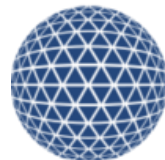




Session-2: Explainable AI Demo on HPC – X-Vent Case Study (Study A)

Miloš Nenadović, Valentina Janev, Dejan Paunović,

On behalf of the IntelliLung consortium



InfAI[®]
Institut für Angewandte Informatik



**Universitätsklinikum
Carl Gustav Carus**
DIE DRESDNER.



AI-based Solution

Decision support system (DSS): Learn from deviations as well as the patient's condition in an attempt to improve the combination of settings and accomplish lung support with reduced risk of damage.

Availability: An awake system that does not get tired or need sleep.

Capacity: A system that can memorize and take into account the history of the patient condition.



an attempt to improve the combination of settings and accomplish lung support with reduced risk of damage!

Methodological Challenges for Building AI-DSS

- Challenges:
 - Limited data for mechanical ventilation.
 - AI training must be offline.
 - No suitable simulation model for exploration.
- Aim:
 - Accurate prediction model for mechanically ventilated patients.
 - Explainable AI outputs that clinicians can easily understand.
 - Improvement in AI recommendations.

How can we improve the decision-support AI agent to reduce high mortality rates in ICUs when patients are mechanically ventilated?

Background

Model-free RL

- Directly learns the policy from acquired static data.
- Can only optimize within the distribution of seen actions.
- May overestimate for out-of-distribution actions.

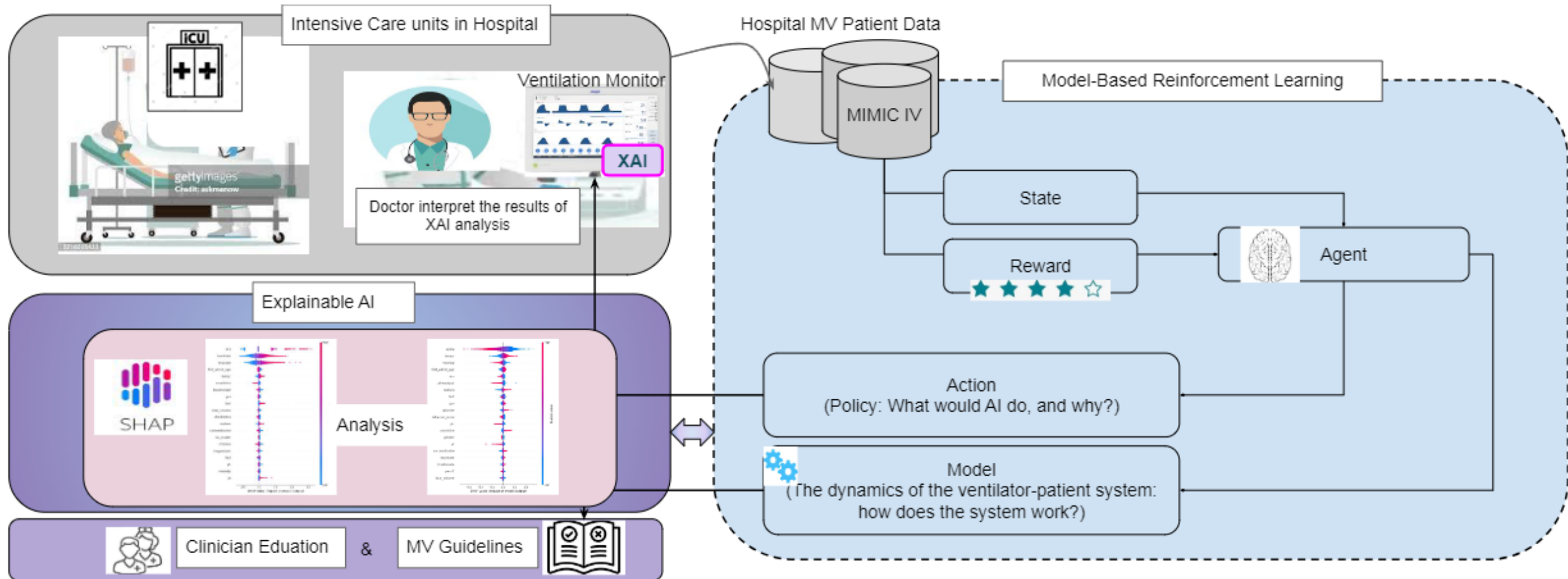
$$Q^{MF}(s, a) = \mathbb{E}[R(s, a) + \gamma \max_{a'} Q^{MF}(s', a')].$$

Model-based RL

- Builds a model using training data to predict future states and rewards.
- Can simulate new data trajectories and improve generalization.
- Efficient with data, but highly dependent on the model.

$$Q^{MB}(s, a) = \mathbb{E}_{(s', r) \sim \mathcal{D} \oplus \mathcal{M}(s, a)} [r + \gamma \max_{a'} Q^{MB}(s', a')].$$

X-Vent in the ICU for mechanical ventilation



Core Algorithms behind X-Vent

Conservative Q-Learning (CQL)

- Penalizes overestimation of Q-values for out-of-distribution actions.
- Minimizes Q-values for actions not seen in the dataset.
- Used when exploration cannot be performed.

Conservative Offline Model-Based Policy Optimization (COMBO)

- Expands the CQL algorithm by introducing a model.
- The model provides additional data trajectories, adding quantity and diversity.

Example Actions

Three values influencing oxygenation and CO₂ elimination:

- **Tidal Volume (V_t):** The amount of air a person breathes in or out during a normal breath.
- **Positive End-Expiratory Pressure (PEEP):** The pressure in the lungs at the end of an exhale, used to keep airways open.
- **Inspiratory Oxygen Fraction (FiO₂):** The percentage of oxygen in the air that a patient breathes in during inhalation.

Performance Evaluation

Estimation using FQE

- an Off-Policy Evaluation method.
- iteratively computes state-action pair Q-values.
- Policy performance is estimated by using the average value of the initial states.

$$Q_k = \arg \min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n (f(x_i, a_i) - y_i)^2$$

	Sparse reward	Heuristic reward
Baseline	0.729 ± 0.002	0.743 ± 0.005
X-Vent	0.793 ± 0.004	0.776 ± 0.006

SHAP: From Feature Importance to Cooperative Game Theory

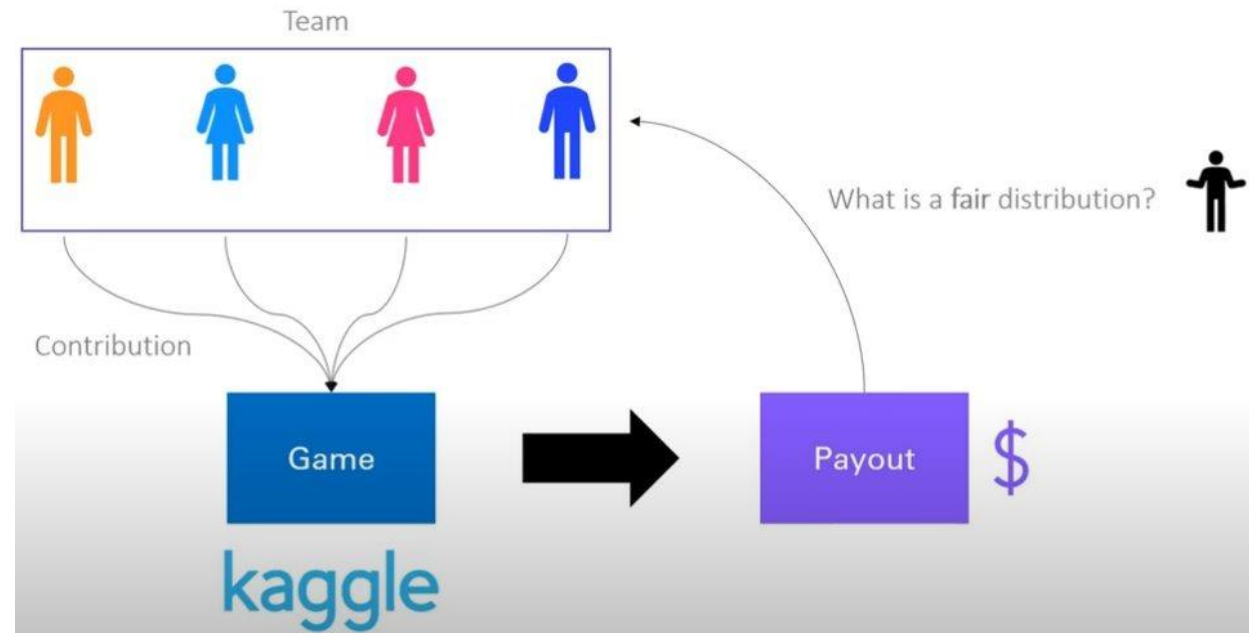
- SHAP = SHapley Additive exPlanations
- Problem: Decompose nonlinear model prediction $f(x)$ into feature-level contributions.
- Heuristic methods (gradients, permutation, LIME) lack fairness & consistency guarantees.

Game-theoretic mapping:

- Players \rightarrow Features
- Coalition \rightarrow Feature subset S
- Payout \rightarrow Model prediction $f(x)$

Additive explanation model:

- $f(x) = \phi_0 + \sum \phi_i$
- $\phi_0 = E[f(x)]$ (baseline)



Shapley Values: Formal Definition & Axioms

Shapley value for feature i :

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

Interpretation: Average marginal contribution across all permutations.

Axioms:

- Efficiency (Local Accuracy): $\sum \phi_i + \phi_0 = f(x)$
- Symmetry
- Zero contribution if no effect
- Additivity

Explainability Strategy using SHAP

- Model: $\pi(a \mid s)$ or deterministic policy $a = f(s)$.
- Input: Physiological state vector s .
- Output: Treatment action.

SHAP computes $\phi_i(s)$: contribution of variable i to action decision.

Interpretation:

- $\phi_i > 0 \rightarrow$ pushes action upward
 - $\phi_i < 0 \rightarrow$ suppresses action
 - $|\phi_i| \rightarrow$ influence magnitude
-
- Explains learned policy behavior, not causality.

Explainability Strategy using SHAP

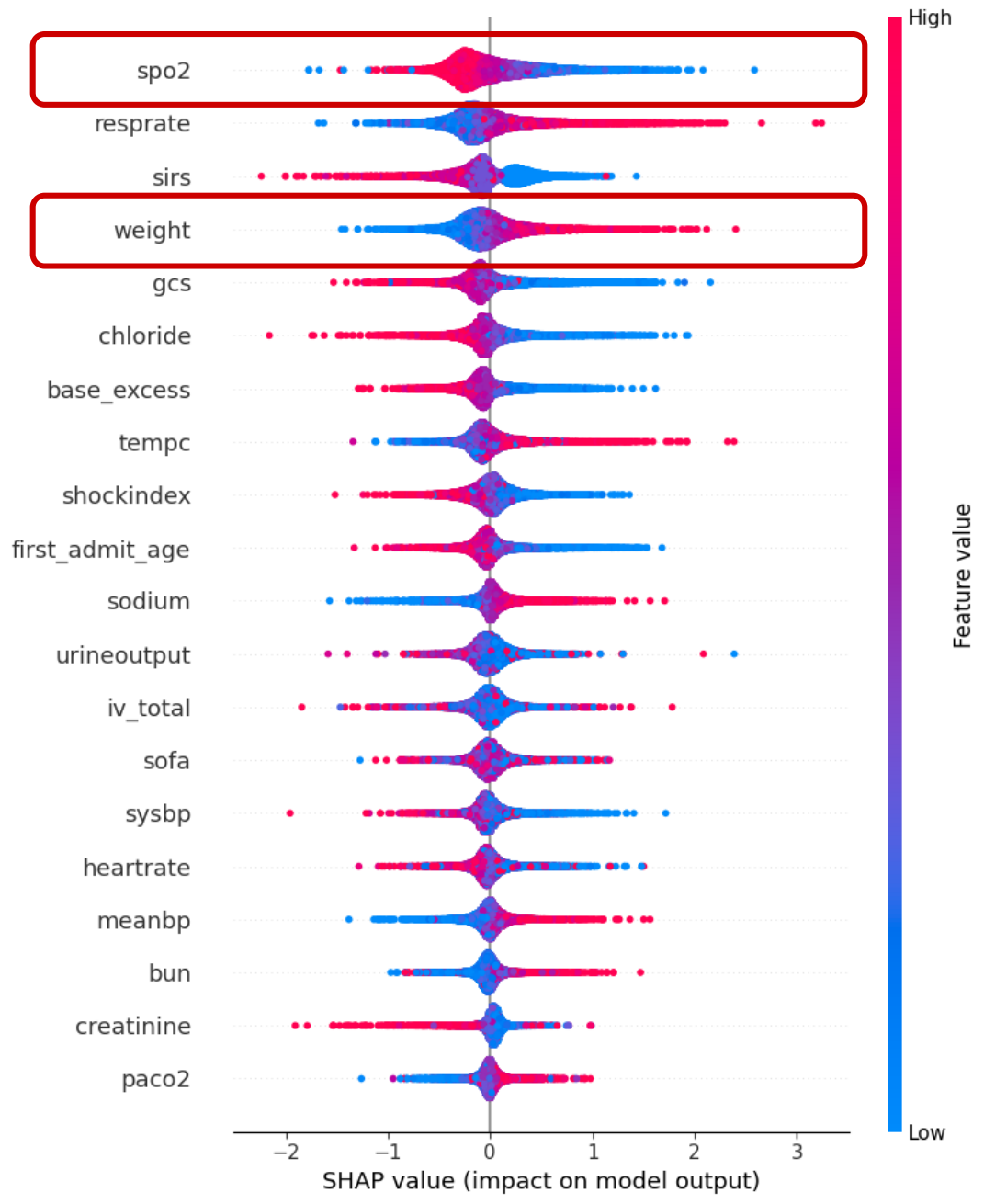
Policy-Level Explanation

- SHAP applied to policy network outputs
- Explains feature contribution to selected actions (PEEP, Vt, FiO₂)
- Reveals directional influence patterns for ventilator settings

Model-Level Explanation

- SHAP applied to learned dynamics model
- Explains predicted physiological transitions
- Supports verification of model behavior

Explainable AI for PEEP



Peripheral oxygen saturation (SpO₂) and body weight play significant roles in determining the choice of PEEP.

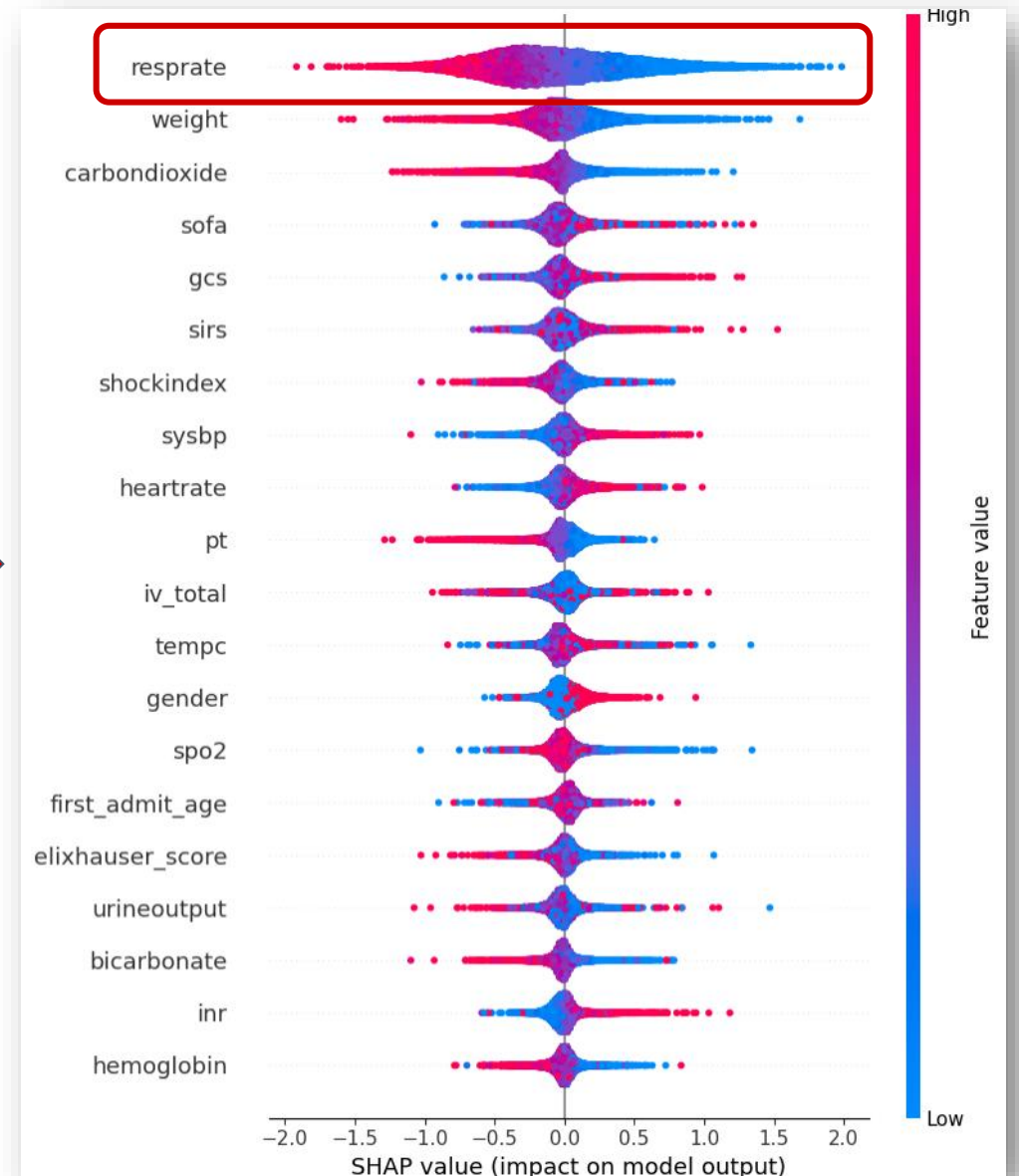
Lower oxygenation is combated by higher PEEP.

Higher body weight, meaning more collapsed lung tissue requires increased PEEP.

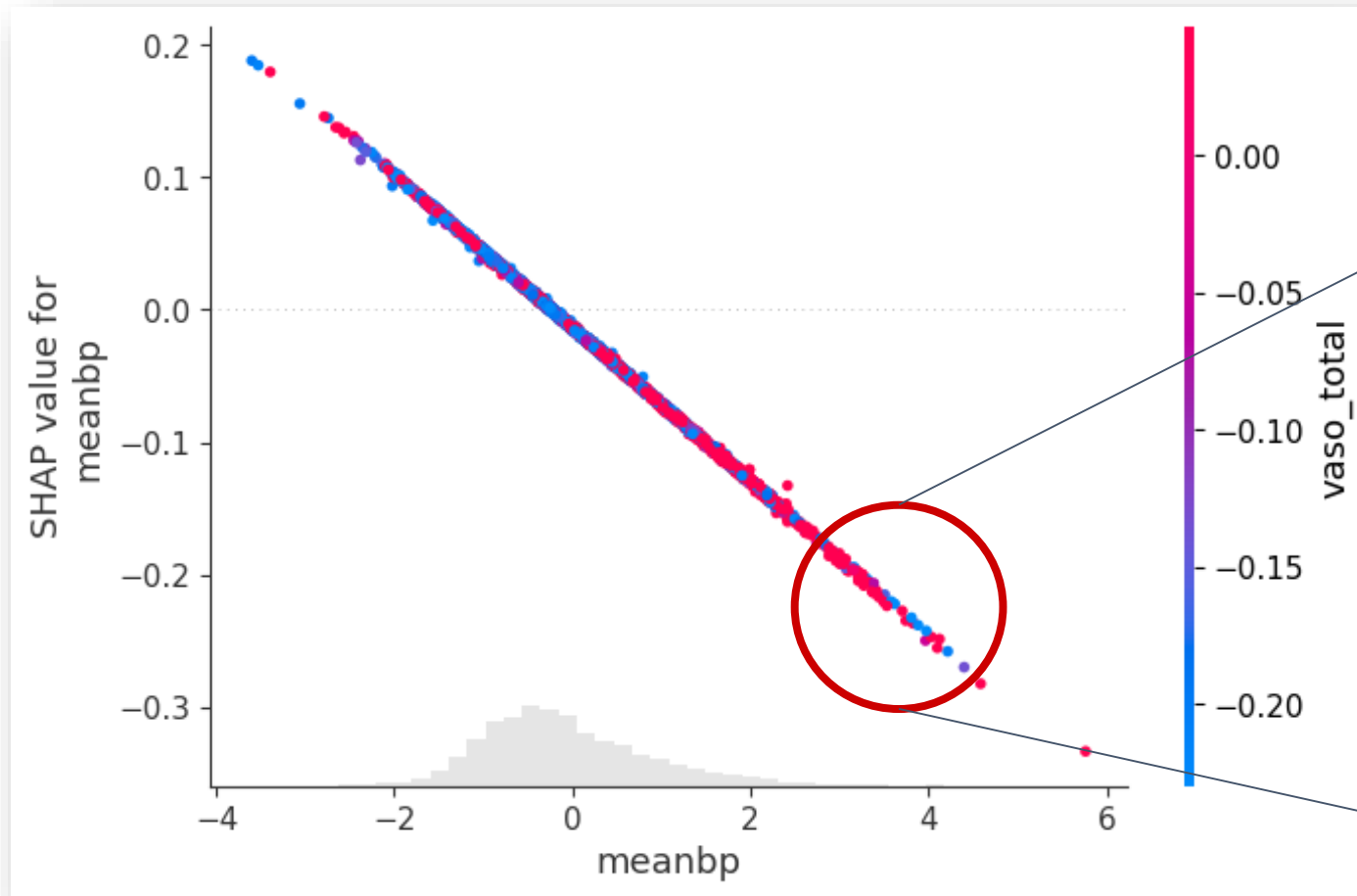
Explainable AI for Vt

Vt and respiratory rate play a crucial role in eliminating CO₂.

A reduction in Vt must be balanced by an increase in respiratory rate.



Explainable AI for Blood Pressure



Inverse relationship between linearly connected blood pressures caused by the use of medication. High values lead to **subsequent decrease in blood pressure!**

Key Takeaways

- Offline RL is essential in safety-critical medical domains.
- Reliable off-policy evaluation is computationally intensive.
- Model-based approaches increase data efficiency but require careful regularization.
- Explainability tools help inspect — not yet fully validate — policy behavior.



- This remains a research prototype and not a deployed clinical system.



The vision of the team behind the IntelliLung project, encapsulated in the motto "Each Breath Counts"!

Thank You – Questions?

Project Website



Universitätsklinikum
Carl Gustav Carus
DIE DRESDNER.



Miloš Nenadović, Valentina Janev, Dejan Paunović,
On behalf of the IntelliLung consortium