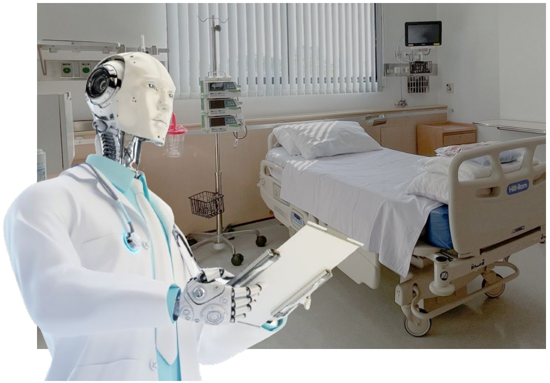
Master Thesis Presentation

Offline Reinforcement Learning with Self-Supervised State Representations for Hemodynamic Support at the ICU

Thomas Bellucci

M Artificial Intelligence (VU)

5 Sept. 2022 - 5 Feb. 2023



Collaboration between…

Amsterdam UMC
Universitair Medische Centra

QDA

Quantitative Data
Analytics Group

VU

Big picture idea

"Use ***Deep Reinforcement Learning*** (DRL) to learn ***optimal control strategies*** for delivering ***treatments*** to ***critically-ill patients*** at the Intensive Care Unit (ICU)"

Why RL?

Many severe conditions treated at ICU:

- Pneumonia (“longontsteking”)
- COVID-19
- Cardiac infarction (“hartinfarct”)
- ***Sepsis*** (severe infection w/ organ failure)

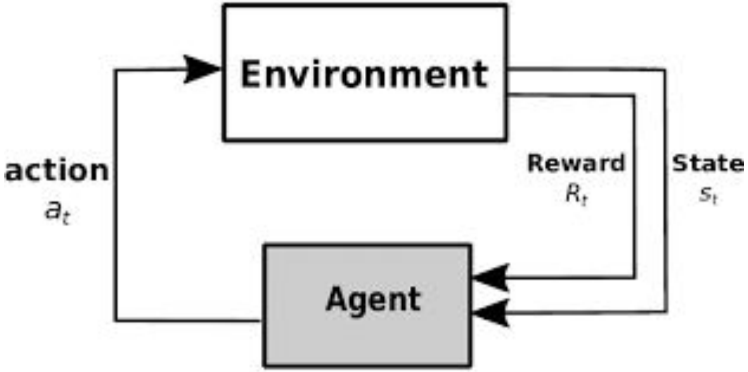
No consesus on best treatment practice

- How to optimally treat patients often remains unclear

Why RL?

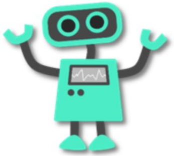
Many severe conditions treated at ICU:

- Pneumonia (“longontsteking”)
- COVID-19
- Cardiac infarction (“hartinfarct”)
- ***Sepsis*** (severe infection w/ organ failure)



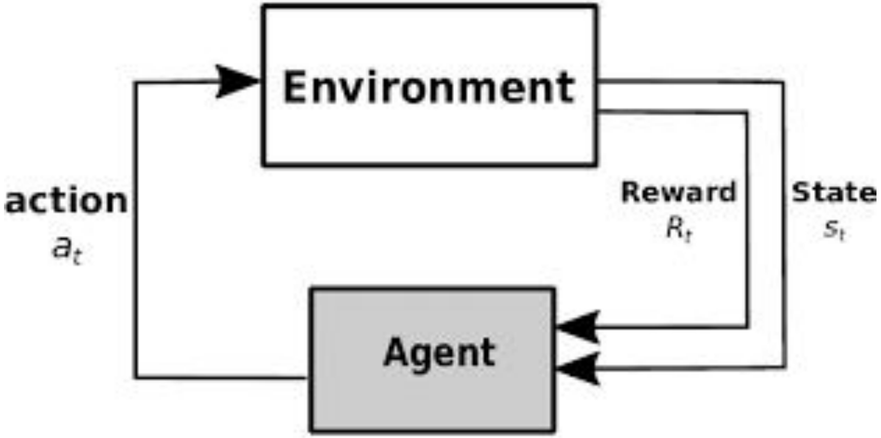
No consesus on best treatment practice

- How to optimally treat patients often remains unclear

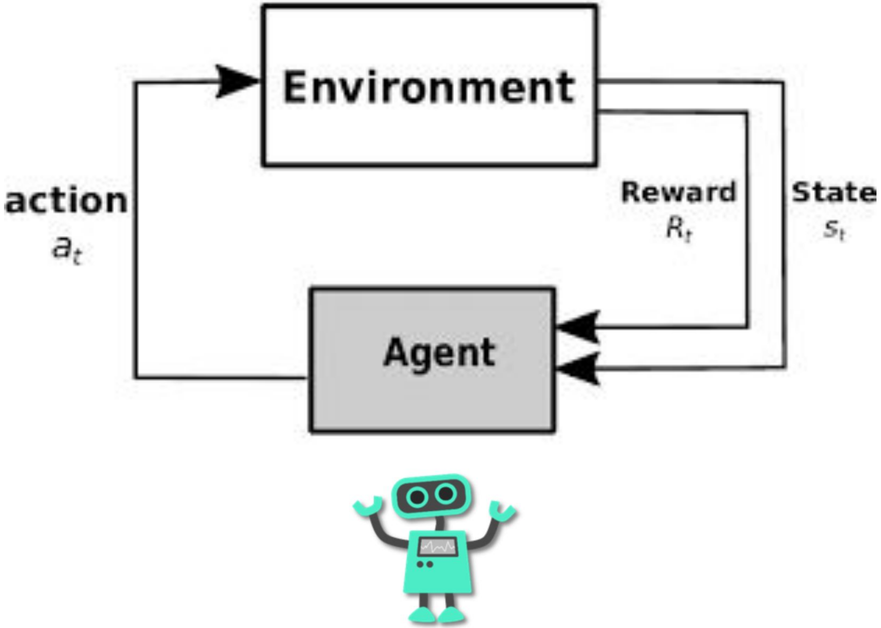


Question: “Can we train RL agent to find an optimal treatment delivery strategy for, e.g. sepsis?”

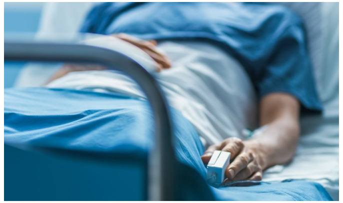
Treatment as MDP

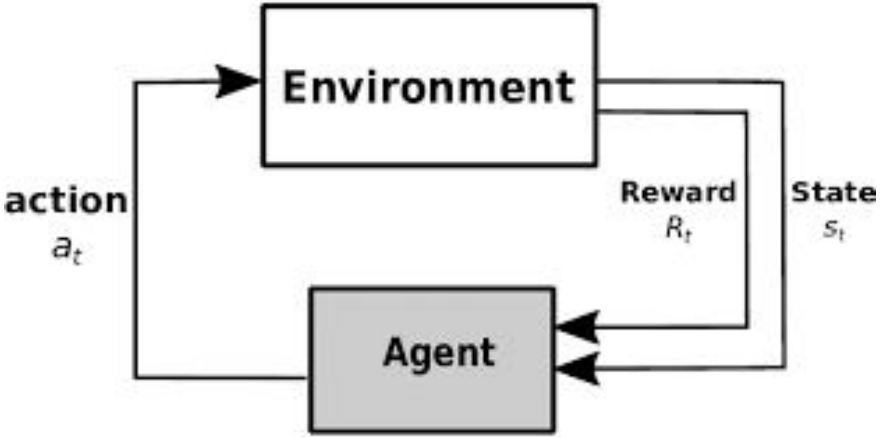


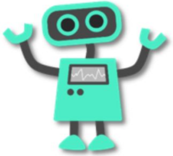
Treatment as MDP



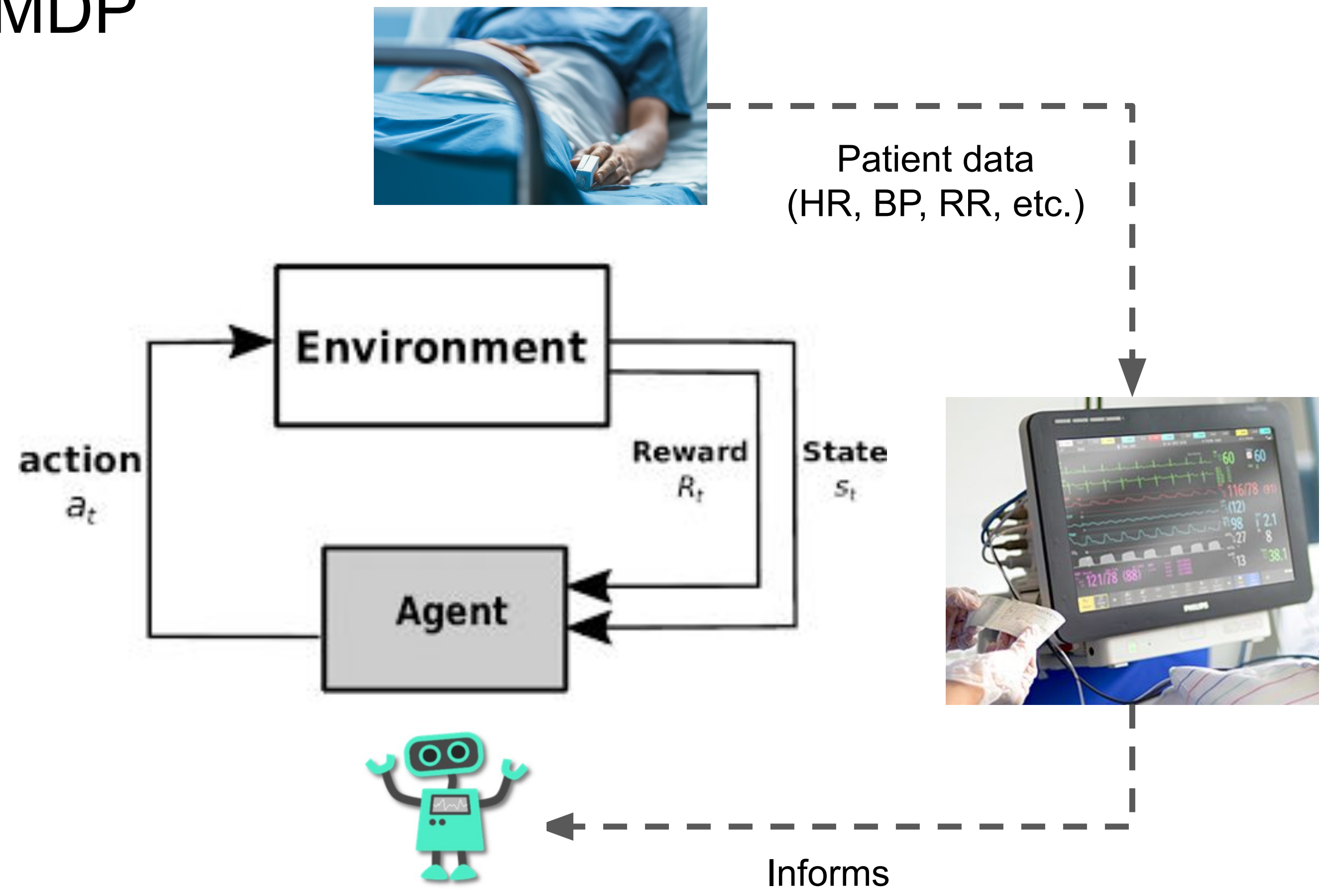
Treatment as MDP



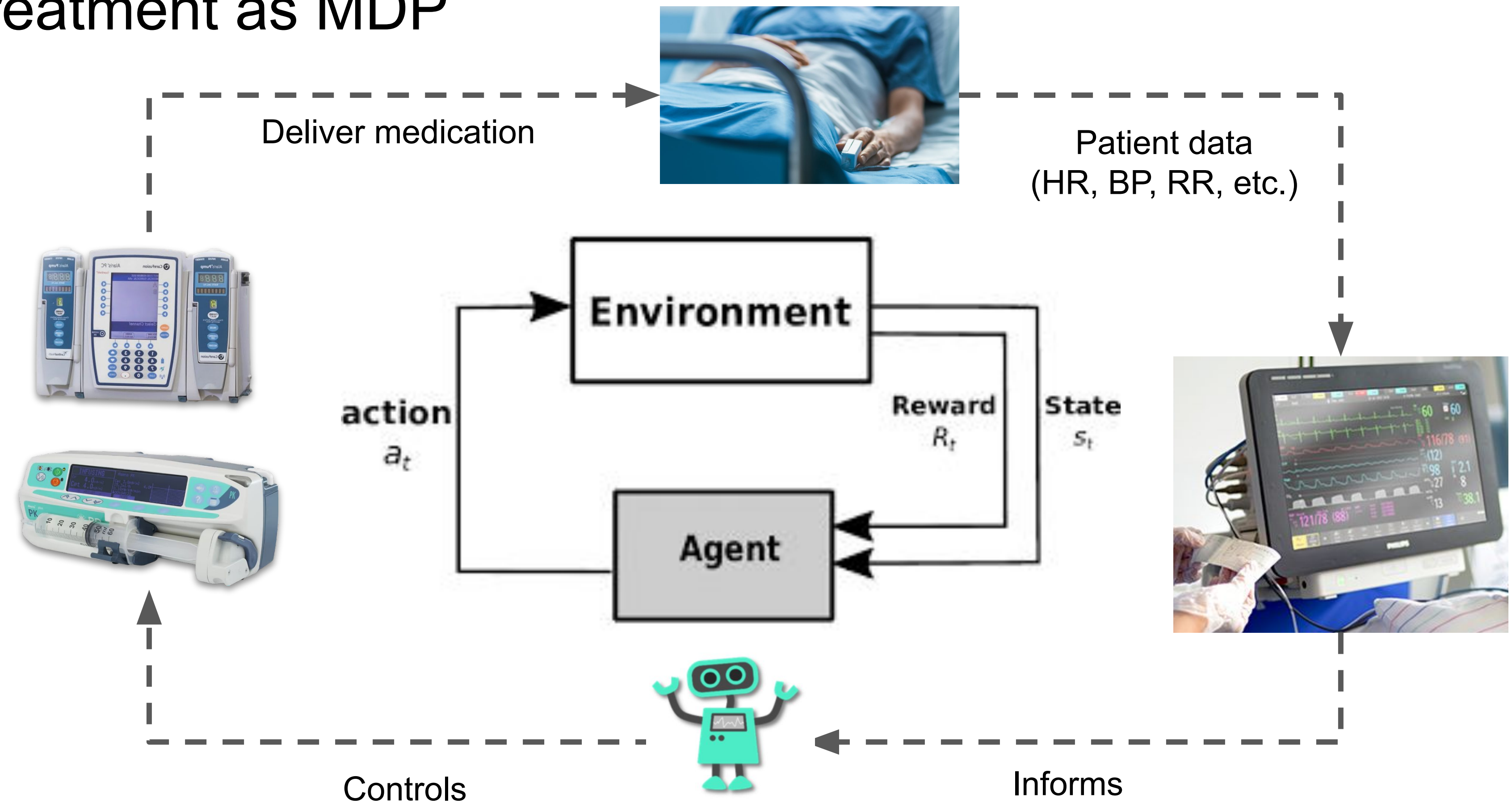




Treatment as MDP



Treatment as MDP



Open problems

Goal: Train agent to choose treatments (actions) which maximize likelihood of patient survival

But:

1. **RL is an "online trial-and-error" paradigm**
   - Cannot train directly from interaction(s) with patients (cf. "learning-on-the-job")
   - High-fidelity "patient sims" do not exist

2. **How to define the states/rewards/actions**?
   - States must include all information possibly relevant from the patient's medical history

Method

Model: Dueling Double Deep Q-Learning + MLP controller

Data: Offline dataset gathered by UMC physicians (±12.000 patients)

- Patient data over time + treatment doses + discharge info

States: Learnt through *self-supervised forward modeling* (see [Lesort et al., 2018])

- No need to elicit knowledge from medical experts
- Train RNN/transformer encoder to extract useful features from patient history to predict $\mathbf{s}_{t+1}$
- i.e. Forward(Enc($\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_t$), $\mathbf{a}_t$) → $\mathbf{s}_{t+1}$

Actions: 2 medications, each 5 levels → 25 discrete actions

Reward: Did patient survive (within ICU)?

Evaluation and Conclusions

Evaluated agent using held-out offline data (±1500 patients)

- Off-policy policy evaluation (OPE)
- Feature attribution (how do measurements influence treatment)
- Manual inspection of model actions which deviate severely from physician

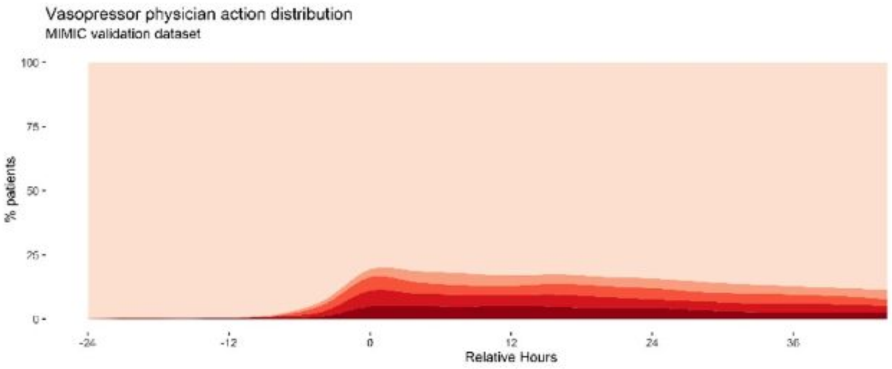
Results: Expected mortality rate ↓**6%** rel. to physician’s policy (baseline **>11%**)

Explanations: Model paid most attention to known biomarkers for disease severity

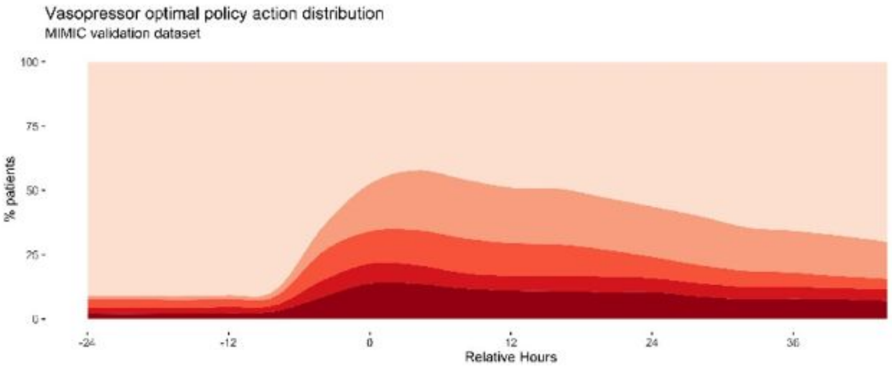
Confirmed recently-discovered strategies:

- Higher doses of vasopressors
- Increasing then decreasing vasopressors increases chances of survival

Evaluation and Conclusions



Physician treatments (in data)



Agent’s treatments

Thanks

Questions?