CSC 2541: Machine Learning for Healthcare

Lecture 3: Clinical Time Series Modelling

Professor Marzyeh Ghassemi, PhD
University of Toronto, CS/Med
Vector Institute
Course Reminders!

• Submit the **weekly reflection questions** to MarkUs!

• Sign up for a **paper presentation slot**!

• Homework 1 due next week!

• Think about your projects!
Logistics

• Course website:  
  https://cs2541-ml4h2020.github.io

• Piazza:  
  https://piazza.com/utoronto.ca/winter2020/csc2541

• Grading:  
  • 20% Homework (3 problem sets)  
  • 10% Weekly reflections on Markus (5 questions)  
  • 10% Paper presentation done in-class (sign-up after the first lecture)  
  • 60% course project (an eight-page write up)
Schedule

Jan 9, 2020, Lecture 1: Why is healthcare unique?
Jan 16, 2020, Lecture 2: Supervised Learning for Classification, Risk Scores and Survival

Jan 23, 2020, Lecture 3: Clinical Time Series Modelling
Jan 30, 2020, Lecture 4: Causal inference with Health Data --- Dr. Shalmali Joshi (Vector)
  Problem Set 1 (Jan 31 at 11:59pm)
Feb 6, 2020, Lecture 5: Fairness, Ethics, and Healthcare
  Project proposals (Feb 6 at 5pm)
Feb 13, 2020, Lecture 6: Deep Learning in Medical Imaging -- Dr. Joseph Paul Cohen (MILA)
  Problem Set 2 (Feb 14 at 11:59pm)
Feb 20, 2020, Lecture 7: Clinical NLP and Audio -- Dr. Tristan Naumann (MSR)
Feb 27, 2020, Lecture 8: Clinical Reinforcement Learning
Mar 5, 2020, Lecture 9: Interpretability / Humans-In-The-Loop --- Dr. Rajesh Ranganath (NYU)
  Problem Set 3 (Mar 6 at 11:59pm)
Mar 12, 2020, Lecture 10: Disease Progression Modelling/Transfer Learning -- Irene Chen (MIT)
Mar 19, 2020, Project Sessions/Lecture
Mar 26, 2020, Course Presentations
April 4, 2020, Course Presentations
  Project Report (Apr 3 at 11:59pm)
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
Problem: Hospital decision-making / care planning

Observe Patient Data

“Real-time” Prediction

Of \{Drug / Mortality / Condition\}

By Gap Time
Problem: Hospital decision-making / care planning

**Observe** Patient Data

“Real-time” **Prediction**

Of {Drug/Mortality/Condition}

By Gap **Time**
How Do We Handle **Time**?

- An image gives a snapshot of an object, but a video dictates form!

- We want to model patient risks/treatments/outcomes as they **live**.

- Strategies:
  - Amortize - Make features out of mean, min, max, etc.
  - Stack - Inputs of fixed size, and concatenate.
  - Deal - Use a method that addresses dynamics.

- Focus on dealing in this lecture.
Outline

1. What’s Time Got To Do With It?
   a. **Missingness**
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
What is Missingness?
Missing Data Details

Data can be missing according to several regimes:

- Missing completely at random (MCAR)
- Missing at random (MAR)
- Missing not at random (MNAR)
Missing Data Details

Data can be missing according to several regimes:

- **Missing completely at random (MCAR)**
  - The observed pattern of missingness is independent from the observed or missing values.
- **Missing at random (MAR)**
- **Missing not at random (MNAR)**
Missing Data Details

Data can be missing according to several regimes:

- **Missing completely at random (MCAR)**
  - The observed pattern of missingness is independent from the observed or missing values.

- **Missing at random (MAR)**
  - The observed pattern of missingness is independent from the missing values (but may depend on the observed values).

- **Missing not at random (MNAR)**
Missing Data Details

Data can be missing according to several regimes:

- **Missing completely at random (MCAR)**
  - The observed pattern of missingness is independent from the observed or missing values.

- **Missing at random (MAR)**
  - The observed pattern of missingness is independent from the missing values (but may depend on the observed values).

- **Missing not at random (MNAR)**
  - All bets are off.
Missing Data Details

Data can be missing according to several regimes:

- **Missing completely at random (MCAR)**
  - The observed pattern of missingness is independent from the observed or missing values.

- **Missing at random (MAR)**
  - The observed pattern of missingness is independent from the missing values (but may depend on the observed values).

- **Missing not at random (MNAR)**
  - All bets are off.
Missing Data is Confounding

predict

Pre

Patient 1:

Post

Patient 2:

24h VENT 24h
How do we handle missing data?
Imputation

1. Statistical Timeseries Forecasting: ARMA/ARIMA/ARIMAX, etc.
2. Easy Baselines: Constant infilling, Sample & Hold (+ indicators), Interpolation
3. Traditional Imputation: MICE/3D-MICE, MissForest, Matrix/Tensor Completion
4. Gaussian Processes
5. Advanced neural methods (GRU-D, GANs, etc.)
Figure 2: Example trajectories of six vital signs for a single admission, following imputation using Gaussian processes. Twelve vital signs are jointly modeled by the GP.

GANs for Imputation

GAIN: Missing Data Imputation using Generative Adversarial Nets

Jinsung Yoon\textsuperscript{1}\textsuperscript{*}  James Jordon\textsuperscript{2}\textsuperscript{*}  Mihaela van der Schaar\textsuperscript{1,2,3}
GANs for Imputation

Right: https://thispersondoesnotexist.com/
GAIN: Generative Adversarial Imputation

Figure 1. The architecture of GAIN
Imputation Papers

2. GRU-D: https://www.nature.com/articles/s41598-018-24271-9

<table>
<thead>
<tr>
<th>Model</th>
<th>Classification</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>AUPRC</td>
</tr>
<tr>
<td>Log/LinReg</td>
<td>0.772 ± 0.013</td>
<td>0.303 ± 0.018</td>
</tr>
<tr>
<td>SVM</td>
<td>0.671 ± 0.005</td>
<td>0.300 ± 0.011</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.829 ± 0.007</td>
<td>0.345 ± 0.007</td>
</tr>
<tr>
<td>RF</td>
<td>0.826 ± 0.008</td>
<td>0.356 ± 0.010</td>
</tr>
<tr>
<td>GRU-M</td>
<td>0.831 ± 0.007</td>
<td>0.376 ± 0.022</td>
</tr>
<tr>
<td>GRU-F</td>
<td>0.821 ± 0.007</td>
<td>0.360 ± 0.013</td>
</tr>
<tr>
<td>GRU-S</td>
<td>0.843 ± 0.007</td>
<td>0.376 ± 0.014</td>
</tr>
<tr>
<td>GRU-D</td>
<td>0.835 ± 0.013</td>
<td>0.359 ± 0.025</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.853 ± 0.007</td>
<td>0.418 ± 0.022</td>
</tr>
</tbody>
</table>
Opportunities

1. Improved imputation methods. How do forecasting, GP, or adversarial methods compare to GRU-D/interpolation prediction network? Can we incorporate uncertainty offered by GPs usefully into downstream tasks? Can we make other models offer uncertainty?

2. Can we model the decision process by which clinicians choose what to measure and what to omit? How would this be helpful in downstream tasks? Can this help account for the MNAR nature of healthcare missingness?

3. Can we control for the confounding effects of missingness? Can we learn a model on underlying physiology from retrospective, care-byproduct data?
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
Representation: Why do we care?
Representations define a notion of “similarity”

Closer in “Conceptual Space”

Closer in “Pixel Space”
Representations learn a notion of similarity

Figure 1: Examples of the kernel $k_{j,c}(x, z)$ in (1) with $c = 5$ on three features evaluated on adult ICU population: Hematocrit, Lactic Acid, and Patient Age.
Representations can stabilize changing data

Figure 1: Performance of RF classifiers using Item-Id and Clinically Aggregated representations on mortality (top) and LOS prediction (bottom). Error bars indicate ± standard error.
Representations can stabilize changing data

Representations can join disparate modalities

Figure 1: The overall experimental pipeline. EA: embedding alignment; Adv: adversarial training.

DeepCluster: Why bother with labels?

Fig. 1: Illustration of the proposed method: we iteratively cluster deep features and use the cluster assignments as pseudo-labels to learn the parameters of the convnet.

Representation Learning in Action: Multitask Learning

Multi-task Prediction of Disease Onsets from Longitudinal Lab Tests

Narges Razavian, Jake Marcus, David Sontag
Courant Institute of Mathematical Sciences, New York University

Multitask Learning and Benchmarking with Clinical Time Series Data

Hrayar Harutyunyan\textsuperscript{1}, Hrant Khachatrian\textsuperscript{2,3}, David C. Kale\textsuperscript{1}, Greg Ver Steeg\textsuperscript{1}, and Aram Galstyan\textsuperscript{1}

MoleculeNet: a benchmark for molecular machine learning\textsuperscript{†}

Zhenqin Wu,\textsuperscript{a} Bharath Ramsundar,\textsuperscript{b} Evan N. Feinberg,\textsuperscript{c} Joseph Gomes,\textsuperscript{d}\textsuperscript{a} Caleb Geniesse,\textsuperscript{c} Aneesh S. Pappu,\textsuperscript{b} Karl Leswing\textsuperscript{d} and Vijay Pande\textsuperscript{*a}
Representation Learning in Action: Clustering

Representation Learning in Action: Clustering

Figure 3: tSNE on context vectors of test dataset from BSS model colored by (a) red: positive examples and blue: negative examples, (b) average systemic diastolic blood pressure; and (c) average central venous pressure.

**Representation Learning in Action: Anomaly Detection**

**Fig. 1.** Anomaly detection framework. The preprocessing step includes extraction and flattening of the retinal area, patch extraction and intensity normalization. Generative adversarial training is performed on healthy data and testing is performed on both, unseen healthy cases and anomalous data.

Fig. 2. (a) Deep convolutional generative adversarial network. (b) t-SNE embedding of normal (blue) and anomalous (red) images on the feature representation of the last convolution layer (orange in (a)) of the discriminator.
Key Points for Healthcare

- Representations can normalize.
- Generalization to unseen tasks is critical (e.g., patient subtyping).
- Representations can aid in interpretability.
- Representations can span many modalities.
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
Remember This? Topics Improves Mortality Prediction

- **Forward-facing ICU mortality** prediction with notes.
- **Latent** representations **add** predictive power.
- **Topics** enable accurately **assess risk** from **notes**.
Add Information About Evolution of Signals

- Learn a new latent representation to evaluate multi-dimensional function similarity ($\theta$).

MTGP models capture movements within and between signals. Transform signals into MTGP hyperparameter representation. Compare patient similarly in the new representation.
Learning Single Task Gaussian Processes (STGP)

- Model each signal as a GP task with mean and covariance functions.
  \[
  \tilde{y}_n = g(\tilde{x}_n) \sim \mathcal{GP}\left(m(\tilde{x}_n), k(\tilde{x}_n, \tilde{x}_n')\right)
  \]

- GP’s commonly used to predict at new indices.
  \[
  p(y^*|x^*, x, y) \sim \mathcal{N}\left(m(y^*), \text{var}(y^*)\right)
  \]
  \[
  m(y^*) = K(x, x^*)^\top K(x, x)^{-1} y
  \]
  \[
  \text{var}(y^*) = K(x^*, x^*) - K(x, x^*)^\top K(x, x)^{-1} K(x, x^*)
  \]

- Learn the parameters (θ) of the kernel from data.
  \[
  \text{NLML} = -\log p(y|x, \theta)
  = \frac{1}{2} \log|K| + \frac{1}{2} y^\top K^{-1} y + \frac{n}{2} \log(2\pi)
  \]
Single vs. Multi-task Gaussian Processes

• Assume we have $m$ sets of:
  • Inputs $X_i$
  • Temporal covariance hyperparameters $\theta_{i_t}$
  • Estimated functions $f^i$
  • Noise terms $\sigma^i$
  • Outcomes $y^i$

• We can train $m$ single-task Gaussian process (STGP) (a) or a multi-task Gaussian process (MTGP) to relate the $m$ tasks through all prior variables, with the tasks’ labels $l$ and similarity matrix $\theta_c$ (b).
Learning MTGPs As Representations

- Use an MTGP representation to relate $m$ inputs through $K_t$ and $K_c$.

\[ K_{MT}(X_n, l, \theta_c, \theta_t) = K_c(l, \theta_c) \otimes K_t(X_n, \theta_t) \]

Movement **between** signals

\[ K_c = LL^T \]

\[ L = \begin{bmatrix} \theta_{c,1} & 0 & \ldots & 0 \\ \theta_{c,2} & \theta_{c,3} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{c,k-m+1} & \theta_{c,k-m+2} & \ldots & \theta_{c,k} \end{bmatrix} \]

Movement **within** a signal

\[ K_t = \theta_t^2 \exp \left\{ -\frac{\| x - x' \|^2}{2\theta_t^2} \right\} \]


[2] Carl Rasmussen's minimize.m was used for gradient-based optimization of the marginal likelihood. 45
Estimating Signal in Traumatic Brain Injury Patients

- Intracranial pressure (ICP) and mean arterial blood pressure (ABP) are important indicators of cerebrovascular autoregulation (CA) in traumatic Brain Injury (TBI) patients.

- CA sustains adequate cerebral blood flow\(^1\) and impairment risks secondary brain damage and mortality.\(^2\)

- CA is assessed using a sliding window Pearson’s correlation between the ICP and ABP – the Pressure-Reactivity Index (PRx)\(^3\).

---

TBI Estimation Methodology

• PRx isn’t calculated when either signal is contaminated - evaluate STGPs/MTGPs for interpolation, and MTGPs for PRx estimation.

• Collected data from 35 TBI patients with 24+ hours of ICP and ABP recordings sampled every 10 seconds.

• Selected 30 ten-minute windows where ICP/ABP were free from artifacts and missing values from each patient recording; randomly introduced artificial gaps in both signals (x’s).
MTGP Representations Improve Signal Forecasting and Outcome Prediction

**Performance on Signal Forecasting**

<table>
<thead>
<tr>
<th>Signal</th>
<th>Measure</th>
<th>STGP</th>
<th>MTGP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICP</td>
<td>RMSE</td>
<td>0.91</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>MSLL</td>
<td>0.6</td>
<td>0.45</td>
</tr>
<tr>
<td>ABP</td>
<td>RMSE</td>
<td>2.77</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>MSLL</td>
<td>0.65</td>
<td>0.55</td>
</tr>
</tbody>
</table>

- MTGPs outperform STGPs in signal reconstruction.
- Automatically estimate cerebrovascular autoregulation.

**Performance on Mortality Prediction**

<table>
<thead>
<tr>
<th>Features</th>
<th>Hospital Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave. Topics</td>
<td>0.759</td>
</tr>
<tr>
<td>SAPS-I + MTGP</td>
<td>0.775</td>
</tr>
<tr>
<td>Ave. Topics + MTGP</td>
<td>0.788</td>
</tr>
<tr>
<td>SAPS-I + Ave. Topics + MTGP</td>
<td>0.812</td>
</tr>
</tbody>
</table>

- MTGP hyperparameter representations improve short-term mortality prediction.

* Final cohort consisted of 10,202 patients, with 313,461 notes.
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion
Can We Predict Interventions?

- 34,148 ICU patients from MIMIC-III
- 5 static variables (gender, age, etc.)
- 29 time-varying vitals and labs (oxygen saturation, lactate, etc.)
- All clinical notes for each patient stay
Raw Physiology vs “Words” Embedding

<table>
<thead>
<tr>
<th>patient</th>
<th>hours in</th>
<th>glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>NaN</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>NaN</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>101.2344</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>patient</th>
<th>hours in</th>
<th>glucose_-2</th>
<th>glucose_-1</th>
<th>glucose_0</th>
<th>glucose_1</th>
<th>glucose_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- Many values are missing!
### Raw Physiology vs “Words” Embedding

#### Numerical

<table>
<thead>
<tr>
<th>patient</th>
<th>hours in</th>
<th>glucose</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>NaN</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>NaN</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>101.2344</td>
</tr>
</tbody>
</table>

#### Physiological Words

<table>
<thead>
<tr>
<th>patient</th>
<th>hours in</th>
<th>glucose_{-2}</th>
<th>glucose_{-1}</th>
<th>glucose_0</th>
<th>glucose_1</th>
<th>glucose_{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

-1

- Many values are missing!
- Z-score existing variables, rounding to the nearest int.
Raw Physiology vs “Words” Embedding

- Many values are missing!
- Z-score existing variables, rounding to the nearest int.
- Convert each z-score into its own binary column.
Raw Physiology vs “Words” Embedding

- Many values are missing!
- Z-score existing variables, rounding to the nearest int.
- Convert each z-score into its own binary column.

A row of all zeros indicates a missing value at that hour.
Many Ways to Model, What Do We Learn?

**LSTM**

- 2 Layer/512 node LSTM with sequential hourly data; at end of window, use the final hidden state to predict output.

**SSAM**

- Learn model parameters over patients with variational EM.
- Infer hourly distribution over hidden states with HMM DP (fwd alg.).
- Logistic regression (with label-balanced cost function)

**CNN**

- CNN for temporal convolutions at 3/4/5 hours, max-pool, combine the outputs, and run through 2 fully connected layers for prediction.
- Predict onset in advance
Many Ways to Model, What Do We Learn?

SSAM

2 Layer/512 node LSTM with sequential hourly data; at end of window, use the final hidden state to predict output.

CNN for temporal convolutions at 3/4/5 hours, max-pool, combine the outputs, and run through 2 fully connected layers for prediction.
Many Ways to Model, What Do We Learn?

SSAM

Covered last week!

LSTM

2 Layer/512 node LSTM with sequential hourly data; at end of window, use the final hidden state to predict output.

CNN

CNN for temporal convolutions at 3/4/5 hours, max-pool, combine the outputs, and run through 2 fully connected layers for prediction.
RNNs on Sequences

To model sequences, we need:

1. To deal with variable-length sequences
2. To maintain sequence order
3. To keep track of long-term dependencies
4. To share parameters across the sequence

Let’s turn to recurrent neural networks.
Example Network
Example Network

Let's take a look at this one hidden unit.
RNNS remember their previous state:

\[ x_0 : \text{"it"} \]

\[ W \rightarrow s_0 \rightarrow s_1 \]

\[ U \]

\[ x_0 : \text{vector representing first word} \]
\[ s_0 : \text{cell state at } t = 0 \text{ (some initialization)} \]
\[ s_1 : \text{cell state at } t = 1 \]

\[ s_1 = \tanh(Wx_0 + Us_0) \]

\[ t = 0 \]

\[ W, U : \text{weight matrices} \]
RNNS remember their previous state:

$x_1$: “was”

\(s_1\): cell state at \(t = 1\)

\(s_2\): cell state at \(t = 2\)

\[s_2 = \tanh(Wx_1 + Us_1)\]

\(x_1\): vector representing second word

\(W, U\): weight matrices
“Unfolding” the RNN across time:
“Unfolding” the RNN across time:

\[
\begin{align*}
W & x_0 \quad s_0 \quad U \\
W & x_1 \quad s_1 \quad U \\
W & x_2 \quad s_2 \quad U \\
\cdots
\end{align*}
\]

notice that we use the same parameters, W and U
“Unfolding” the RNN across time:

$s_n$ can contain information from all past timesteps.
Why do LSTMs help?

1. Forget gate allows information to pass through unchanged
2. Cell state is separate from what’s outputted
3. $s_j$ depends on $s_{j-1}$ through addition!
   → derivatives don’t expand into a long product!
Predict Onsets of Interventions

- Delay prediction by 6-hour gap time.

- Attempt to predict onest, weaning, staying off, staying on.
NNs Do Well; Improved Representation Helps

<table>
<thead>
<tr>
<th>Task</th>
<th>Model</th>
<th>VENT</th>
<th>NI-VENT</th>
<th>VASO</th>
<th>COL BOL</th>
<th>CRYST BOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset AUC</td>
<td>Baseline</td>
<td>0.60</td>
<td>0.66</td>
<td>0.43</td>
<td>0.65</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>LSTM Raw</td>
<td>0.61</td>
<td>0.75</td>
<td>0.77</td>
<td>0.52</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>LSTM Words</td>
<td>0.75</td>
<td>0.76</td>
<td>0.76</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>0.62</td>
<td>0.73</td>
<td>0.77</td>
<td>0.70</td>
<td>0.09</td>
</tr>
<tr>
<td>Wean AUC</td>
<td>Baseline</td>
<td>0.83</td>
<td>0.71</td>
<td>0.74</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Raw</td>
<td>0.90</td>
<td>0.80</td>
<td>0.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Words</td>
<td>0.90</td>
<td>0.81</td>
<td>0.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>0.91</td>
<td>0.80</td>
<td>0.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stay On AUC</td>
<td>Baseline</td>
<td>0.50</td>
<td>0.79</td>
<td>0.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Raw</td>
<td>0.96</td>
<td>0.86</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Words</td>
<td>0.97</td>
<td>0.86</td>
<td>0.95</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>0.96</td>
<td>0.86</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stay Off AUC</td>
<td>Baseline</td>
<td>0.94</td>
<td>0.71</td>
<td>0.93</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Raw</td>
<td>0.95</td>
<td>0.86</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Words</td>
<td>0.97</td>
<td>0.86</td>
<td>0.95</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>0.95</td>
<td>0.86</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Micro AUC</td>
<td>Baseline</td>
<td>0.72</td>
<td>0.72</td>
<td>0.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Raw</td>
<td>0.86</td>
<td>0.82</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LSTM Words</td>
<td>0.90</td>
<td>0.82</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CNN</td>
<td>0.86</td>
<td>0.81</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Representations with “physiological words” for missingness significantly increased AUC for interventions with the lowest proportion of examples.

Deep models perform well in general, but words are important for ventilation tasks.
Physiological data were more important for the more invasive interventions.

Clinical note topics were more important for less invasive tasks.
Convolutional Filters Target Short-term Trajectories

Most differentiated features of 10 real patient trajectories that are highest/lowest activating for each task.

**Ventilation**
- Higher diastolic blood pressure, respiratory rate, and heart rate, and lower oxygen saturation: **Hyperventilation**

**Vasopressor**
- Decreased systolic blood pressure, heart rate and oxygen saturation rate:  
  **Altered peripheral perfusion or stress hyperglycemia**

**Non-inv. Vent**
- Decreased creatinine, phosphate, oxygen saturation and blood urea nitrogen:  
  **Neuromuscular respiratory failure**
Convoluotional Filters Target Short-term Trajectories

- “Hallucinations” give insight into underlying properties of the network.
- The trajectories are made to maximize the output of the model, (do not correspond to physiologically plausible trajectories).

**Blood pressure** drops are maximally activating for **vasopressor onset**.

**Respiratory rate** decreasing is maximally activating for **ventilation onset**.
Outline

1. What’s Time Got To Do With It?
   a. Missingness
   b. Representation

2. Case Study 1: MTGPs for Mortality Prediction and TBI

3. Case Study 2: RNNs/CNNs for Intervention Onset Prediction

4. Project Discussion