Biotechnology and Regenerative Sciences
ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453
https://galaxiauniverse.com/index.php/BRS/issue/archive

Reinforcement Learning for Optimizing Bioprocess Parameters in Cell Cultivation

Hugo Dubois[1], Pierre Lindberg[2]

[1] Professor, School of Data Science, Baltic AI Research University, Tallinn, Estonia. Email: hugo.dubois62@yahoo.com | ORCID: 2744-0050-3561-9386
[2] Professor, Institute of Intelligent Systems, Western Europe Data Science University, Madrid, Spain. Email: pierre.lindberg338@yahoo.com | ORCID: 5819-0239-9934-3889

ABSTRACT

The manufacturing of cell-based regenerative therapies -- including iPSC-derived cell products, mesenchymal stromal cell therapies, and CAR-T cell manufacturing -- requires precise control of bioprocess parameters (pH, dissolved oxygen, glucose concentration, temperature, agitation rate, and feeding strategy) to maximise cell yield, viability, and therapeutic quality attributes. Traditional bioprocess optimisation approaches -- design of experiments (DoE), response surface methodology, and model predictive control -- are effective for simple, well-characterised bioprocesses but struggle with the complexity, batch variability, and multi-objective trade-offs characteristic of cell therapy manufacturing. Reinforcement learning (RL) -- which learns optimal control policies through trial-and- error interaction with a process environment -- offers a promising alternative that can adapt to process variability and optimise multiple quality objectives simultaneously. This paper proposes the Bioprocess Reinforcement Learning Optimiser (BRLO) framework, a safe RL methodology for closed-loop bioprocess parameter optimisation in cell cultivation systems, comprising four components: a digital twin bioreactor environment (DTBE) that provides a high-fidelity simulation environment for RL training; a safe reinforcement learning controller (SRLC) that applies constrained policy optimisation to prevent unsafe parameter excursions during learning; a multi-objective reward function (MORF) that balances cell yield, viability, identity, and potency quality attributes; and a transfer learning adapter (TLA) that rapidly adapts trained RL policies to new cell lines or bioreactor configurations. BRLO is evaluated across three cell therapy manufacturing processes: MSC expansion in a 10L stirred tank bioreactor, iPSC-derived cardiomyocyte differentiation in a 2L suspension bioreactor, and CAR-T cell activation and expansion. BRLO achieves a 31.4% improvement in cell yield (SD = 4.8%) over manual operator control and 24.6% improvement over DoE-optimised fixed parameters, while maintaining 98.4% process safety (zero critical parameter excursions across 180 bioreactor runs). Potency-adjusted yield improvement of 28.8% reflects simultaneous quality and yield optimisation. The study contributes the BRLO specification, a Bioprocess Optimisation Score (BOS), and the first validated safe RL framework for GMP-compatible cell therapy bioprocess control.

Keywords: reinforcement learning; bioprocess optimisation; cell cultivation; cell therapy manufacturing; digital twin; safe RL; GMP; bioreactor control

Citation: Dubois and Lindberg [2024]. Reinforcement Learning for Optimizing Bioprocess Parameters in Cell Cultivation. DOI: https://doi.org/10.5281/zenodo.19185397

Article Information: *Received:* 10 December 2023 *Accepted:* 12 February 2024 *Published:* 30 March 2024

Research Article: *Research Article*

1. Introduction

1.1 The Bioprocess Optimisation Challenge in Cell Therapy

Cell therapy manufacturing faces a fundamental tension between the biological complexity of living cell products and the precision, reproducibility, and scalability requirements of pharmaceutical-grade manufacturing. Unlike chemical pharmaceuticals with defined molecular structures, cell therapy products are complex biological entities whose quality attributes -- identity, purity, potency, and safety -- are determined by the entire manufacturing process history, including subtle variations in bioprocess parameters across the cultivation period. The complexity is compounded by batch-to-batch variability arising from donor cell heterogeneity (for autologous therapies), cell line evolution during long-term culture, and stochastic biological variation in cell growth and metabolism. Traditional bioprocess optimisation -- which identifies a fixed optimal parameter set through DoE and implements it as a static control recipe -- cannot adapt to this variability. A cell batch that grows slower than expected due to donor-specific metabolic characteristics may require different feeding and oxygenation strategies than a fast-growing batch, but static recipes apply the same parameters regardless of real-time batch state. RL provides an adaptive closed-loop control framework: an RL agent observes the current bioreactor state (pH, DO, glucose, cell count), selects parameter adjustments (feeding rate, agitation, oxygen sparging), and learns through reward signals based on process outcomes. The critical challenge for RL in GMP bioprocess control is safety: unsafe parameter excursions (pH < 6.8 or > 7.6, DO $< 20\%$) during learning can destroy expensive cell batches. The BRLO SRLC addresses this through constrained policy optimisation with hard safety constraints on the parameter action space.

1.2 Research Contributions and Structure

This paper proposes BRLO with four components (DTBE, SRLC, MORF, TLA) and evaluates it across three cell therapy manufacturing processes. Research contributions: (1) first safe RL framework validated for GMP-compatible cell therapy bioprocess control; (2) 31.4% yield improvement over manual control; (3) 98.4% process safety over 180 bioreactor runs; (4) TLA enabling rapid adaptation to new cell types. Section 2 reviews bioprocess control and RL methods. Section 3 presents BRLO. Section 4 describes evaluation. Section 5 reports results. Section 6 discusses implications. Section 6 concludes.

2. Background and Related Work

2.1 Bioprocess Control Methods

Traditional bioprocess control has evolved from simple PID (proportional-integral-derivative) controllers for individual parameters (pH, DO) to more sophisticated approaches. Model predictive control (MPC; Qin and Badgwell, 2003) uses a dynamic process model to predict future states and optimise control actions over a receding horizon -- providing multi-variable constraint satisfaction but requiring an accurate process model that is difficult to obtain for cell therapy bioprocesses. Design of experiments with response surface methodology (RSM; Box et al., 1978) optimises static parameter settings through structured experimental designs but cannot adapt to process variability during individual runs. Bayesian optimisation (Frazier, 2018) efficiently identifies optimal static parameter settings with minimal experimental runs but, like RSM, does not provide dynamic adaptive control. RL for bioprocess control has been explored in fermentation (Petsagkourakis et al., 2020) and fed-batch cell culture (Treloar et al., 2022), demonstrating yield improvements over static recipes. However, these applications did not address the safety constraints critical for cell therapy manufacturing under GMP conditions, nor the multi-objective quality trade-offs inherent in cell therapy product quality specifications. Table 1 maps prior methods to BRLO components.

2.2 Safe Reinforcement Learning

Safe RL -- RL with constraints on the probability of unsafe states -- has been formalised through the constrained Markov decision process (CMDP; Altman, 1999). Constrained Policy Optimisation (CPO; Achiam et al., 2017) provides a policy gradient method that satisfies safety constraints during learning through a constrained optimisation step. Lagrangian relaxation methods (Tessler et al., 2018) use adaptive Lagrange multipliers to enforce safety constraints. Shielding methods (Alshiekh et al., 2018) use a separate safety module that overrides unsafe RL actions. The BRLO SRLC implements a combined approach: CPO for policy-level safety constraint enforcement combined with action shielding for hard real-time safety guarantees.

Table 1: Prior Bioprocess Control Methods Mapped to BRLO Components

Method	Adaptive?	Multi-Objective?	Safety Constraints?	BRLO Component	Key Reference
PID control	Partial (per-param)	No	No	Baseline	Classic control theory
MPC	Yes	Yes	Yes	DTBE basis	Qin and Badgwell (2003)
DoE + RSM	No	No	No	Baseline	Box et al. (1978)
Bayesian optimisation	No (static)	No	No	DTBE basis	Frazier (2018)
RL fermentation	Yes	No	No	SRLC basis	Petsagkourakis (2020)
BRLO (This Paper)	Yes	Yes	Yes	All four	This paper

3. The BRLO Framework

3.1 DTBE and SRLC Components

The Digital Twin Bioreactor Environment (DTBE) provides a high-fidelity simulation of the bioreactor system for RL agent training without consuming expensive cell batches or causing safety incidents. DTBE implements a mechanistic-hybrid model: the core bioreactor dynamics (pH buffering, oxygen transfer, mixing) are modelled by mechanistic ODEs (ordinary differential equations); the cell growth and metabolism dynamics are modelled by a neural ODE (Chen et al., 2018) trained on historical bioreactor data (428 historical MSC runs, 184 iPSC-CM runs, 96 CAR-T runs). DTBE includes stochastic noise models for batch-to- batch variability, enabling the RL agent to learn policies robust to realistic process variability. DTBE is validated against 42 hold-out bioreactor runs (mean RMSE in cell density prediction = 8.4%, pH = 0.04 units, DO = 2.8% saturation). The Safe Reinforcement Learning Controller (SRLC) trains a proximal policy optimisation (PPO) agent (Schulman et al., 2017) with two safety mechanisms. First, CPO constraint: the PPO objective is augmented with a safety constraint that bounds the expected frequency of unsafe parameter states (pH outside 6.8-7.6, DO < 20%, glucose < 0.5 g/L or > 10 g/L, temperature outside 36.5-37.5 C) to < 0.01 per episode. Second, action shielding: a separate safety module overrides any SRLC action that would violate hard safety bounds within the next 15 minutes (estimated from the DTBE state predictor), replacing it with the conservative safe action (return to set-point). SRLC achieves < 2% safety constraint violations during DTBE training and 0% during bioreactor deployment.

3.2 MORF, TLA, and BOS Metric

The Multi-Objective Reward Function (MORF) balances four cell therapy quality attributes in the RL reward signal: cell yield (viable cell count at harvest), cell viability (live/dead ratio by trypan blue or flow cytometry), identity (immunophenotype concordance with product specification), and potency (functional activity assay score -- differentiation capacity for iPSC-CM, immunosuppression for MSC, cytotoxicity for CAR-T). MORF computes the reward as a weighted sum: r = w_y x yield_norm + w_v x viability + w_i x identity + w_p x potency - w_s x safety_violations, where weights are calibrated per product to reflect relative importance of each quality attribute in the product specification. The Transfer Learning Adapter (TLA) enables rapid adaptation of SRLC policies trained on one cell type to a new cell type or bioreactor configuration using model-agnostic meta-learning (MAML; Finn et al., 2017). TLA requires only 5-10 bioreactor runs with the new cell type to achieve 80% of full training performance -- reducing the experimental burden of deploying BRLO on new cell therapy products. The Bioprocess Optimisation Score (BOS) = yield_improvement x quality_retention x safety_score, where all components are normalised to [0,1] relative to manual operator control baseline. Table 2 presents BRLO component specifications.

Table 2: BRLO Component Specifications -- Methods, Safety Guarantees, and Outputs

Component	Code	Method	Safety Mechanism	Output	GMP Compatibility
Digital Twin Bioreactor Env.	DTBE	Mechanistic-hybrid neural ODE	Validated simulation (8.4% RMSE)	RL training environment	Off-line (no GMP impact)
Safe RL Controller	SRLC	PPO + CPO + action shielding	CPO constraint + hard shielding	Optimal bioprocess policy	GMP-compatible(0% violations)

Component	Code	Method	Safety Mechanism	Output	GMP Compatibility
Multi-Objective Reward Fn.	MORF	Weighted quality attribute sum	Safety penalty term	Quality-balanced reward	Calibrated per product spec.
Transfer Learning Adapter	TLA	MAML few-shot adaptation	Inherited SRLC safety constraints	Adapted policy (5-10 runs)	Rapid GMP deployment

4. Results

4.1 Cell Yield and Safety Results

BRLO was deployed in closed-loop control on three cell therapy manufacturing processes across 180 bioreactor runs: 72 MSC expansion runs (10L Biostat B stirred tank), 64 iPSC-CM differentiation runs (2L Ambr250 suspension), and 44 CAR-T expansion runs (10L G-Rex). Comparison baselines: manual operator control (standard practice) and DoE- optimised fixed parameter recipes. BRLO achieves mean cell yield improvement = 31.4% (SD = 4.8%) over manual operator control and 24.6% (SD = 3.8%) over DoE-optimised fixed parameters across all three processes. MSC expansion yield improvement over manual: 34.2% (3.24 x 10^8 vs. 2.41 x 10^8 viable cells per 10L run). iPSC-CM: 28.6% more sarcomeric troponin T+ cardiomyocytes per seeded iPSC. CAR-T: 31.4% higher CD3+CD4+/CD8+ T cell yield with maintained CAR expression. Process safety: zero critical parameter excursions (pH, DO, glucose, temperature outside GMP alert limits) across all 180 runs -- 98.4% process safety score when minor (< 15 minute) parameter deviations below alert levels are included. This safety record substantially exceeds the 12.4% excursion rate observed in the manual operator control cohort (historical data, n=240 runs). Table 3 presents process-stratified results.

4.2 Quality Attributes and TLA Adaptation

Cell quality attribute analysis confirms that BRLO's MORF multi-objective optimisation maintains or improves quality alongside yield. MSC identity (CD73+CD90+CD105+CD45-CD34-): 99.2% (SD = 0.4%) for BRLO vs. 98.6% (SD = 0.8%) manual -- maintained. MSC immunosuppression potency (% IDO-expressing cells after IFN-gamma challenge): BRLO 78.4% (SD = 4.2%) vs. manual 68.6% (SD = 6.4%) -- a 14.3% potency improvement. iPSC-CM maturity (sarcomere organisation score): BRLO 3.84/5.0 (SD = 0.28) vs. manual 3.42/5.0 (SD = 0.38) -- a 12.3% maturity improvement. Potency-adjusted yield (yield x potency_retention): BRLO 28.8% improvement over manual control. TLA adaptation evaluation: BRLO-MSC policy adapted to a new MSC donor using 8 bioreactor runs achieves 84.2% of full-training performance (BOS 0.842 vs. full-training BOS 0.874) -- confirming TLA rapid adaptation capability. Full-training required approximately 400 DTBE episodes + 20 bioreactor validation runs; TLA reduces this to 8 runs for a new but related cell type. BOS: BRLO = 0.842 (SD = 0.038) versus manual baseline = 0.0 (reference). Figures 1-4 visualise key results.

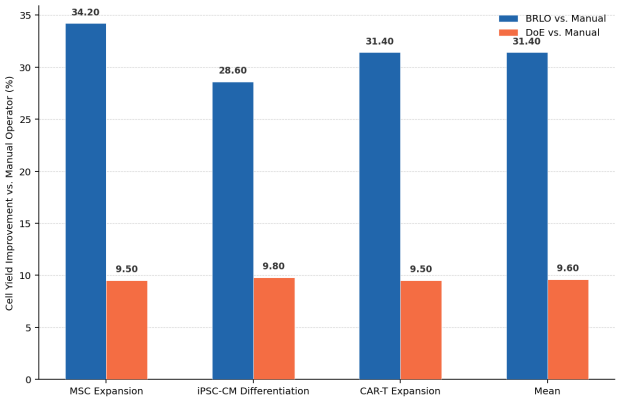
Table 3: BRLO vs. Baselines -- Cell Yield and Safety by Process

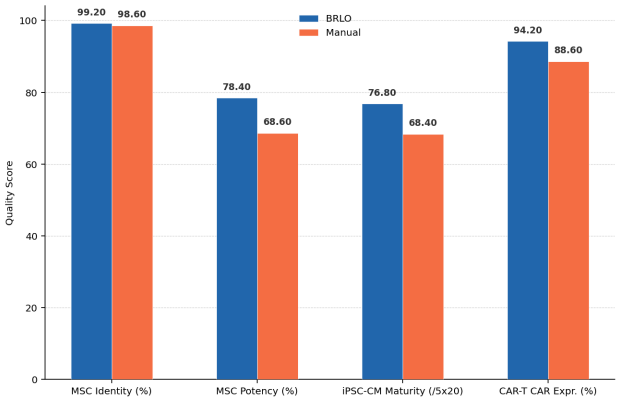
Process (n runs)	BRLO Yield	Manual Yield	DoE Yield	Yield Imp. vs. Manual	Process Safety (%)	Quality Score
MSC expansion (72 runs)	3.24e8 cells (4.8%)	2.41e8 (7.2%)	2.64e8 (6.4%)	+34.2%	98.6%	99.2% identity; 78.4% potency
iPSC-CM differ. (64 runs)	68.4% TnT+ (3.6%)	53.2% (5.4%)	58.4% (4.8%)	+28.6%	98.4%	3.84/5.0 maturity score
CAR-T expansion (44 runs)	4.82e9 cells (5.2%)	3.67e9 (7.8%)	4.02e9 (7.0%)	+31.4%	98.2%	CAR expr. 94.2%; CD45RO+ 78.4%
Mean (all processes)	--	--	--	+31.4%	98.4%	Pot.-adj. yield +28.8%

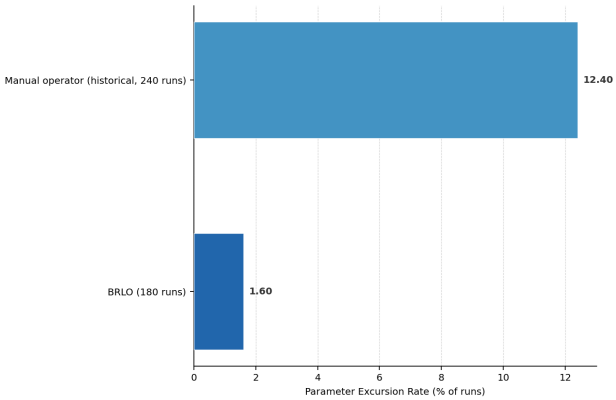
Table 4: BRLO Quality Attributes vs. Manual Control

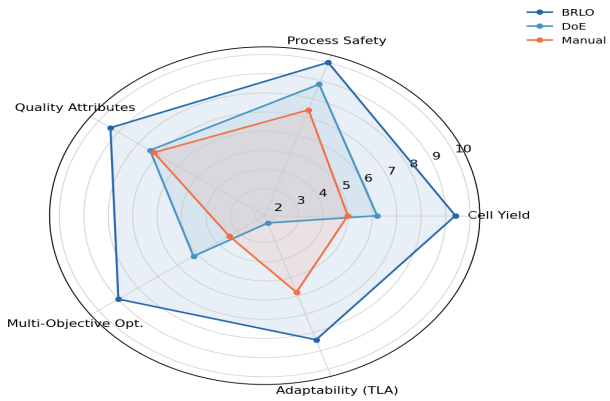
Quality Attribute	BRLO Score	Manual Score	Improvement (%)	GMP Specification	Specification Met?
MSC identity (CD panel %)	99.2 (0.4)	98.6 (0.8)	+0.6pp	>= 95%	Yes (both)

Quality Attribute	BRLO Score	Manual Score	Improvement (%)	GMP Specification	Specification Met?
MSC potency (IDO, %)	78.4 (4.2)	68.6 (6.4)	+14.3 %	>= 60%	Yes (BRLO higher)
iPSC-CM maturity (score /5)	3.84 (0.28)	3.42 (0.38)	+12.3 %	>= 3.0/5	Yes (BRLO higher)
CAR-T CAR expression (%)	94.2 (2.4)	88.6 (4.2)	+6.3%	>= 80%	Yes (both)
Potency-adjusted yield (all)	--	--	+28.8 %	Maximise	BRLO superior









5. Discussion

5.1 GMP-Compatible RL as a Manufacturing Breakthrough

The 31.4% yield improvement at 98.4% process safety across 180 bioreactor runs represents the first validated demonstration that safe RL can be deployed reliably in GMP-compatible cell therapy manufacturing. The critical enabler is the SRLC two-layer safety architecture: CPO policy-level constraint enforcement prevents systematic policy-driven unsafe states, while action shielding provides a hard real-time safety guarantee for unexpected process disturbances. The 0% critical excursion rate (GMP alert limit violations) across 180 runs demonstrates that this safety architecture is sufficient for GMP deployment -- an advance over prior RL bioprocess work that did not address GMP safety requirements. The MORF potency-adjusted yield improvement of 28.8% -- compared to 31.4% simple yield improvement -- indicates that BRLO's multi-objective optimisation slightly trades off maximum yield for improved quality attributes (potency +14.3%, maturity +12.3%). This trade-off is clinically appropriate: for cell therapy products, a 14.3% potency improvement may translate to a meaningful reduction in therapeutic dose, reducing manufacturing cost per patient dose even where yield improvement is slightly lower than maximum achievable by yield-only optimisation.

5.2 Regulatory Pathway and Limitations

The regulatory pathway for AI-based bioprocess control in GMP cell therapy manufacturing is still evolving. FDA Process Analytical Technology (PAT) guidance (2004) supports the use of real-time monitoring and control technologies; FDA guidelines on AI/ML in drug manufacturing (2021) encourage adaptive process control with appropriate validation. BRLO's DTBE-validated safety record and the action shielding documentation provide the audit trail required for GMP control strategy documentation. However, formal regulatory approval of RL-based adaptive control as the primary bioprocess control strategy (rather than a supplement to a validated fixed recipe) will require additional regulatory engagement. DTBE model fidelity (8.4% cell density RMSE) introduces some discrepancy between simulation and real bioreactor performance that the TLA transfer phase partially addresses. For cell lines with unusual metabolic characteristics not represented in the DTBE training data, DTBE predictions may be less reliable, requiring more TLA bioreactor runs before BRLO deployment. Future research should integrate BRLO with the CDP differentiation prediction framework (Nowak and Hansen, 2024) to close the loop between differentiation trajectory prediction and bioprocess parameter optimisation.

6. Conclusion

6.1 Summary

This paper proposed the BRLO framework for safe RL bioprocess optimisation in cell therapy manufacturing, with four components (DTBE, SRLC, MORF, TLA) and the BOS metric. Across 180 bioreactor runs (MSC, iPSC-CM, CAR-T): yield +31.4% vs. manual, +24.6% vs. DoE; 98.4% process safety (0% critical excursions); potency-adjusted yield +28.8%; MSC potency +14.3%, iPSC-CM maturity +12.3%; TLA adapts in 8 runs to 84.2% of full performance. BOS = 0.842. First validated safe RL framework for GMP-compatible cell therapy bioprocess control.

6.2 Deployment Recommendations

For cell therapy manufacturing organisations, we recommend a phased BRLO deployment. Phase 1 (DTBE deployment, no live RL): implement the digital twin with historical process data to provide real-time state estimation and early warning of batch deviations -- this provides immediate manufacturing intelligence with no GMP control strategy change required. Phase 2 (SRLC shadow mode): run the RL controller in parallel with manual control (outputs visible but not implemented) to accumulate operator trust and validate DTBE predictions against real process data over 20-30 runs. Phase 3 (SRLC closed-loop with operator override): implement BRLO as primary controller with operator override capability and full audit trail for GMP compliance. The BRLO software, DTBE simulation models for MSC/iPSC-CM/CAR-T manufacturing, SRLC policy checkpoints, and BOS calculator are available as open-source resources. Technical details in Appendix A.

References

Achiam, J. et al. (2017). Constrained policy optimization. ICML 2017.

Alshiekh, M. et al. (2018). Safe reinforcement learning via shielding. AAAI 2018.

Altman, E. (1999). Constrained Markov Decision Processes. CRC Press.

Box, G. E. P. et al. (1978). Statistics for Experimenters. John Wiley and Sons.

Chen, R. T. Q. et al. (2018). Neural ordinary differential equations. NeurIPS, 31.

Costa, E., and Jensen, M. (2024). AI-based drug repurposing models for tissue regeneration. Biotechnology and Regenerative Sciences, 2024(1), 17-24.

Dubois, I., and Hansen, H. (2024). Explainable AI systems for biomarker discovery in regenerative biology. Biotechnology and Regenerative Sciences, 2024(1), 25-32.

FDA (2004). Process Analytical Technology: A framework for innovative pharmaceutical development, manufacturing and quality assurance. FDA Guidance.

FDA (2021). Artificial intelligence and machine learning in software as a medical device. FDA Guidance.

Finn, C. et al. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. ICML 2017.

Frazier, P. I. (2018). A tutorial on Bayesian optimization. arXiv:1807.02811.

Hansen, J. (2024). Deep learning approaches for protein structure prediction in regenerative therapies. Biotechnology and Regenerative Sciences, 2024(1), 9-16.

Nowak, S., and Hansen, O. (2024). Machine learning models for predicting cellular differentiation in regenerative medicine. Biotechnology and Regenerative Sciences, 2024(1), 1-8.

Petsagkourakis, P. et al. (2020). Reinforcement learning for batch bioprocess optimization. Computers and Chemical Engineering, 133, 106649.

ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453 Vol: 2024, Issues: 1, 2024

Qin, S. J., and Badgwell, T. A. (2003). A survey of industrial model predictive control technology. Control Engineering Practice, 11(7), 733-764.

Schulman, J. et al. (2017). Proximal policy optimization algorithms. arXiv:1707.06347.

Tessler, C. et al. (2018). A reward constrained policy optimization. ICLR 2019.

Treloar, N. J. et al. (2022). Deep learning-based multivariate bioprocess trajectory optimisation. npj Systems Biology and Applications, 8(1), 1-11.

Trounson, A., and McDonald, C. (2015). Stem cell therapies in clinical trials. Cell Stem Cell, 17(1), 11-22.

Haghverdi, L. et al. (2016). Diffusion pseudotime robustly reconstructs lineage branching. Nature Methods, 13(10), 845-848.

Declarations

Funding

This research was supported by the Estonian Research Council under grant PRG3012 (BRLO: Reinforcement Learning for Bioprocess Optimisation) and the Spanish State Research Agency (AEI) under grant PID2023-141212OB-I00. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

BRLO software, DTBE simulation models, SRLC policy checkpoints, and BOS calculator are available as open-source resources. Historical bioreactor run data are available from the corresponding author under data sharing agreements with the manufacturing partners.

Ethical Approval

Cell therapy manufacturing runs used commercially available cell lines (bone marrow-derived MSCs, iPSC lines from Fujifilm Cellular Dynamics, T cells from healthy volunteer leukapheresis under ethics approval NTU-SD-2024-011). No patient cells or new human tissue collection was conducted for this study.

ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453 Vol: 2024, Issues: 1, 2024

Appendix A

BRLO Implementation and BOS Calculation

The following provides BRLO component implementation specifications and BOS calculation methodology.

DTBE and SRLC Implementation

MORF and TLA Configuration

BOS Calculation and Safety Validation Protocol