Reinforcement Learning for ICU Treatment Planning

Goureesh Chandra	31	TVE22CS069
Ivin Mathew Kurian	33	TVE22CS075
Muhammed Farhan	39	TVE22CS094
Rethin Francis	72	LTVE22CS149

Advisor: Prof. Divya S K

College of Engineering, Trivandrum
Dept. of Computer Science & Engineering

September 18, 2025

Contents

Introduction

Doctors often face tough choices when treating critically ill patients.

Our project builds an **AI system** that learns from hospital data to suggest better treatment plans.

The goal is to support doctors, make care safer, and improve patient outcomes.

Problem Statement

Making timely and personalized treatment decisions for patients with complex, evolving conditions is a significant clinical challenge.

- **High-Stakes Decisions** – Clinicians must manage highly dynamic patient trajectories where standardized protocols are often insufficient.
- **Fragmented Data** – Patient health records are often scattered and incomplete, hindering a comprehensive view for effective treatment.
- **Learning Constraints** – Existing AI methods are often limited to mimicking historical data and cannot safely explore potentially better treatment strategies.

Objective

The objective of this project is to develop a novel AI framework using model-based reinforcement learning to provide personalized and safer clinical decision support.

- **Patient Simulator ("Digital Twin")** – Learn and simulate patient physiological dynamics from historical clinical data.
- **AI-Driven Policy Optimization** – Use the simulator as a safe environment to train a reinforcement learning agent to discover optimal and personalized treatment strategies.
- **Clinical Decision Support** – Deliver data-driven treatment recommendations to support clinicians, improve patient outcomes, and enhance safety.

Literature Review

Sl. no.	Name of Publication	Journal / Conference	Author(s)	Date of publication	Remarks
1	**Challenges with reinforcement learning model transportability for sepsis treatment in emergency care**	Critical Care Medicine	A. Komorowski et al.	2021	**Advantage:** Explores transportability of RL models from ICU to ED, addressing real-world sepsis care needs. **Disadvantage:** ICU-trained models struggle due to missing ED data, limiting applicability.
2	**Deep learning prediction models based on EHR trajectories: A systematic review**	Journal of Biomedical Informatics	Ali Amirahmadi, Mattias Ohlsson, and Kobra Etminani	2023	**Advantage:** Provides comprehensive review of deep learning for EHR trajectory prediction, highlighting RNNs and attention. **Disadvantage:** Challenges include data insufficiency, explainability, and handling complex dependencies.
3	**Deep offline reinforcement learning for real-world treatment optimization applications**	arXiv Preprint	Mila Nambiar et al.	2023	**Advantage:** Introduces transition sampling to improve offline RL, achieving better safety and constraint satisfaction in sepsis/diabetes. **Disadvantage:** Offline RL limited by suboptimal retrospective data and strict safety needs.
4	**Learning transformer based world models with contrastive predictive coding**	NeurIPS	Anon. Authors	2021	**Advantage:** TWISTER achieves SOTA on Atari 100k without search, showing the power of contrastive objectives in Transformers. **Disadvantage:** Prior Transformer models underperform due to weak next-state objectives.

Literature Review (contd.)

Sl. no.	Name of Publication	Journal / Conference	Author(s)	Date of publication	Remarks
5	**Mastering Memory Tasks with World Models**	ICML	Anon. Authors	2023	**Advantage:** R2I enables state-of-the-art memory/credit assignment, outperforming DreamerV3 in BSuite and POPGym. **Disadvantage:** MBRL agents still struggle with long-term dependencies and recalling distant observations.
6	**Predicting the impact of treatments over time with uncertainty aware neural differential equations**	AISTATS	Edward De Brouwer, Javier González Hernández, and Stephanie Hyland	2022	**Advantage:** CF-ODE provides accurate treatment impact predictions with uncertainty estimates. **Disadvantage:** Counterfactual prediction limited due to confounding in observational data.
7	**Reinforcement Learning in Healthcare - Optimizing treatment strategies dynamic resource allocation and adaptive clinical decision making**	Int. J. of Computer Applications Tech.	Hassan Ali	2022	**Advantage:** Broad survey of RL in healthcare, covering treatment optimization, resource allocation, and adaptive decision-making. **Disadvantage:** Does not provide detailed shortcomings since it is a review, not an experiment.
8	**Safe Reinforcement Learning for Sepsis Treatment**	IEEE Int. Conf. on Healthcare Informatics	Yan Jia et al.	2020	**Advantage:** Introduces safe RL approaches for sepsis care, with focus on safety constraints. **Disadvantage:** Early-stage proceedings paper, lacks in-depth limitation analysis.

Literature Review (contd.)

Sl. no.	Name of Publication	Journal / Conference	Author(s)	Date of publication	Remarks
9	**Smart Imitator: Learning from Imperfect Clinical Decisions**	arXiv Preprint	Anon. Authors	2021	**Advantage:** Combines imitation learning with RL to learn from imperfect clinical data, outperforming baselines on sepsis/diabetes. **Disadvantage:** Struggles with high-risk patient treatment due to data scarcity.
10	**Towards Safe Mechanical Ventilation Treatment Using Deep Offline Reinforcement Learning**	AAAI-23	Flemming Kondrup et al.	2023	**Advantage:** Proposes DeepVent, a CQL-based safe offline DRL agent, optimizing ventilation parameters for 90-day survival. **Disadvantage:** Explicit limitations not detailed, focus mainly on safety improvements.

Research Gaps

- **Incomplete and irregular EHR data** – missing values, noise, and inconsistent sampling limit effective learning.
- **Over-reliance on offline RL** – many systems copy historical clinician decisions instead of exploring new strategies.
- **Limited patient personalization** – existing models often treat all patients similarly, ignoring individual differences.
- **Safety concerns in real-world testing** – direct deployment on patients is risky without reliable simulators.
- **Generalization challenges** – models trained on one dataset (e.g., MIMIC) may not transfer well to other hospitals.

Proposed Solution Methodology

This approach is inspired by DreamerV3 and its clinical adaptation in medDreamer.

Two sequential stages:

1. Train the **Patient Simulator (World Model)** to predict patient outcomes.
2. Train the **RL Agent (Policy)** using the simulator to discover optimal treatments.

Algorithm 1: Patient Simulator Training

Goal: Learn patient health dynamics from EHR data to create a reliable "digital twin".

Steps:

1. **Initialization:** Neural network parameters ϕ
   - Encoder
   - Dynamics predictor
   - Prediction heads (reconstruction, rewards, outcomes)
2. **Training Loop:** For multiple epochs:
   1. Sample patient trajectories (observations, actions, rewards)
   2. Encode observations to latent states s_t
   3. Predict next latent state s_{t+1} given s_t and action a_t
   4. Reconstruct observations and predict rewards/outcomes
   5. Compute combined loss and update ϕ via backpropagation
3. **Output:** Trained Patient Simulator for realistic patient trajectories

Algorithm 2: RL Agent Training

Goal: Train an optimal treatment policy safely in the simulated environment.

Steps:

1. Load trained Patient Simulator (weights frozen)
2. Initialize RL Agent: policy network θ (actor), value network ψ (critic)

Phase 1: Clinically Grounded Policy Initialization

- Generate hybrid trajectories (real data + short simulator rollouts)
- Train actor and critic:
    - Critic: evaluate long-term state values
    - Actor: choose actions leading to high-value states

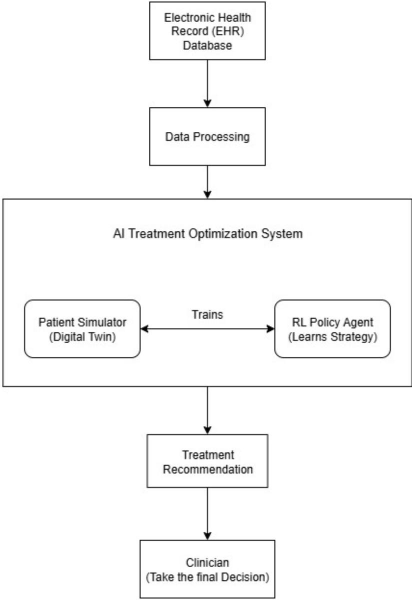
Algorithm 2 (cont.): Strategic Refinement

Phase 2: Strategic Refinement through Long-Horizon Imagination

- Generate fully imagined trajectories using the Patient Simulator
- Fine-tune actor and critic to improve long-term planning

Output: Final trained **Optimal Treatment Policy** that recommends actions given a patient's current state

Design of Proposed Methodology



Design Flow (Contd.)

1. **Electronic Health Record (EHR) Database**
   - Large, de-identified patient records (vitals, labs, diagnoses, meds)
   - Source of "experience" for the AI
   - Example: MIMIC-IV
2. **Data Preprocessing**
   - Clean and structure messy clinical data
   - Handle missing/irregular data
   - Convert into patient timelines/trajectories

Design Flow (Contd.)

3A. Patient Simulator ("Digital Twin")

- Model predicts patient response to treatments
- Provides safe virtual testing ground
- Based on model-based RL (e.g., DreamerV3)

3B. RL Policy Agent

- Learns best actions through trial-and-error in simulator
- Maps patient state $\rightarrow$ treatment decision
- Optimizes for long-term stability/survival

Design Flow (Contd.)

4. Treatment Recommendation

- AI suggests adaptive treatment actions (e.g., dosage)
- Produces a Dynamic Treatment Regime (DTR)

5. Clinician (Final Decision)

- Doctor reviews AI suggestions
- Combines with expertise, ethics, patient context
- Ensures safety and accountability

Resource Allocation

1. **Data Resources**
   - MIMIC-III, MIMIC-IV, eICU datasets.
   - Preprocessing: Python (pandas, NumPy, scikit-learn).
2. **Computational Resources**
   - GPU-enabled system (RTX 4060/4090 or cloud GPUs).
   - Frameworks: PyTorch, Ray RLlib.

Future Work

- Extend framework to other diseases (e.g., diabetes, cancer care).
- Incorporate multi-modal data (genomics, imaging).
- Clinician-in-the-loop evaluation.

Conclusion

- We are developing a **model-based reinforcement learning framework** trained on ICU data for clinical decision support.
- By building a **digital twin of patients**, the system safely explores treatment strategies without direct risk to patients.
- The approach enables **personalized and safer treatment recommendations**, reducing unnecessary interventions and improving patient outcomes.
- Future work includes extending the model to more diverse patient populations and validating in real-world clinical settings.

References

Qianyi Xu, Dilruk Perera, Gousia Habib, and Mengling Feng (2025).
medDreamer: Model-Based Reinforcement Learning with Latent Imagination on Complex EHRs for Clinical Decision Support.
arXiv preprint arXiv:2505.19785.

Arthur Komorowski et al. (2021).
Challenges with reinforcement learning model transportability for sepsis treatment in emergency care.
Critical Care Medicine.

Ali Amirahmadi, Mattias Ohlsson, and Kobra Etminani (2023).
Deep learning prediction models based on EHR trajectories: A systematic review.
Journal of Biomedical Informatics.

Mila Nambiar et al. (2023).
Deep offline reinforcement learning for real-world treatment optimization applications.
arXiv preprint arXiv:2302.07549.

Anon. Authors (2023).
Mastering Memory Tasks with World Models.
ICML.

Edward De Brouwer, Javier González Hernández, and Stephanie Hyland (2022).
Predicting the impact of treatments over time with uncertainty aware neural differential equations.
AISTATS.

Flemming Kondrup et al. (2023).
Towards Safe Mechanical Ventilation Treatment Using Deep Offline Reinforcement Learning.
AAAI-23.

Thank You!