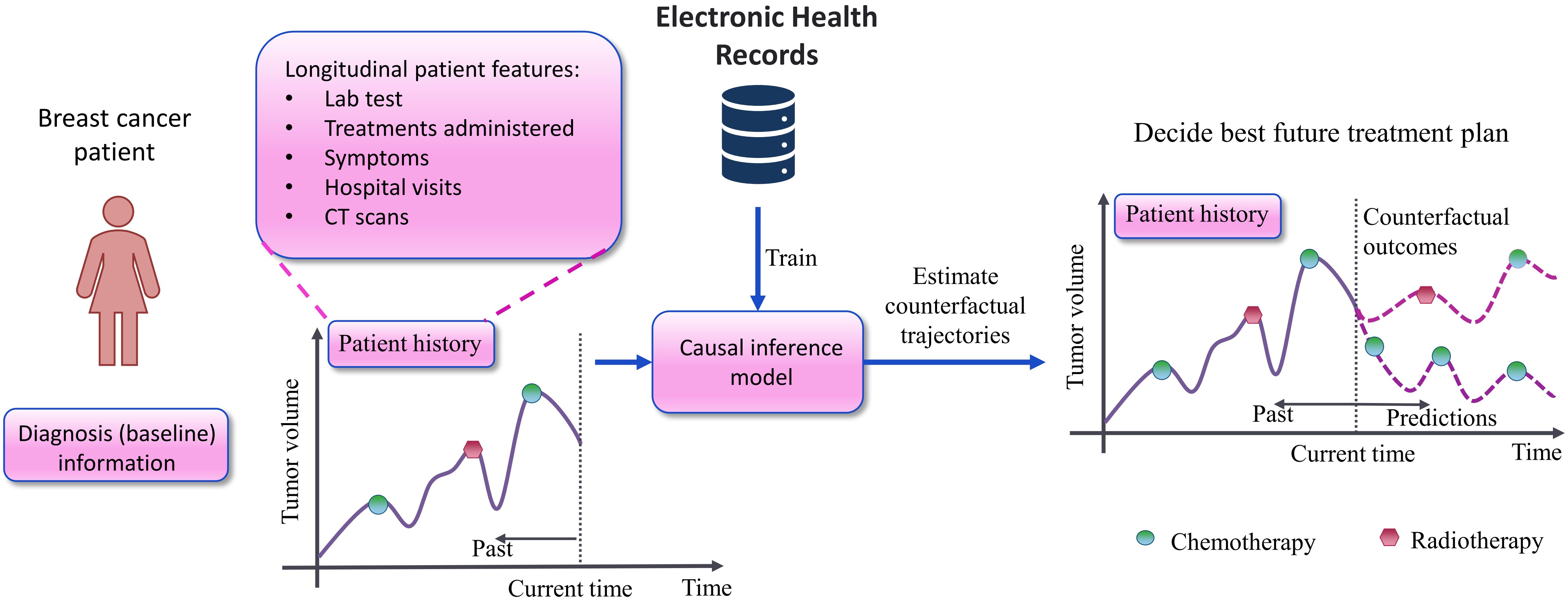
UNIVERSITY OF CAMBRIDGE

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

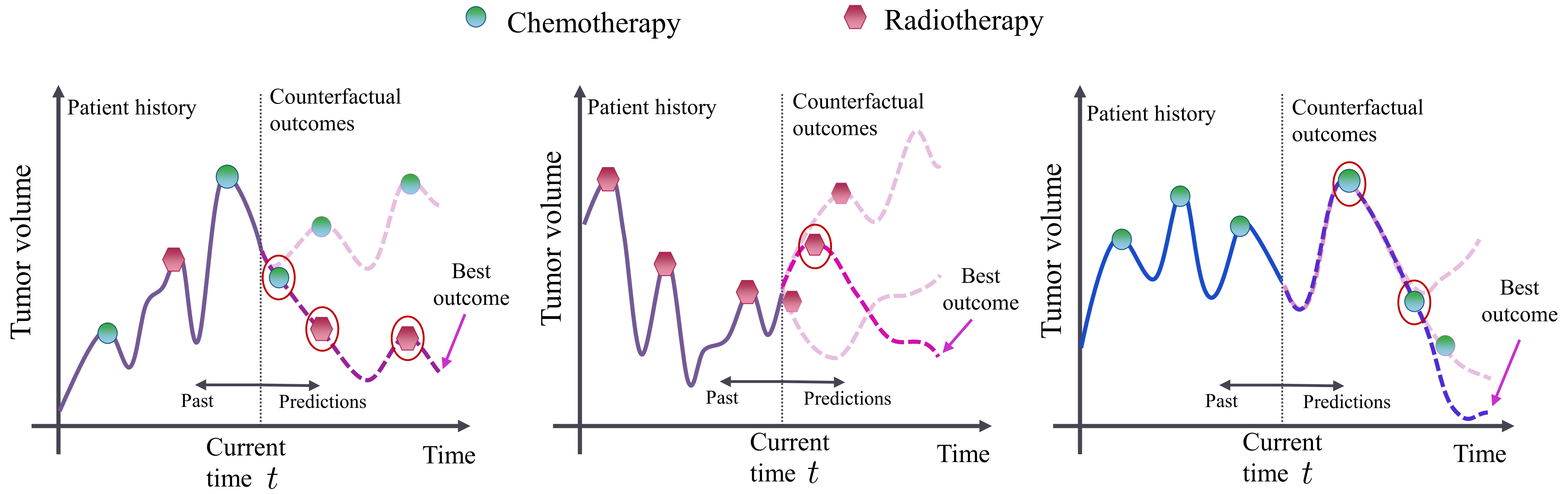
- Reproducibility and visualization

Individualized treatment effects over time



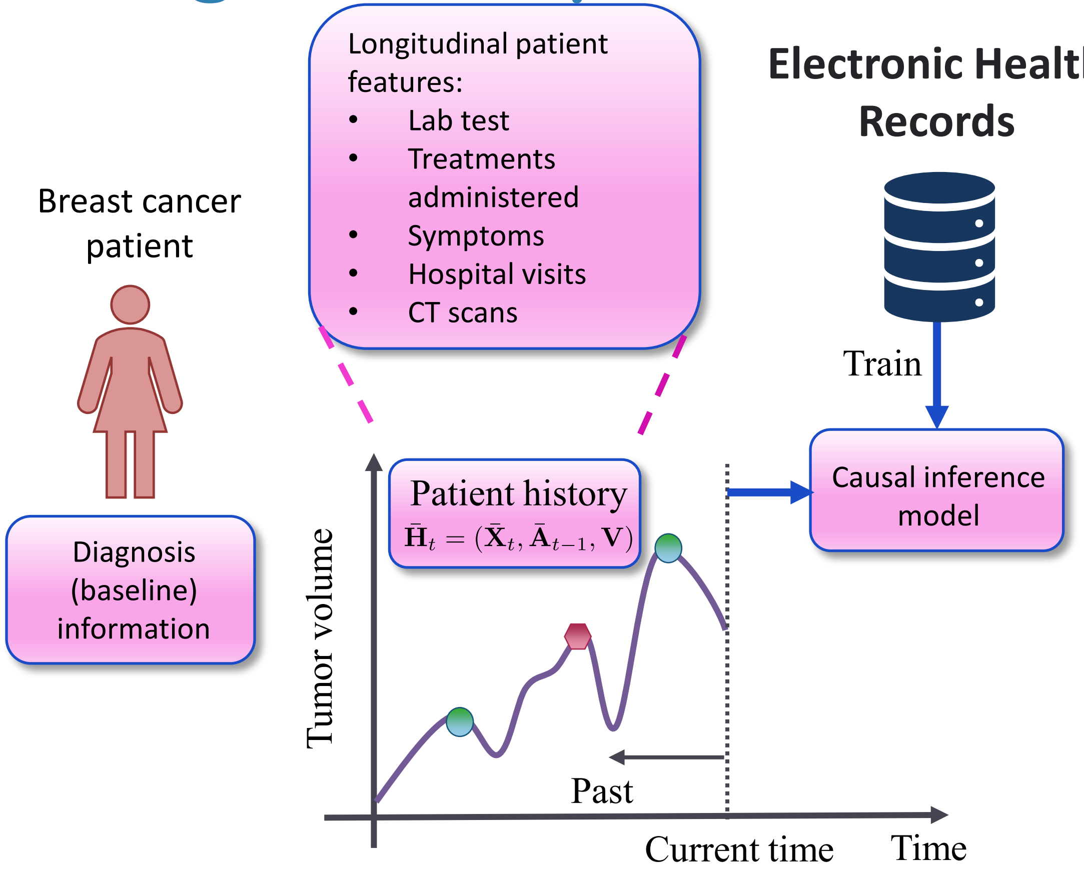
Individualized treatment effects over time

How to treat? | **When to give treatment?** | **When to stop treatment?**



(a) Decide treatment plan | **(b) Decide optimal time of treatment** | **(c) Decide when to stop treatment**

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: $\mathbf{V}$

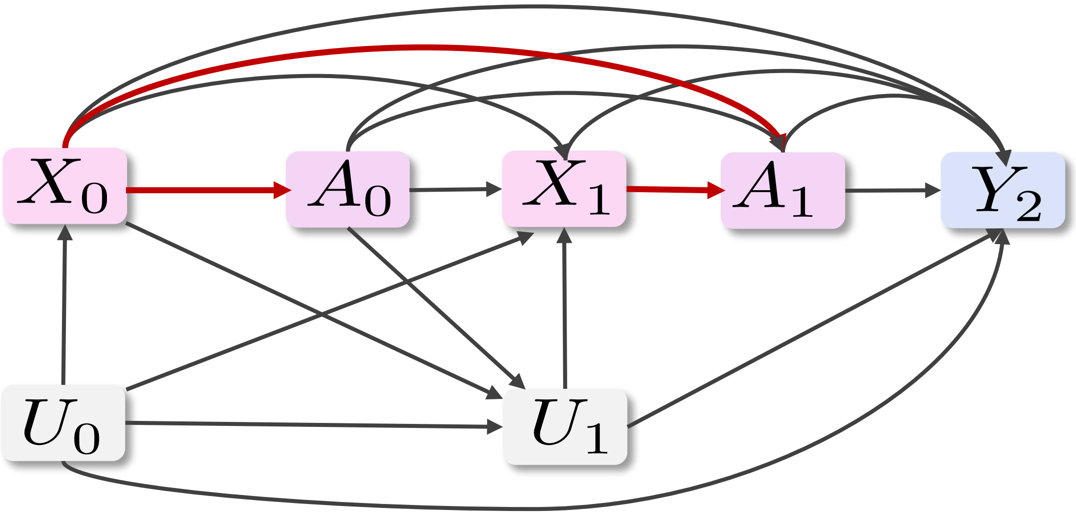
Patient history: $\bar{\mathbf{H}}_t = (\bar{\mathbf{X}}_t, \bar{\mathbf{A}}_{t-1}, \mathbf{V})$

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



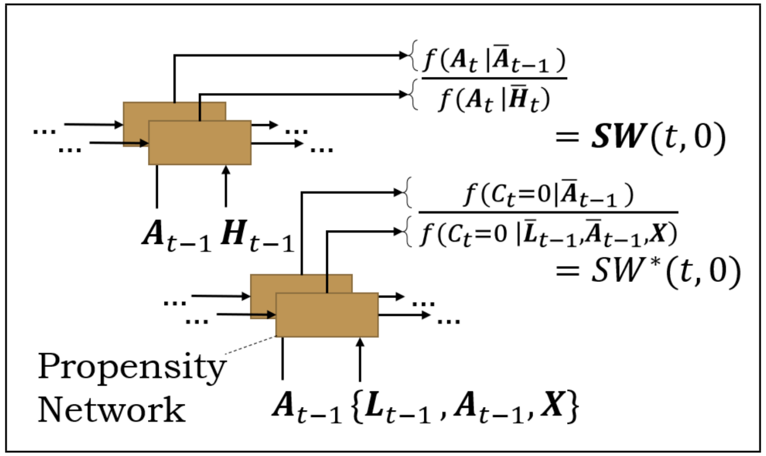
Bias from time-dependent confounders.

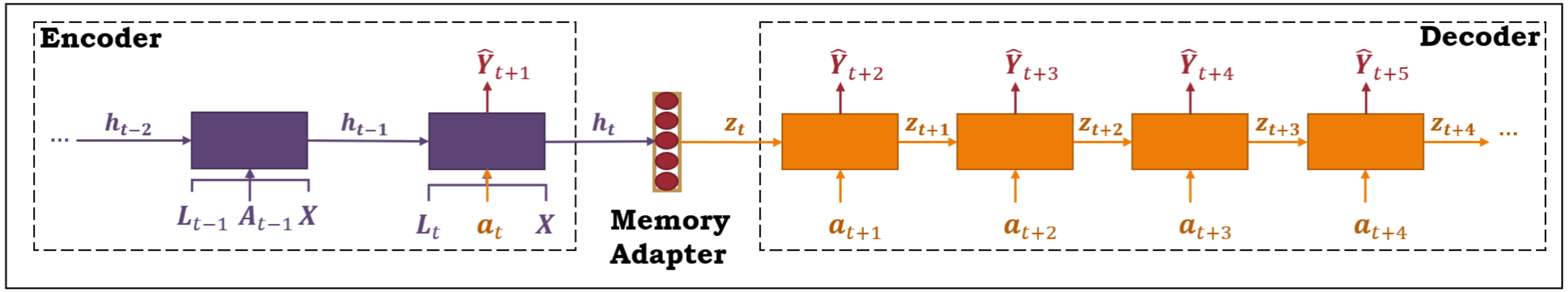
Handling time-dependent confounding bias

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

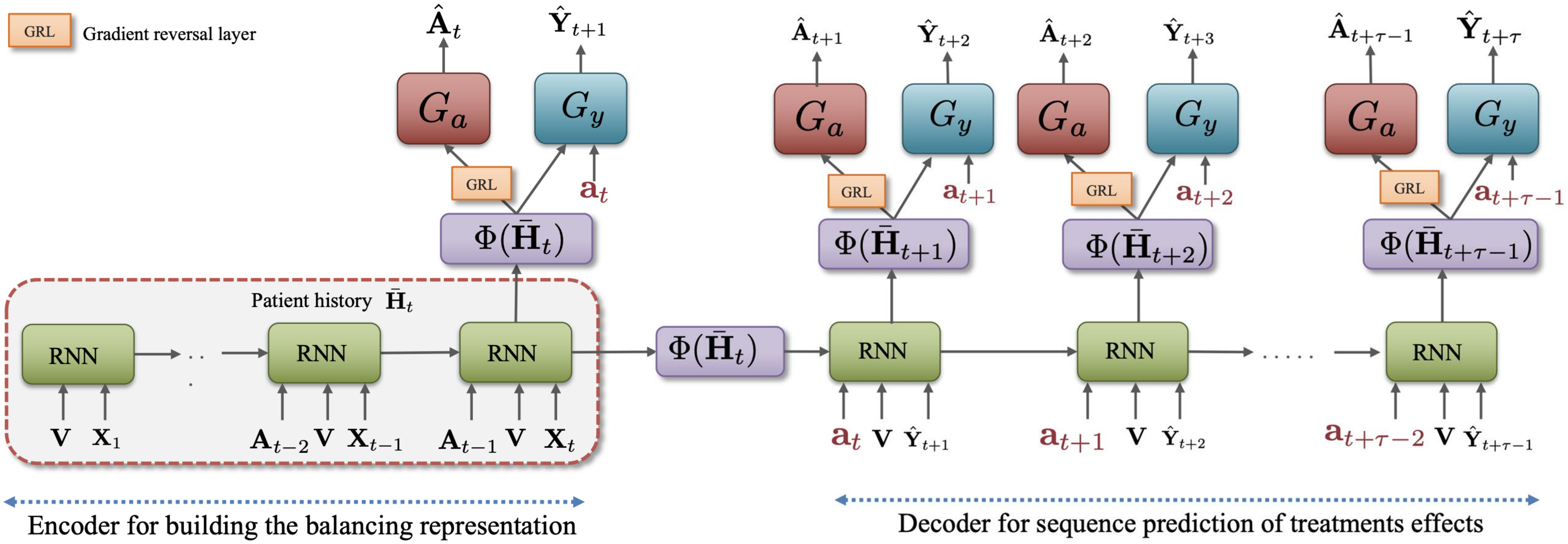
→ Counterfactual recurrent network [Bica, Alaa, Jordon, van der Schaar, ICLR 2020]

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

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) 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

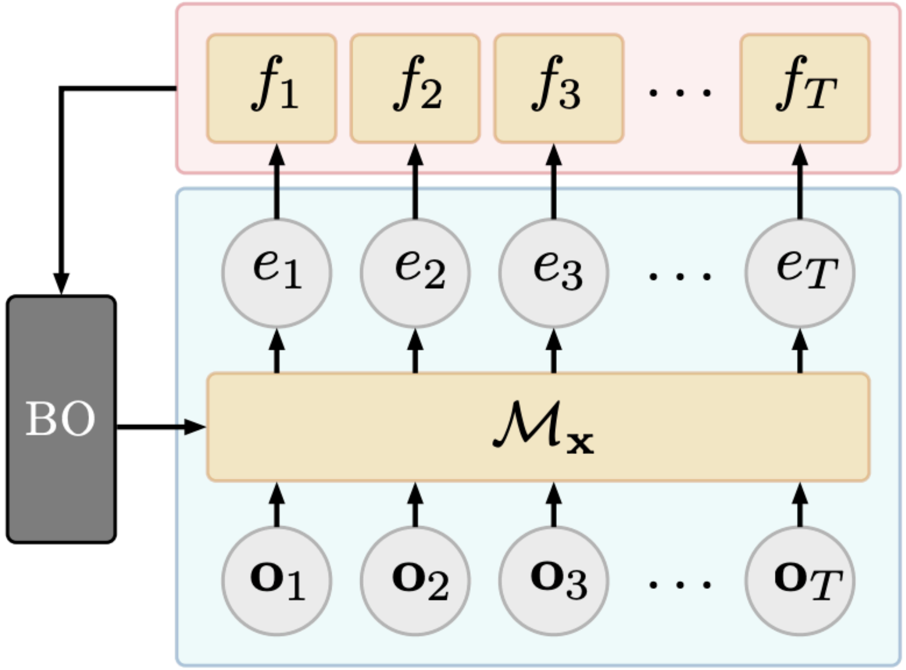
Objective: Model performance at each time step

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 ☹

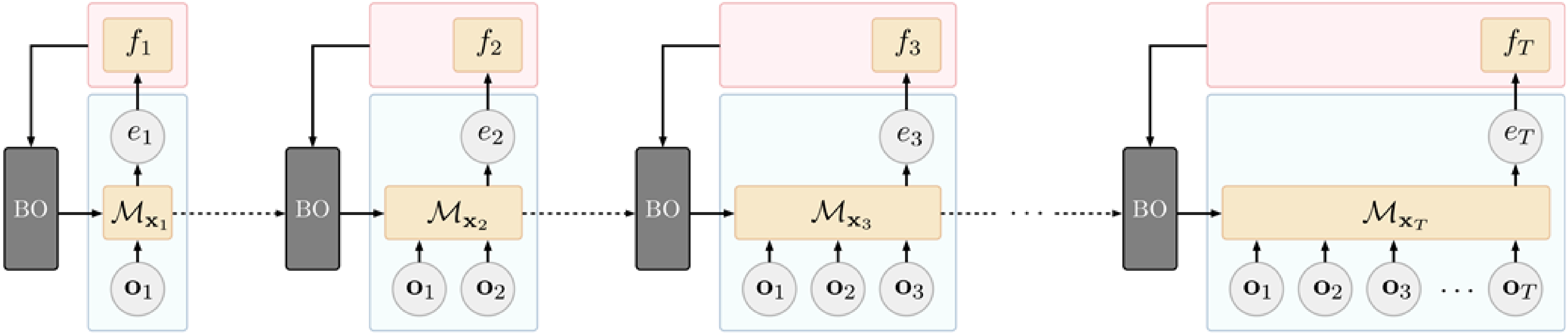
Other solutions?

Black box functions



Multi-Objective Bayesian Optimization (MOBO)

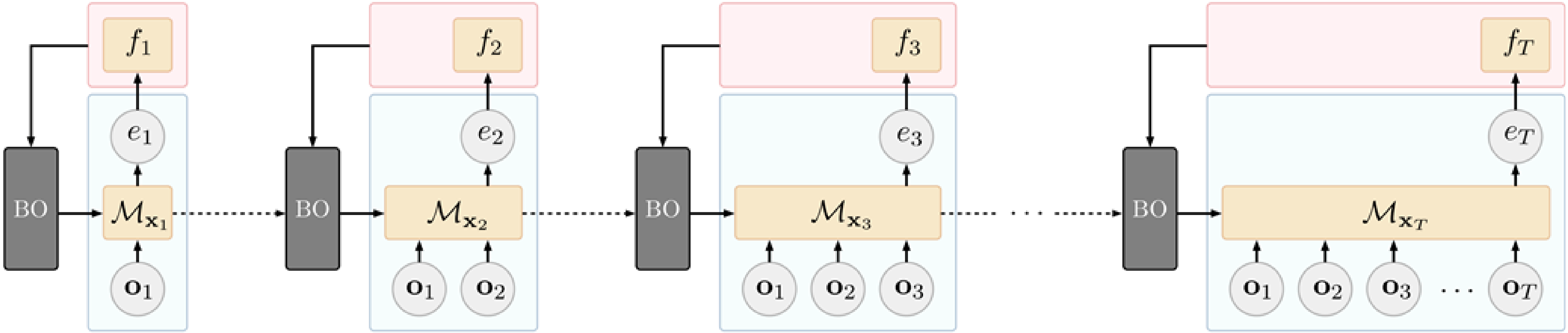
Apply BO sequentially across time-steps as multi-task? No!



Multi-Task Bayesian Optimization (MTBO)

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

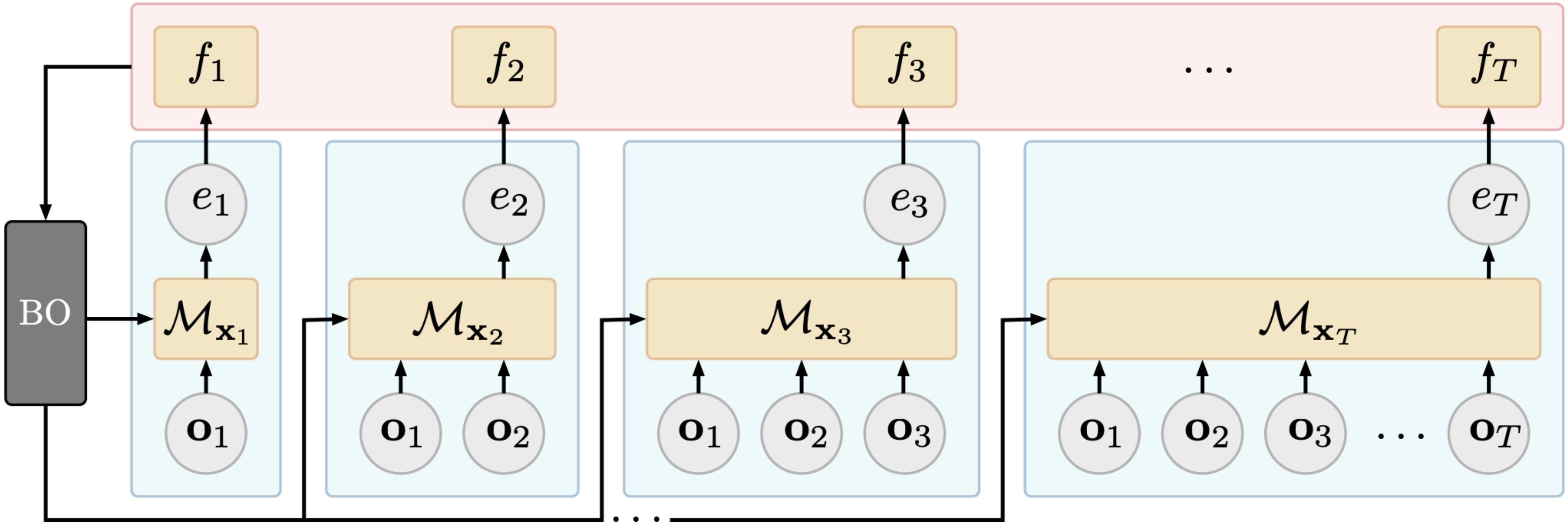
Apply BO sequentially across time-steps as multi-task? No!



- ☹ **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

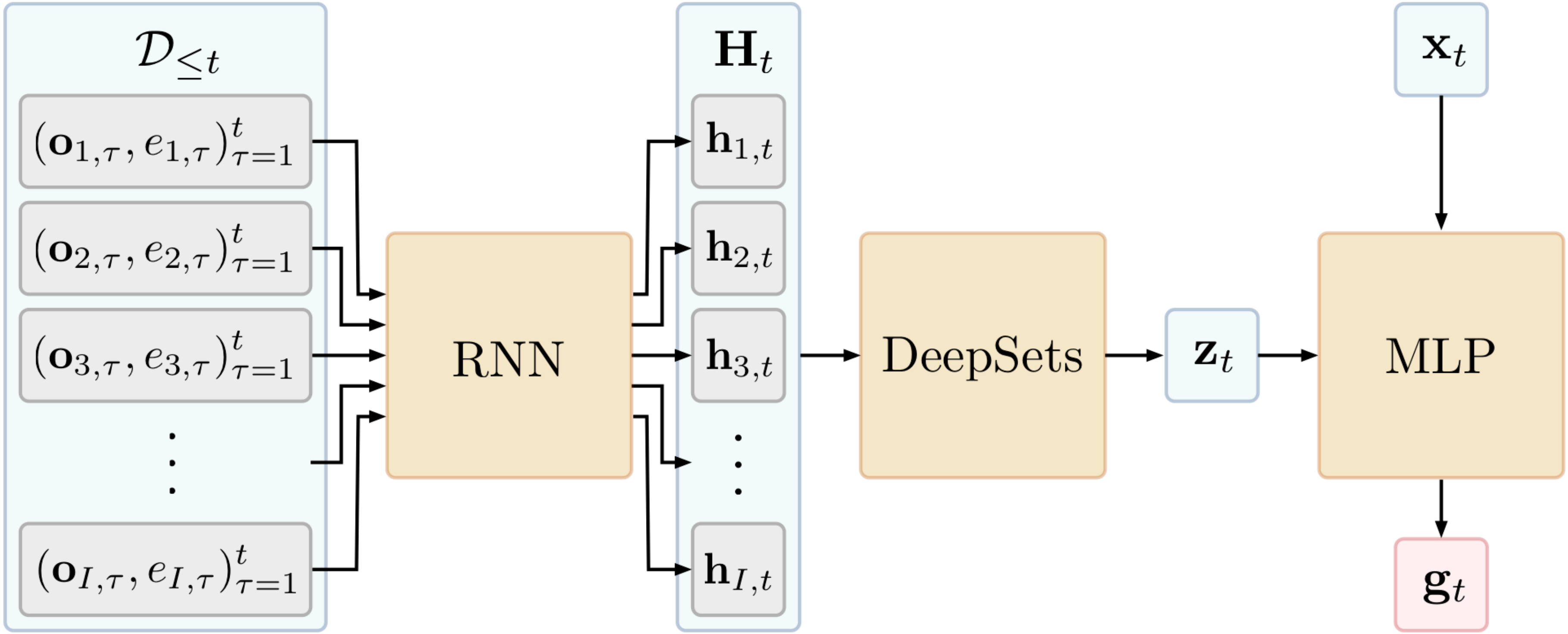


Solve the multiple black-box function optimization problem jointly and efficiently by learning and exploiting correlations among black-box functions using deep kernel learning

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

- Reproducibility and visualization

Engagement sessions: Inspiration Exchange

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

Inspiration Exchange

Themed discussion sessions specifically for <u>machine learning students</u> (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.

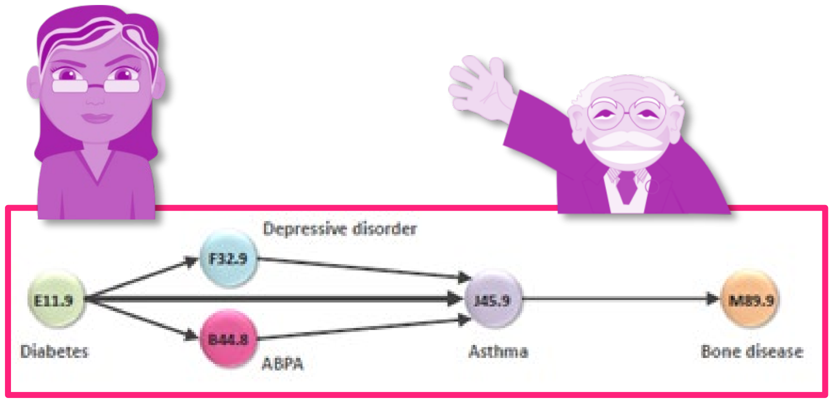
Time-series: a multi-faceted problem

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

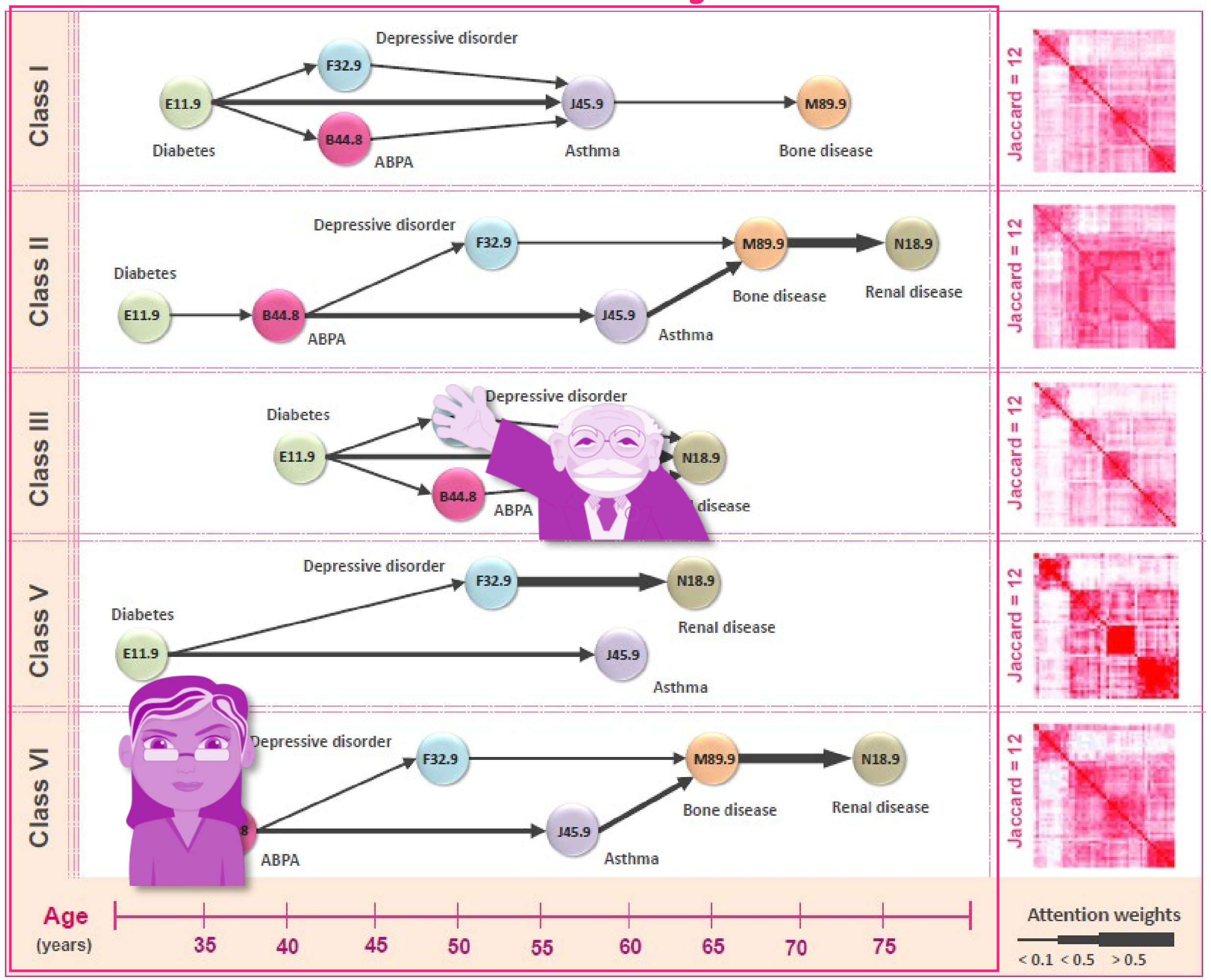
- Reproducibility and visualization

Morbidity networks: personalized and dynamic

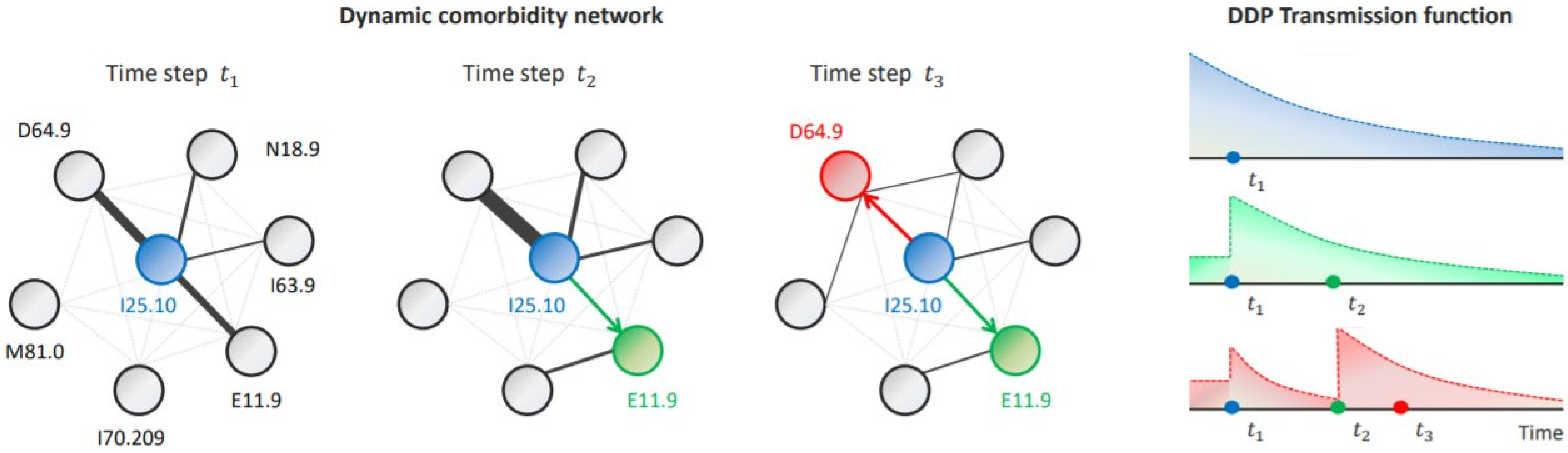
Population-level morbidity network



Personalized morbidity networks



Deep Diffusion Processes (DDP) [Qian, Alaa, vdS, AISTATS 2020]



DDP models temporal relationships between comorbid disease onsets expressed through a dynamic graph

DDP comprises events modelled as a multidimensional point process, with an intensity function parameterized by the edges of a dynamic weighted graph.

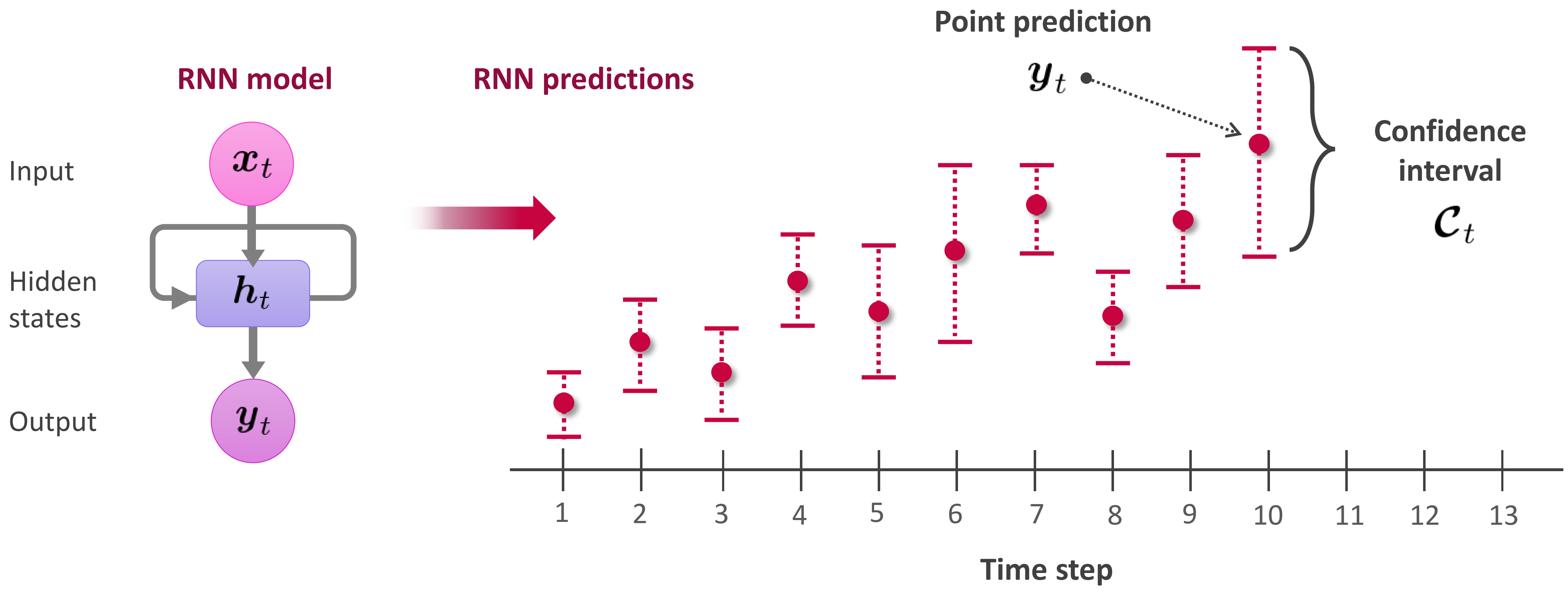
Time-series: a multi-faceted problem

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

- Reproducibility and visualization

Objective: sequential confidence intervals for RNNs

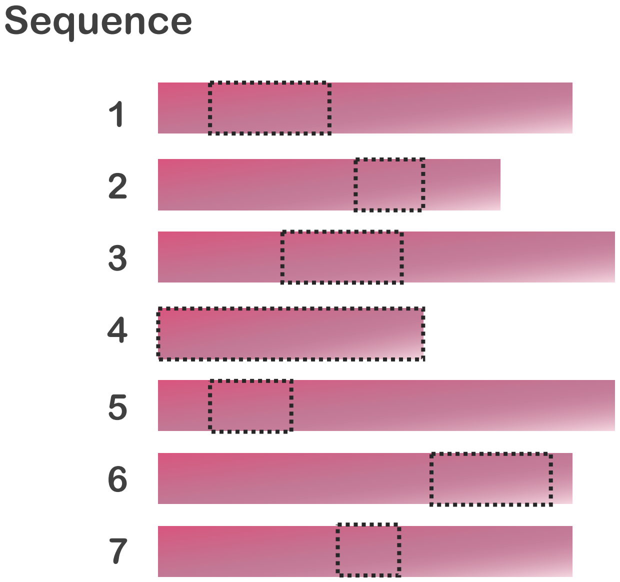
Predictive intervals for Recurrent Neural Networks (RNNs).



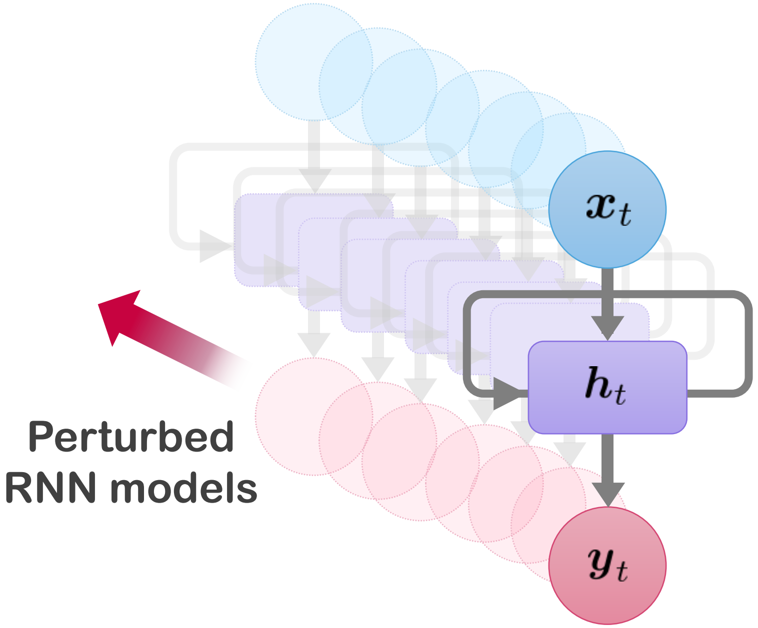
Frequentist Uncertainty in Recurrent Neural Networks via Blockwise Influence Functions [Alaa & vdS, ICML 2020]

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

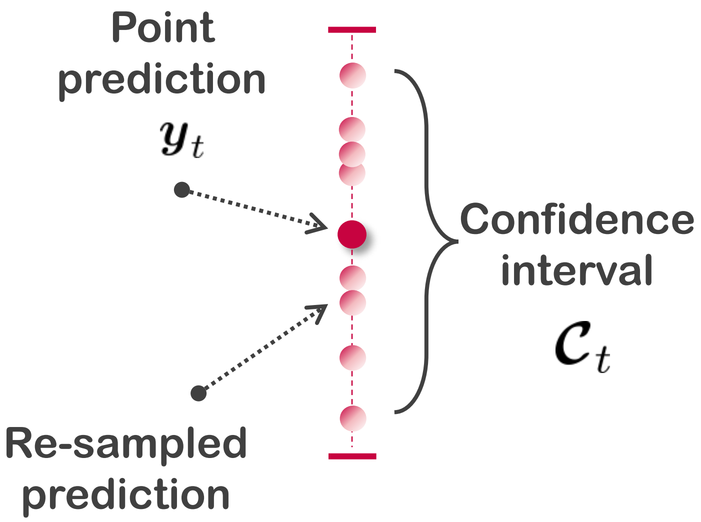
Block deletion



RNN model re-training



RNN prediction re-sampling



van_der_Schaar \ LAB

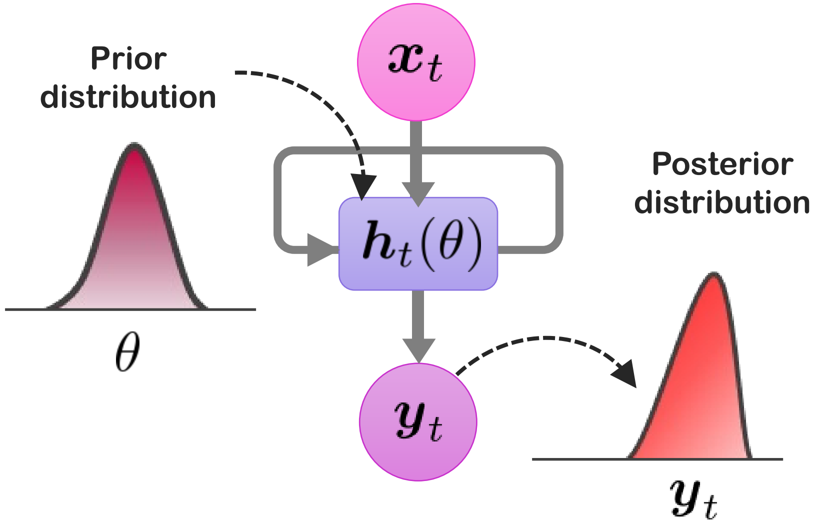
vanderschaar-lab.com

UNIVERSITY OF CAMBRIDGE

Previous work

Bayesian RNNs

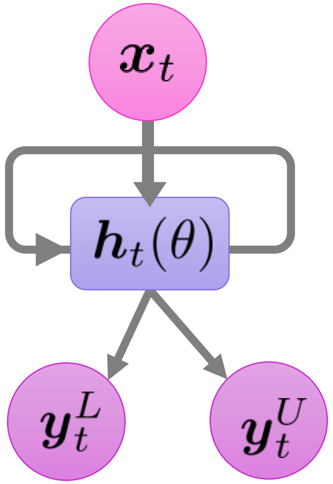
Prior over RNN parameters
Uncertainty = credible intervals



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

Quantile RNNs

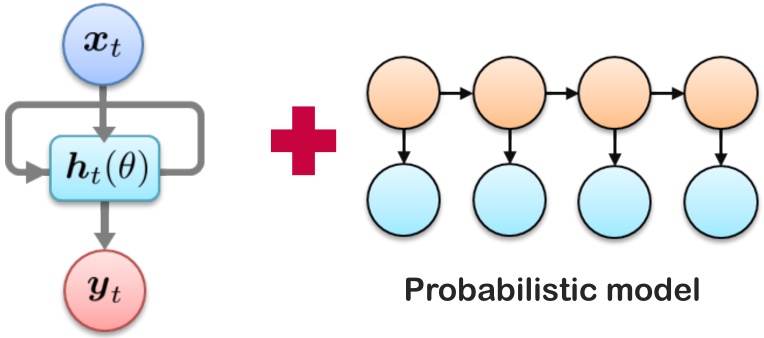
Explicitly train a multi-output RNN to predict intervals



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

Probabilistic RNNs

Combine RNNs with variants of state-space models



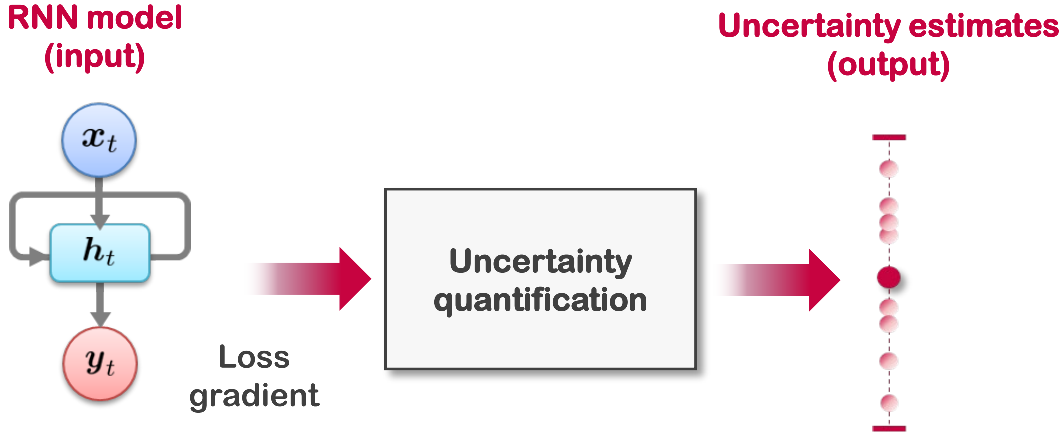
Attentive state-space model (Alaa & van der Schaar, 2019)

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

How is our approach different?

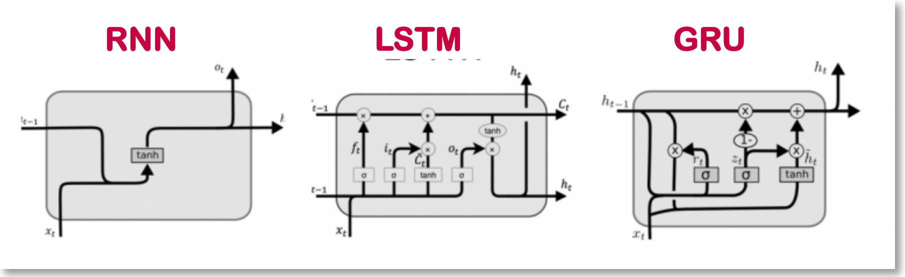
Post-hoc application

- 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

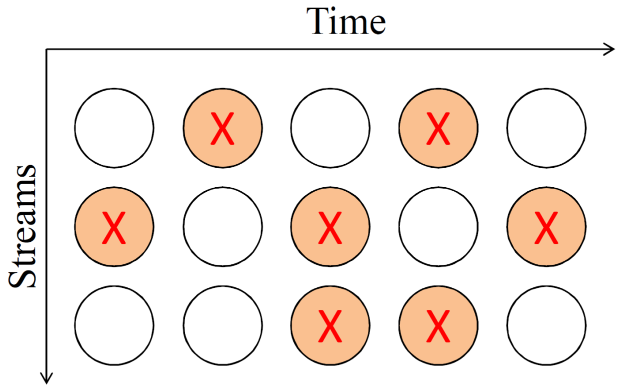
[Stankevičiūtė, Alaa, vdS, NeurIPS 2021]

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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) Synthetic data generation

- Reproducibility and visualization

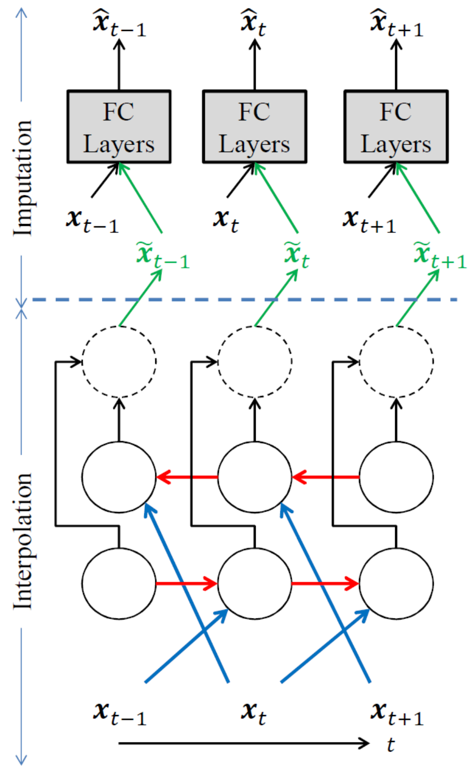
Multi-directional RNN (M-RNN) [Yoon, Zame, vdS, TBME 2018]

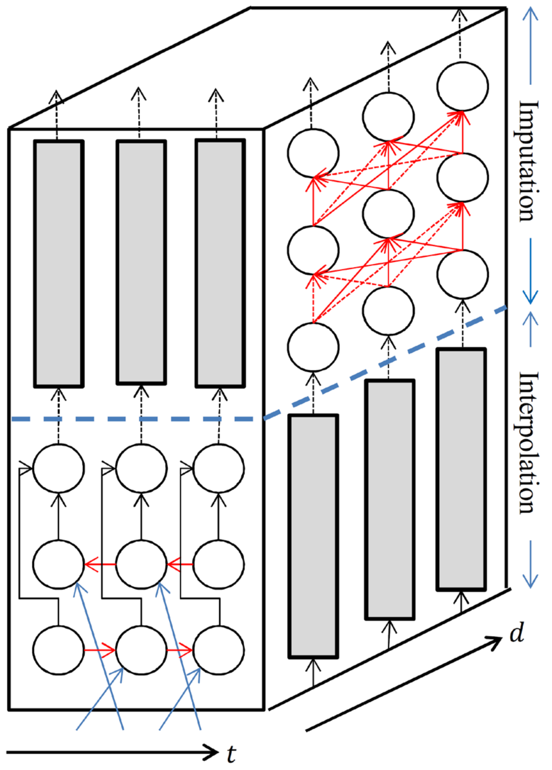


Temporal data streams

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

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

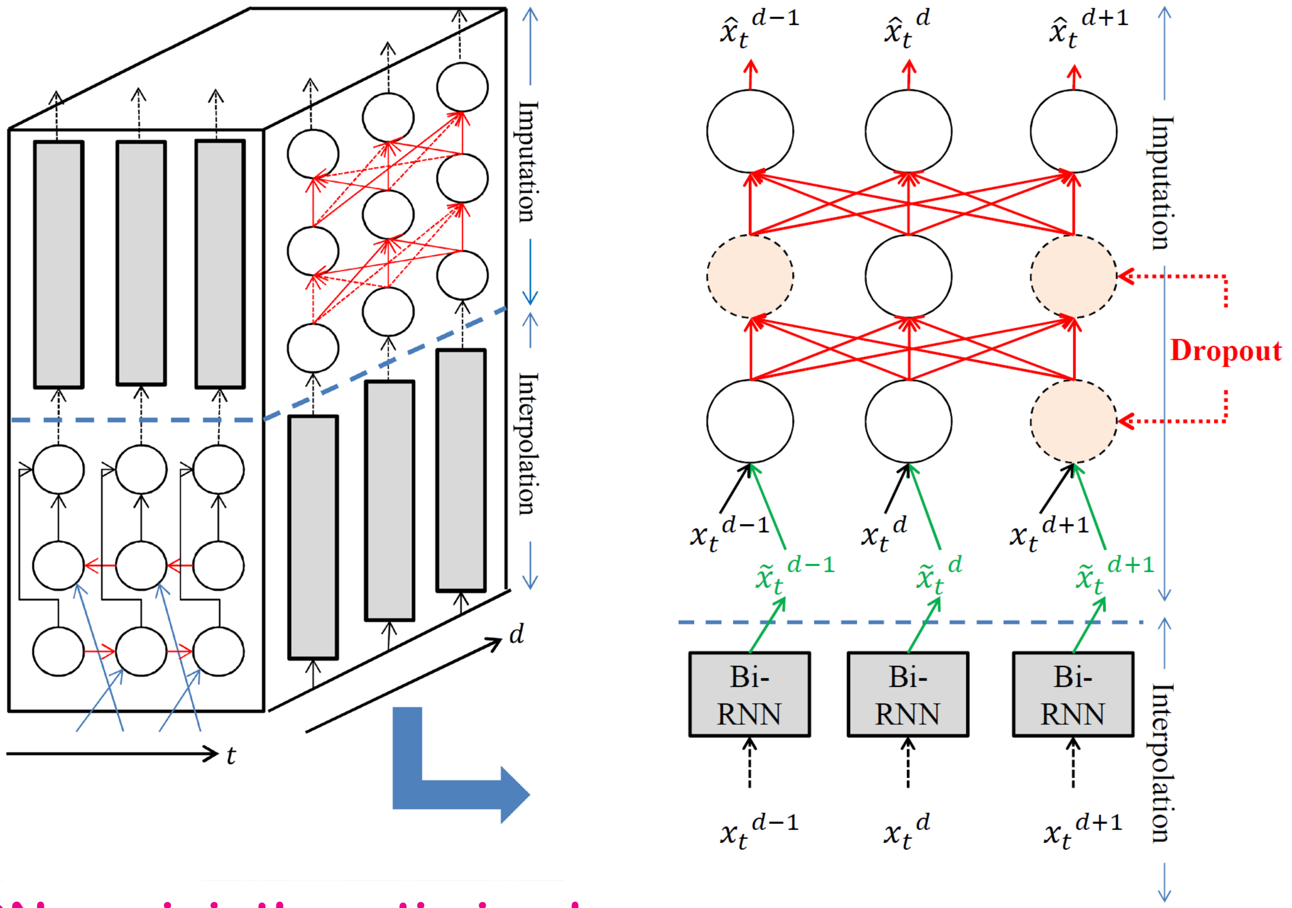




…ward direction and advanced in the backward direction

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

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



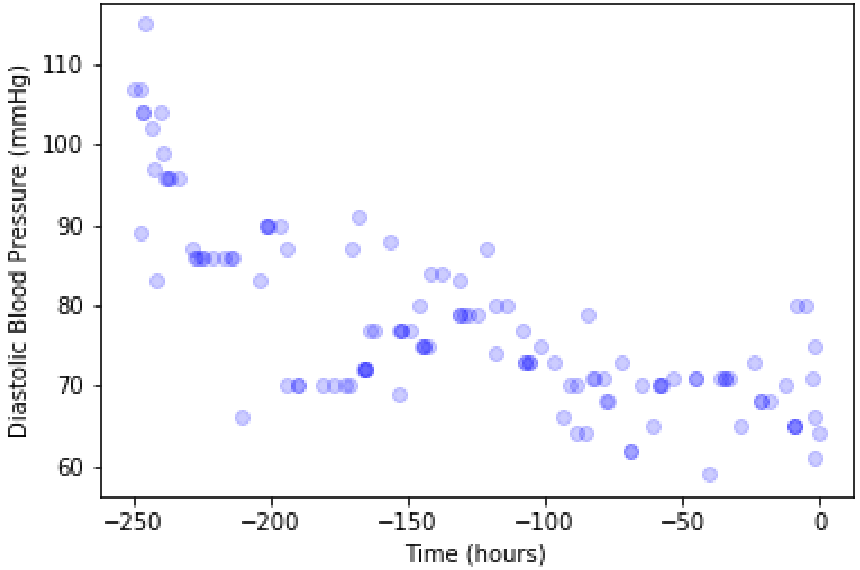
Bi-RNN and FCN are jointly optimized

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



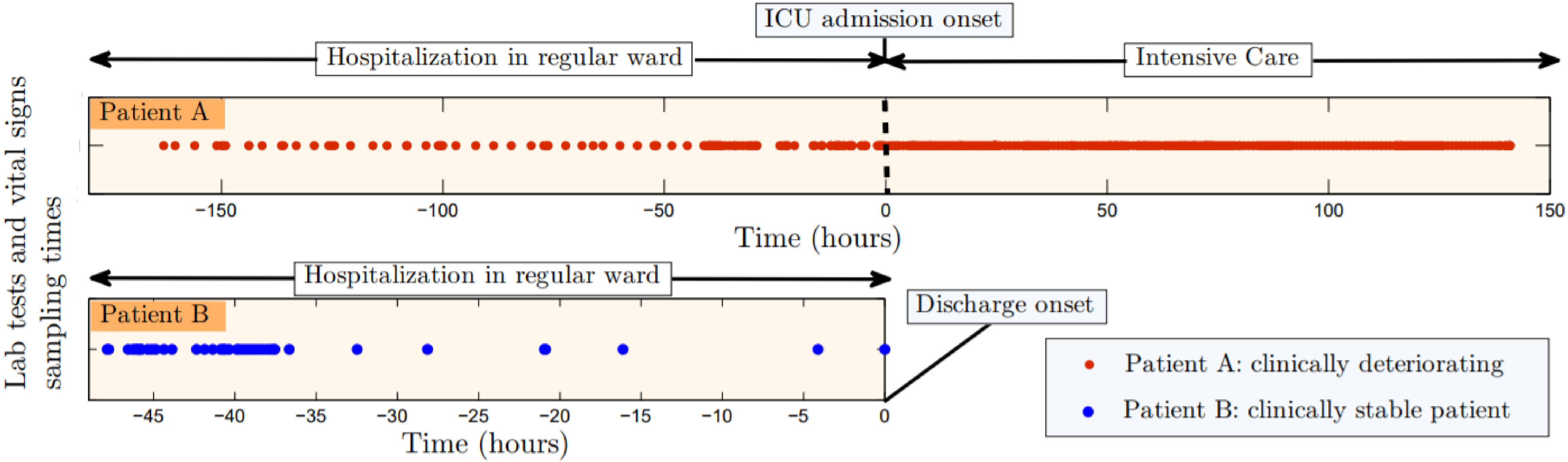
Informative sampling:

Time-varying sampling frequency

Model a patient's trajectory as a <u>marked point process</u> 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



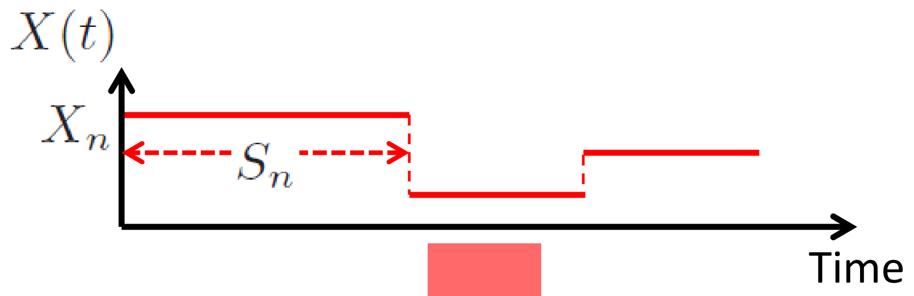
Elements of the probabilistic model (II): the observation process

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

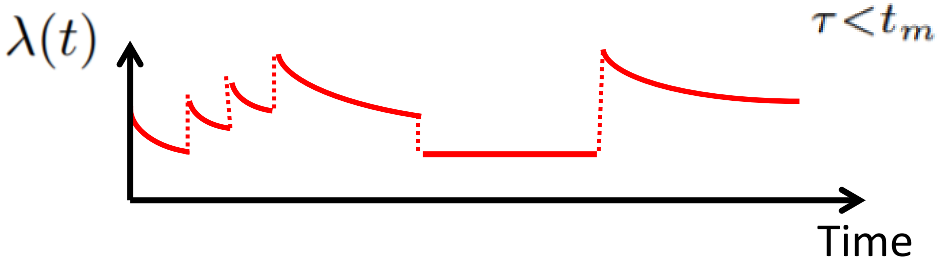
Captures impact of patient's health state on clinicians' sampling behavior



…with a self-exciting Triggering kernel

Captures dependence between observation events

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



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) Discovery and understanding of event data
10) Uncertainty estimation
11) Missing data and informatively missing data
12) 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

- **But we need data to develop analytics and facilitate reproducible research**

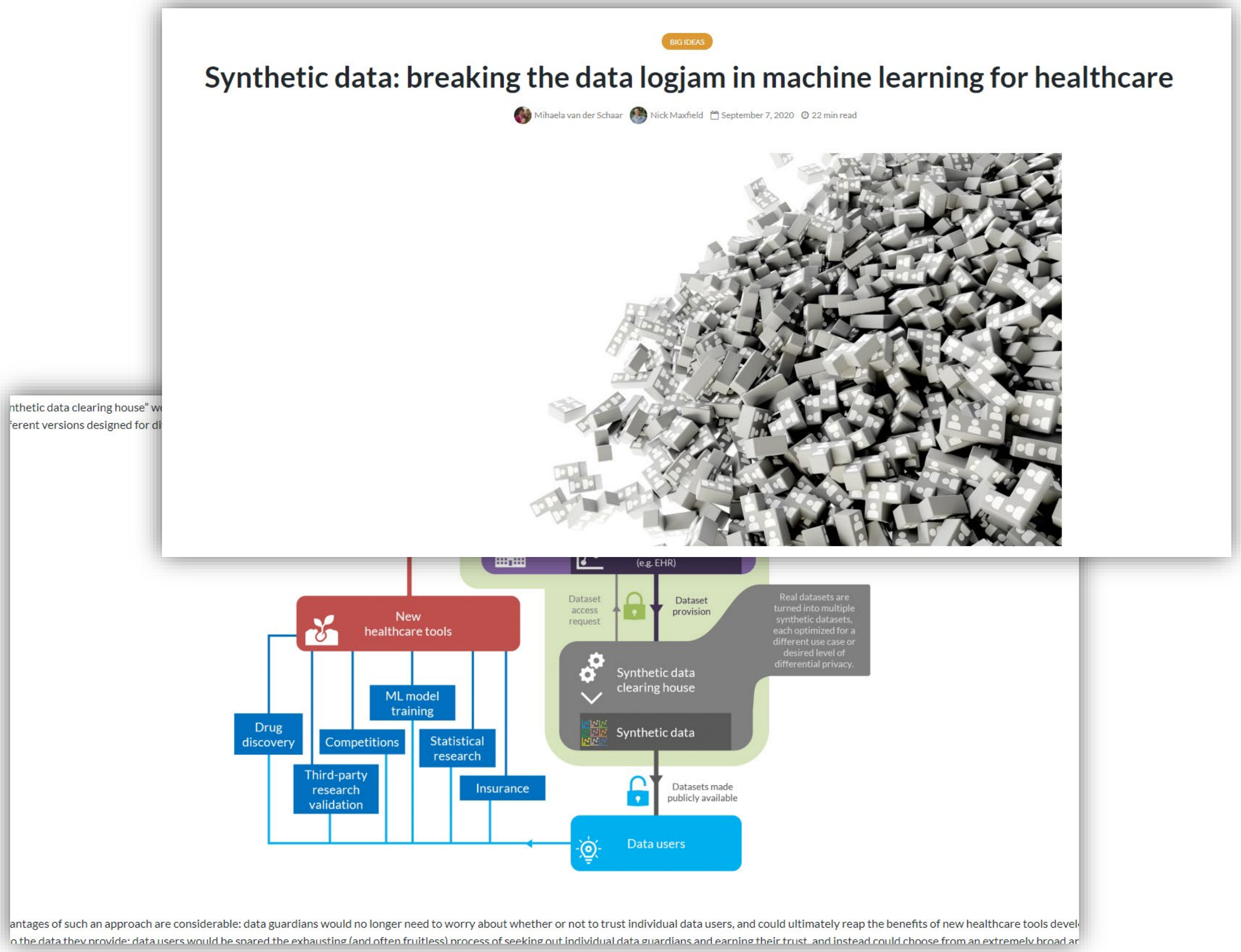
Overview of our work on synthetic data

Introduction for a variety of audiences

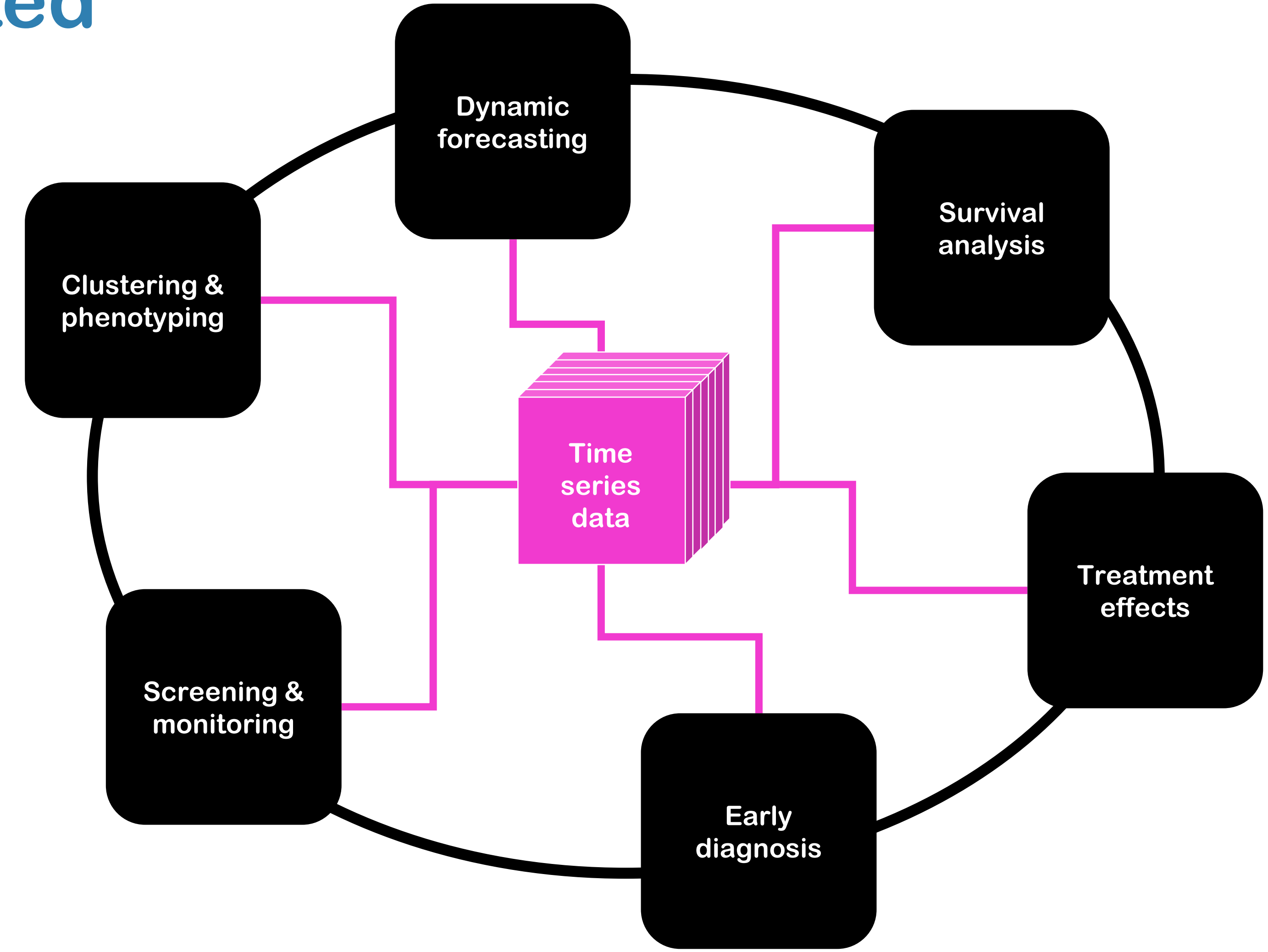
Outlines the importance of synthetic data; explores and summarizes recent cutting-edge synthetic data approaches and methods

Links to a range of additional resources
- vision, papers, software

vanderschaar-lab.com/
→ Research pillars
→ Synthetic data



New Frontiers: Healthcare problems (and models) are interconnected



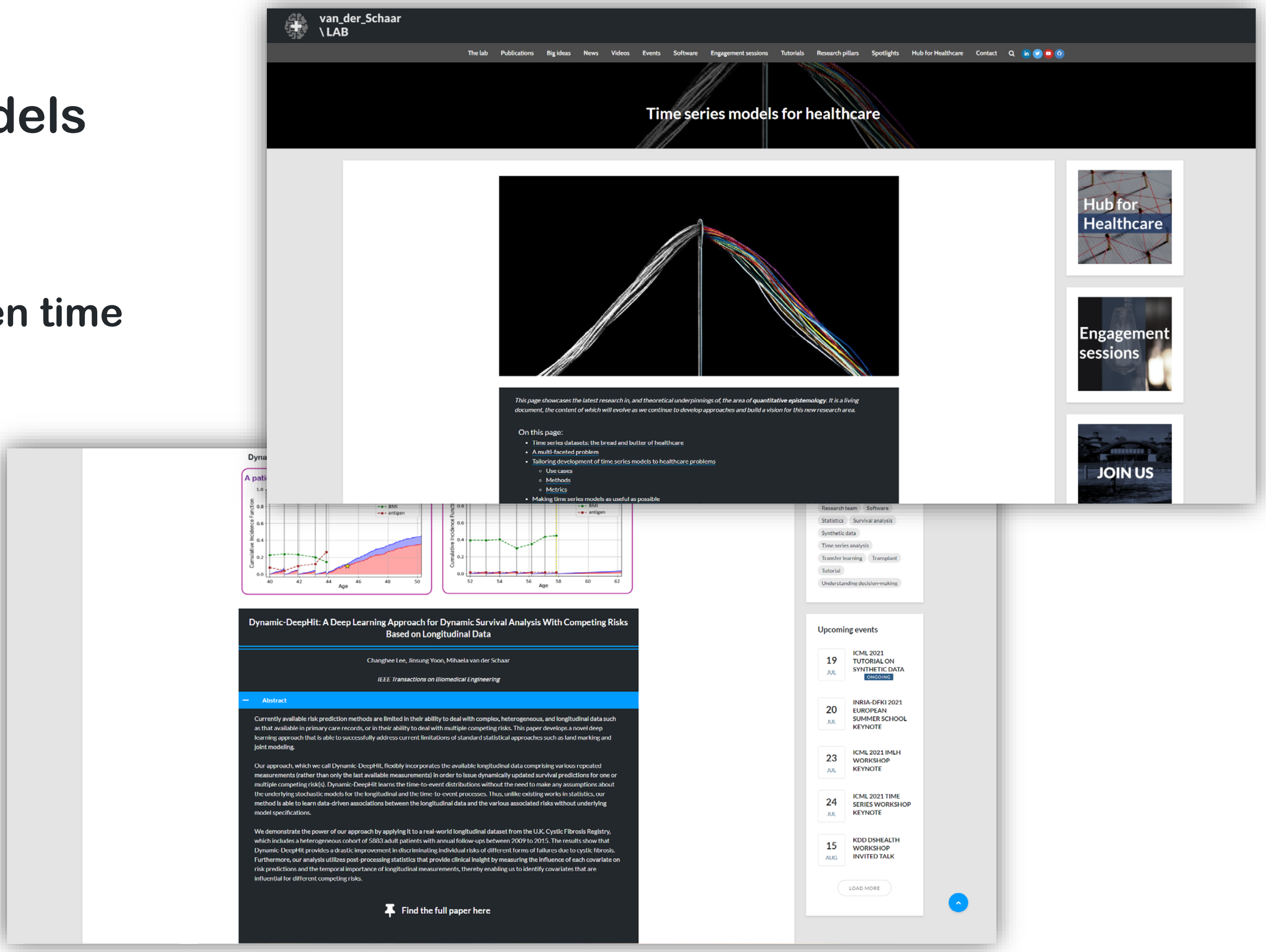
More information and updates

Overview of our work on time series models

Introduction for a variety of audiences

Explores the “cross-sectional” interactions between time series and many other areas of research

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



	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.