Adverse Event Prediction by Telemonitoring and Deep Learning

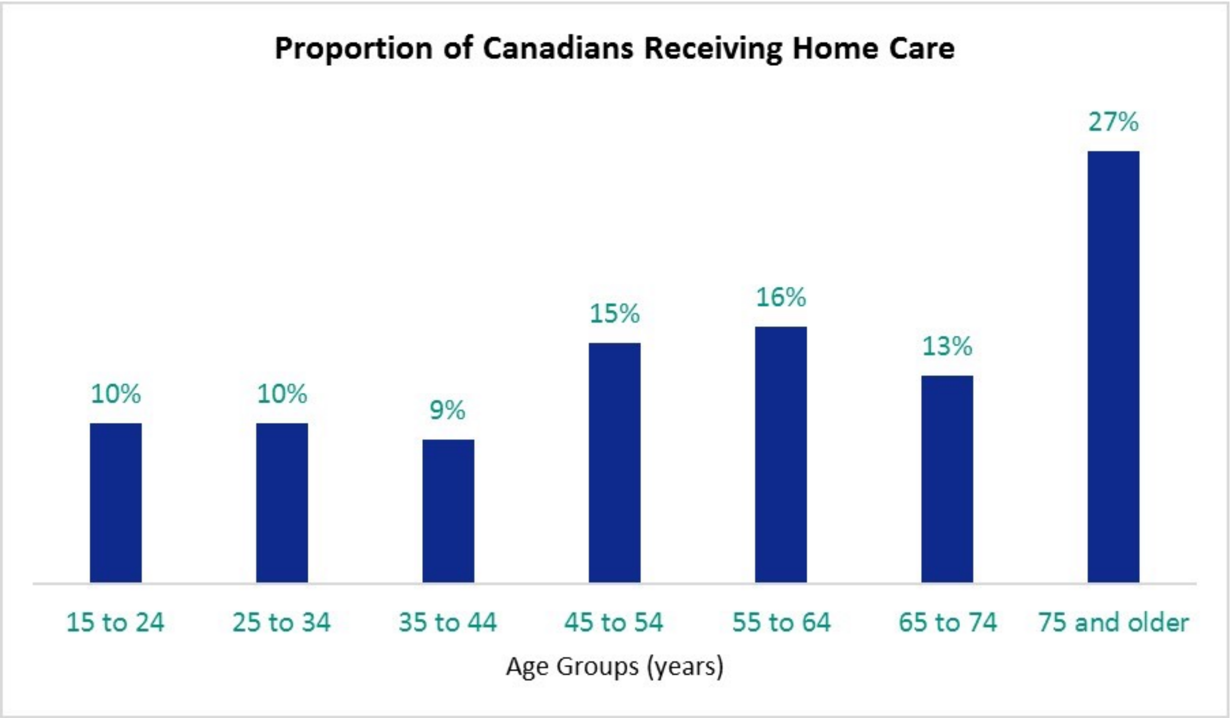
Antoine Prouvost,[2,3] Andrea Lodi,[2,3]
Louis-Martin Rousseau,[2] Jonathan Valle[1]

[1] AlayaCare
[2] DS4DM
[3] MILA

HCSE Montréal, 30th of May 2019

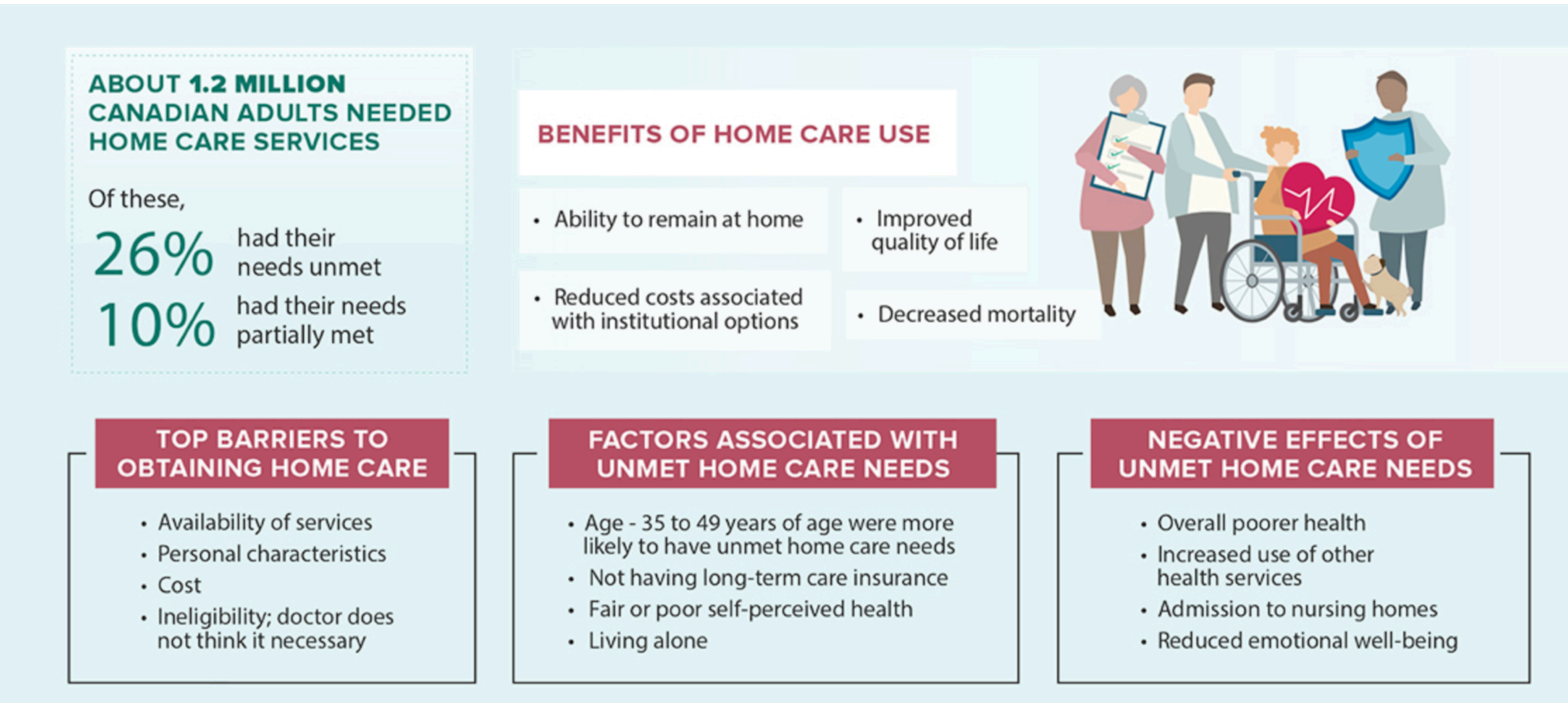
Home Healthcare

- 2 million 15y+ Canadians received some form of Home Care
- long-term health condition, disability or aging needs.
- Mostly seniors



Information published by Stats Can (2012)

Home Healthcare



Information published by Stats Can (2018)

Setting

- Every patient has different homecare needs
  → Personalization

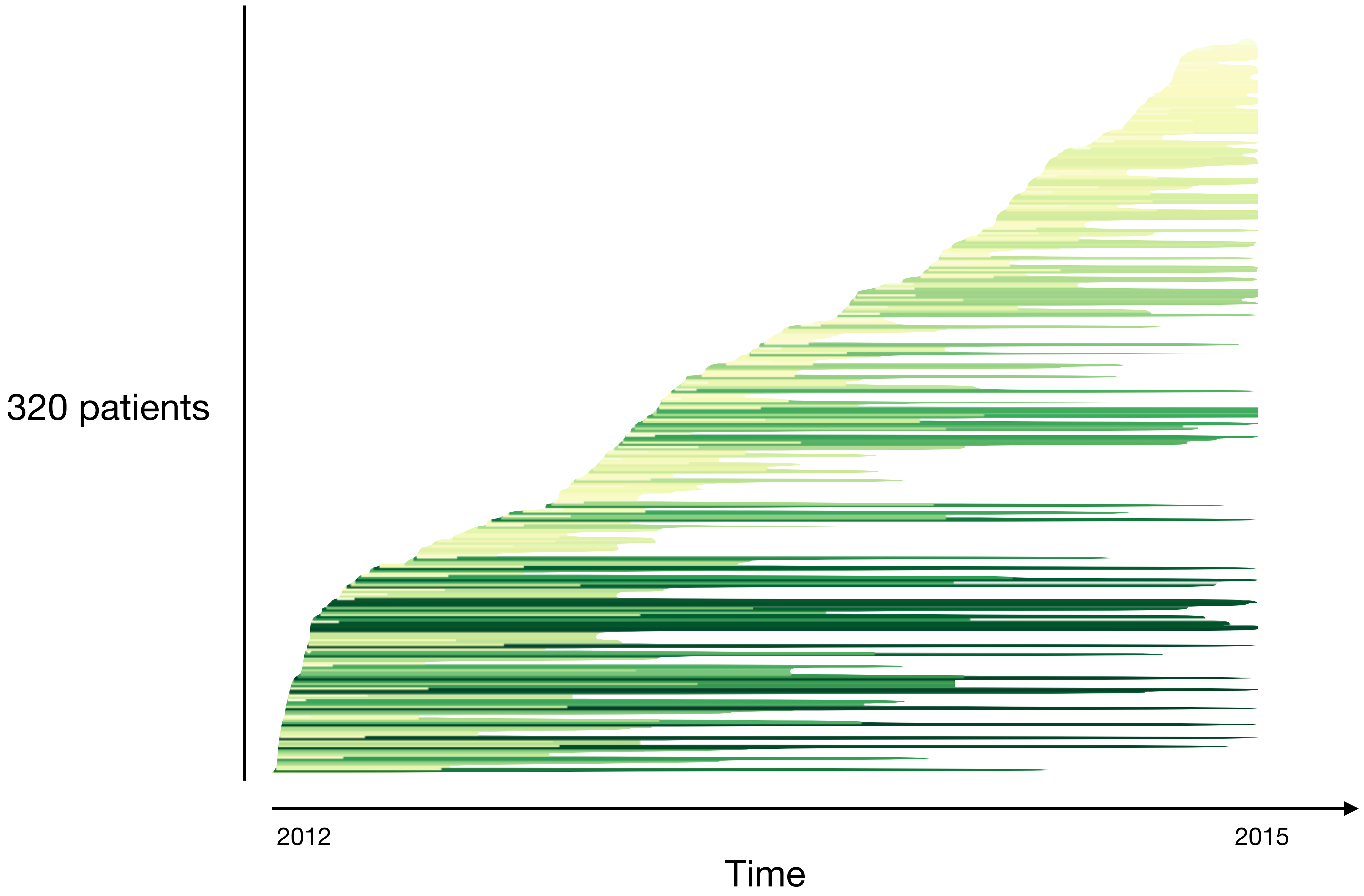
- Patient need monitoring + digitalization
  → Streams of data to process

→ Using ML to evaluate patient risk and need for care worker

Objective

- Use machine learning to help select patients on which to focus attention
- Nurses can then initiate priority contact (phone call, visit)
- All of this on a daily basis

Data



Data

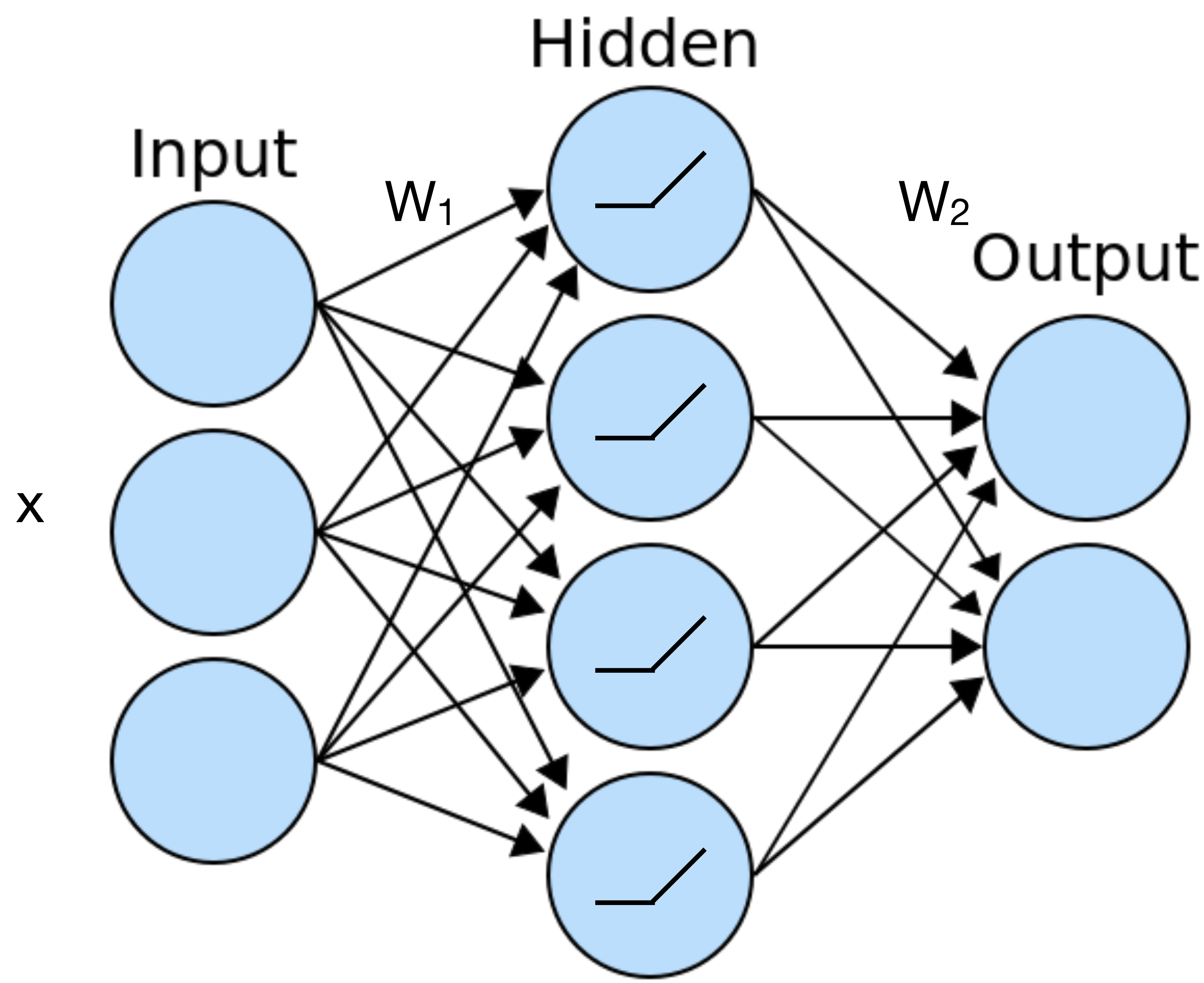
- Clinical patient information
  - age
  - sex
  - initial medical record (ICD)
- Daily measure of vital signs
  - Blood glucose
  - Heart measures (systolic, diastolic, heart beat)
  - Oximeter (SpO_2, heart rate)
  - Weight
- Observed adverse event and type

- 36.25% of patients experienced at least one adverse event while on the HT program
- < 1% of daily adverse event
- Missing values

Challenges

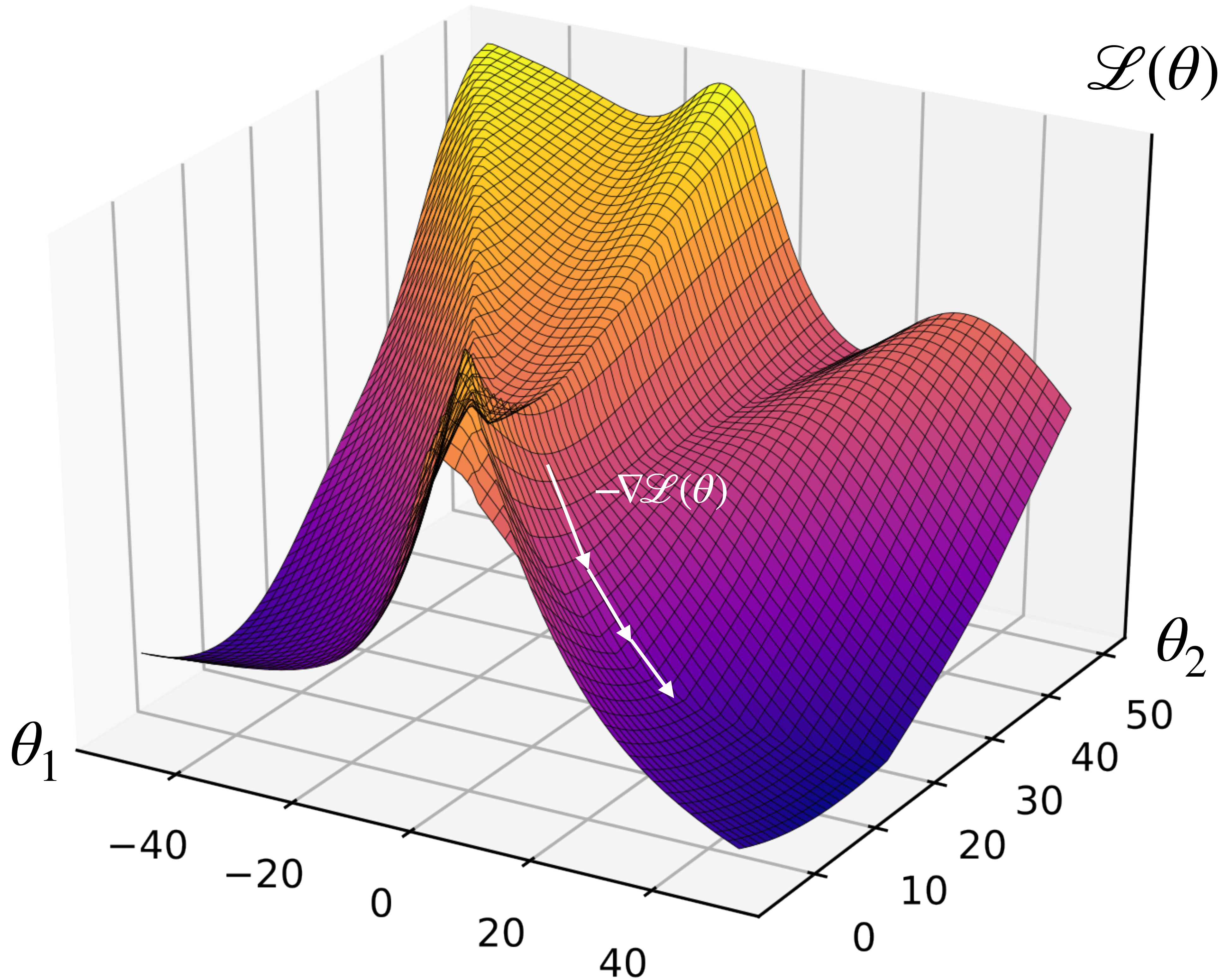
- Supervised learning setting with less than 1% of events
- Missing values
- Restricted dataset
- Noisy data
- Multi-structure problem:
  static, recurrent information, variable size of medical records

Neural Network



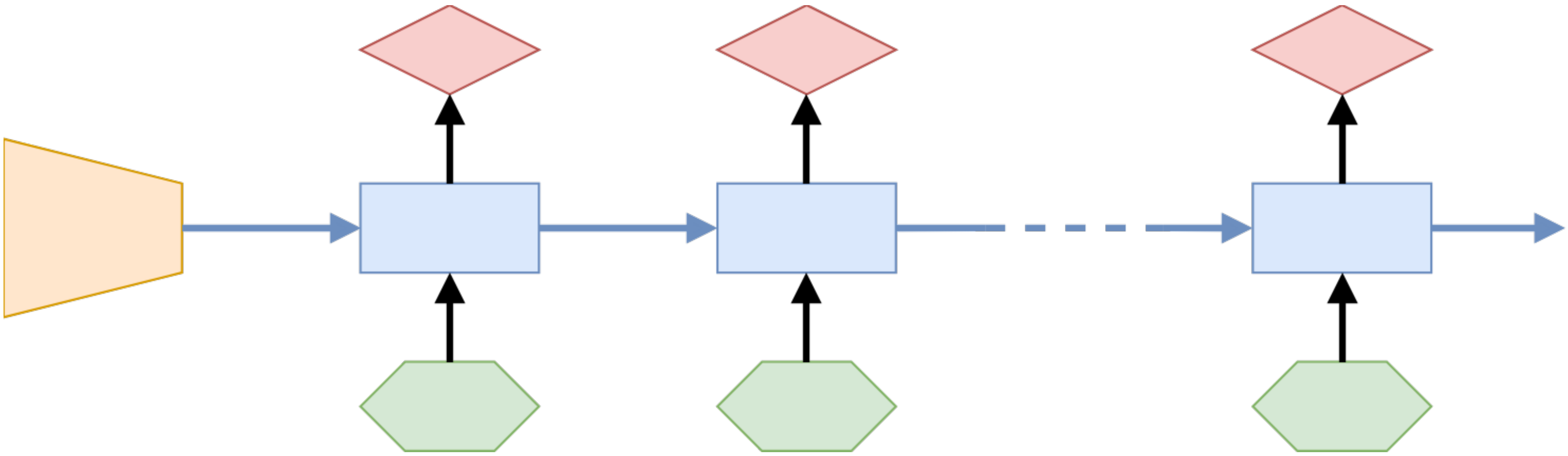
$$\min_{\theta := W_1, W_2} \quad \mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(f_\theta(x_i), y_i)$$

Gradient Descent

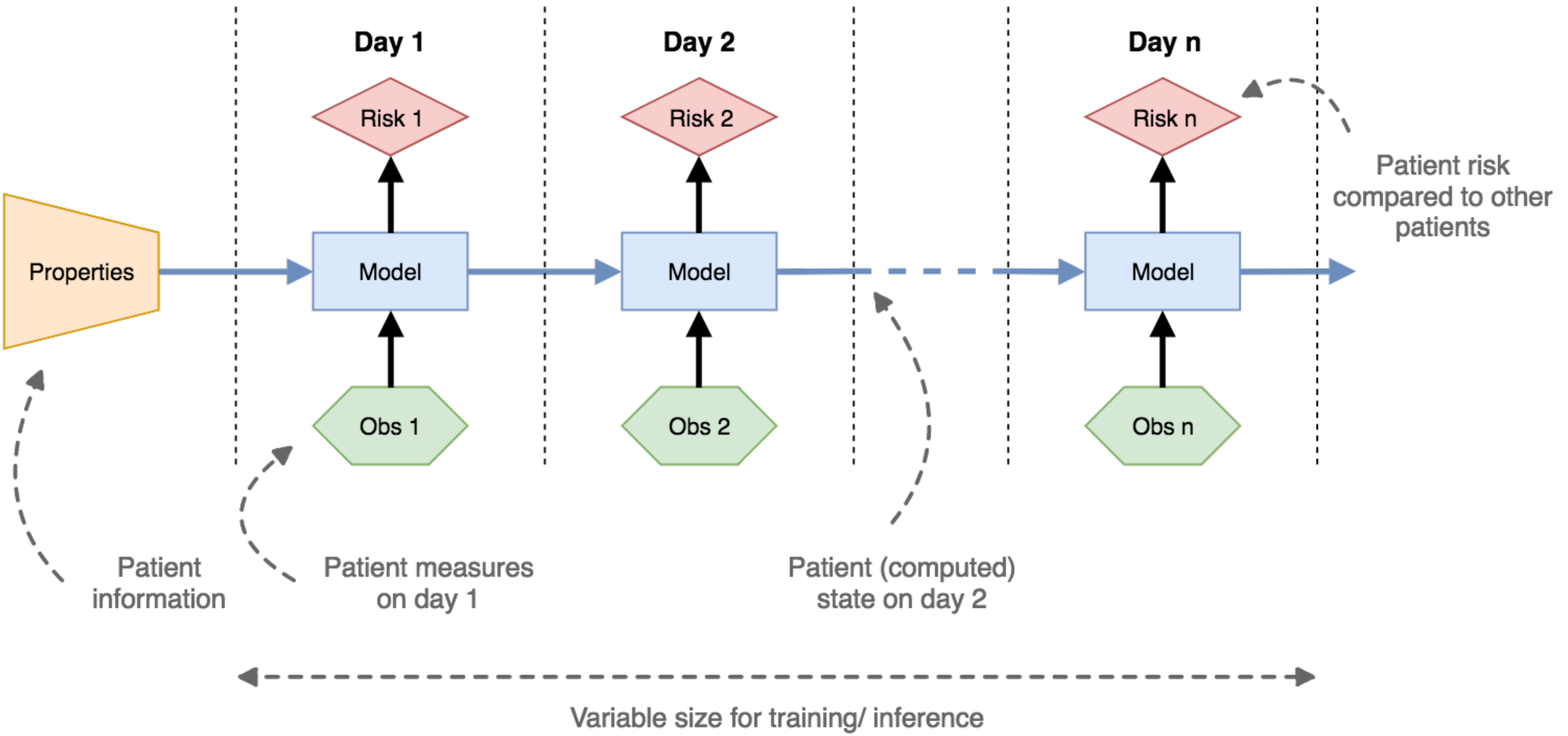


Model

- Using recurrent neural networks to address the multi-structure of the problem



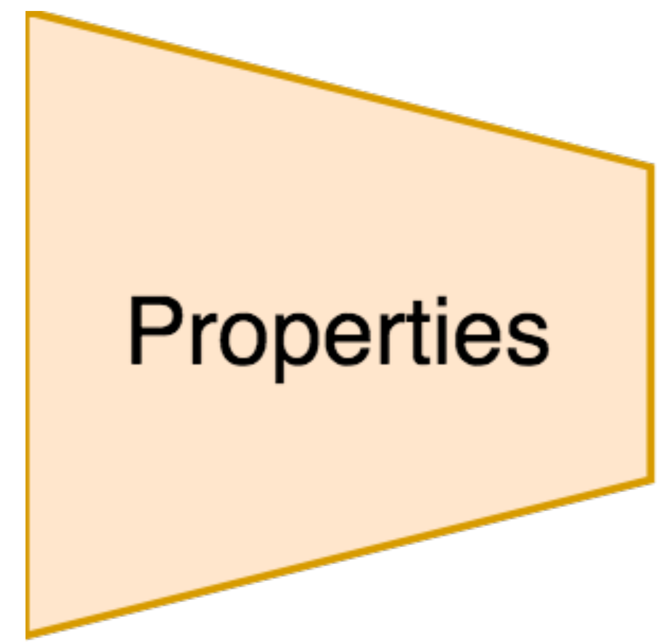
Model



Day 1
Day 2
Day n
Risk 1
Risk 2
Risk n
Properties
Model
Model
Model
Obs 1
Obs 2
Obs n
Patient risk compared to other patients
Patient information
Patient measures on day 1
Patient (computed) state on day 2
Variable size for training/ inference

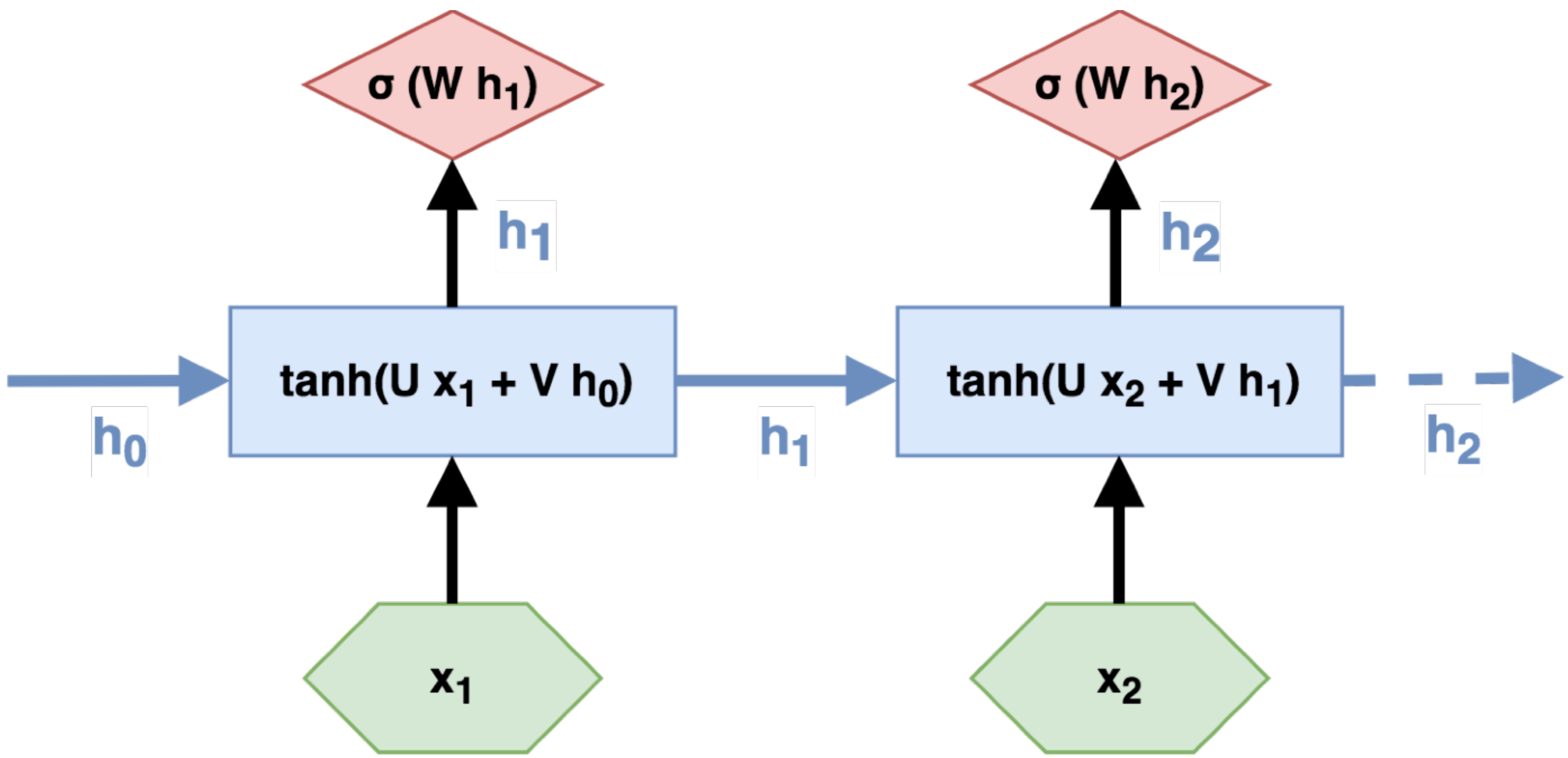
Patient Information

- Patient diseases (ICD) embedded into a high dimensional space
- In that space, similar disease, drugs and symptoms are close in a mathematical distance
- We use the pre-trained model from [1]
- Variable # of ICD is addressed with another RNN



[1] Y. Choi, C. Y.-I. Chiu, and D. Sontag, “Learning Low-Dimensional Representations of Medical Concepts,” AMIA Jt Summits Transl Sci Proc, vol. 2016, pp. 41–50, Jul. 2016.

Recurrent Modeling



- Parameter sharing enable sequence modelling and easier training
- Better with long term dependencies: LSTM [2], GRU [3]

[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[3] K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," arXiv:1406.1078 [cs, stat], Jun. 2014.

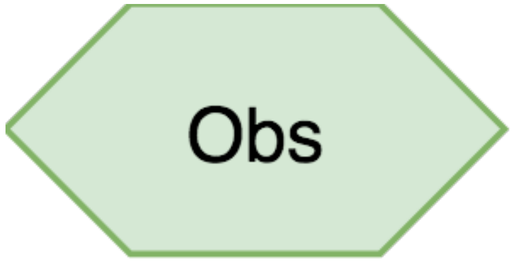
Loss function

- Risk computed by according to survival analysis: higher risk means shorter expected time to adverse event
- Avoid predicting time to next event -> more data efficient
- Allow for ranking patient by order of risk
- C-index

[4] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," BMC Medical Research Methodology, vol. 18, p. 24, Feb. 2018.

Missing data

- Missing features/ data point are a problem for simple neural networks
- Can also be a source of information:
  Patient not following treatment
- Dummy variable for availability

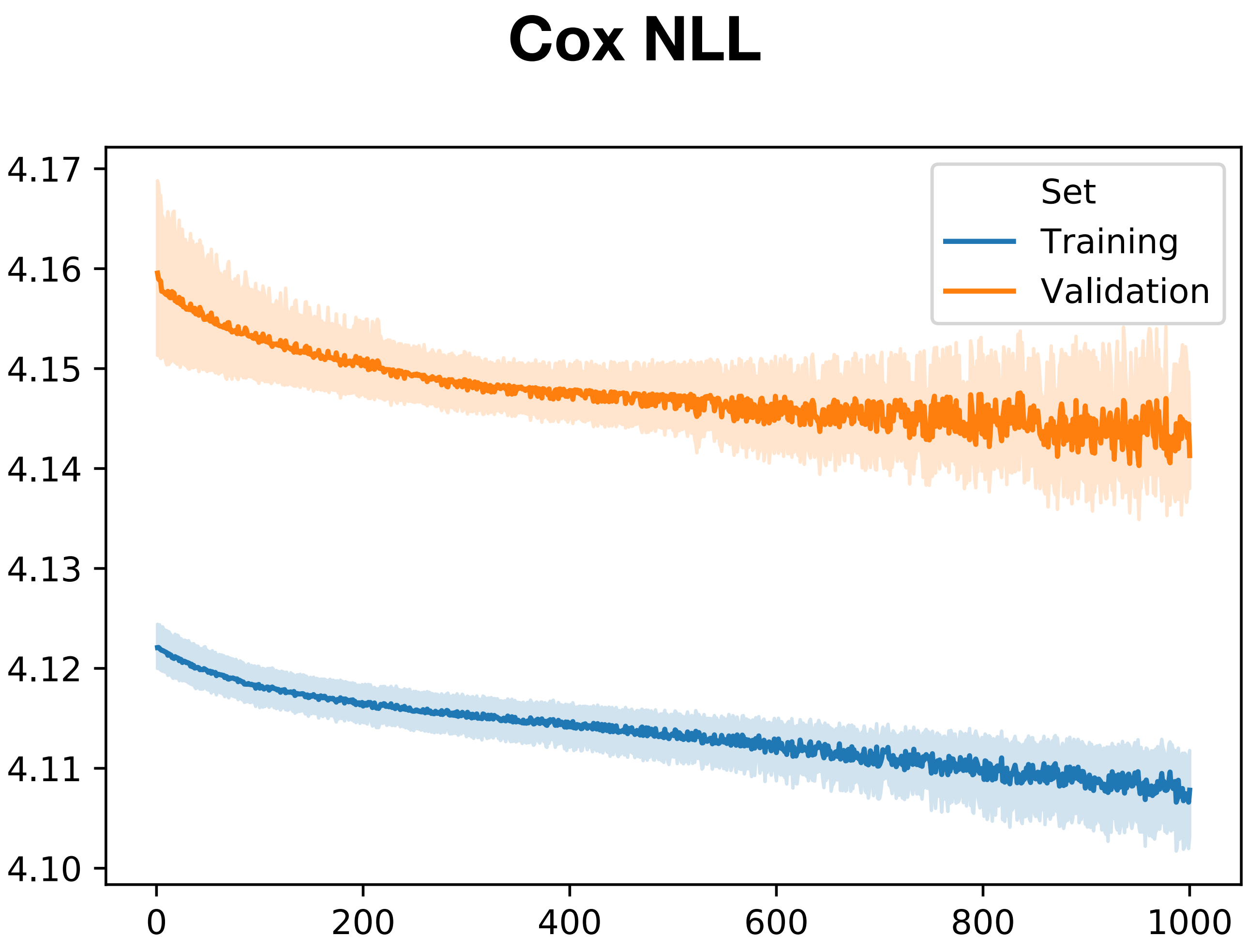


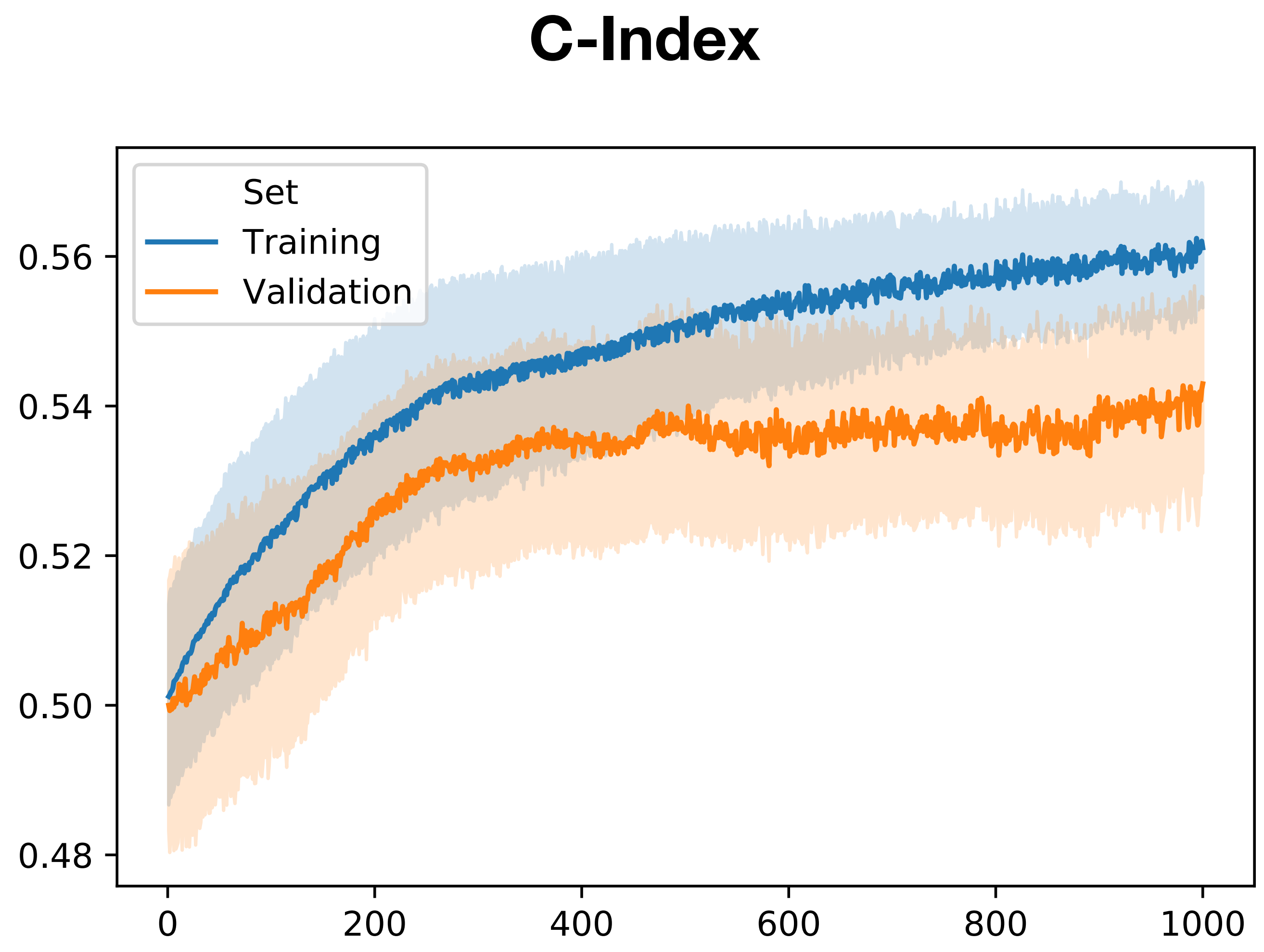
[5] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, “Recurrent Neural Networks for Multivariate Time Series with Missing Values,” Scientific Reports, vol. 8, no. 1, May 2018.

Experimental

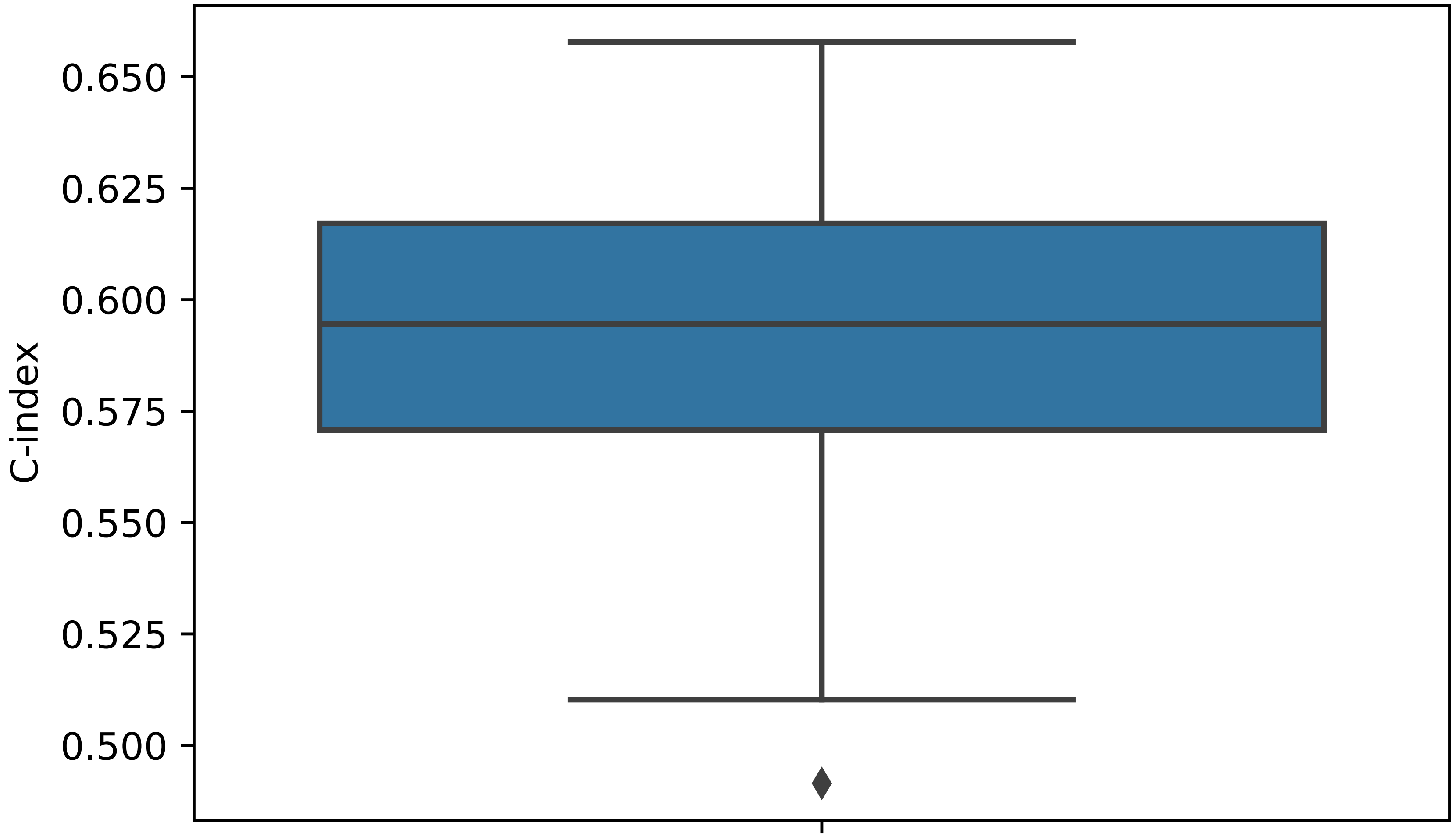
- Hyperparameters:
  - Optimization: learning rate, batch_size, tbptt size, …
  - Network architecture: depth, width, GRU vs LSTM, static features, …
- Random search over 100 configurations
- Manual selection and retraining with averaging

Experimental





Experimental



50% → Random
100% → Perfect

Nurses, linear baseline not significantly different from random

Limitations

- Static features where not helping (but are also harder to train)
- Not a fair comparison to other methodologies
- Difficult to explain

Questions ?