Machine Learning-Based Optimization of Healthcare Building Environments for Improving Indoor Thermal Comfort, Daylight Exposure, and Patient Well-Being

Herberto Galiano
Department of Computer Science, University of North Texas, Denton, TX, USA.
herbertomail@unt.edu

Sigfrid Gustavsson
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
sigfrid.gustavsson@uab.edu

Mieszko Bany
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
mieszkowork@oregonstate.edu

Abstract

Healthcare buildings must support thermal comfort, daylight exposure, and patient well-being simultaneously, yet design and operation often optimize these outcomes in separate models and at different time scales. Static setpoints and rule-based shading can satisfy average energy targets while producing local discomfort, glare, or insufficient daylight in patient rooms. This manuscript develops a machine-learning-based optimization framework that connects room-level environmental predictors to clinical-adjacent well-being indicators and uses constrained multi-objective optimization to select design and control variables. The framework couples a physics-informed feature representation with a Gaussian-process or ensemble surrogate, models uncertainty from weather, occupancy, and model discrepancy, and solves a weighted objective subject to operative temperature, daylight autonomy, useful daylight illuminance, and glare constraints. Patient well-being enters as a transparent proxy composed of view access, daylight sufficiency, thermal stability, and controllability, rather than as a validated clinical endpoint. A robust expected-value formulation and a Bayesian optimization acquisition rule reduce the risk of selecting designs that perform well only under nominal conditions. The main contribution is a reproducible methodological architecture that translates healthcare-specific comfort and well-being constraints into an optimization problem that can be evaluated with simulation, monitored post-occupancy, and updated through digital-twin calibration. The discussion identifies trade-offs among solar control, daylight, and thermal load; failure modes such as surrogate extrapolation, occupancy shift, and glare metric misspecification; and governance requirements for clinical environments. The framework provides a technically defensible basis for prospective evaluation before deployment and clarifies which claims require empirical validation.

Keywords

machine learning; indoor thermal comfort; daylight autonomy; healthcare buildings; multi-objective optimization; patient well-being

1. Introduction

Healthcare environments concentrate competing environmental requirements within a single room. Patients may have limited mobility, altered thermoregulation, medication-related sensitivity, and restricted control over clothing and window operation, so nominal comfort ranges derived for office occupants are not automatically transferable. Thermal comfort assessment has conventionally relied on predicted mean vote and operative temperature criteria [1], with standards defining acceptable indoor conditions for occupied spaces [2]. Adaptive comfort research further shows that acceptable temperatures depend on outdoor conditions, expectation, and available control, which complicates fixed setpoints in hospitals and clinics [3]. Daylight adds a separate performance dimension: dynamic metrics such as daylight autonomy and useful daylight illuminate replace single-point daylight factors because they describe temporal sufficiency and excess [4,5]. Visual comfort indices capture glare and luminance contrast, which are especially relevant when patients face windows or lie in fixed orientations [6]. Beyond comfort, access to window views has been associated with recovery-related outcomes [7], and reviews of healing environments identify daylight, nature contact, noise, and control as psychologically mediated design factors [8].

Despite this evidence base, healthcare environmental design remains fragmented. Thermal models, daylight simulations, and well-being checklists are often used sequentially, with the output of one analysis becoming an informal constraint on another. This separation obscures trade-offs: larger apertures increase daylight but can raise cooling loads and glare; deeper shading reduces solar gains but may suppress useful daylight; tighter temperature control can improve stability while increasing energy use. Machine learning offers a way to approximate these coupled relationships quickly enough for iterative design and operational decision-making, yet purely data-driven models can extrapolate poorly and provide little assurance about clinical-adjacent outcomes. The central gap is therefore not the absence of predictive algorithms but the absence of a constrained optimization architecture that treats healthcare thermal comfort, daylight exposure, and well-being as coupled, uncertainty-aware objectives rather than independent performance scores.

This manuscript develops such an architecture. The contribution is a methodological framework for machine-learning-based optimization of healthcare building environments that links physics-informed features, surrogate prediction, constrained multi-objective optimization, and robust decision rules. It does not report a deployed clinical trial or a completed field evaluation. Instead, it specifies how patient well-being can enter an optimization problem as a transparent proxy, how thermal and daylight constraints can be encoded, how surrogate uncertainty can be propagated, and how failure modes in healthcare settings should shape model validation. The framework is intended to support prospective simulation studies, digital-twin calibration, and post-occupancy monitoring without overstating what can be inferred from model outputs alone.

2. Environmental and Clinical Performance Objectives

2.1 Thermal comfort and adaptive constraints

Thermal comfort in patient rooms must be represented with variables that can be measured or simulated at room scale. The predicted mean vote is a function of air temperature, mean radiant temperature, air speed, relative humidity, metabolic rate, and clothing insulation. For an optimization model, the relevant relationship can be written as a parameterized comfort function:

$$PMV = f(T_a, T_r, v_a, RH, M, I_c l)$$

Here T_a is air temperature, T_r is mean radiant temperature, v_a is air speed, RH is relative humidity, M is metabolic rate, and I_cl is clothing insulation. The function f is established comfort theory rather than a newly fitted model, and its output must be interpreted with healthcare-specific caution because patient metabolic rates and clothing differ from office populations. Reviews of indoor comfort factors show that thermal sensation interacts with visual and acoustic conditions, so a room that satisfies a single PMV threshold may still be unacceptable when glare or noise is present [9]. The optimization framework therefore treats thermal comfort as a constraint set with local room-level checks rather than as a single averaged score.

2.2 Daylight exposure and glare

Daylight performance requires time-series evaluation. Two complementary metrics are daylight autonomy and useful daylight illuminance. For a grid of sensor points and an annual occupancy schedule, daylight autonomy can be expressed as the fraction of occupied hours in which illuminance meets or exceeds a target:

$$DA = (1/N) \sum I(E_i \geq E_t)$$

Here N is the number of occupied time steps, E_i is illuminance at sensor point i, E_t is the target illuminance, and I is an indicator function equal to one when the condition holds and zero otherwise. Useful daylight illuminance refines this by penalizing both under-illumination and excessive illuminance that may produce visual discomfort:

$$UDI = (1/N) \sum I(E_low \leq E_i \leq E_high)$$

E_low and E_high define the acceptable illuminance band. These metrics matter in healthcare settings because a room can have high average daylight while still suffering long periods of glare or low morning exposure. Shading and aperture control alter both cooling demand and lighting conditions, so daylight cannot be optimized independently of solar control [10]. A robust formulation should evaluate glare alongside illuminance rather than treating high daylight as unambiguously beneficial.

2.3 Patient well-being as a constrained proxy

Patient well-being is not directly observable from a building model. The framework therefore uses a proxy composed of measurable or simulatable design attributes: view access, daylight sufficiency, thermal stability, privacy, and controllability. This proxy is not a clinical outcome measure and should not be interpreted as one. Its purpose is to prevent optimization from selecting rooms that meet thermal and daylight thresholds while ignoring features associated with patient experience. The proxy is deliberately transparent so that clinicians, facilities staff, and designers can inspect which attributes drive a recommendation. Data-driven building performance prediction provides a practical basis for estimating some of these attributes from geometry, schedules, and environmental variables [11], but the well-being component requires explicit normative choices about weights and constraints.

3. Machine-Learning Surrogate and Optimization Workflow

3.1 Data and feature representation

The input vector must distinguish design variables from operational and exogenous variables. Design variables include aperture size, shading depth, glazing properties, room orientation, and layout. Operational variables include setpoints, shading schedules, and lighting controls. Exogenous variables include weather, occupancy, and clinical usage patterns. Geometry and massing choices influence solar access and density, and generative optimization has been proposed for building-block performance involving density, solar, and structural objectives [12]. In a healthcare context, the same logic applies at room and façade scale: layout decisions constrain daylight access before any control system is tuned.

The feature representation should combine physical quantities with informative transformations. Solar azimuth and altitude, direct and diffuse irradiance, view factors, window-to-wall ratio, and shading obstruction angles preserve the mechanisms that govern daylight and heat transfer. Temporal features such as hour, season, and occupancy state allow the surrogate to represent non-stationary conditions. Model predictive control studies show that building dynamics are strongly time-dependent and that prediction horizons matter for control performance [13]. Reinforcement learning research further illustrates that control policies must account for delayed effects and competing rewards [14]. These findings motivate a surrogate that predicts not only annual averages but also time-series behavior relevant to constraints.

3.2 Surrogate model and uncertainty

A surrogate maps design and operational inputs to predicted thermal, daylight, and well-being indicators. A general form is:

$$\hat{y} = g_\theta(x, z) + \varepsilon, \varepsilon \sim N(0, \sigma^2(x))$$

Here ŷ is the predicted performance vector, x is the design and control vector, z is the exogenous condition vector, θ denotes model parameters, and ε is a heteroscedastic error term. Gaussian-process regression is attractive when simulation budgets are limited because it provides posterior predictive variance and can represent smooth response surfaces [15,16]. Ensemble trees and neural networks may be preferable for high-dimensional or highly non-smooth response surfaces, but they require separate uncertainty quantification. In either case, the surrogate should be trained on a design of experiments that covers the feasible healthcare design space, and extrapolation beyond that space should be flagged rather than silently accepted.

3.3 Constrained multi-objective optimization

The optimization problem combines thermal discomfort, daylight deficiency or excess, well-being proxy loss, and a regularization term that penalizes unnecessarily complex or costly solutions. A weighted formulation is:

$$J(x) = w_T C_T(x) + w_D C_D(x) + w_H C_H(x) + \lambda \Omega(x)$$

C_T is thermal discomfort, C_D is daylight performance loss, C_H is well-being proxy loss, w_T, w_D, and w_H are nonnegative weights, Ω is a complexity or cost penalty, and λ controls its influence. The weights are not universal constants; they encode healthcare priorities and should be treated as decision parameters subject to stakeholder review. Machine learning across the building life cycle has been used for prediction, fault detection, and

control support [17], but optimization requires explicit constraints so that trade-offs remain clinically and operationally defensible.

4. Decision Formulation and Constraint Handling

4.1 Objective and constraints

The optimization must satisfy hard constraints that represent unacceptable conditions. Operative temperature should remain within a clinical range, useful daylight should not fall below a minimum in patient-occupied zones, and glare probability should not exceed a specified threshold. A compact constraint set is:

$$T_min \le T_op(x,t) \le T_max, E_i(x) \ge E_min, DGP(x) \le DGP_max$$

T_op is operative temperature at time t, E_i is illuminance at sensor point i, E_min is the minimum acceptable illuminance, DGP is daylight glare probability, and DGP_max is the maximum tolerated glare. These constraints can be evaluated directly in simulation or through the surrogate with uncertainty margins. In healthcare settings, constraint violations are not interchangeable with objective penalties: a room with insufficient daylight or persistent overheating should be treated as infeasible even if its aggregate score is favorable. This distinction prevents the optimizer from trading a serious local failure for a small improvement in an average metric.

4.2 Robust optimization and acquisition

Nominal optimization can select designs that perform well only for one weather year or occupancy schedule. A robust formulation penalizes variability across uncertain conditions:

$$J_robust(x) = E_\theta[J(x;\theta)] + \kappa \sqrt{Var_\theta[J(x;\theta)]}$$

E_θ is the expectation over uncertain parameters θ, Var_θ is the corresponding variance, and κ controls risk aversion. The uncertain parameters may include weather sequences, occupant behavior, internal gains, and surrogate coefficients. A larger κ favors designs with stable performance, which is appropriate when vulnerable patients occupy the same rooms across seasons. During search, Bayesian optimization can propose the next candidate by balancing predicted performance and uncertainty:

$$x_next = argmin_x \, \mu_J(x) + \beta \, \sigma_J(x)$$

μ_J and σ_J are the posterior mean and standard deviation of the objective, and β controls exploration. This acquisition rule is useful when each candidate requires an expensive simulation. It should not be used to justify clinical claims; its role is to allocate computational effort efficiently while maintaining an explicit record of uncertainty.

5. Technical Discussion

The proposed framework exposes several trade-offs that are easily hidden when thermal and daylight analyses are conducted separately. Increasing aperture area can improve daylight autonomy but may increase solar gains, cooling demand, and glare risk. External shading can reduce peak solar load but may suppress useful daylight in winter or under overcast skies. Automated shading can resolve some of these conflicts, yet control algorithms must be evaluated against failure modes such as oscillation, sensor drift, and occupants overriding controls. Evidence on green buildings and health supports a broad association between environmental quality and occupant outcomes [18], and housing and health guidance emphasizes indoor temperature, ventilation, and safety as public health concerns [19]. These

sources justify treating environmental quality as a health-relevant design domain, but they do not establish that any particular optimization weight produces clinical benefit.

A second issue is the well-being proxy. View access, daylight, thermal stability, privacy, and controllability are plausible design attributes, but their relative importance varies by patient population, length of stay, and clinical function. The proxy should therefore be reported as a transparent decision construct, not as a measured patient outcome. Adaptive comfort research shows that expectation and control influence acceptable conditions [20], which suggests that controllability should be represented as a feature rather than omitted from the model. At the same time, not all patients can exercise control, so the optimization must retain conservative default conditions for rooms where blinds, thermostats, or windows are inaccessible.

A third issue is surrogate validity. Machine-learning models are vulnerable to distribution shift when a hospital undergoes renovation, changes clinical use, or experiences unusual weather. The framework should include out-of-distribution checks, constraint margins, and periodic recalibration against simulation or monitored data. A digital twin can support this process by comparing predicted and observed room conditions, but the comparison must distinguish sensor error, occupancy uncertainty, and model discrepancy. If a surrogate is used to rank design alternatives, the ranking should be tested for stability under plausible changes in weights and uncertainty parameters. If a surrogate is used for operational control, fail-safe rules should override the model when sensor data are missing or when thermal or glare constraints are violated.

The framework also has governance implications. Healthcare spaces are regulated environments, and optimization outputs cannot substitute for clinical risk assessment, infection control review, or facilities engineering approval. A defensible workflow records the objective weights, constraint thresholds, training domain, uncertainty assumptions, and approval decisions. This record allows later reviewers to determine whether a recommendation was based on evidence, professional judgment, or an arbitrary model setting. The most useful scientific contribution of machine learning in this domain is therefore not a single optimized room but an auditable decision process that makes trade-offs explicit and testable.

6. Conclusion

This manuscript has developed a machine-learning-based optimization framework for healthcare building environments that treats indoor thermal comfort, daylight exposure, and patient well-being as coupled performance objectives. The framework combines established comfort and daylight metrics with a surrogate model that predicts room-level outcomes, a constrained multi-objective formulation that prevents unacceptable local conditions, and a robust optimization rule that penalizes sensitivity to weather, occupancy, and model uncertainty. Patient well-being is represented through a transparent proxy rather than a validated clinical endpoint, which keeps the method honest about what building models can and cannot establish. The main technical value lies in the explicit connection among physics-informed features, surrogate uncertainty, constraint handling, and decision auditability. The main practical value lies in giving design teams a way to compare alternatives before construction and to monitor whether intended performance persists after occupancy. Future work should evaluate the framework prospectively with simulation-based case studies, calibrated digital twins, and pre-specified outcome measures; until such evidence exists, optimization results should be presented as decision support rather than clinical proof.

References

1. Fanger, P. O. (1970). Thermal comfort: Analysis and applications in environmental engineering. Danish Technical Press.

2. ASHRAE. (2020). ANSI/ASHRAE Standard 55-2020: Thermal environmental conditions for human occupancy. ASHRAE.

3. Brager, G. S., & de Dear, R. J. (1998). Thermal adaptation in the built environment: A literature review. Energy and Buildings, 27(1), 83–96.

4. Reinhart, C. F., Mardaljevic, J., & Rogers, Z. (2006). Dynamic daylight performance metrics for sustainable building design. LEUKOS, 3(1), 7–31.

5. Nabil, A., & Mardaljevic, J. (2006). Useful daylight illuminances: A replacement for daylight factors. Energy and Buildings, 38(7), 905–913.

6. Carlucci, S., Causone, F., De Rosa, F., & Pagliano, L. (2015). A review of indices for assessing visual comfort with a view to their use in optimization processes to support building integrated design. Renewable and Sustainable Energy Reviews, 47, 1016–1033.

7. Ulrich, R. S. (1984). View through a window may influence recovery from surgery. Science, 224(4647), 420–421.

8. Dijkstra, K., Pieterse, M. E., & Pruyn, A. T. H. (2006). Physical environmental stimuli that turn healthcare facilities into healing environments through psychologically mediated effects: Systematic review. Journal of Advanced Nursing, 56(2), 166–181.

9. Frontczak, M., & Wargocki, P. (2011). Literature survey on how different factors influence human comfort in indoor environments. Building and Environment, 46(4), 922–937.

10. Tzempelikos, A., & Athienitis, A. K. (2007). The impact of shading design and control on building cooling and lighting demand. Solar Energy, 81(3), 369–382.

11. Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192–1205.

12. Tabasi, S. F., Luo, D., Rafizadeh, H. R., Le Nguyen, K., & Banihashemi, S. (2025). Generative optimization of building blocks for density, solar and structural performance. Journal of Building Engineering, 111, 113307.

13. Killian, M., & Kozek, M. (2016). Ten questions concerning model predictive control for energy efficient buildings. Building and Environment, 105, 403–412.

14. Wang, Z., & Hong, T. (2020). Reinforcement learning for building controls: The opportunities and challenges. Applied Energy, 269, 115036.

15. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

16. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

17. Hong, T., Wang, Z., Luo, X., & Zhang, W. (2020). State-of-the-art on research and applications of machine learning in the building life cycle. Energy and Buildings, 212, 109831.

18. Allen, J. G., MacNaughton, P., Laurent, J. G. C., Flanigan, S. S., Eitland, E. S., & Spengler, J. D. (2015). Green buildings and health. Current Environmental Health Reports, 2(3), 250–258.

19. World Health Organization. (2018). WHO housing and health guidelines. World Health Organization.

20. de Dear, R., & Brager, G. S. (2002). Thermal comfort in naturally ventilated buildings: Revisions to ASHRAE Standard 55. Energy and Buildings, 34(6), 549–561.