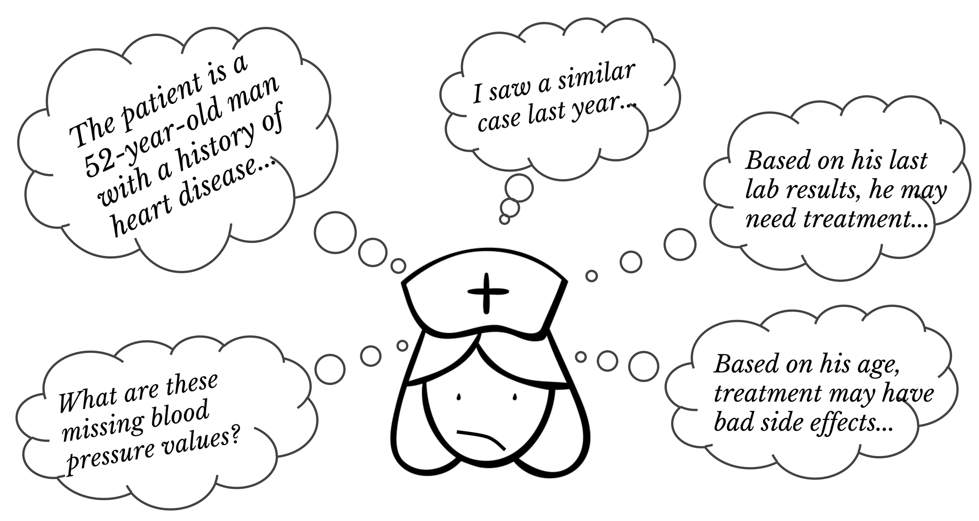
Clinical Event Prediction and Understanding with Neural Networks

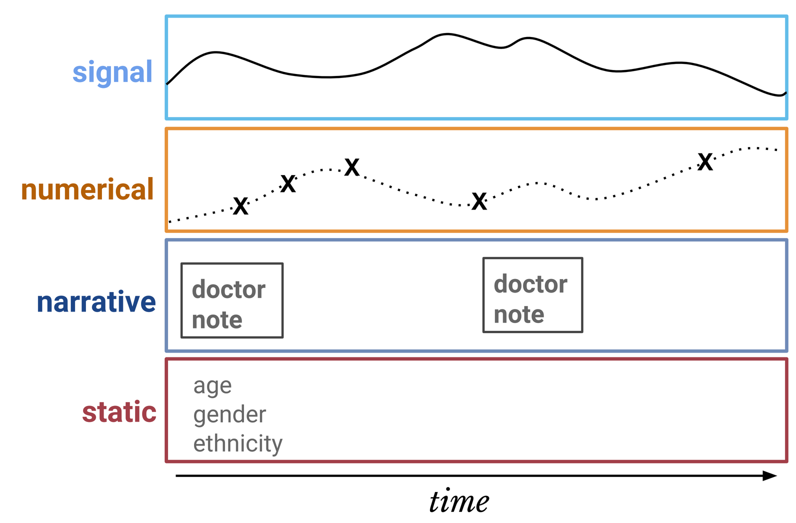
Harini Suresh, Nathan Hunt, Alistair Johnson, Leo Anthony Celi,
Peter Szolovits, Marzyeh Ghassemi
Massachusetts Institute of Technology

The Setting: Currently in the ICU...



- Many standard treatments can be detrimental
- Important to administer, but only when necessary

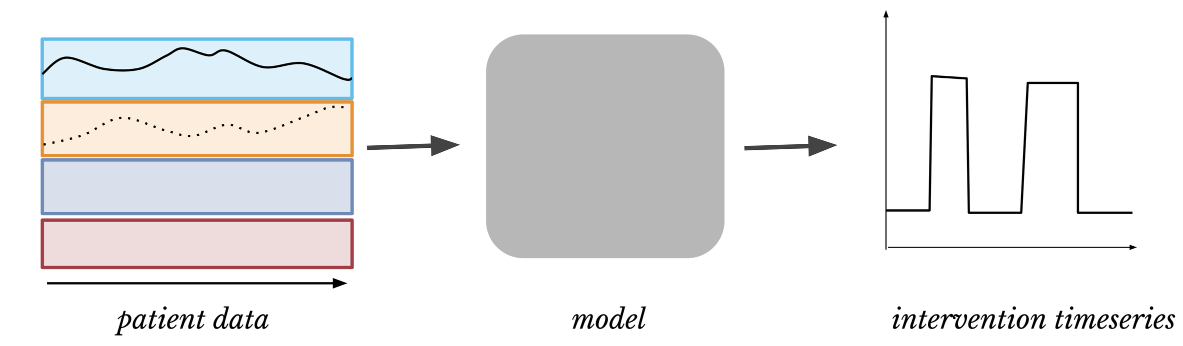
The Problem: Patient Timeseries Data



- Lots of data modalities
- Irregularly sampled and varying length
- Complex and time-dependent relationships

The Solution: Latent Representations of Patient State

Patients react differently to interventions based on their **underlying physiological state**.



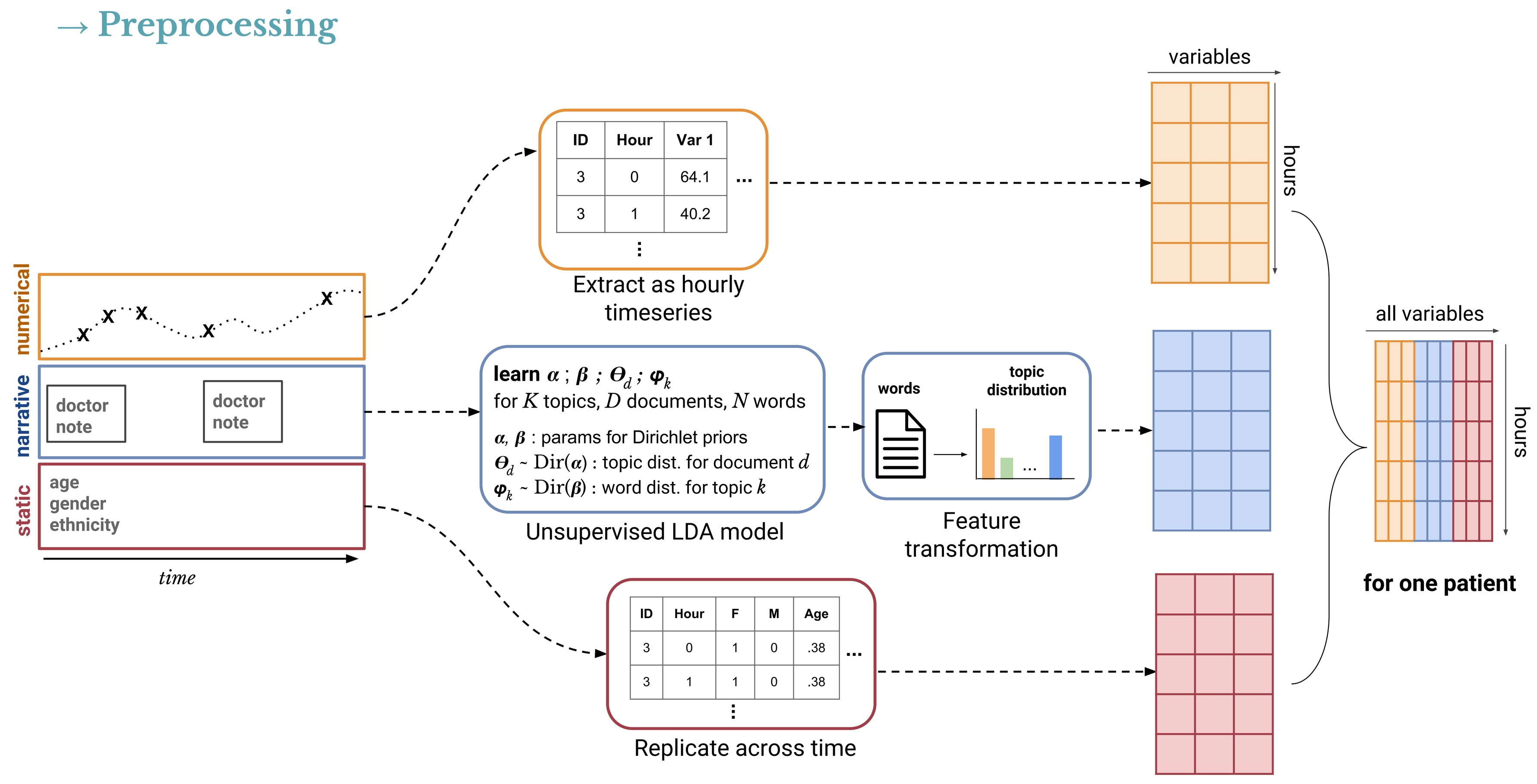
- Use **multimodal** data
- Predict all parts of patient waveform
- State embedded in neural network model is more **robust** and **generalizable**
- Avoids building separate models for many different tasks (i.e. onset vs weaning)

The Approach: Outcome Classification with Deep Networks

→ Data

34,148 ICU stays from Beth Israel Deaconess Hospital (MIMIC III) containing lab results, physiological signals, timestamps, and static information

→ Preprocessing



→ Physiological Words

To create a physiological word, features become categorical variables representing standard deviations from the mean versus numerical values that are forward-filled, mean-imputed and normalized.

raw X

patient	hours in	glucose
3	1	NaN
3	2	NaN
3	3	101.2344
...	...	...

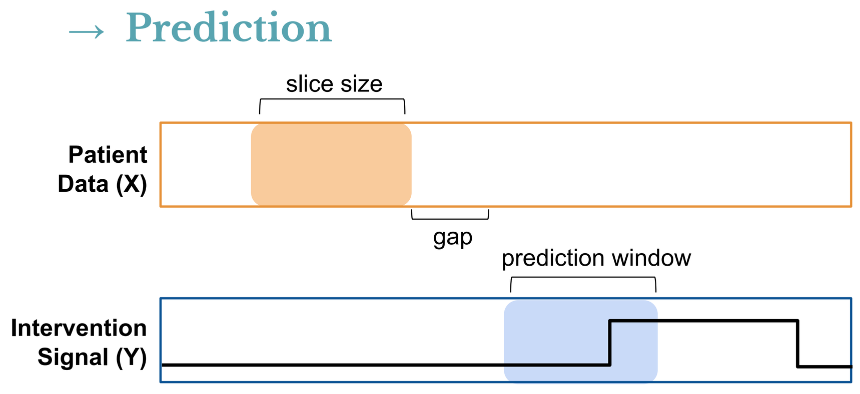
→

phys_words

patient	hours in	glucose_-2	glucose_-1	glucose_0	glucose_1	glucose_2
3	1	0	0	0	0	0
3	2	0	0	0	0	0
3	3	0	1	0	0	0
...	...	...	...	...	...	...

The categorical *phys_words* features are better able to deal with missing data and avoid noise introduced by imputation.

→ Prediction



→ Outcomes

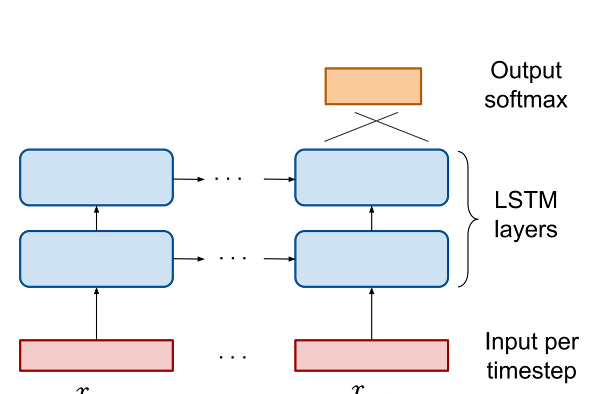
1. Onset of intervention
2. Stay on intervention
3. Stay off intervention
4. Weaning from intervention

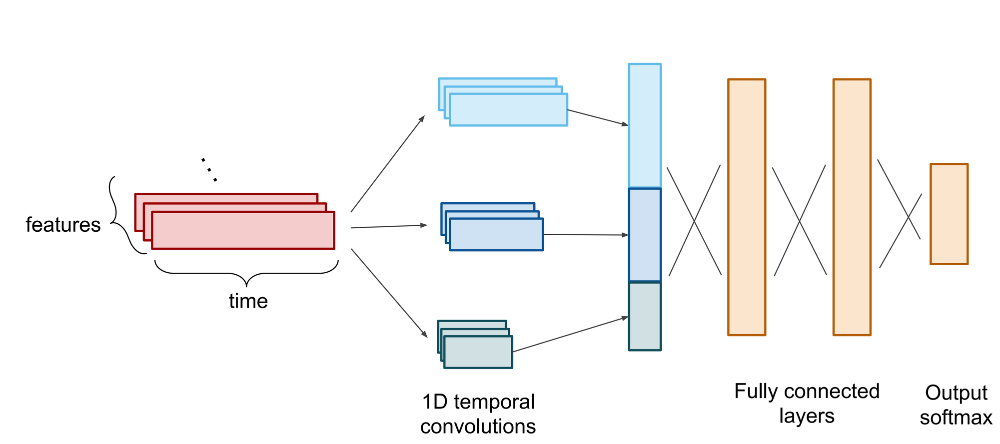
Given patient data timeseries in a fixed-length **slice**, predict outcome in a **prediction window** after a **gap time**.

→ Models

LSTM: 2-layers of 512 nodes each, with output prediction as softmax after last timestep.

CNN: 3 sizes of 1D temporal convolutions with 64 filters each, over input with features as channels. Followed by fully connected layers before output softmax.





The Results: AUCs and Interpretability

Outcome AUCs: compare the AUCs for each outcome along with the macro average across classes.

area-under-ROC

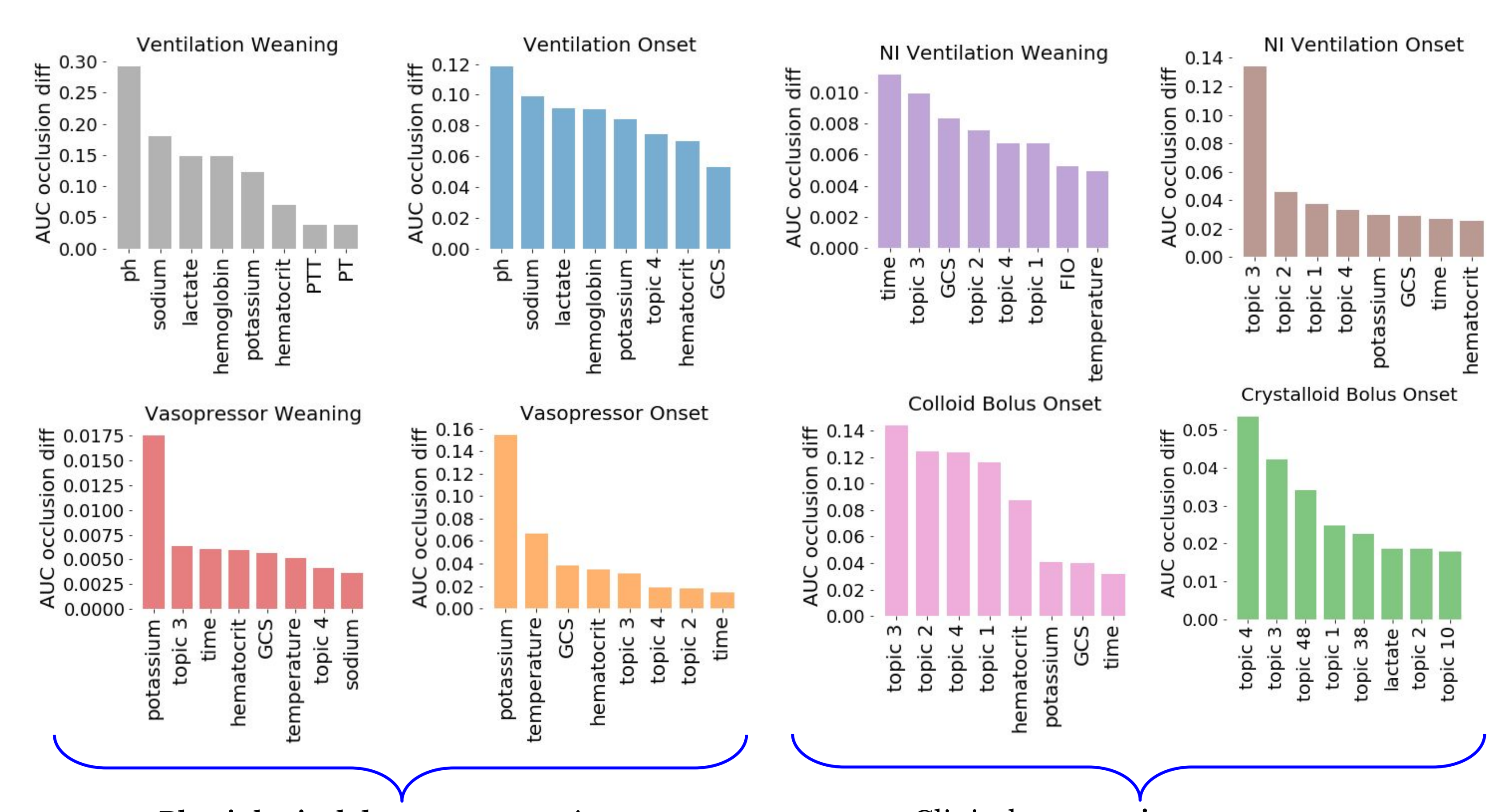
Task	Model	Intervention Type: VENT	NI-VENT	VASO	COL BOL	CRYS BOL
Onset AUC	Baseline	0.60	0.66	0.43	0.65	0.67
	LSTM Raw	0.61	0.75	**0.77**	0.52	0.70
	LSTM Words	$\boxed{\textbf{0.75}}$	**0.76**	0.76	$\boxed{\textbf{0.72}}$	**0.71**
	CNN	0.62	0.73	**0.77**	0.70	0.69
Wean AUC	Baseline	0.83	0.71	0.74	-	-
	LSTM Raw	0.90	0.80	**0.91**	-	-
	LSTM Words	0.90	**0.81**	**0.91**	-	-
	CNN	**0.91**	0.80	**0.91**	-	-
Stay On AUC	Baseline	0.50	0.79	0.55	-	-
	LSTM Raw	0.96	**0.86**	**0.96**	-	-
	LSTM Words	**0.97**	**0.86**	0.95	-	-
	CNN	0.96	**0.86**	**0.96**	-	-
Stay Off AUC	Baseline	0.94	0.71	0.93	-	-
	LSTM Raw	0.95	**0.86**	**0.96**	-	-
	LSTM Words	**0.97**	**0.86**	0.95	-	-
	CNN	0.95	**0.86**	**0.96**	-	-
Macro AUC	Baseline	0.72	0.72	0.66	-	-
	LSTM Raw	0.86	**0.82**	**0.90**	-	-
	LSTM Words	$\boxed{\textbf{0.90}}$	**0.82**	0.89	-	-
	CNN	0.86	0.81	**0.90**	-	-

$^{*}\ \mathrm{AUC}_{macro} = \frac{1}{n}\sum_{\lambda=1}^{n}\mathrm{AUC}(tp_\lambda, fp_\lambda, tn_\lambda, fn_\lambda)$

Physiological words significantly increased AUC for interventions with the lowest proportion of examples.

Deep models perform well in general, but words are important for ventilation tasks.

Feature Importances through Occlusion: compare how predictions change after replacing a given feature with random noise, in order to compare feature importance.

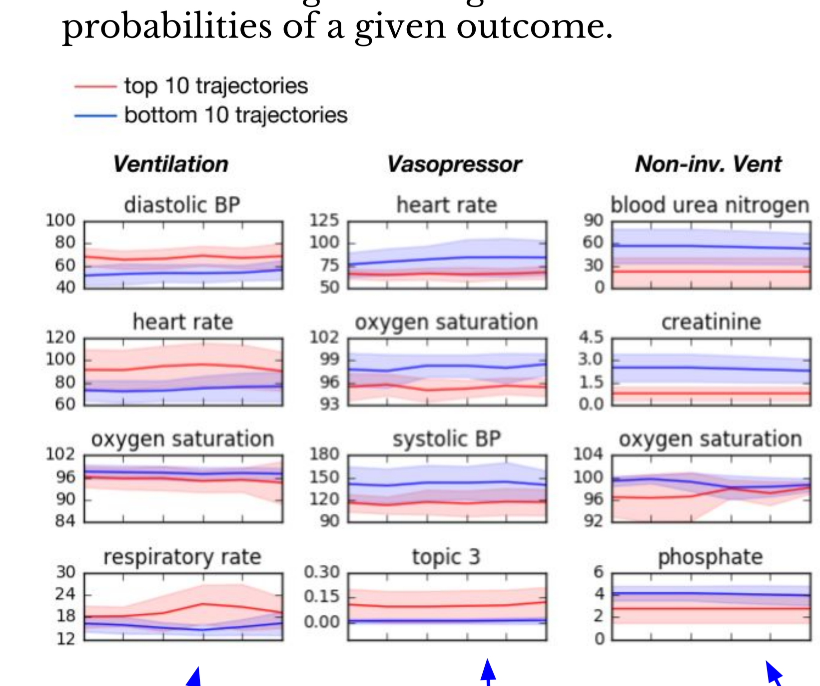


Physiological data were more important for the more **invasive** interventions.

Clinical **note topics** were more important for **less invasive** tasks.

Trend Importance through Visualizing CNN Filters:

1) plot the real patient trajectories to which the model assigns the highest and lowest probabilities of a given outcome.

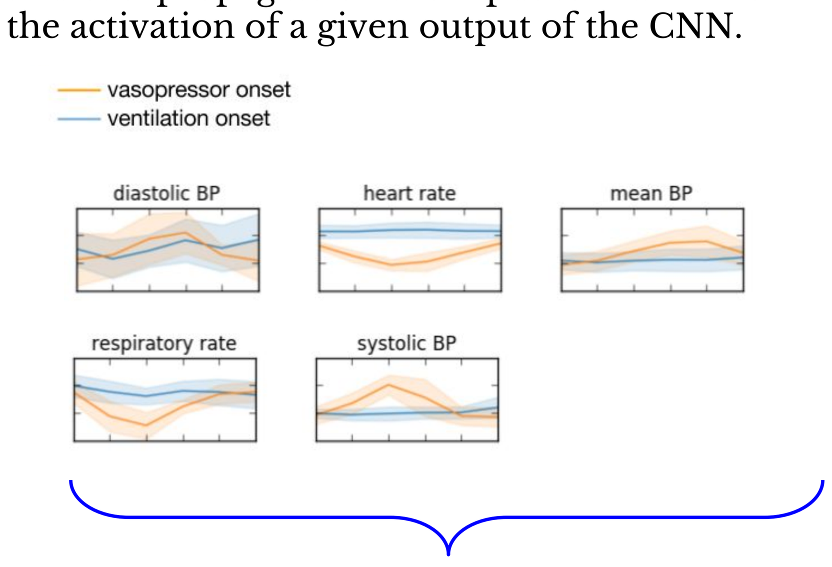


Higher diastolic blood pressure, respiratory rate, and heart rate, and lower oxygen saturation: **hyperventilation**

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

Decreased creatinine, phosphate, oxygen saturation and blood urea nitrogen: **Neuromuscular respiratory failure**

2) start with an input image of random noise, and backpropagate on the input to maximize the activation of a given output of the CNN.



- **Blood pressure** drops are maximally activating for vasopressor onset.
- **Respiratory rate** decreasing is maximally activating for ventilation onset.