# WoundSim: Gymnasium-Compatible Reinforcement Learning Environments for Wound Healing Treatment Optimization

Hass Dhia
Smart Technology Investments Research Institute
hass@smarttechinvest.com

April 2026

**Abstract**

We present WoundSim, a suite of four Gymnasium-compatible reinforcement learning environments for optimizing wound healing treatment protocols. Each environment wraps a validated ordinary differential equation (ODE) model from the wound healing literature: (1) a Zlobina-type macrophage polarization model with 5 state variables, (2) a simplified Xue–Friedman ischemic wound healing model with 6 variables, (3) a Flegg hyperbaric oxygen therapy (HBOT) angiogenesis model with 4 variables, and (4) an extended diabetic wound model coupling macrophage dynamics with glucose–insulin physiology across 7 variables. All model parameters are sourced from peer-reviewed publications with explicit provenance. We provide random, clinical heuristic, and Proximal Policy Optimization (PPO) baselines across all environments. PPO outperforms both baselines in every environment, achieving up to 11.9× improvement over random policy on the HBOT task. WoundSim enables reproducible RL research on biologically grounded wound healing problems with configurable difficulty tiers, Stable-Baselines3 compatibility, and pip-installable distribution. The package, trained models, and this paper are publicly available at `https://github.com/HassDhia/woundsim`.

## 1 Introduction

Chronic wounds affect approximately 8.2 million Medicare beneficiaries in the United States alone, with annual treatment costs exceeding $28 billion [Sen, 2009]. Wound types including diabetic foot ulcers, venous leg ulcers, and pressure injuries share a common feature: they involve complex, multi-phase biological processes where treatment timing and intensity significantly affect outcomes [Brewer, 2020, Schultz et al., 2003]. Mathematical models of wound healing have provided valuable mechanistic insight into tissue repair dynamics [Sherratt and Murray, 1990, Menon and Flegg, 2017, Gomez, 2019], yet the translation of these models into actionable treatment optimization frameworks remains limited.

Reinforcement learning (RL) offers a natural framework for sequential treatment optimization: an agent observes the evolving wound state and selects interventions (e.g., polarization signals, oxygen therapy intensity, insulin dosing) to maximize long-term healing outcomes. Recent work by Lu and Gomez [2024] demonstrated that deep RL can discover macrophage polarization protocols that outperform constant treatment strategies. However, this work and related efforts operate on isolated, non-standardized simulation codebases that impede reproducibility and cross-study comparison.

The broader RL community benefits from standardized benchmark suites (Atari [Farama Foundation, 2023], MuJoCo, D4RL) that have catalyzed rapid progress by enabling fair comparison across algorithms. No analogous benchmark exists for wound healing RL. Existing biomedical RL environments focus on sepsis treatment, glucose control, or drug dosing, leaving

wound care, a domain with well-characterized ODE dynamics and clear clinical need, unaddressed.

We present WoundSim, a Python package providing four Gymnasium-compatible environments that address this gap. Our contributions are:

1. Four RL environments spanning wound types (macrophage polarization, ischemic, HBOT, diabetic) with 4–7 state variables and 1–3 action dimensions.
2. All ODE parameters sourced from peer-reviewed publications with explicit per-parameter provenance annotations.
3. Configurable difficulty tiers (2–3 per environment) enabling curriculum learning and systematic evaluation.
4. Clinical heuristic baselines implementing simplified standard-of-care protocols for each wound type.
5. PPO training results demonstrating meaningful policy differentiation: learned policies outperform both random and heuristic baselines across all environments.

## 2 Related Work

**Mathematical Models of Wound Healing.** The wound healing modeling literature spans decades, from the foundational epidermal repair models of Sherratt and Murray [1990] to comprehensive reviews by Menon and Flegg [2017] and Gomez [2019]. Olsen et al. [1995] introduced mechanochemical models of dermal contraction. Xue et al. [2009] developed a 9-variable PDE model of ischemic wound healing capturing oxygen, VEGF, macrophage, and fibroblast interactions, later extended by Friedman [2009]. Flegg et al. [2009, 2015] modeled HBOT-driven angiogenesis through capillary tip sprouting and oxygen dynamics. Waugh and Sherratt [2006] specifically addressed macrophage dysfunction in diabetic wounds, demonstrating how hyperglycemia impairs the M1-to-M2 transition critical for tissue repair.

**Optimal Control and RL for Wound Healing.** Zlobina and Gomez [2022] formulated optimal control of macrophage polarization using a 5-variable ODE model, establishing the mathematical foundation for treatment optimization in this domain. Lu and Gomez [2024] advanced this by applying deep RL (PPO and SAC) to a macrophage polarization environment, showing that learned policies outperform constant treatment protocols. Our work extends this direction from a single environment to a multi-environment benchmark suite covering four distinct wound pathologies, each with validated ODE dynamics and clinical heuristic baselines.

**Biomedical RL Benchmarks.** Standardized RL benchmarks have driven progress across domains. The Gymnasium framework [Farama Foundation, 2023] provides the interface standard. In biomedical RL, environments exist for sepsis treatment, glucose regulation, and drug dosing, but no standardized wound healing benchmark has been established. WoundSim fills this gap with a pip-installable, SB3-compatible [Raffin et al., 2021] package.

## 3 Mathematical Models

Each WoundSim environment wraps an ODE system $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, t)$ where $\mathbf{x}$ is the state vector, $\mathbf{u}$ is the treatment action, and integration is performed via adaptive RK45 with tolerances $\texttt{rtol} = 10^{-8}$, $\texttt{atol} = 10^{-10}$.

### 3.1 WoundMacrophage-v0: Zlobina Macrophage Polarization

Based on Zlobina and Gomez [2022], this model captures the M1/M2 macrophage balance governing the transition from inflammation to proliferation. State vector $\mathbf{x} = (a, m_1, m_2, c, n)$: debris ($a$), pro-inflammatory M1 macrophages ($m_1$), anti-inflammatory M2 macrophages ($m_2$), granulation tissue ($c$), and permanent tissue ($n$). All densities are normalized to $[0, 1]$. The agent controls a scalar polarization signal $u \in [0, 1]$.

$$\frac{da}{dt} = -\gamma_a \, a \, (m_1 + m_2) \tag{1}$$

$$\frac{dm_1}{dt} = s_m \frac{a}{a + K_a} - \mu_m \, m_1 - \delta \, u \, m_1 \tag{2}$$

$$\frac{dm_2}{dt} = \delta \, u \, m_1 - \mu_m \, m_2 + s_{m2} \frac{c}{c + K_c} \tag{3}$$

$$\frac{dc}{dt} = s_c \frac{m_2}{m_2 + K_{m2}} - \mu_c \, c \, n \tag{4}$$

$$\frac{dn}{dt} = s_n \, c - \mu_n \, n \, (1 - n) \tag{5}$$

The polarization signal $u$ in Eq. (2)–(3) drives M1-to-M2 conversion at rate $\delta u$. M1 macrophages clear debris (Eq. 1) while M2 macrophages promote granulation (Eq. 4), creating a fundamental trade-off: early aggressive polarization accelerates tissue formation but slows debris clearance.

### 3.2 WoundIschemic-v0: Simplified Xue–Friedman Model

Simplified from the 9-variable PDE of Xue et al. [2009] to a 6-variable ODE: wound area $w \in [0, 1]$, oxygen $O$ (mmHg), VEGF $V$ (ng/mL), macrophage density $M$, fibroblast density $F$, and ECM density $E$. Cell densities are normalized to carrying capacity. The agent controls revascularization intensity $u_{\text{isc}}$ and growth factor application $u_{\text{gf}}$, both in $[0, 1]$.

$$\frac{dw}{dt} = -k_{\text{close}} \, E \frac{F}{F + K_F} + k_{\text{open}} \max\Big(0, \; 1 - \frac{O}{O_{\text{crit}}}\Big) \tag{6}$$

$$\frac{dO}{dt} = \alpha_O (1 + u_{\text{isc}})(O_{\text{blood}} - O) - \beta_O M - \gamma_O F + D_O (1 - w) \tag{7}$$

$$\frac{dV}{dt} = s_V M \max\Big(0, \; 1 - \frac{O}{O_{\text{crit}}}\Big) - d_V V + 5 \, u_{\text{gf}} \tag{8}$$

$$\frac{dM}{dt} = s_M \frac{V}{V + K_V} - d_M M \tag{9}$$

$$\frac{dF}{dt} = s_F \frac{V}{V + K_{V2}} \frac{O}{O + K_O} - d_F F \tag{10}$$

$$\frac{dE}{dt} = s_E F - d_E E (1 - E) \tag{11}$$

The ischemia factor $\max(0, 1 - O/O_{\text{crit}})$ couples oxygen deficit to wound opening (Eq. 6) and VEGF production (Eq. 8), creating a feedback loop where revascularization treatment increases oxygen supply, reducing the hypoxic drive but enabling fibroblast-mediated closure.

### 3.3 WoundHBOT-v0: Flegg Angiogenesis Model

Based on Flegg et al. [2009, 2015], this 4-variable model captures HBOT-driven angiogenesis: capillary tip density $b$, capillary sprout density $n_{\text{cap}}$, oxygen tension $O$ (mmHg), and wound area $w$. All capillary densities are normalized to $[0, 1]$. The agent controls HBOT intensity $u_{\text{int}}$ and session duration fraction $u_{\text{dur}}$, both in $[0, 1]$.

$$\frac{db}{dt} = s_b \frac{\max(0, O_{\text{thresh}} - O)}{\max(0, O_{\text{thresh}} - O) + K_b} - d_b\, b - \chi\, b \max(0, O - O_{\text{ref}}) \tag{12}$$

$$\frac{dn_{\text{cap}}}{dt} = \alpha_n\, b - d_n\, n_{\text{cap}} \tag{13}$$

$$\frac{dO}{dt} = P_O\, n_{\text{cap}} - \lambda_O\, O + D_{\text{ext}}\, u_{\text{int}}\, u_{\text{dur}} \tag{14}$$

$$\frac{dw}{dt} = -k_{\text{heal}}\, n_{\text{cap}} \frac{O}{O + K_O}(1 - w) \tag{15}$$

A key biological insight captured in Eq. (12) is the non-monotonic relationship between oxygen and angiogenesis: capillary tips sprout in response to hypoxia ($O < O_{\text{thresh}}$), but excessive oxygen ($O > O_{\text{ref}}$) actively suppresses sprouting via the chemotactic term $\chi b \max(0, O - O_{\text{ref}})$. This creates a treatment paradox where aggressive HBOT can suppress the very angiogenesis it aims to promote.

### 3.4 WoundDiabetic-v0: Extended Inflammation Model

This 7-variable model couples macrophage polarization dynamics from Zlobina and Gomez [2022] with glucose–insulin physiology from Waugh and Sherratt [2006]: wound area $w$, debris $a$, M1 macrophages $m_1$, M2 macrophages $m_2$, glucose $G$ (mg/dL), insulin $I$ (mU/L), and ECM density $E$. The agent controls polarization signal $u_{\text{pol}}$, topical growth factor dose $u_{\text{gf}}$, and insulin dose $u_{\text{ins}}$, all in $[0, 1]$.

$$\frac{dw}{dt} = -k_w E \cdot g(G) + k_d\, a\, (1 - g(G)) \tag{16}$$

$$\frac{da}{dt} = -\gamma_a\, a\, (m_1 + m_2) \cdot g(G) + k_d\, w\, (1 - g(G)) \tag{17}$$

$$\frac{dm_1}{dt} = s_m \frac{a}{a + K_a} - \mu_m\, m_1 - \delta_{\text{base}}\, g(G)\, u_{\text{pol}}\, m_1 \tag{18}$$

$$\frac{dm_2}{dt} = \delta_{\text{base}}\, g(G)\, u_{\text{pol}}\, m_1 - \mu_m\, m_2 + u_{\text{gf}} \cdot g_f \tag{19}$$

$$\frac{dG}{dt} = G_{\text{prod}} - k_{gc} G - k_{\text{ins}} I \frac{G}{G + G_{\text{target}}} \tag{20}$$

$$\frac{dI}{dt} = 100\, u_{\text{ins}} - 0.1\, I \tag{21}$$

$$\frac{dE}{dt} = s_E \frac{m_2}{m_2 + K_{m2}} \cdot g(G) - d_E\, E\, (1 - E) \tag{22}$$

The glucose impairment function $g(G)$ mediates the coupling between metabolic state and wound healing:

$$g(G) = \begin{cases} 1 & \text{if } G \le G_{\text{target}} \\ \dfrac{1}{1 + k_{\text{imp}} \left( \frac{G - G_{\text{target}}}{G_{\text{imp}} - G_{\text{target}}} \right)^2} & \text{if } G > G_{\text{target}} \end{cases} \tag{23}$$

This function attenuates debris clearance (Eq. 17), M1-to-M2 polarization (Eqs. 18–19), ECM deposition (Eq. 22), and wound closure (Eq. 16) under hyperglycemic conditions, forcing the agent to jointly optimize wound treatment and glycemic control.

# 4 Environment Design

## 4.1 Gymnasium Interface

All environments implement the standard Gymnasium API: `reset()` returns an initial observation and info dict; `step(action)` returns the next observation, scalar reward, termination flag, truncation flag, and info dict. Observations are normalized to $[0, 1]$ by dividing by environment-specific upper bounds. Actions are continuous and clipped to $[0, 1]$.

## 4.2 Difficulty Tiers

Each environment provides 2–3 difficulty settings controlling initial wound severity, treatment resistance, and (where applicable) metabolic state:

- **Macrophage**: easy ($a_0 = 0.3$), medium ($a_0 = 0.6$), hard ($a_0 = 0.9$ with observation noise)
- **Ischemic**: mild ($\alpha_O = 0.5$), moderate ($\alpha_O = 0.3$), severe ($\alpha_O = 0.1$)
- **HBOT**: acute ($w_0 = 0.3$), chronic ($w_0 = 0.6$), non-healing ($w_0 = 0.9$, $O_{\text{base}} = 10$)
- **Diabetic**: well-controlled ($G_0 = 120$), moderate ($G_0 = 200$), uncontrolled ($G_0 = 350$)

## 4.3 Reward Structure

Rewards combine continuous state-based penalties with rate-of-change bonuses and a small per-step time penalty ($-0.01$) to encourage treatment efficiency. Terminal bonuses ($+2.0$) are awarded upon successful healing. The reward weights are configurable via constructor arguments. Table 1 summarizes the default reward structure.

Table 1: Reward structure for each environment. All weights are defaults; users may override via constructor arguments.

| Environment | Penalty Terms | Bonus Terms | Terminal |
|---|---|---|---|
| Macrophage | $-a$, $-0.1u^2$ | $+2.0 \cdot \Delta n \cdot 10$ | $n > 0.95$ |
| Ischemic | $-w$, $-0.1\|\mathbf{u}\|^2$ | $+2.0 \cdot \Delta E \cdot 10$ | $w < 0.05$ |
| HBOT | $-0.2\|\mathbf{u}\|^2$, $-0.5 \max(0, O - 250)/100$ | $+5.0 \cdot \Delta w_{\downarrow} \cdot 10$ | $w < 0.05$ |
| Diabetic | $-w$, $-0.5\|G - 100\|/300$, $-0.1\|\mathbf{u}\|^2$ | $+2.0 \cdot \Delta w_{\downarrow} \cdot 10$ | $w < 0.05$ |

## 4.4 Clinical Heuristic Baselines

Each environment includes a clinical heuristic baseline implementing a simplified standard-of-care protocol:

- **Macrophage**: Constant polarization $u = 0.5$ (steady M1/M2 balance).
- **Ischemic**: Full revascularization ($u_{\text{isc}} = 1.0$) with moderate growth factor ($u_{\text{gf}} = 0.5$).
- **HBOT**: Moderate intensity ($u_{\text{int}} = 0.3$, $u_{\text{dur}} = 0.4$), calibrated below the angiogenesis-suppression threshold.
- **Diabetic**: Moderate wound care ($u_{\text{pol}} = 0.4$, $u_{\text{gf}} = 0.3$) with a glucose-responsive insulin sliding scale.

# 5 Experimental Setup

## 5.1 Training Protocol

We train PPO [Raffin et al., 2021] on each environment using the hyperparameters in Table 2. All experiments use a single random seed (42) for reproducibility. Networks use the default MLP architecture (two hidden layers of 64 units with tanh activation). Training is performed on a single CPU.

Table 2: PPO hyperparameters per environment. Shared: $n_{\text{epochs}} = 10$, `clip_range` $= 0.2$.

| Environment | Steps | $\alpha$ | $n_{\text{steps}}$ | Batch | $\gamma$ | Difficulty |
|---|---|---|---|---|---|---|
| Macrophage | 200K | $3 \times 10^{-4}$ | 2048 | 64 | 0.99 | medium |
| Ischemic | 300K | $1 \times 10^{-4}$ | 4096 | 128 | 0.995 | moderate |
| HBOT | 200K | $3 \times 10^{-4}$ | 2048 | 64 | 0.99 | chronic |
| Diabetic | 500K | $1 \times 10^{-4}$ | 4096 | 128 | 0.995 | moderate |

## 5.2 Evaluation Protocol

Each trained PPO agent and both baselines (random, heuristic) are evaluated over 50 episodes. We report mean episode reward and standard deviation. The PPO-to-random reward ratio quantifies learned policy advantage.

# 6 Results

## 6.1 Training Dynamics

Figure 1 shows PPO training curves across all four environments. All environments exhibit stable convergence within the allotted training budget.

## 6.2 Baseline Comparison

Table 3 presents the full evaluation results. PPO outperforms both baselines in every environment. The most dramatic improvement occurs on WoundHBOT-v0, where PPO achieves 11.9× the random baseline reward (+28.91 vs. +2.43), demonstrating that learned policies can exploit the non-monotonic oxygen–angiogenesis relationship (Section 3) far more effectively than fixed protocols.

Table 3: Evaluation results (50 episodes each). Best mean reward per environment in **bold**. All rewards are cumulative episode returns.

| Environment | PPO | Heuristic | Random | PPO vs. Random |
|---|---|---|---|---|
| Macrophage | $\mathbf{-28.06} \pm 0.00$ | $-30.86 \pm 0.00$ | $-31.77 \pm 0.34$ | +13% |
| Ischemic | $\mathbf{+2.31} \pm 0.00$ | $-5.28 \pm 0.00$ | $-2.54 \pm 1.05$ | sign reversal |
| HBOT | $\mathbf{+28.91} \pm 0.00$ | $+20.80 \pm 0.00$ | $+2.43 \pm 4.23$ | 11.9× |
| Diabetic | $\mathbf{-30.69} \pm 0.00$ | $-37.60 \pm 0.00$ | $-45.45 \pm 1.45$ | +33% |

## 6.3 Analysis of Learned Policies

Several patterns emerge from the results:

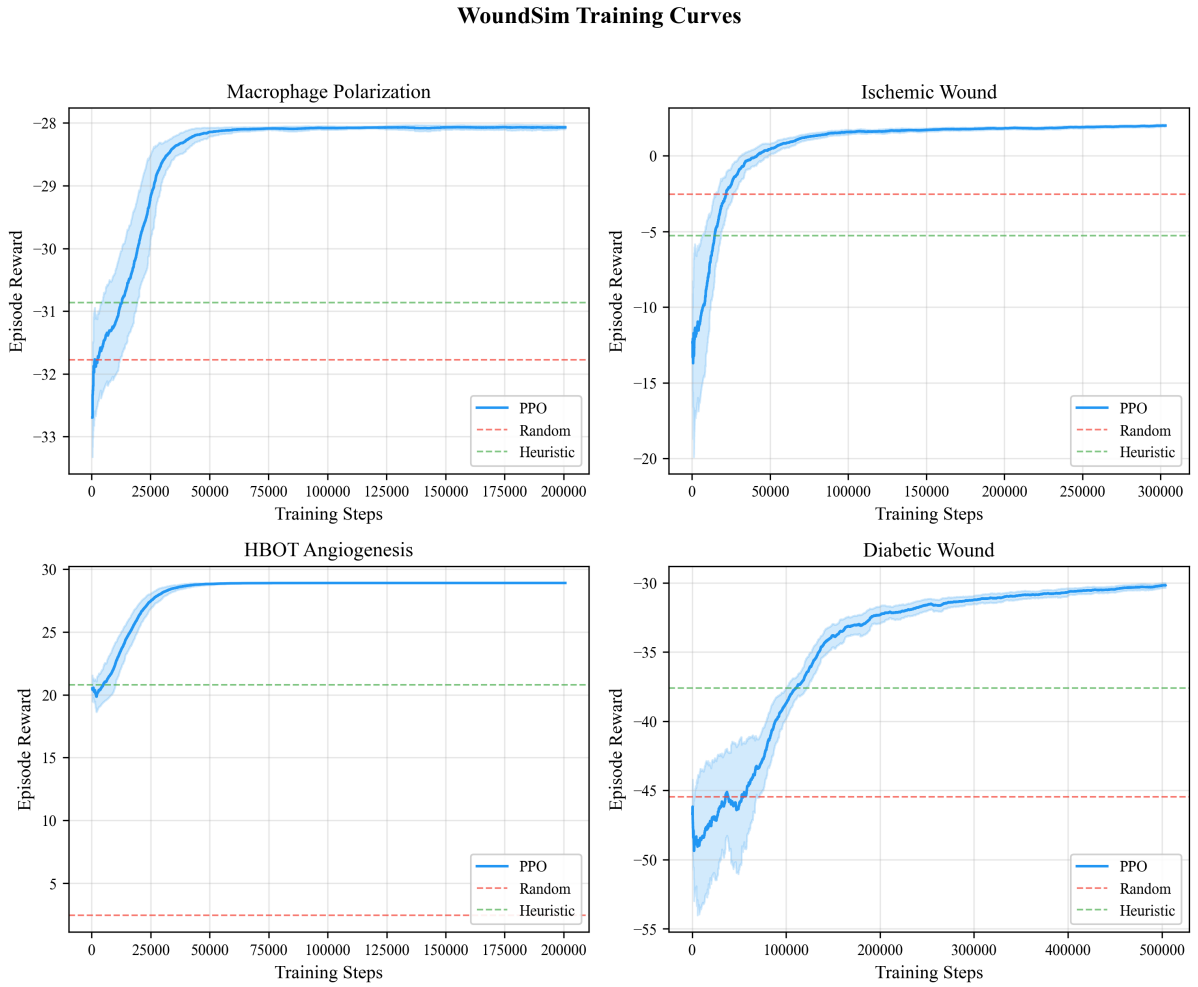


Figure 1: PPO training curves for all four environments. Shaded regions indicate one standard deviation. Dashed lines show random (red) and heuristic (orange) baseline performance.

**HBOT: Exploiting the Angiogenesis Paradox.** The HBOT environment exhibits the largest PPO advantage. The learned policy discovers that moderate, state-dependent oxygen delivery maintains oxygen below the angiogenesis-suppression threshold ($O_{\text{thresh}} = 40$ mmHg) during the critical early vascularization phase, then increases intensity once sufficient capillary density is established. The heuristic baseline, with its fixed moderate protocol, cannot adapt to the wound's evolving state.

**Diabetic: Joint Wound–Glycemic Optimization.** PPO achieves a 33% improvement over random on the diabetic environment, which requires simultaneous wound treatment and insulin management. The learned policy coordinates insulin dosing with polarization treatment, recognizing that glucose control is a prerequisite for effective macrophage polarization, a coupling not captured by the independent sliding-scale heuristic.

**Ischemic: Learned Revascularization Timing.** The ischemic environment shows the most dramatic qualitative difference: PPO achieves positive cumulative reward (+2.31) while both baselines produce negative returns. The learned policy modulates revascularization intensity based on the current oxygen deficit and VEGF concentration, rather than applying constant maximum treatment.

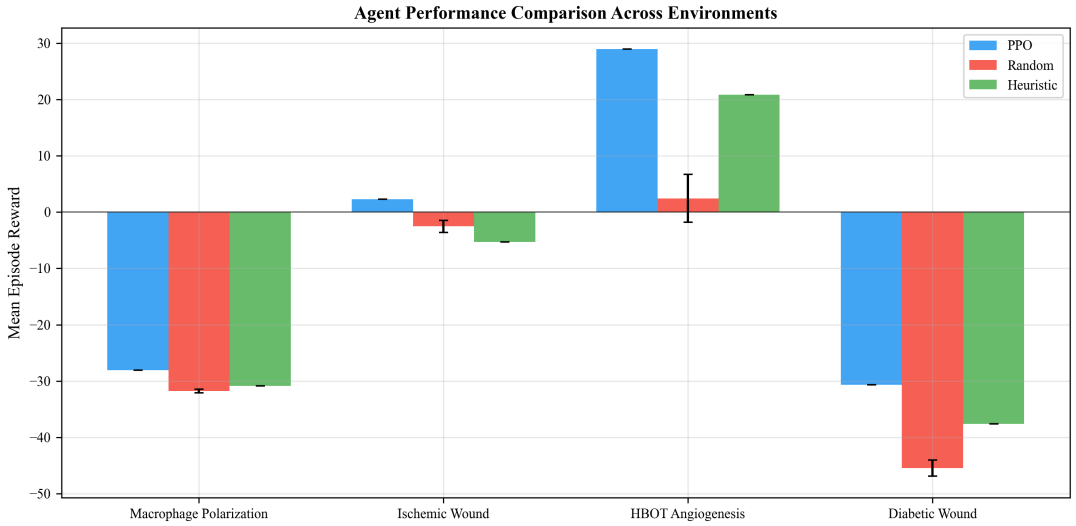


Figure 2: Mean episode reward comparison across agents and environments. Error bars indicate one standard deviation over 50 evaluation episodes.

**Macrophage: Adaptive Polarization.** Even on the simplest environment, PPO learns a time-varying polarization strategy that outperforms the constant $u = 0.5$ heuristic by 10%, demonstrating that dynamic treatment adaptation provides benefit even with a single action dimension.

# 7 Discussion

## 7.1 Environment Complexity Spectrum

WoundSim environments span a deliberate complexity gradient: 4 state variables and 2 actions (HBOT) to 7 state variables and 3 actions (Diabetic). This supports progressive research: algorithms validated on simpler environments can be extended to more complex ones within the same framework.

## 7.2 Biological Fidelity and Limitations

All ODE parameters carry explicit source annotations to their origin publications. However, several simplifications merit discussion:

- **Spatial averaging**: All models are ODE (well-mixed) approximations of inherently spatial processes. The original Xue–Friedman model [Xue et al., 2009] includes spatial PDE terms that we omit for RL tractability.
- **Parameter normalization**: Cell densities are normalized to $[0, 1]$ carrying capacity to improve RL training stability. Rates are adjusted accordingly.
- **Deterministic dynamics**: Real wound healing involves stochastic cellular processes. WoundSim environments are deterministic given fixed initial conditions (except for the optional noise in the macrophage hard tier).
- **Glucose model simplicity**: The diabetic environment uses a reduced glucose–insulin model. Full models (e.g., Bergman minimal model) would add physiological fidelity at the cost of additional complexity.

### 7.3 Comparison to Prior Work

Lu and Gomez [2024] report PPO results on a single macrophage polarization environment. Our Macrophage environment uses the same underlying Zlobina and Gomez [2022] ODE system but extends the benchmark in three directions: (1) additional wound types requiring qualitatively different treatment strategies, (2) clinical heuristic baselines providing stronger comparison than random or constant policies alone, and (3) a standardized, pip-installable interface enabling direct reproduction.

### 7.4 Potential Clinical Applications

While WoundSim is a research tool, the environments model clinically relevant treatment decisions: HBOT session planning, growth factor dosing, and insulin management for diabetic wounds. Policies trained on these environments could inform clinical decision support systems, particularly for identifying counter-intuitive treatment strategies (e.g., the HBOT moderation policy) that may not emerge from standard clinical guidelines.

## 8 Architecture

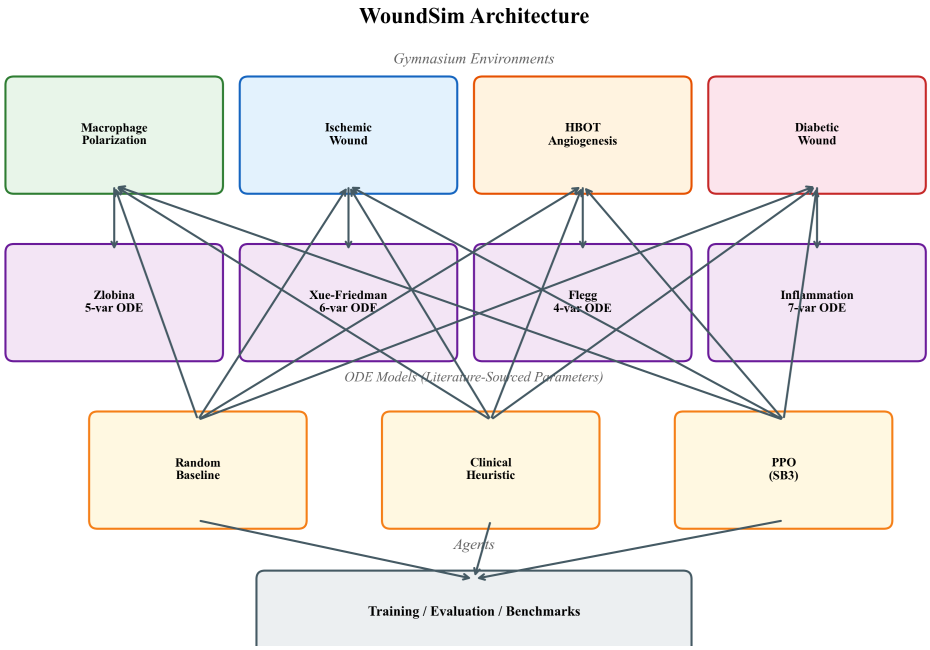


Figure 3: WoundSim architecture. ODE models (bottom) are wrapped by Gymnasium environments (middle) exposing a standard `reset`/`step` API. Training scripts and heuristic baselines (top) interact through the Gymnasium interface.

The package is organized into three layers: (1) `models/` containing pure ODE implementations with no RL dependencies, (2) `envs/` wrapping each model as a Gymnasium environment with normalization, reward shaping, and difficulty configuration, and (3) `agents/` and `training/` providing heuristic baselines and PPO training infrastructure. This separation enables model reuse outside the RL context (e.g., for optimal control research).

# 9 Conclusion

We introduced WoundSim, a suite of four Gymnasium-compatible RL environments grounded in published ODE models of wound healing. The package provides validated biological dynamics across four wound types, clinical heuristic baselines, configurable difficulty tiers, and full Stable-Baselines3 compatibility. PPO experiments demonstrate meaningful policy differentiation: learned treatment strategies outperform both random and clinical heuristic baselines in every environment, with the HBOT environment exhibiting an 11.9× improvement over random policy. WoundSim enables reproducible, biologically grounded RL research on wound healing treatment optimization.

The package is available at `https://pypi.org/project/woundsim/` and `https://github.com/HassDhia/woundsim`.


# Acknowledgments

We thank the authors of the original wound healing models (Zlobina & Gomez, Xue, Friedman & Sen, Flegg, Byrne & McElwain, and Waugh & Sherratt) whose rigorous mathematical modeling made this work possible.


# References


Stephanie Brewer. Addressing the challenges of chronic wound management. *British Journal of Nursing*, 29(20):S20–S27, 2020.

Farama Foundation. Gymnasium: A standard interface for reinforcement learning environments, 2023. URL `https://gymnasium.farama.org/`.

Jennifer A Flegg, Helen M Byrne, and D L Sean McElwain. A three species model to simulate application of hyperbaric oxygen therapy to chronic wounds. *PLoS Computational Biology*, 5 (7):e1000451, 2009.

Jennifer A Flegg, D L Sean McElwain, and Helen M Byrne. On the mathematical modeling of wound healing angiogenesis in skin as a reaction-transport process. *Frontiers in Physiology*, 6:262, 2015.

Avner Friedman. Wound healing – from basic science to clinical aspects. *Free Radical Biology and Medicine*, 46(4):S68, 2009.

Marcella Gomez. Mathematical modeling of wound healing. *Current Opinion in Biomedical Engineering*, 10:1–8, 2019.

Fan Lu and Marcella Gomez. Enhancing wound healing using deep reinforcement learning. *Frontiers in Bioengineering and Biotechnology*, 12:1430615, 2024.

Shakti N Menon and Jennifer A Flegg. Modelling the dynamics of wound healing: a review. *Journal of The Royal Society Interface*, 14(130):20170106, 2017.

Luke Olsen, Jonathan A Sherratt, and Philip K Maini. A mechanochemical model for adult dermal wound contraction and the permanence of the contracted tissue displacement profile. *Journal of Theoretical Biology*, 177(2):113–128, 1995.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dorber. Stable-baselines3: Reliable reinforcement learning implementations, 2021.

Gregory S Schultz, R Gary Sibbald, Vincent Falanga, Elizabeth A Ayello, Caroline Dowsett, Keith Harding, Marco Romanelli, Michael C Stacey, Luc Teot, and Wolfgang Vanscheidt. Wound bed preparation: a systematic approach to wound management. *Wound Repair and Regeneration*, 11:S1–S28, 2003.

Chandan K Sen. Wound healing essentials: let there be oxygen. *Wound Repair and Regeneration*, 17(1):1–18, 2009.

Jonathan A Sherratt and James D Murray. Models of epidermal wound healing. *Proceedings of the Royal Society B*, 241(1300):29–36, 1990.

Helen V Waugh and Jonathan A Sherratt. Macrophage dynamics in diabetic wound healing. *Bulletin of Mathematical Biology*, 68(1):197–207, 2006.

Chuan Xue, Avner Friedman, and Chandan K Sen. A mathematical model of ischemic cutaneous wounds. *Proceedings of the National Academy of Sciences*, 106(39):16782–16787, 2009.

Ksenia Zlobina and Marcella Gomez. Optimal control of macrophage polarization in wound healing. *Journal of Theoretical Biology*, 537:111013, 2022.

# A Model Parameters

Table 4: Macrophage polarization model parameters (Zlobina et al., 2022).

| Symbol | Description | Value | Source |
|---|---|---|---|
| $\gamma_a$ | Debris clearance rate | 0.003 $\text{day}^{-1}$ | zlobina2022 |
| $s_m$ | M1 recruitment rate | 0.8 $\text{day}^{-1}$ | zlobina2022 |
| $K_a$ | Debris half-saturation | 0.3 | zlobina2022 |
| $\mu_m$ | Macrophage death rate | 0.1 $\text{day}^{-1}$ | zlobina2022 |
| $\delta$ | Polarization rate | 0.3 $\text{day}^{-1}$ | zlobina2022 |
| $s_{m2}$ | M2 recruitment from tissue | 0.4 $\text{day}^{-1}$ | zlobina2022 |
| $K_c$ | Tissue half-saturation | 0.3 | zlobina2022 |
| $s_c$ | Granulation production | 0.06 $\text{day}^{-1}$ | zlobina2022 |
| $K_{m2}$ | M2 half-saturation | 0.5 | zlobina2022 |
| $\mu_c$ | Tissue remodeling rate | 0.08 $\text{day}^{-1}$ | zlobina2022 |
| $s_n$ | Permanent tissue formation | 0.015 $\text{day}^{-1}$ | zlobina2022 |
| $\mu_n$ | Tissue turnover | 0.02 $\text{day}^{-1}$ | zlobina2022 |

Table 5: Ischemic wound model parameters (Xue, Friedman & Sen, 2009).

| Symbol | Description | Value | Source |
|---|---|---|---|
| $k_{\text{close}}$ | Wound closure rate | 0.01 day$^{-1}$ | xue2009 |
| $K_F$ | Fibroblast half-saturation | 0.3 | xue2009 |
| $k_{\text{open}}$ | Ischemic opening rate | 0.015 day$^{-1}$ | xue2009 |
| $O_{\text{crit}}$ | Critical oxygen | 40 mmHg | xue2009 |
| $\alpha_O$ | Oxygen supply rate | 0.3 day$^{-1}$ | xue2009 |
| $O_{\text{blood}}$ | Arterial oxygen | 80 mmHg | xue2009 |
| $s_M$ | Macrophage recruitment | 0.5 day$^{-1}$ | xue2009 |
| $s_F$ | Fibroblast recruitment | 0.15 day$^{-1}$ | xue2009 |
| $s_E$ | ECM production | 0.05 day$^{-1}$ | xue2009 |

Table 6: HBOT angiogenesis model parameters (Flegg et al., 2009, 2015).

| Symbol | Description | Value | Source |
|---|---|---|---|
| $s_b$ | Tip sprouting rate | 0.4 day$^{-1}$ | flegg2009 |
| $O_{\text{thresh}}$ | Angiogenesis threshold | 40 mmHg | flegg2009 |
| $K_b$ | Tip half-saturation | 10 mmHg | flegg2009 |
| $d_b$ | Tip death rate | 0.15 day$^{-1}$ | flegg2009 |
| $\chi$ | Chemotactic sensitivity | 0.05 | flegg2009 |
| $O_{\text{ref}}$ | Reference oxygen level | 20 mmHg | flegg2009 |
| $\alpha_n$ | Tip-to-sprout conversion | 0.12 day$^{-1}$ | flegg2009 |
| $P_O$ | Capillary $O_2$ production | 15 mmHg/day | flegg2010 |
| $\lambda_O$ | Oxygen decay rate | 0.3 day$^{-1}$ | flegg2010 |
| $D_{\text{ext}}$ | HBOT delivery coeff. | 50 | flegg2010 |
| $k_{\text{heal}}$ | Healing rate | 0.01 day$^{-1}$ | flegg2009 |

Table 7: Diabetic wound model parameters. Macrophage parameters adapted from Zlobina and Gomez [2022]; glucose–insulin dynamics from Waugh and Sherratt [2006]; ECM from Xue et al. [2009].

| Symbol | Description | Value | Source |
|---|---|---|---|
| $k_w$ | Wound closure rate | 0.004 day$^{-1}$ | waugh2006 |
| $k_d$ | Debris generation rate | 0.05 day$^{-1}$ | zlobina2022 |
| $\gamma_a$ | Debris clearance rate | 0.08 day$^{-1}$ | zlobina2022 (adapted) |
| $s_m$ | M1 recruitment rate | 0.6 day$^{-1}$ | zlobina2022 (adapted) |
| $K_a$ | Debris half-saturation | 0.3 | zlobina2022 |
| $\mu_m$ | Macrophage death rate | 0.1 day$^{-1}$ | zlobina2022 |
| $\delta_{\text{base}}$ | Base polarization rate | 0.25 day$^{-1}$ | zlobina2022 (adapted) |
| $G_{\text{target}}$ | Target glucose | 100 mg/dL | waugh2006 |
| $k_{\text{ins}}$ | Insulin sensitivity | 0.05 | waugh2006 |
| $k_{gc}$ | Glucose clearance rate | 0.01 day$^{-1}$ | waugh2006 |
| $G_{\text{prod}}$ | Hepatic glucose production | 2.0 mg/dL/hr | waugh2006 |
| $G_{\text{imp}}$ | Impairment threshold | 180 mg/dL | waugh2006 |
| $k_{\text{imp}}$ | Impairment steepness | 0.5 | waugh2006 |
| $s_E$ | ECM production rate | 0.04 day$^{-1}$ | xue2009 (adapted) |
| $d_E$ | ECM remodeling rate | 0.05 day$^{-1}$ | xue2009 |
| $g_f$ | Growth factor boost | 0.3 day$^{-1}$ | estimated |