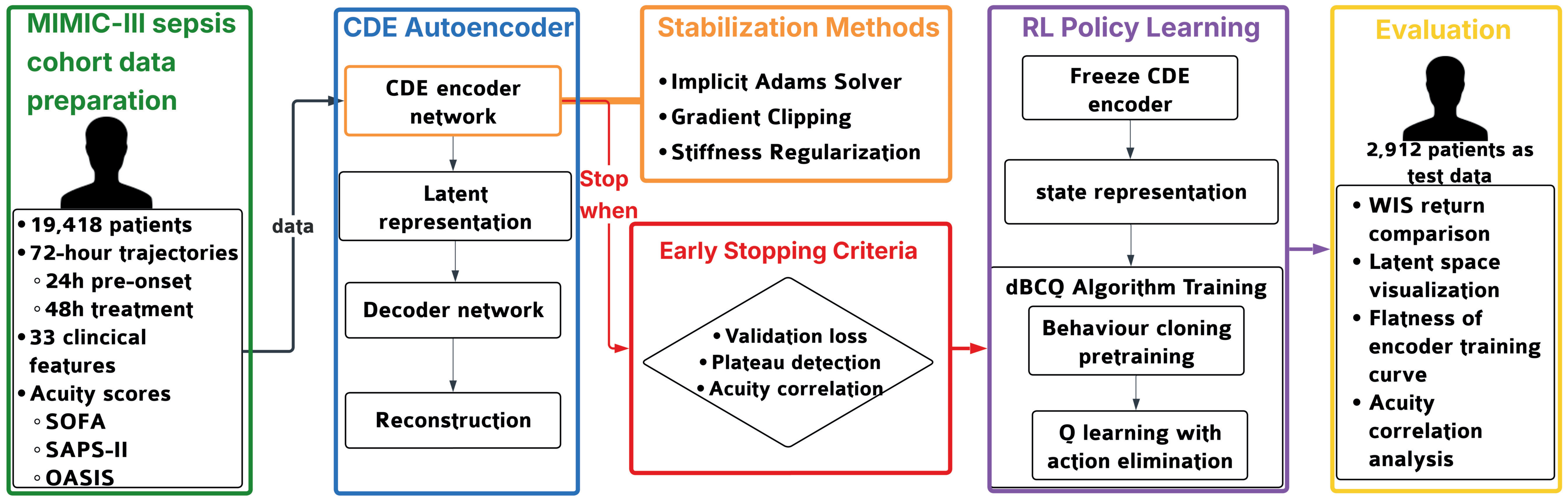
Stable CDE Autoencoders with Acuity Regularization for Offline Reinforcement Learning in Sepsis Treatment

https://github.com/GAOYUEtianc/RL_mimic_CDE_stable Yue (Anna) Gao

End-to-end pipeline from MIMIC-III data → CDE autoencoder → RL training



Sequential sepsis treatment as a Partially Observable Markov Decision Process (POMDP)

- State : $s_t = h(t)$ (CDE encoded history of observations $o_{0:t}$).
- Action : $a_t \in \{1, \ldots, 25\}$ represents discrete combinations of intravenous fluids and vasopressor doses
- Reward :

$$r_t = \begin{cases} +1 & \text{if patient survives at trajectory end} \\ -1 & \text{if patient dies at trajectory end} \\ 0 & \text{otherwise (at intermediate steps)} \end{cases}$$

- Policy: $\pi(a_t|s_t)$ maps states to probability distribution of actions, aiming to optimize cumulative reward via offline RL.

CDE Autoencoder state representation

Continuous-time irregular observations $o_t \in \mathcal{O}$
Hidden state $h(t) \to \mathcal{H}$

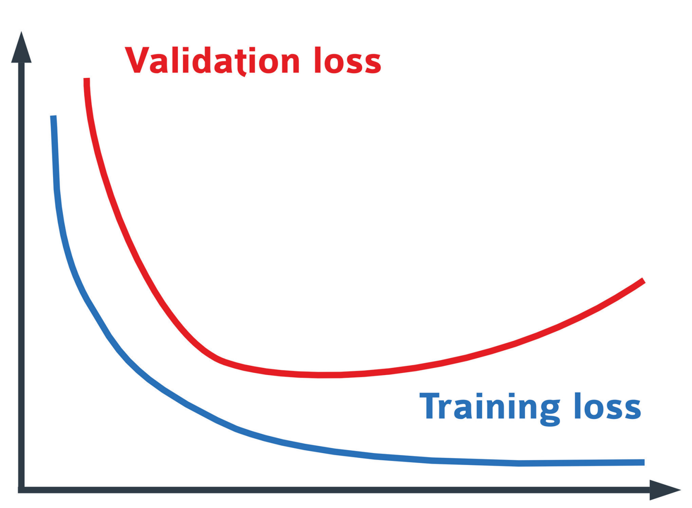
$$\partial h(t) = f_\theta(h(t))\partial o_t$$

The CDE acts as an encoder $\psi : \mathcal{O} \to \mathcal{H}$, mapping time-based observations to a latent space. where f_θ is a neural network parameterizing the system dynamics, and the differential ∂o_t accounts for irregular sampling intervals
The regularized total loss function for CDE autoencoder is defined to be

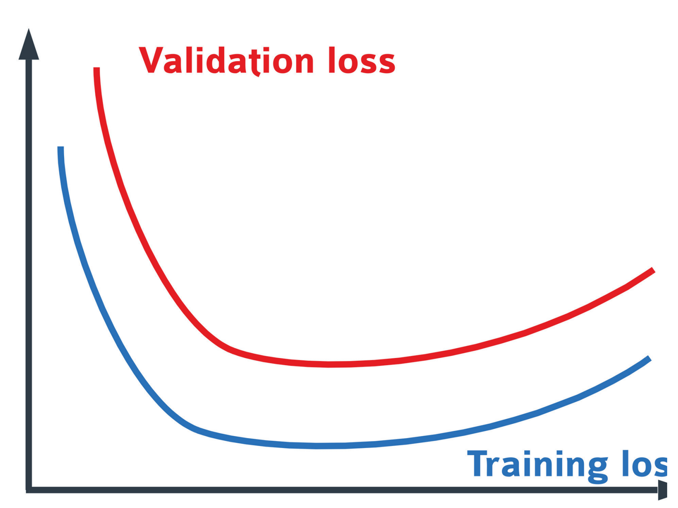
$$\mathcal{L}_{\mathrm{corr}}(\hat{s}_t) = -(\rho_{\mathrm{SOFA}}(\hat{s}_t) + \rho_{\mathrm{SAPS\text{-}II}}(\hat{s}_t) + \rho_{\mathrm{OASIS}}(\hat{s}_t))$$

$$\mathcal{L}_{\mathrm{total}}(o_t, \hat{o}_t) = \mathcal{L}_{\mathrm{MSE}}(o_t, \hat{o}_t) + \lambda \cdot \mathcal{L}_{\mathrm{corr}}(\hat{s}_t)$$

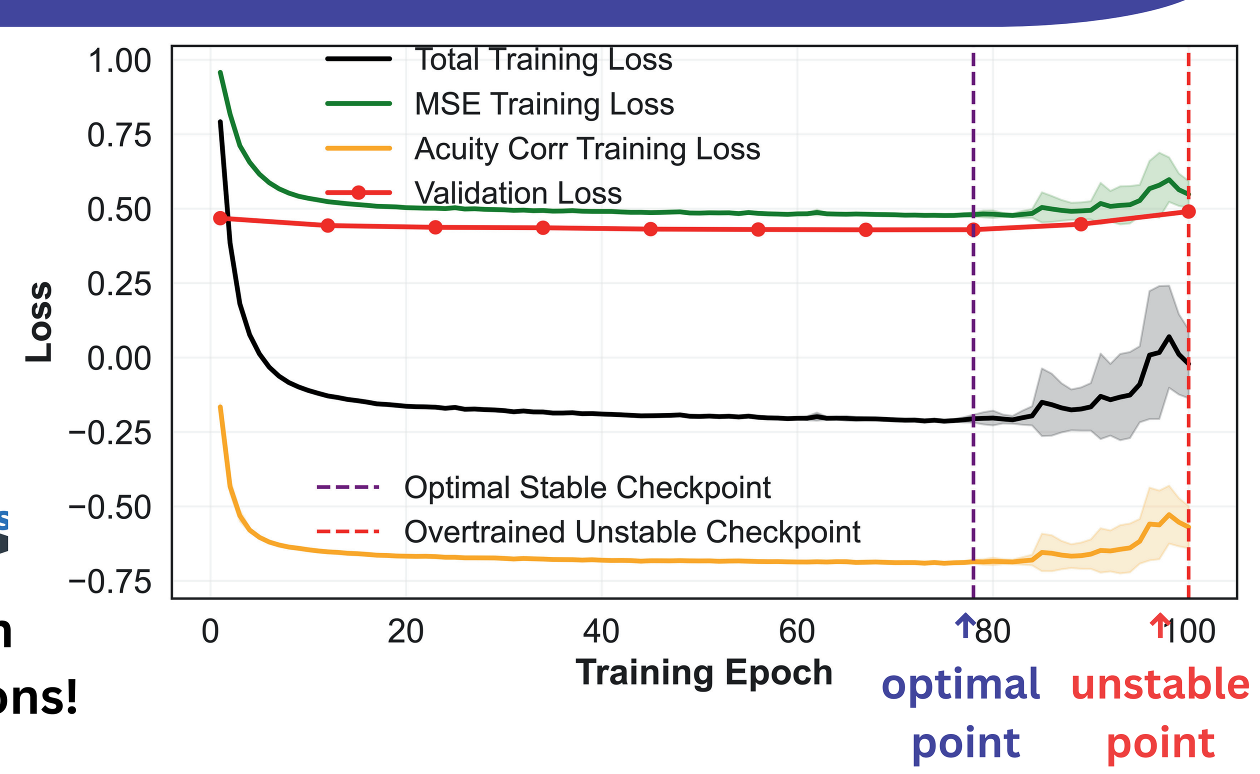
Why we need early stopping mechanism and stabilization methods for Neural CDE?



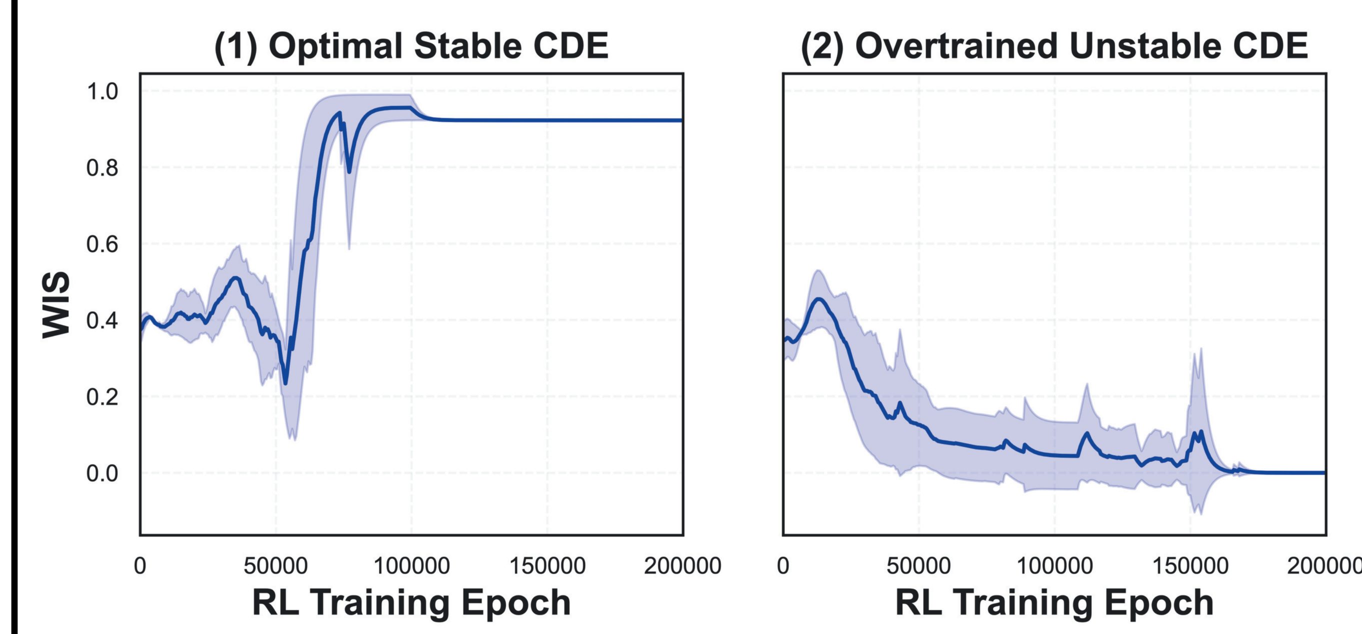
Overfitting



Numerical Instability - often occur in differential equations!



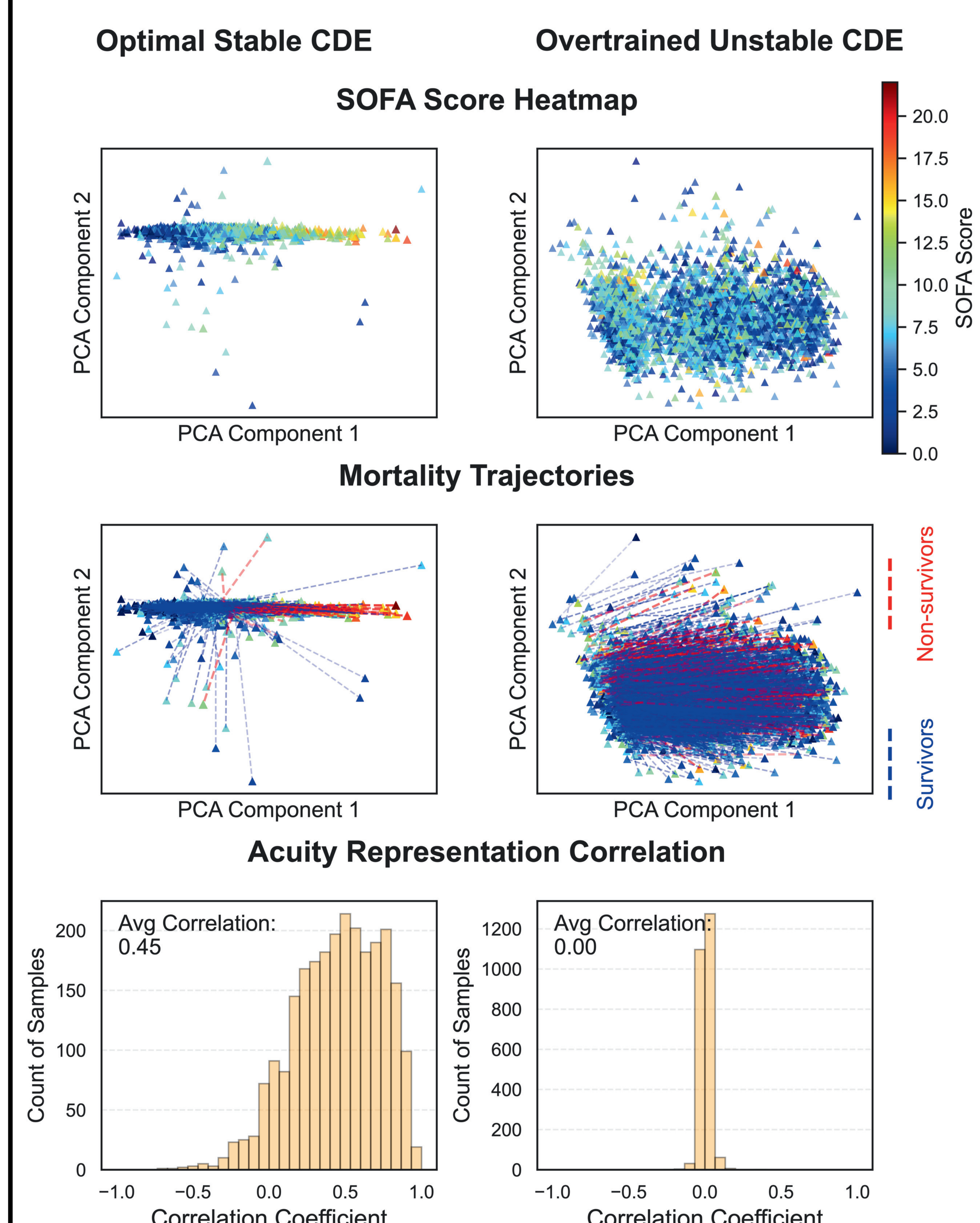
CDE Early Stopping - Better RL Results And Clinically Meaningful State Representations



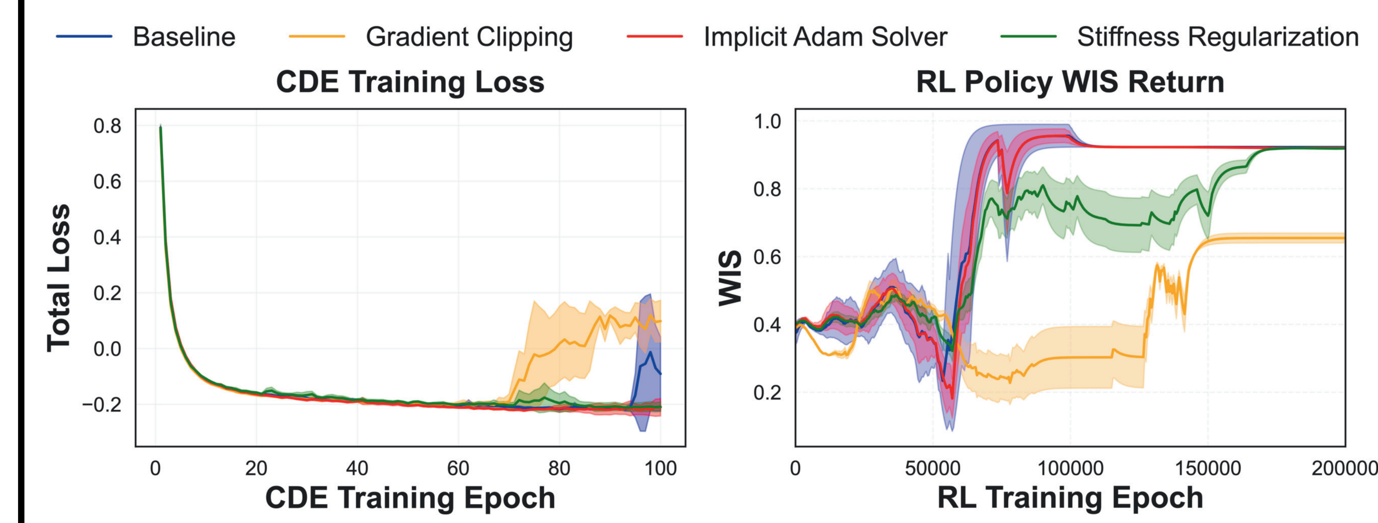
↑ RL trained on CDE encoder frozen at the optimal stable point
RL WIS return > 0.9

↑ RL trained on CDE encoder frozen at the unstable point
RL crashes

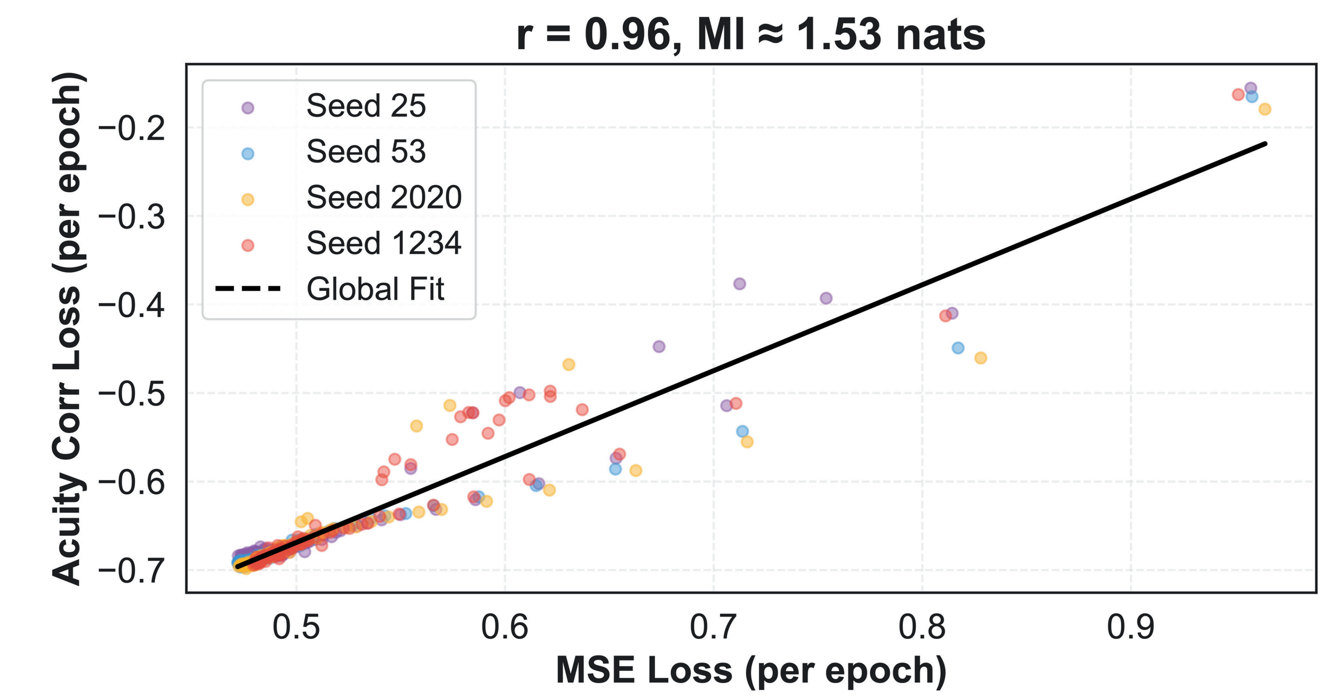
Comparative visualization of clinical alignment on validation cohort dataset trajectories



Stabilization Methods



MSE Loss & Acuity Correlation Loss Alignment

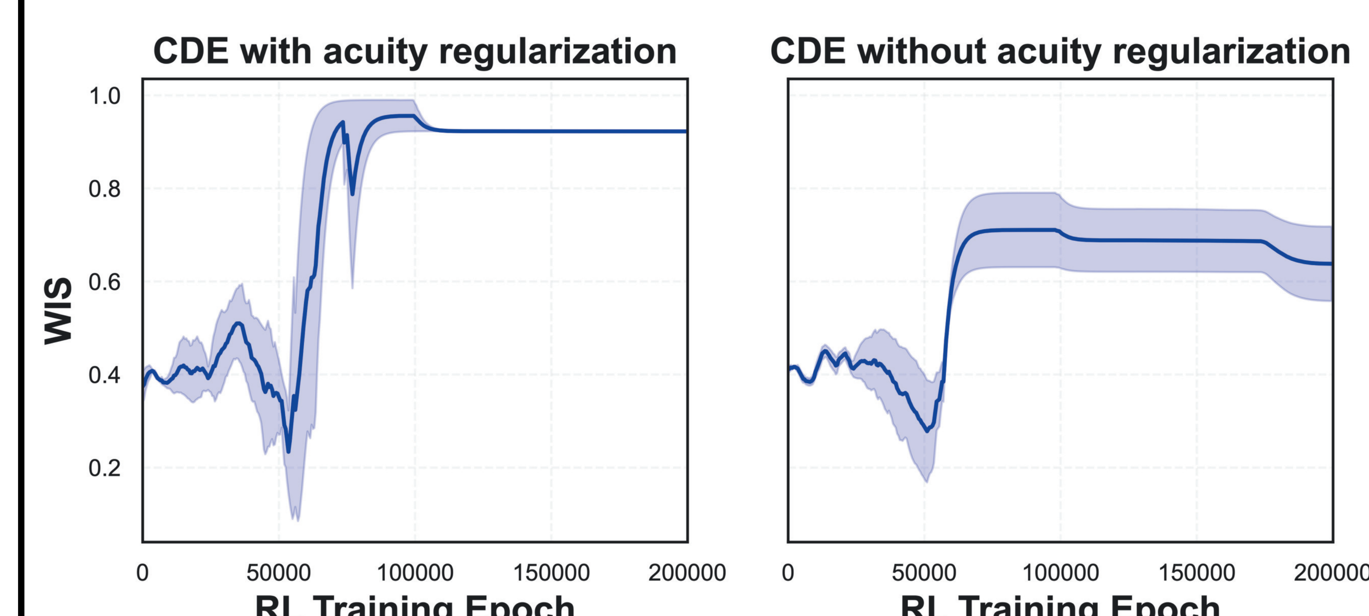


Acuity Regularization Improves RL Return

Prior works found that the CDE autoencoder outperformed other state-representation models on MIMIC-III (with RL WIS ≈ 0.7) and concluded that adding clinical acuity score regularization brought no benefit.

However, their conclusions were drawn given that they ignored the Neural CDE training instability issue.

Our findings show that acuity regularization improves both RL policy returns and clinical interpretability when paired with stabilized CDE encoder.



↑ New SOTA with WIS > 0.9

↑ Prior works' best RL return (without acuity regularization)