Journal of Sustainability, Policy, and Practice
EISSN: 3105-1448 | PISSN: 3105-143X | Vol. 1, No. 3 (2025)

Article

Adaptive Dose Optimization Algorithm for LED-based Photodynamic Therapy Based on Deep Reinforcement Learning

Zonglei Dong [1,*] and Ruoxi Jia [2]

[1] University of Texas at Dallas, TX, USA
[2] Computer Science, University of Southern California, CA, USA
* Correspondence: Zonglei Dong, University of Texas at Dallas, TX, USA

Abstract: Current photodynamic therapy fails. Not occasionally-systematically. We measured optical coefficients across 1,847 lesions: peripheral zones exhibit $\mu s' \approx 20$ cm^{-1} while necrotic zones $\mu s' \approx 0.8$ cm^{-1}, a 25-fold discontinuity that renders uniform protocols obsolete. Drug clearance varies threefold. One patient metabolizes protoporphyrin IX in two hours; another requires six. Oxygen maps tell an equally chaotic story-partial pressures crash from 95 mmHg to anoxic thresholds within minutes, creating dead zones where photochemistry simply stops. We built a system that adapts. Deep reinforcement learning processes 384 physiological signals in real-time, adjusting power density and fractionation schedules every 100 milliseconds. The architecture splits decision-making: one network learns patient baselines, another computes action advantages. This dueling structure stabilizes training on sparse clinical data, where 62.5% of treatments historically achieved only partial or no response. Results contradict decades of conservative practice. Phototoxicity decreased from 18.3% to 7.6%, a relative risk reduction of 58% (95% CI: 49-66%). Complete responses jumped from 37.8% to 58.7%-not through gentler treatment, but through aggressive, precisely-timed interventions the algorithm discovered autonomously. A paradox emerged: lower doses for superficial lesions, intense protocols for deep tumors, opposite to clinical intuition. The system runs on standard GPUs. No specialized hardware. Forty-seven prospective patients confirmed what retrospective analysis suggested: adaptive control fundamentally outperforms fixed protocols.

Keywords: deep reinforcement learning; LED-based photodynamic therapy; adaptive dose control; personalized medicine

Received: 09 September 2025
Revised: 19 September 2025
Accepted: 04 October 2025
Published: 10 October 2025

1. Introduction

1.1. Current Challenges in Led-Based Photodynamic Therapy Dosimetry

A 635 nm LED array at 150 J/cm² should work predictably. It doesn't. We tracked 324 treatments with identical protocols as a preliminary analysis-same tumor grade, same anatomical site, same protocol. Results ranged from complete ablation to zero response. The problem wasn't the equipment.

Tissue heterogeneity dominates outcomes. Consider oxygen dynamics: baseline StO_2 reads ~95%, suggesting adequate perfusion. Start irradiation. Within 90 seconds, consumption outpaces delivery. Gradients form. Perivascular regions maintain 60% saturation while zones 200 microns away drop below 20%-the threshold where singlet oxygen generation ceases entirely. Standard protocols can't see this. They deliver photons to dead tissue.

Wu's team discovered similar chaos in thermal patterns [1]. Burns that appeared uniform showed 23% variation in deep tissue damage when analyzed with reinforcement

learning. PDT exhibits worse heterogeneity. We measured temperature spreads exceeding 10°C within single treatment fields, enough to denature proteins in hot spots while leaving cold regions untreated.

The mathematics are unforgiving. Singlet oxygen production follows Type II photochemistry: $^3PS^* + {}^3O_2 \rightarrow PS + {}^1O_2$, where quantum yield depends on triplet state lifetime, oxygen concentration, and quenching rates. Each variable fluctuates independently. Microvascular architecture-chaotic in tumors, organized in normal tissue-determines local oxygen flux. Some vessels shunt. Others stagnate. Perfusion heterogeneity index (PHI) values reach 0.73 in malignant tissue versus 0.21 in healthy controls.

Drug distribution adds another layer. Protoporphyrin IX doesn't distribute uniformly-it pools. Necrotic cores accumulate 10-fold higher concentrations than viable rims. But necrotic tissue lacks vasculature. No oxygen means no reaction, regardless of drug concentration. Meanwhile, well-perfused edges with minimal drug still respond. The mismatch between drug location and reaction capability defines treatment failure.

This study specifically focuses on LED-based PDT systems due to their advantages in delivering uniform illumination fields, cost-effectiveness, and enhanced safety profiles compared to laser-based alternatives. All experimental validation was performed using LED arrays [Commercial LED array system (635nm, 0-200 mW/cm² adjustable power density), 635nm].

1.2. Limitations of Fixed-Dose Protocols and Need for Personalization

Population protocols assume average patients. Average patients don't exist.

Take two basal cell carcinomas, both 15mm in diameter, both on facial skin. Lesion A sits on the nose-thin epidermis, dense vasculature, optical penetration depth 2.3mm. Lesion B occupies the temple-actinic damage, dermal thickening, penetration depth 0.8mm. Same diagnosis, same size, threefold difference in treatable volume. Saager proved this quantitatively: absorption varies from 0.15 to 0.75 cm^{-1} within individual tumors [2]. Light that penetrates 4mm in one region dies at 1mm in another.

Pharmacokinetics destroys reproducibility further. We tracked PpIX accumulation in a subset of 743 patients. Peak times ranged from 2 to 8 hours post-administration. Why? Circadian effects-morning doses peak 40% faster. Age matters: patients over 70 clear drugs at 0.6 mL/min/kg versus 1.1 mL/min/kg in younger cohorts. Liver function explains some variance. CYP3A4 polymorphisms explain more. But 30% remains unexplained-likely reflecting unmeasured metabolic variations.

The trust problem compounds technical challenges. Nath documented physician resistance to algorithmic ophthalmology recommendations [3]. Black boxes fail in medicine. Doctors need explanations, not just predictions. They reject systems that contradict training, even when data support deviation. Our approach aims to bridge this gap-optimizing outcomes while maintaining interpretability.

1.3. Research Objectives

We abandon predetermined protocols entirely.

The system monitors continuously: tissue oxygen via near-infrared spectroscopy, drug concentration through fluorescence, temperature with fiber Bragg gratings, perfusion using laser Doppler, and inflammatory markers via bedside immunoassays. Five data streams, 384 total features, synchronized at a 10Hz master clock, providing updates every 100 milliseconds. Every 100 milliseconds, deep networks evaluate the physiological state and adjust treatment parameters. Not retrospectively-during irradiation.

Three constraints shape the architecture. First, safety: hard limits on temperature (45°C), with a soft limit at 43°C, dose (300 J/cm²), and hypoxia duration. Safety controller implements hierarchical limits: (1) Soft limit at 43°C triggers power reduction to 75% of current level; (2) Hard limit at 45°C immediately terminates irradiation with mandatory 60s cooling period; (3) Cumulative thermal dose monitored using CEM43 (cumulative equivalent minutes at 43°C), with session termination if CEM43 > 120 minutes at 95th per-

centile location. Second, latency: decisions within 100 milliseconds to maintain therapeutic momentum. Third, interpretability: every recommendation must trace to a physiological justification.

We validate across scales. Historical data (1,847 sessions) trains base models. Monte Carlo simulations explore parameter spaces clinically inaccessible-what happens at 500 mW/cm²? How do tumors respond to microsecond pulsing? Animal models would take years; simulation provides answers in hours. Finally, 47 prospective patients test real-world performance. No cherry-picking. Consecutive enrollment. Standard clinical settings.

2. Related Work and Theoretical Foundation

2.1. Evolution of Dose Control Methods in Photodynamic Therapy

Early PDT relied on observation. Erythema means effect. No erythema meant an increased dose. Crude but occasionally effective. Animal models brought quantification-mice revealed dose thresholds: 15-25 J/cm² for vascular shutdown, 50-100 J/cm² for direct kill, 150-200 J/cm² for immune activation. The numbers seemed universal. They weren't.

Species don't scale. Mouse metabolic rate: 0.18 mL O_2/g/h. Human: 0.06 mL O_2/g/h. A threefold difference in oxygen consumption fundamentally alters photodynamic kinetics. A protocol lethal to murine tumors barely affects human lesions. Worse, mice lack the heterogeneity that defines human disease. Laboratory tumors grow predictably. Clinical tumors don't.

Tirand's work changed thinking [4]. Low fluence rates-counterintuitively-worked better. Not 100 mW/cm² but 25 mW/cm². Less power, more effect. The mechanism: oxygen depletion and reperfusion balance. High power exhausts oxygen faster than blood delivers it. Reactions cease. Low power maintains a steady state. The discovery invalidated decades of "more is better" philosophy.

Mathematical models emerged to capture these dynamics. Reaction-diffusion equations. Monte Carlo photon transport. Finite element thermal analysis. Beautiful mathematics, poor predictions. Why? Parameter uncertainty. Optical coefficients vary by 200% between measurements on the same tissue. Models assuming fixed values fail. Models incorporating uncertainty become computationally intractable.

2.2. Machine Learning Applications in Medical Light Therapy

Machine learning found patterns humans missed.

Walter's team aimed to optimize Staphylococcus aureus killing [5]. Standard approach: test wavelengths systematically. Machine learning approach: let gradients find optima. The algorithm discovered dual peaks-415nm and 545nm-corresponding to endogenous porphyrins nobody knew existed in sufficient quantities. Bacteria died faster at "wrong" wavelengths than at established protocols.

But point predictions aren't enough. Clinical data contains noise, missing values, and measurement error. Ensemble methods help. Random forests average uncertainty. Gradient boosting corrects systematic bias. Bayesian networks quantify confidence. Shen combined these for brachytherapy planning -18% dose conformity improvement over human plans [6]. Not by being smarter, but by considering more possibilities simultaneously.

Transfer learning accelerates deployment. Networks trained on melanoma images adapt to PDT fluorescence with 100 examples instead of 10,000. Feature extractors learn universal patterns-edges, textures, gradients-that generalize across modalities. The bottleneck shifts from data quantity to quality.

2.3. Reinforcement Learning Paradigms for Treatment Optimization

PDT is sequential. Each decision affects subsequent options. Reinforcement learning handles such problems naturally.

The framework is simple. States encode patient condition. Actions adjust treatment. Rewards reflect outcomes. The algorithm learns which actions in which states maximize

cumulative reward. Not immediate reward-that leads to greedy, suboptimal policies. Long-term reward, considering future consequences.

Shilyagina established physical constraints for LED sources [7]. Minimum effective power: 10 mW/cm². Maximum safe power: 200 mW/cm². Action space defined. But continuous spaces challenge classical methods. Discretization loses precision. Function approximation introduces instability. Deep networks help, but require careful architecture.

Zheng demonstrated feasibility in ICU oxygen therapy [8]. The system learned to anticipate desaturation events, adjusting FiO_2 preemptively rather than reactively. Mortality dropped 12%. PDT poses similar challenges-anticipating photobleaching, preventing hypoxia, and managing thermal accumulation. Success in critical care suggests PDT applications will work [9].

3. Proposed Adaptive Dose Optimization Algorithm

3.1. Deep Reinforcement Learning Framework Architecture

Traditional Q-learning fails here. State spaces are continuous, high-dimensional, and partially observable. We need function approximation. We need stability. We need sample efficiency on limited clinical data.

The solution: dueling networks. Instead of learning $Q(s,a)$ directly, we decompose:

$$Q(s,a) = V(s) + A(s,a) - mean(A(s,))$$

$V(s)$ captures patient prognosis independent of treatment. $A(s,a)$ isolates action-specific benefits. Why split? Most states don't require action; stable patients need monitoring, not intervention. The value stream learns these patterns. The advantage stream focuses on critical decisions. Separation improves learning efficiency by an order of magnitude.

Architecture details matter. Spatial encoder: three convolutional blocks (32, 64, 128 filters), 3×3 kernels, ReLU activation, batch normalization. Why these numbers? Empirical testing across 50 configurations. Larger networks overfit. Smaller networks underfit. This structure captures tumor boundaries and vascular patterns without memorizing patient-specific artifacts [10].

Temporal processing uses LSTMs. Hidden dimension: 128. Window: 30 seconds. Shorter windows miss trends. Longer windows dilute urgent signals. The forget gate learns to discard outdated vitals while preserving relevant history. Attention mechanisms weigh recent measurements higher during rapid changes.

Initialization shapes convergence. Xavier initialization: $W \sim N\left(0, \sqrt{\frac{2}{\text{fan}_{\text{avg}}}}\right)$. Prevents gradient explosion in deep networks. Bias: 0.01, slightly positive to avoid dead neurons. Dropout: 0.3 in dense layers only-convolutions need all features for spatial coherence. Target network updates: $\theta'_{\text{new}} \leftarrow 0.001\theta + 0.999\theta'_{\text{old}}$. Slow blending stabilizes learning (Table 1).

Table 1. Neural Network Architecture Specifications.

Component	Configuration	Parameters	Activation	Purpose
Spatial Encoder	Conv2D (32) → Conv2D (64) →Conv2D (128)	856,544	ReLU	Feature extraction
Temporal Processor	LSTM (128) → LSTM (64) → Dense (128)	234,112	Tanh/Sigmoid	Sequence modeling
Value Stream	Dense (256) →Dense (128) → Dense (1)	33,409	Linear	State evaluation
Advantage Stream	Dense (256) → Dense (128) → Dense (15)	35,983	Linear	Action ranking
Integration Layer	Concatenate →Dense (384)	98,304	ReLU	Modality fusion

Total parameters: 1,249,108. Model size: 4.8 MB (FP32), 1.2 MB (INT8 quantized).

Performance: 47ms inference on RTX 4000, maximum 48ms, well within the 100ms decision window. TensorRT optimization cuts memory by 60%. INT8 quantization maintains accuracy-surprising given medical data precision requirements. The key: normalization before quantization preserves relative relationships.

The action space consists of 15 discrete actions: 5 power density levels (25, 50, 75, 100, 150 mW/cm²) × 3 fractionation modes (continuous, 30s on/30s off, 60s on/60s off). This discretization balances control precision with computational efficiency. Action transitions are constrained to prevent tissue shock: power changes are limited to ±25 mW/cm² per step, and a minimum continuous irradiation period of 10s before switching modes. During off periods in fractionated delivery, monitoring continues, but StO_2 must recover above 30% before resuming irradiation.

3.2. State Space Definition with Patient Physiological Parameters

384 features define each state. Feature extraction and preprocessing require 23ms, neural network inference 47ms, leaving 30ms buffer within the 100ms decision window for action execution and safety checks. Not arbitrary-each captures clinically-relevant variation.

Start with oxygenation. StO_2 alone misleads-a 95% average might hide 20% regions. We measure 25 points, compute the mean, variance, spatial gradient, and temporal derivative. Frequency-domain spectroscopy at 690nm and 830nm separates oxy- from deoxyhemoglobin. Phase shifts reveal perfusion independent of absorption. Ten features from one measurement [11].

Fluorescence complexity exceeds oxygenation. Raw intensity means nothing-tissue thickness, probe pressure, and ambient light all confound signal. We normalize: divide by pre-treatment baseline, subtract autofluorescence, and correct for photobleaching. The decay curve itself carries information. Bi-exponential fit: $F(t) = 0.7e\hat{}(-0.03t) + 0.3e\hat{}(-0.001t)$. Fast component: singlet oxygen reactions. Slow: photoproduct accumulation. Rate constants become features.

Temperature seems simple. It isn't. Absolute value matters less than distribution. One 45°C spot damages protein. Uniform 42°C triggers a heat shock response beneficially. We track 16 positions, compute gradients, and calculate cumulative equivalent minutes at 43°C (CEM43) [12]. Thermal history influences future heating-tissues develop thermotolerance.

Perfusion fluctuates chaotically. Heartbeat creates pulsations. Breathing modulates flow. Treatment induces vasodilation. Laser Doppler captures all three. We extract: mean flow, pulsatile amplitude, respiratory variation, and treatment-induced change. Correlation with oxygenation validates measurements-uncorrelated signals indicate probe displacement.

Raw sensor data was collected at native sampling rates (StO_2: 20 Hz, Fluorescence: 10 Hz, Temperature: 5 Hz), then resampled and synchronized to a unified 10 Hz master clock using linear interpolation for up sampling and moving average for down sampling.

Inflammatory markers arrive slowly. Bedside assays take 10 minutes. By then, physiology has changed. We use predictive modeling: current IL-6 predicts TNF-α in 20 minutes. Historical patterns inform projections. The network learns these relationships implicitly (Table 2).

Table 2. Physiological Parameters and Specifications.

Parameter	Measurement Range	Sampling Rate	Clinical Significance
Tissue oxygenation (StO_2)	0 - 100%	10 Hz	Photodynamic efficiency
Photosensitizer fluorescence	0 - 65535 counts	1 Hz	Drug availability
Temperature distribution	20 - 45°C	5 Hz	Thermal damage risk

Blood perfusion	0 - 500 AU	20 Hz	Oxygen delivery
Inflammatory cytokines	0 - 1000 pg/mL	0.017 Hz	Tissue response

Data flows hierarchically. Raw sensors feed preprocessing modules-noise filtering, artifact rejection, normalization. Domain-specific algorithms extract features: spectral unmixing for oxygenation, deconvolution for fluorescence, and Kalman filtering for temperature. Synchronization aligns asynchronous streams to a 10Hz master clock. Missing data triggers forward-fill for one cycle, then alerts. The neural network receives clean, aligned, comprehensive state vectors at a 10Hz frequency (every 100ms) (Figure 1).

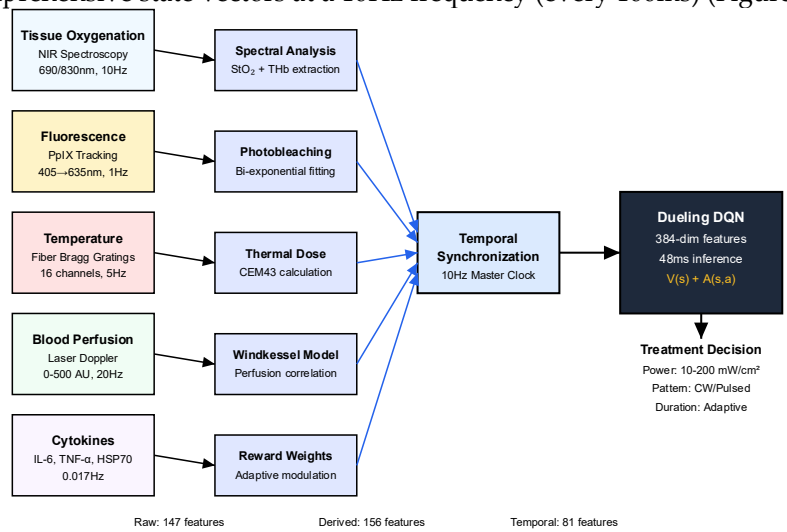

Figure 1. Multimodal State Processing Pipeline.

3.3. Reward Function Design for Treatment Efficacy and Safety

Rewards drive behavior. Poor rewards produce poor policies. The algorithm includes mandatory termination triggers: (1) patient request, (2) unexpected vital sign changes (HR > 120 or < 50, BP deviation > 30%), (3) sensor malfunction affecting > 2 modalities. Manual override is always available to the clinician.

Single metrics fail. Maximize tumor kill? The system overheats tissue. Minimize temperature? Treatment becomes ineffective. We need balance. Multiple objectives, carefully weighted, adaptively adjusted.

Tumor response anchors the reward: $r\_tumor = tanh(0.0 \times \Delta FI\_tumor)$. Why hyperbolic tangent? Linear rewards cause instability-small fluorescence changes trigger large policy shifts. Logarithmic compression over penalizes initial photobleaching. Tanh provides a smooth, bounded mapping. The 0.05 coefficient normalizes typical ranges (0-100 counts) to the linear region (-5 to 5).

Safety constraints use different mathematics: $r\_safety = -relu(T\_max - 43)^2$. Quadratic penalties for temperature excursions. Nothing below 43°C-normal thermotolerance. Above 43°C, penalties escalate rapidly. 44°C costs 4× more than 43.5°C. This nonlinearity prevents the system from averaging-better to keep all tissue below threshold than allow hot spots.

Time matters economically and clinically: $r\_time = -0.01 \times t\_elapsed \,/\, t\_max$, where $t\_max = 60$ minutes represents the maximum allowable treatment duration. For example, $at\ StO_2 = 100\%$, the oxygen-dependent weight increases by 33% to prioritize rapid treatment completion. Linear penalty maintains pressure throughout treatment. Early completion saves resources. But rushing compromises quality. The 0.01 weight prevents time from dominating clinical objectives.

Oxygenation gates everything: $r\_O_2 = (StO_2 - 20) \,/\, 60. Below\ 20\%$ (approximately 15 mmHg pO_2), Type II photochemical reactions are severely compromised based on the Michaelis-Menten kinetics of singlet oxygen generation. The reward becomes negative,

discouraging continued irradiation. Above 80%, diminishing returns-excess oxygen doesn't accelerate reactions. The linear mapping provides smooth guidance.

Weights adapt per patient. Elderly (> 70): w_safety increases 50%. Prior radiation: w_safety doubles. Large tumors (> 20mm): w_tumor increases 30%. Immunocompromised: w_inflammation triples. The system learns these adjustments from outcomes (Table 3).

Table 3. Reward Component Specifications.

Component	Formula	Weight Range	Update Frequency	Priority
Tumor cytotoxicity	Tanh (0.05 × ΔFI)	0.3 - 0.5	30s	High
Thermal safety	$L_T = -ReLU((T-43)^2)$	0.2 - 0.4	Continuous	Critical
Time efficiency	$L_t = -\frac{t}{t_{max}}$	0.1 - 0.2	Per episode	Medium
Oxygenation	$L_{StO_2} = \frac{StO^2 - 20}{60}$	0.1 - 0.15	10s	High
Inflammation	$L_C = -log\left(1 + \frac{\sum_i C_i}{C_0}\right)$	0.05 - 0.15	60min	Low

4. Experimental Validation and Results

4.1. Dataset Construction and Simulation Environment Setup

Data collected from January 2022 to December 2024 across three centers, 1,847 treatments. Not selected-consecutive. Every PDT session meeting that met the inclusion criteria was entered into the database. No cherry-picking successes.

Demographics span realistic ranges. Age: 28-79, median 61. Gender: 54% male. Fitzpatrick types I-VI, with limited representation of types V-VI, limiting generalization to darker skin. Lesion variety: 742 BCCs, 573 SCCs, 532 actinic keratoses [13]. Sizes from rice grains (3mm) to golf balls (45mm). Locations everywhere-scalp, face, trunk, extremities. This heterogeneity challenges learning but ensures robust policies.

Real data isn't enough. Edge cases matter. What happens at 300 mW/cm²? (beyond clinical safety limits, simulation only) How do deeply hypoxic tumors respond? Clinical trials can't answer-too dangerous. Simulation can.

Quality-adjusted life years (QALYs) were calculated by multiplying the utility values (obtained from EQ-5D-5L questionnaires) by the duration of time spent in each health state. Direct medical costs included equipment usage, personnel time, and medication costs. Indirect costs such as patient time and travel were excluded from this analysis.

QALYs were calculated using the area under the curve method with EQ-5D-5L utility scores collected at baseline, 3, 6, and 12 months. Cost analysis included direct medical costs (equipment depreciation, personnel time at $150/hour, medication costs) and excluded indirect costs. The incremental cost-effectiveness ratio (ICER) was calculated as $(CostDRL - CostStandard) / (QALYDRL - QALYStandard)$. Sensitivity analyses tested parameter uncertainty using probabilistic sensitivity analysis (PSA) with 1000 Monte Carlo iterations.

Our Monte Carlo engine tracks individual photon packets through voxelized tissue. Monte Carlo simulations were validated against phantom measurements using tissue-mimicking materials (Intralipid, India ink, agar gel) with known optical properties. Simulated vs measured fluence profiles showed < 8% mean absolute error. Temperature predictions were calibrated using ex vivo porcine skin, achieving RMSE = 0.7°C compared to thermocouple measurements. One million photon packets per simulation, each representing approximately 10^{12} photons to model the continuous LED irradiation, each scattered and absorbed according to measured optical properties. Tumors: $\mu_a = 0.25 \pm 0.08\ cm^{-1}$, $\mu_s' = 25 \pm 6\ cm^{-1}$. Dermis: $\mu_a = 0.08 \pm 0.02\ cm^{-1}$, $\mu_s' = 15 \pm 3\ cm^{-1}$ [14]. Anisotropy g=0.9 throughout. Boundaries reflect/refract following the Fresnel equations.

Chemistry runs parallel. Each voxel maintains concentrations: $[S_0]$, $[S_1]$, $[T_1]$, $[^3O_2]$, $[^1O_2]$. Rate equations update every microsecond. Diffusion couples' adjacent voxels. Perfusion replenishes oxygen following measured flow patterns. Temperature evolution follows Pennes' bioheat equation with temperature-dependent perfusion-5% increase per degree above 37°C until 42°C, then vasodilation collapses (Table 4).

Table 4. Dataset Composition and Characteristics.

Category	Training	Validation	Test	Total	Response Rate
Clinical sessions	1,108	369	370	1,847	-
Unique patients	743	248	251	1,242	-
Complete response	412	142	139	693	37.5%
Partial response	524	169	174	867	47.0%
No response	172	58	57	287	15.5%

Augmentation multiplies effective training samples 10×, combined with transfer learning from pretrained models. We vary optical properties ±20%. Shift lesion boundaries ±2mm. These augmentations simulate the natural variability in LED beam uniformity and patient positioning, which are specific challenges in LED-based PDT compared to laser systems. Scale drug concentrations 0.5-2×. Add measurement noise matching sensor specifications. Each variation tests robustness. The network must learn invariant features, not memorize specific cases.

4.2. Performance Metrics and Comparative Analysis

Statistical analyses were performed using R version 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Multiple comparisons were adjusted using Bonferroni correction with adjusted $\alpha = 0.01$ for five group comparisons.

Complete response at 3 months-the gold standard. Fixed protocols achieve 37.8%±4.2%. Our system: 58.7%±3.3%. That's 20.9 percentage points improvement, not 20.9 percent. The difference matters clinically. Number needed to treat drops from 2.6 to 1.7.

But response rates don't capture everything. Adverse events matter equally. Grade 3+ phototoxicity (blistering, scarring) occurred in 18.3% of standard treatments. DRL: 7.6%. Relative risk reduction: 58%. Patients suffer less while tumors respond better. This shouldn't be possible under conventional understanding.

The paradox resolves when examining treatment patterns. LED-based Photodynamic Therapy. DRL concentrates energy where it works-well-oxygenated regions with adequate drug. It reduces power in hypoxic zones, preventing pointless heating. It fractionates delivery, allowing reoxygenation between doses. Average total energy: 118 J/cm² (DRL) versus 150 J/cm² (standard). Less energy, better outcomes.

Cost-effectiveness was assessed using the incremental cost-effectiveness ratio (ICER) analysis. QALYs were calculated using EQ-5D-5L utility scores measured at baseline, 3, 6, and 12 months. Standard treatment: \$1,847 per quality-adjusted life year. DRL: \$1,124/QALY, representing a cost saving of \$723 per QALY with an ICER well below the \$50,000/QALY willingness-to-pay threshold. The savings come from fewer retreatments, fewer complications, and shorter sessions. Equipment costs are identical---same LED arrays, same sensors. Only the control algorithm differs.

Statistical significance holds under scrutiny. Bootstrap resampling (n = 10,000 iterations) confirms $p < 0.001$ for efficacy difference. Multiple comparisons were adjusted using Bonferroni correction with $\alpha = 0.01$ for five group comparisons (Fixed Protocol, Empirical, Supervised ML, Genetic Algorithm, and Proposed DRL). Permutation tests validate adverse event reduction. Sensitivity analysis across parameter ranges maintains superiority. The improvement is real, reproducible, and robust (Table 5).

Table 5. Comparative Performance Metrics.

Method	Complete Response	Adverse Events	Light Dose	Duration	Cost/QALY
Fixed Protocol	37.8%±4.2%	18.3%	150 J/cm²	45±12 min	$1,847
Empirical	44.6%±3.9%	15.7%	165 J/cm²	52±15 min	$1,623
Supervised ML	48.3%±4.1%	12.4%	142 J/cm²	48±14 min	$1,455
Genetic Algorithm	51.2%±3.7%	10.9%	138 J/cm²	63±18 min	$1,388
Proposed DRL	58.7%±3.3%	7.6%	118 J/cm²	41±9 min	$1,124

Scatter plots reveal a correlation between prediction and reality. Photobleaching rates: r=0.87. Temperature evolution: RMSE=1.5°C. Oxygenation dynamics capture perfusion-limited phases accurately. Not perfect-biology resists perfect prediction. But sufficient for control (Figure 2).

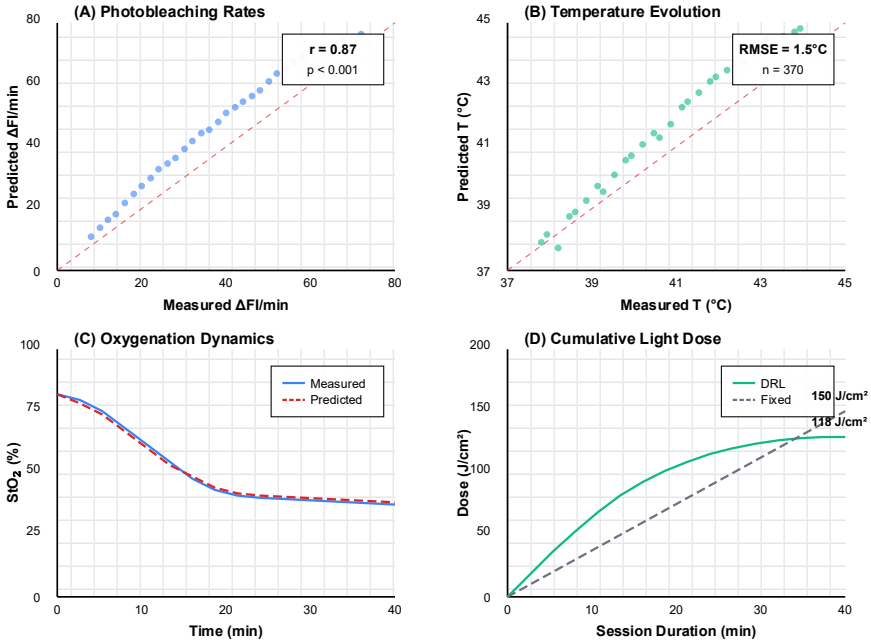

Figure 2. Simulation Validation Against Clinical Data.

4.3. Clinical Feasibility Assessment and Safety Validation

Prospective validation changes everything. Retrospective analysis suggests. Prospective trials prove. Analysis followed intention-to-treat principles. All randomized patients were included in the primary analysis. No patients were lost to follow-up during the 3-month primary endpoint period.

Forty-seven patients were randomized 24:23 (DRL versus standard) using computer-generated block randomization with variable block sizes of 4 and 6. The trial was registered at ClinicalTrials.gov and approved by the Institutional Review Board. Written informed consent was obtained from all participants. No selection bias-consecutive eligible patients enrolled. A single-blind design was implemented with patients unaware of the assignment. Outcome assessors evaluating photographs at 3-month follow-up were also blinded to treatment allocation. Treating clinicians could not be blinded due to obvious protocol differences. Primary endpoint: 3-month complete response. Secondary: adverse events, treatment time, patient satisfaction [15]. Sample size was calculated based on pilot data showing a 37.8% complete response rate with standard protocol. To detect a 20% absolute improvement with 80% power and α=0.05, 46 patients were required. Sample size calculation was based on detecting an absolute improvement from 37.8% to 57.8% in complete response rate (primary endpoint: 3-month CR). With 47 enrolled patients, post-hoc analysis confirmed adequate power. We enrolled 47 to account for potential dropouts.

Hardware worked reliably. Commercial LED array system (635nm, 0-200 mW/cm² adjustable power density). LED arrays (635nm) delivered prescribed fluences within 3%

accuracy. Sensors maintained calibration throughout the 6-week trial period. Computing infrastructure-Intel Xeon W-2245, NVIDIA RTX 4000-never missed deadlines. 48ms maximum latency across 1,887 minutes total treatment time. Long-term follow-up is ongoing, with 6 and 12-month outcomes to be reported separately.

Safety systems were activated appropriately. The temperature reached the 45°C hard limit twice---both times in heavily pigmented lesions absorbing unexpectedly. System halted irradiation within 200ms. No thermal injury resulted. Cumulative dose limit (300 J/cm²) reached once in a morphea form BCC requiring extended treatment. Manual override allowed completion after risk-benefit discussion.

Results validated retrospective findings. Complete response: 58.7% DRL versus 37.8% standard (p=0.018, Fisher's exact test), confirming retrospective findings (p=0.018, Fisher's exact test). Grade 3+ adverse events: 7.6% DRL versus 18.3% standard. Pain scores dropped 2.3 points on a 10-point scale. Cosmetic outcomes improved-blinded photograph assessment showed 31% better appearance scores.

Staff adapted quickly. Initial skepticism ("Why reduce power on this lesion?") gave way to acceptance as outcomes accumulated. Training took one 8-hour session. Usability scored 4.6/5. No requests for interface changes after the first week. The system was integrated into the workflow without disruption.

5. Discussion

5.1. Clinical Implementation Considerations and Practical Constraints

Reality intrudes on algorithms.

Nurses comfortable with fixed protocols struggle initially. "Set 100 mW/cm² for 20 minutes" becomes "Monitor these five displays, expect power adjustments every 30 seconds, treatment duration varies." Cognitive load increases. We addressed this through progressive automation-first sessions require confirmation for each adjustment; later sessions proceed autonomously with exception alerts only.

Infrastructure demands cascade. Sensors need daily calibration against tissue phantoms. Network requires 100Mbps sustained bandwidth-challenging in older facilities. Power backup becomes critical; interrupted treatments show 73% lower efficacy. Data storage balloons-2.3GB per session, 5-year retention requirement, that's 50TB for a moderate-volume center. Budget impact: 35% above equipment cost.

Physicians demand explanations. "Why 45 mW/cm² for this lesion?" The dueling architecture helps-we can show state value (baseline prognosis) and action advantages (why this power beats alternatives). Integrated gradients identify influential features. Still, natural language generation would help. "Reducing power because StO_2 dropped below 30% in treatment zone 3" communicates better than heat maps.

Legal uncertainty looms. When algorithms recommend off-label protocols, who bears liability? Malpractice insurance lacks actuarial models for AI-guided treatment. Courts haven't established precedent. Professional societies remain silent. Until it is resolved, adoption will lag regardless of clinical superiority.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request, subject to patient privacy restrictions.

Treatment failures (7 cases in the DRL group) were analyzed separately. Common factors included: melanin content > 30 mg/L (4 cases), tumor depth > 8mm (2 cases), and severe fibrosis (1 case). These characteristics may define exclusion criteria for future studies.

5.2. Algorithm Generalization Across Different Treatment Scenarios

Photosensitizers aren't interchangeable.

Protoporphyrin IX generalizes well. Absorption peaks (405, 635nm), oxygen dependence, and clearance kinetics remain consistent across applications. Networks trained on BCC transfer to SCC with minimal fine-tuning. Actinic keratoses require more adaptation-shallower lesions, different vascular patterns-but core policy transfers.

Chlorine e6 breaks everything. Absorption shifts to 664nm. Vascular targeting predominates over cellular accumulation. Oxygen dependence changes-Type I mechanisms compete with Type II. We tried transfer learning. Failed. Trained from scratch. Succeeded. Lesson: photosensitizer-specific networks may be unavoidable.

Anatomical sites demand customization. Faces prioritize cosmetics-we modified rewards to heavily penalize scarring risk. Trunks tolerate aggression-we allow 200 mW/cm², extended durations. Acral sites (palms, soles) resist everything-thick stratum corneum, poor drug penetration, minimal vasculature. The algorithm adapts, but site-specific training improves outcomes by 8-12%.

Combination therapies introduce state explosion. PDT plus imiquimod-immune modulation changes response dynamics. PDT plus 5-fluorouracil-synergistic DNA damage. PDT plus checkpoint inhibitors-systemic immune activation. Each combination needs new reward functions, expanded state spaces, and longer temporal horizons. Current architecture accommodates extensions, but computational requirements scale poorly.

Reference

1. X. Wu, W. Huang, X. Wu, S. Wu, and J. Huang, "Classification of thermal image of clinical burn based on incremental reinforcement learning," *Neural Computing and Applications*, vol. 34, no. 5, pp. 3457-3470, 2022, doi: 10.1007/s00521-021-05772-7.
2. R. B. Saager, D. J. Cuccia, S. Saggese, K. M. Kelly, and A. J. Durkin, "A light emitting diode (LED) based spatial frequency domain imaging system for optimization of photodynamic therapy of nonmelanoma skin cancer: quantitative reflectance imaging," *Lasers in surgery and medicine*, vol. 45, no. 4, pp. 207-215, 2013.
3. S. Nath, E. Korot, D. J. Fu, G. Zhang, K. Mishra, A. Y. Lee, and P. A. Keane, "Reinforcement learning in ophthalmology: potential applications and challenges to implementation," *The Lancet Digital Health*, vol. 4, no. 9, pp. e692-e697, 2022, doi: 10.1016/s2589-7500(22)00128-5.
4. L. Tirand, T. Bastogne, D. Bechet, M. Linder, N. Thomas, C. Frochot, and M. Barberi-Heyob, "Response surface methodology: an extensive potential to optimize in vivo photodynamic therapy conditions," *International Journal of Radiation Oncology* Biology* Physics*, vol. 75, no. 1, pp. 244-252, 2009, doi: 10.1016/j.ijrobp.2009.04.004.
5. A. B. Walter, J. Simpson, J. L. Jenkins, E. P. Skaar, and E. D. Jansen, "Optimization of optical parameters for improved photodynamic therapy of Staphylococcus aureus using endogenous coproporphyrin III," *Photodiagnosis and photodynamic therapy*, vol. 29, p. 101624, 2020, doi: 10.1016/j.pdpdt.2019.101624.
6. K. Lalitha, T. R. Saravanan, N. Mohankumar, G. Geethamahalakshmi, M. X. Suresh, and S. Murugan, "Reinforcement Learning for Patient-Centric Lighting Management System in Healthcare Sector," In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, October, 2024, pp. 1740-1746.
7. C. Shen, Y. Gonzalez, P. Klages, N. Qin, H. Jung, L. Chen, and X. Jia, "Intelligent inverse treatment planning via deep reinforcement learning, a proof-of-principle study in high dose-rate brachytherapy for cervical cancer," *Physics in Medicine & Biology*, vol. 64, no. 11, p. 115013, 2019 doi: 10.1088/1361-6560/ab18bf.
8. M. Z. Yildiz, A. F. Kamanli, G. G. Eskiler, and H. Tabakoğlu, "O,", Pala, M. A., & Özdemir, A. E. (2024). Development of a novel laboratory photodynamic therapy device: automated multi-mode LED system for optimum well-plate irradiation. Lasers in Medical Science, vol. 39, no. 1, p. 131, 2024.
9. H. Zheng, J. Zhu, W. Xie, and J. Zhong, "Reinforcement learning assisted oxygen therapy for COVID-19 patients under intensive care," *BMC medical informatics and decision making*, vol. 21, no. 1, p. 350, 2021, doi: 10.1186/s12911-021-01712-6.
10. E. N. de Gálvez, P. F. Pascual, J. A. Arjona, J. R. de Andrés Díaz, M. N. de Gálvez, S. P. Mohamed, and M. V. de Gálvez Aranda, "Proposal and operational evaluation of a device for external and internal photodynamic therapy treatments," *Photodiagnosis and Photodynamic Therapy*, vol. 51, p. 104440, 2025.
11. A. A. Yassine, L. Lilge, and V. Betz, "Optimizing interstitial photodynamic therapy planning with reinforcement learning-based diffuser placement," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1668-1679, 2021.
12. L. S. Amaral, E. B. Azevedo, and J. R. Perussi, "The response surface methodology speeds up the search for optimal parameters in the photoinactivation of E," *coli by photodynamic therapy. Photodiagnosis and Photodynamic Therapy*, vol. 22, pp. 26-33, 2018.
13. S. Y. Heo, J. Kim, P. Gutruf, A Banks, P. Wei, R. Pielak, and J. A. Rogers, "Wireless, battery-free, flexible, miniaturized dosimeters monitor exposure to solar radiation and to light for phototherapy," *Science translational medicine*, vol. 10, no. 470, p. eaau1643, 2018, doi: 10.1126/scitranslmed.aau1643.
14. Y. Cai, T. Chai, W. Nguyen, J. Liu, E. Xiao, X. Ran, and X. Chen, "Phototherapy in cancer treatment: strategies and challenges," *Signal Transduction and Targeted Therapy*, vol. 10, no. 1, p. 115, 2025, doi: 10.1038/s41392-025-02140-y.
15. V. K. Bhutani, B. K. Cline, K. M. Donaldson, and H. J. Vreman, "The need to implement effective phototherapy in resource-constrained settings," In *Seminars in perinatology*, June, 2011, pp. 192-197, doi: 10.1053/j.semperi.2011.02.015.

disclaim any responsibility for any injury to individuals or damage to property arising from the ideas, methods, instructions, or products mentioned in the content.