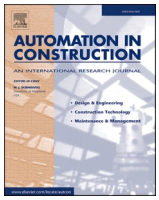
Automation in Construction 181 (2026) 106582

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Automatic evaluation and analysis of indoor visual comfort for sustainable building design using interpretable ensemble learning

Yuxin Zhou [a], Tomohiro Fukuda [a,*], Nobuyoshi Yabuki [a,b]

[a] Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, The University of Osaka, 2-1, Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Advanced Research Laboratories, Tokyo City University, 1-28-1 Tamazutsumi, Setagaya-ku, Tokyo 158-8557, Japan

ARTICLE INFO

Keywords:
Stacking ensemble learning approach
SHAP
Daylight and glare
Performance-driven design
Building performance

ABSTRACT

Sustainable building design increasingly emphasizes daylight access and glare reduction due to their impact on energy efficiency and occupant comfort. However, integrating daylight distribution with dynamic glare risk from an occupant-centered perspective remains a significant challenge. To address this, this paper develops an interpretable Stacking ensemble framework enhanced with SHapley Additive exPlanations (SHAP) method for automated evaluation of indoor visual comfort (IVC). Six ensemble models are optimized through Bayesian optimization and 5-Fold cross-validation. The final Stacking model, which includes ensemble XGBoost, LightGBM, and CatBoost, achieves high predictive accuracy (R^2 = 0.911) and efficient prediction capability. SHAP analysis identifies six key design variables accounting for 80.6 % of the model's contribution, with building forms (46.6–52.7 %) and fenestration features (22.6–24.9 %) as primary factors. The framework provides rapid feedback in early-stage design, supporting data-driven decisions to optimize IVC and integrate performance analysis into occupant-centered design processes.

1. Introduction

Driven by the sustainable development goals (SDGs), the concept of green building has been widely recognized as a way to balance indoor environmental quality with resource efficiency [1]. Daylight is a critical factor in this context: it reduces dependence on artificial lighting to improve energy performance and also affects occupants' mental health, social behavior, and productivity [2,3]. Since people spend about 90 % or more of their time indoors, optimizing building daylighting design has become a key requirement for improving indoor environmental quality [4]. Recent studies have tried to balance daylight performance with energy consumption and thermal comfort [5–7]. However, these approaches often remain fragmented: many focus on specific design components—such as window configurations or shading systems—without adequately capturing the nonlinear interactions among multiple building features [8–10]. Furthermore, some evaluation frameworks emphasize overall indicators, such as energy savings or daylight distribution [11–13]. While useful, these indicators provide limited information on localized, dynamic indoor visual comfort (IVC) parameters, including glare frequency and luminance distribution. As a result, such methods have difficulty supporting design decisions that reflect the actual visual experience of occupants across different building uses and environmental conditions.

Although advancements in generative design and machine learning (ML) technologies offer promising methodologies for performance-driven workflows, a gap remains between computational outcomes and occupant-centered design needs [14]. In current practice, decision-making often relies on static standards and professional experience, which can lead to results that do not fully reflect occupants' preferences or comfort [15]. This mismatch is especially evident in lighting design. Prior studies have shown that visual comfort is as important as energy efficiency and electricity consumption in evaluating building performance [16]. Therefore, there is a need to develop frameworks that move beyond purely technology-driven approaches and explicitly

Abbreviations: AB, AdaBoost (Adaptive Boosting); ASE, Annual sunlight exposure; BPS, Building performance simulation; CB, CatBoost (Categorical Boosting); DGP, Daylight glare probability; DVs, Design variables; EL, Ensemble learning; ET, Extra Tree; GA, Glare autonomy; GA_{window}, Glare autonomy for window direction; IQR, Interquartile range; IVC, Indoor visual comfort; LGBM, LightGBM (Light Gradient Boosting Machine); ML, Machine learning; RF, Random forest; sGA, Spatial glare autonomy; SHAP, SHapley additive exPlanations; UDI, Useful daylight illuminance; WWR, Window-to-wall-ratio; XGB, XGBoost (eXtreme Gradient Boosting).

* Corresponding author.
E-mail address: fukuda.tomohiro.see.eng@osaka-u.ac.jp (T. Fukuda).

https://doi.org/10.1016/j.autcon.2025.106582
Received 7 March 2025; Received in revised form 22 September 2025; Accepted 1 October 2025
Available online 12 October 2025

address issues of IVC. Such frameworks should support decision-making that integrates IVC requirements with sustainability objectives, thereby contributing to occupant-centered green building design.

1.1. Objectives

The aim of this study is to propose a framework to explore the relationship between occupant-centered IVC environment and building features. Specifically, the objectives are as follows: (1) a rapid occupant-centered method for zonal visual comfort evaluation, (2) an explanatory stacking architecture model exposing the effect of design variables on IVC, and (3) validation of the feasibility of this approach across different climatic zones.

1.2. Contributions

This study introduces an automated IVC evaluation framework for sustainable building design by integrating Stacking model with the SHapley Additive exPlanations (SHAP) method. The framework offers enhanced insight into the potential impacts of IVC on occupants. The main contributions of this paper are summarized as follows:

- Interaction between daylight distribution and glare risk: this study proposes a framework that incorporates occupants' viewing directions to evaluate localized glare from windows, particularly accounting for views oriented toward the window as they change with building rotation.
- Stacking ensemble architecture for IVC predictions: this study establishes an optimal stacking architecture by comparing the predictive performance of different ML model. A stacking architecture—composed of XGBoost, LightGBM, and CatBoost—was developed to address the complexities of IVC evaluation, offering enhanced predictive performance and feedback for designers.
- Explanatory analysis using SHAP: this study introduces a "prediction-explanation" mechanism based on the SHAP method, which deconstructs the nonlinear interactions between design variables (DVs) and identifies the key DVs affecting IVC. To validate the generalizability of this framework, testing was conducted across 10 typical cities, yielding highly satisfactory results.

1.3. Structure of the paper

The remainder of this paper is organized as follows. Section 2 provides an overview of current research on IVC challenges and artificial intelligence technologies. Section 3 describes the research methodology, which includes the identification of DVs, the definition of IVC metrics, the development of stacking ensemble approach, and the creation of interpretable evaluation platform for IVC. Section 4 and 5 present and discuss the experimental results, analyzing the performance of various stacking ensemble architectures, elucidating how different training parameters affect the target outcomes, and highlighting several noteworthy—or potentially contentious—practices observed during experimentation. Finally, Section 6 concludes the paper by summarizing its contributions, discussing the main limitations of the study, and outlining directions for future work.

2. Related works

This section provides an overview of the developments and challenges in the field of IVC. It begins by addressing the challenges in current research on indoor daylight distribution and glare control, exploring the relationship between these two factors, and discussing the application of advanced ML techniques to further advance this field.

2.1. Challenges in daylight distribution and glare control for indoor visual comfort

IVC has emerged as a critical factor in assessing indoor environmental quality, closely related to various aspects. Research has shown that IVC is strongly linked to the uniformity of daylight distribution, daylight intensity, and local glare control [17]. To better quantify these factors, the academic community has developed various quantitative evaluation indicators. Illuminance-based daylight metrics such as daylight autonomy (DA), useful daylight illuminance (UDI), and daylight factor (DF) primarily evaluate daylight distribution, whereas luminance-based glare metrics like daylight glare probability (DGP) assess visual comfort by accounting for contrast-related glare effects [18,19]. Carlucci et al. studied 34 light environment evaluation indicators and found that 50 % of the existing systems prioritize glare avoidance, while 26 % focus on the amount of light, emphasizing the tension between lighting efficiency and glare control in light environment design [16]. Moreover, simply increasing daylight performance could increase glare risks, requiring designers to balance illuminance levels, glare control, and light uniformity [20].

These results illustrate the dilemma in sustainable building design: how to improve daylight performance while simultaneously controlling glare, thereby maintaining overall indoor environmental quality. Quantifying glare probability by measurements is more complicated than measuring illuminance [21]. Although DGP is considered the most reliable glare quantification tool, its reliance on specific scene modeling and rendering limits its early-stage application in building performance evaluations [19]. Jones introduced glare autonomy (GA) as an innovative approach to reconstruct the DGP equation, replacing luminance and illuminance terms with line-of-sight vectors for rapid assessment without rendering [20]. This technology shifts the research focus from static illuminance evaluation to discomfort glare for occupants with different viewing directions. Additionally, Wasilewski et al. introduced a method called raytraverse [22], which addresses the spatiotemporal distribution of daylight conditions within architectural spaces. This method is capable of accurately calculating visual comfort metrics such as DGP and unified glare probability (UGP).

2.2. Fragmented frameworks and the need for integrated in indoor visual comfort assessment

The development of lightweight frameworks for IVC assessment has received limited attention compared to the progress in energy consumption prediction. First, some focus on energy use or thermal comfort, combining daylight distribution indicators such as UDI and DF to build evaluation models [23–26], while others solely target glare reduction [27,28]. Few studies have combined daylight distribution and glare risk into a unified framework [10,29]. Additionally, certain frameworks are constrained by specific case studies or climate conditions [10,12,27], while broader frameworks often have difficulty accounting for regional variations [11,23,25]. This fragmentation limits the ability to comprehensively evaluate glare and indoor daylight performance [30], reducing the usefulness of these methods for sustainable building design. Second, most existing frameworks do not consider occupants' line-of-sight factors, such as seating positions and viewing directions, which creates inconsistencies between technical metrics and actual occupant experience [31]. Therefore, a unified and adaptable framework that accounts for both daylight and glare, as well as occupant variability, is necessary for achieving sustainable building design.

Performance-driven workflows are increasingly used in architectural design because they allow multi-dimensional evaluation of building performance [5,14,23]. Understanding the impact of building elements on IVC is essential for improving occupant comfort and satisfaction. Parametric design approaches provide a systematic way to explore different design options in changing environmental conditions [32]. For example, Razmi et al. defined ten DVs for school dormitories,

targeting daylight performance, thermal comfort, and energy demand, and applied optimization algorithms to identify improved spatial configurations [5]. Similarly, Marzouk et al. employed spatial daylight autonomy and annual solar exposure to assess daylight performance in heritage buildings, examining skylight configurations, materials, and glazing types [9]. Although such elements like window size, room size, layout, and daylighting have been studied, less attention has been given to how these elements interact to shape occupants' viewing directions and visual experiences. Therefore, there is a need for an evaluation framework that integrates building features and visual comfort, establishing a dual assessment system that considers both daylight distribution and glare risk.

2.3. Machine learning approaches for enhancing indoor visual comfort

Traditional building performance simulation (BPS) methods face challenges in computation time and complexity. While software such as Radiance provides accurate results, its high computational cost limits the efficiency of design optimization [6,33]. ML-based surrogate models have been proposed to address this issue because they offer fast prediction and generalization [5]. Ensemble learning (EL) models, including random forest and XGBoost, have shown higher prediction accuracy and robustness, particularly when applied to complex and high-dimensional datasets [34]. Moreover, for structured data prediction, ML models have outperformed deep learning models in several cases [35], which supports their application in studies of indoor comfort [33,36–38].

EL combines the strengths of multiple models to reduce overfitting and variance while capturing complex nonlinear relationships. The main methods in EL are Bagging, Boosting, and Stacking [39]. Bagging creates sub-datasets through bootstrap sampling, trains models in parallel, and aggregates their predictions, reducing variance and ensuring diversity. Boosting trains models sequentially, adjusting sample weights with gradient descent to improve performance on difficult data points. For instance, Lee et al. studied solar radiation forecasting and found that their EL models significantly outperformed individual models [40]. Similarly, Yan et al. applied various ML techniques to model indoor lighting and outdoor thermal comfort, with XGBoost yielding the best results [12]. On the other hands, the stacking ensemble architecture have gained popularity for improving predictive accuracy in building performance prediction over the past decade [41]. Stacking is useful for capturing the complex, nonlinear relationships in building performance, where individual models often struggle. Some studies have applied Stacking models to building energy predictions [42,43]. However, the use of Stacking models to IVC assessment is still emerging, with most studies appearing in recent years.

As ML models in building performance research become increasingly complex, ensuring their interpretability has become a critical concern, beyond merely achieving predictive accuracy [44]. Interpretable ML methods are becoming a vital bridge between data-driven predictions and practical design decisions [36,45]. Therefore, models must offer transparent decision logic to build trust and support informed IVC design decisions. To address this, eXplainable Artificial Intelligence (XAI) tools, such as SHAP and LIME, have been increasingly applied to improve model transparency and uncover nonlinear interactions among various environmental and DVs [46–48]. While local interpretation methods like LIME provide useful insights, SHAP offers more robust explanations, especially under sample perturbations [46], making it particularly effective for feature contribution analysis. Thus, model interpretability not only helps clarify causal relationships but also enhances the applicability of research findings in guiding architectural practice, shaping industry standards, and informing policy development [49,50].

3. Methodology

Fig. 1 shows the three steps of the proposed framework, developed to address the challenges of evaluating IVC in sustainable building design. Step 1 involves BPS process and dataset generation. The dataset includes site conditions, DVs, and IVC evaluation metrics (Section 3.1). By sampling a range of building features, a comprehensive dataset was created to reflect performance across diverse scenarios (Section 3.2). Step 2 focuses on ML model selection, training, and evaluation. Various stacking ensemble strategies were compared to balance computational efficiency with predictive accuracy (Section 3.3). Step 3 focuses on analyzing the contribution of DVs (Section 3.4). This step identifies which variables most influence IVC, enhancing the interpretability and transparency of decision-making. Ultimately, this framework offers an automated and interpretable ML-based tool that enables designers to quickly evaluate design scenarios and personalize solutions to specific project needs.

3.1. Parametric modeling and design variables setting

This study developed a parametric modeling framework using two prototype rooms to efficiently evaluate the impact of key DVs on IVC. The prototype rooms were derived from the M3 Building at Suita Campus, The University of Osaka, Japan. The prototypes represent typical office building configurations, enabling a focused exploration of realistic daylighting scenarios (Fig. 2). The 3D scene of the building and its surroundings was modeled using Grasshopper, a widely recognized parametric modeling platform [51]. Digital elevation data from the Shuttle Radar Topography Mission (SRTM) provided a 30-m resolution of the terrain [52], and building height data from OpenStreetMap was used to generate 3D models of the surrounding area [53]. Moreover, typical meteorological year data for Osaka area was used for weather calculation [54]. The indoor light environment was simulated using Ladybug Tools (1.6.0) [55], which operates on the Radiance engine (5.4a)—a highly accurate, Monte Carlo-based ray-tracing tool widely adopted in international lighting standards [56].

This study defined and evaluated several DVs that influence daylight performance. The DVs were defined according to the Comprehensive Assessment System for Built Environment Efficiency (CASBEE) in Japan [57], LEED standard in United States [58], and insights from relevant literature [5,6,12,59], ensuring they align with empirical findings and industry standards. Focusing on 15 key DVs with the most significant impact on IVC, as identified in existing researches, this study avoids including all possible variables. The variation ranges for these DVs, defined in Table 1, were based on the aforementioned standards and reference experiences. Fig. 3 shows the simplified prototype rooms and the associated building elements categorized into four main groups: building forms, fenestration, shading devices, and surface materials, illustrating how building features affect IVC. The relationship between window-to-wall-ratio (WWR) and other fenestration variables is interdependent, meaning adjustments in one may require corresponding changes in others. In other words, increasing the WWR may necessitate adjustments to window dimensions to maintain balanced daylight distribution. Conversely, altering the window's height or length will inherently affect the WWR, as the window's total area relative to the wall area changes.

3.2. Generation of indoor visual comfort dataset

This subsection first establishes an IVC evaluation system, which is developed based on the indoor daylight and glare performance. Then, a comprehensive IVC dataset is constructed by sampling and recording various building features.

3.2.1. Evaluation metrics of indoor visual comfort

This study develops an evaluation framework based on occupant's

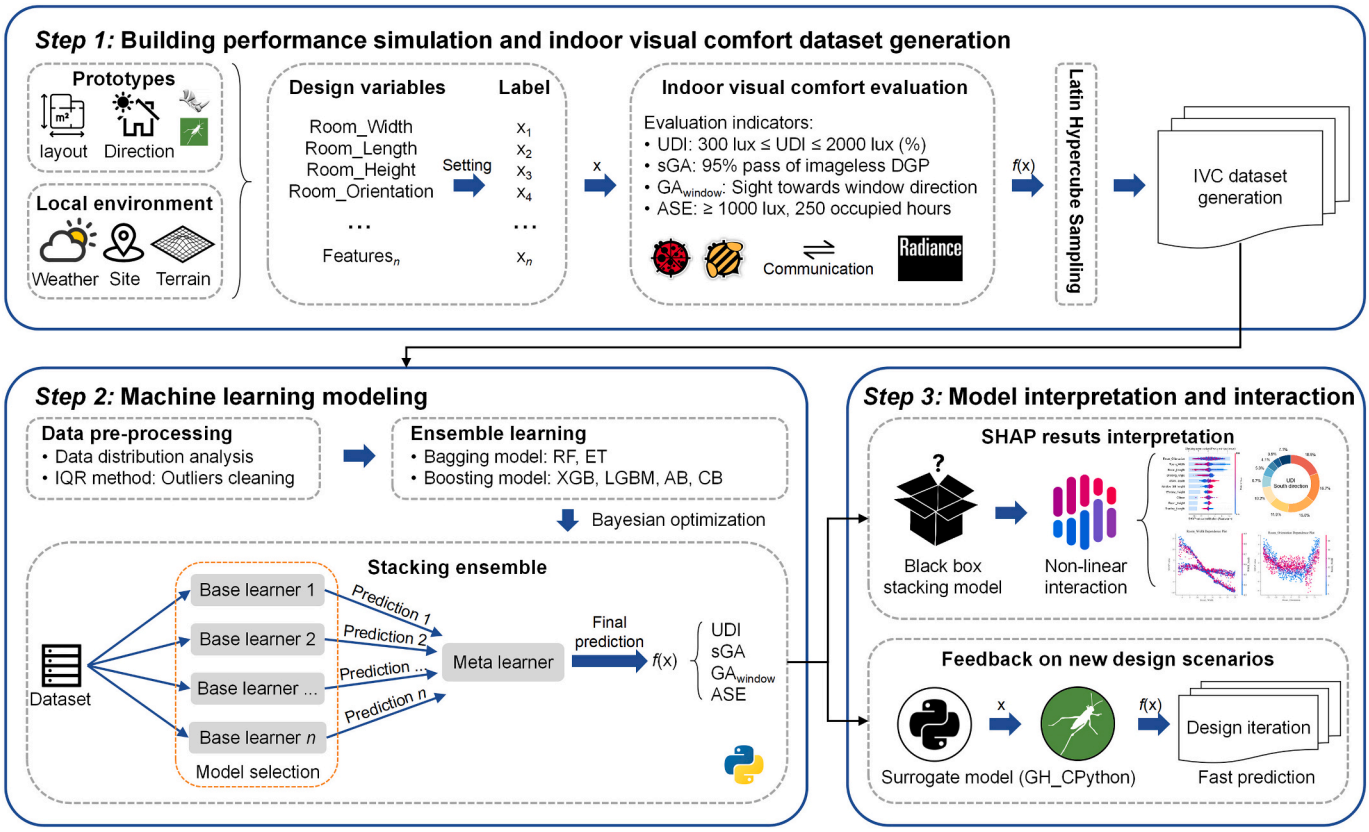


Fig. 1. Research framework.

Fig. 2. Floor plans and photographs of rooms derived from real scenarios in (a) south-facing window openings, and (b) north-facing window openings.

Table 1
Definition of the DVs for IVC environment.

Building features	DVs	Label	Range	Steps	Unit	Description
Building forms	Room_Width	X_1	[5.0, 20.0]	0.1	m	This DV influences spatial layout. The range is set to ensure a balanced room shape that it is neither too narrow nor excessively wide.
	Room_Length	X_2	[5.0, 20.0]	0.1	m	The goal of this DV is same as the Room_Width.
	Room_Height	X_3	[3.0, 5.0]	0.1	m	This DV affects airflow and lighting, and plays a role in indoor comfort.
	Room_Orientation	X_4	[−90, 90]	1	degree	This is a crucial DV for understanding how daylight enters the room and how it can affect IVC and energy use.
Fenestration	WWR_South/North	X_5	[0.30, 0.90]	0.01	–	This DV determines the amount of daylight entering the room and can significantly affect energy efficiency, daylight distribution, and thermal comfort.
	Window_Height	X_6	[2.0, 3.0]	0.1	m	This DV affects the amount of daylight penetration and views to the outside.
	Window_SillHeight	X_7	[0.10. 1.00]	0.01	m	This dimension influences how daylight enters the room, as well as how occupants interact with the window.
	Window_Length	X_8	[0.8, 2.4]	0.1	m	This DV affects the total surface area available for daylight and ventilation.
Shading device	Shading_Length	X_9	[0.00, 2.00]	0.01	m	This DV controls how much direct sunlight enters through the window, thereby affecting energy use and occupant comfort.
	Shading_Angle	X_{10}	[−60, 60]	1	degree	This DV helps to regulate solar gain and visual comfort by adjusting how shading is applied.
Surface material	Reflectance_Wall	X_{11}	[0.50, 0.90]	0.01	–	This DV is important for understanding how much light is reflected into the room, which can enhance daylight availability and visual comfort.
	Reflectance_Ceiling	X_{12}	[0.70, 0.90]	0.01	–	A higher reflectance increases the amount of reflected light, contributing to brighter interiors and improved visual comfort.
	Reflectance_Floor	X_{13}	[0.20, 0.50]	0.01	–	Floor reflectance influences the distribution of light across the room, affecting both lighting efficiency and visual comfort.
	Transmittance_Window	X_{14}	[0.70, 0.95]	0.01	–	This DV determines how much daylight is transmitted through the window, influencing energy efficiency and indoor lighting quality.
	Reflectance_Shading	X_{15}	[0.50, 0.90]	0.01	–	This affects the overall performance of the shading system in terms of both controlling solar gain and improving the distribution of daylight in the space.

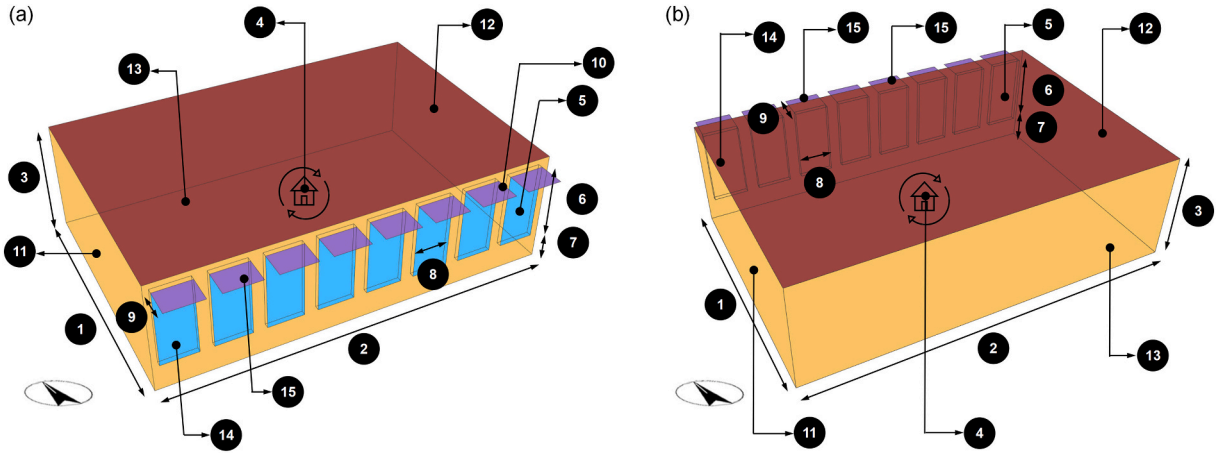


Fig. 3. Prototype rooms used in BPS process, showing the applied DVs: (a) south-facing window openings, and (b) north-facing window openings.

line of sight, integrating both horizontal illuminance distribution and vertical glare risk from window directions. Four indicators—UDI, ASE, GA_{window}, and sGA—form a multi-scale system for IVC assessment. The simulation period spans from 8:00 to 18:00, aligning with the building operation schedule outlined in Japan's green building standards [57].

Useful daylight illuminance (UDI) is a critical metric for evaluating daylight distribution on the horizontal work plane, defined as the percentage of time the illuminance on the work plane stays within a defined effective range during a specific period [60]. UDI has three illuminance ranges, each offering distinct insights into daylight quality: insufficient daylight (UDI-i, illuminance <300 lx), autonomous daylight (UDI-a, 300 lx ≤ illuminance ≤2000 lx), and excessive daylight (UDI-e, illuminance >2000 lx). In this study, a 1 × 1 m grid sensor matrix was used at a work plane height of 0.75 m, with the autonomous illuminance range of UDI (UDI-a at here) set between 300 and 2000 lx, representing the range of daylight that is most suitable for work activities. While UDI-e identifies zones with high illuminance levels that may pose a glare risk, it does not directly quantify directional glare effects. It focuses on the temporal frequency of overexposure (time with illuminance >2000 lx), while glare assessment requires additional metrics.

Annual sunlight exposure (ASE) complements UDI by evaluating the spatial distribution of prolonged high-exposure areas, defined as the percentage of floor area exceeding 1000 lx for at least 250 occupied hours annually [61]. ASE highlights areas prone to glare and thermal discomfort, complementing UDI by providing spatial context for high illuminance, while UDI focuses on temporal frequency. Although UDI and ASE together offer a comprehensive evaluation of daylight quality, their potential redundancy is acknowledged. Both metrics highlight excessive illuminance, but their focus on time versus space provides complementary insights.

The actual glare effects necessitate additional metrics and vertical-plane analysis. Glare autonomy (GA) assesses visual discomfort based on the occupant's line of sight. The brightness and contrast terms in the DGP equation were replaced to create an imageless evaluation model based on multi-step calculations [20]. This method calculates the percentage of glare-free time (DGP ≤ 0.4) from specific viewpoints without

sacrificing accuracy in contrast terms, making it suitable for early-stage design analysis. However, it does not account for dynamic viewing directions, limiting its adaptability to varied occupant behaviors.

The DGP equation is as follows:

$$DGP = \underbrace{5.87 \times 10^{-5} \cdot E_v}_{\text{brightness term}} + \underbrace{0.0918 \cdot log_{10}\left(1 + \sum_{i=1}^{n} \frac{L_{s,i}^2 \omega_{s,i}}{E_v^{1.87} P_i^2}\right)}_{\text{contrast term}} + 0.16 \quad (1)$$

where E_v is the vertical eye illuminance, L_s and ω_s are the luminance and solid angle of glare source, respectively, P is Guth position index, and n is the number of glare sources.

In GA, the brightness and contrast terms are redefined using daylight coefficient-based calculations:

$$E_v^{'} = kD_{total} \bullet S \quad (2)$$

$$L_s = \frac{kd_{direct} s_i}{\omega cos\theta} \quad (3)$$

where the vector D_{total} is daylight coefficients, S is point in time sky luminance vector of all Reinhart sky patches, d_{direct} is the direct component coefficient for sky patch i, s_i is point in time sky luminance value of sky patch i, and k is 179 lm/W represents the luminous efficacy of daylight.

Unlike previous methods [20,62], which rely on uniform directional sampling or fixed glare metrics, this study introduces a dynamic view-tracking approach that adapts to both building rotation and window positioning. The proposed glare autonomy for window direction (GA_{window}) index effectively integrates room orientation and occupant viewing direction for dynamic glare detection: (1) Seated eye-level height: A height of 1.1 m is assumed based on anthropometric data [63,64], which is adopted in glare evaluations. The horizontal field of view is divided into eight segments, with three (spanning 135°) used for glare calculation. (2) View-tracking algorithm: This algorithm dynamically links room orientation to the field of view (Fig. 4). As the building rotates, the occupant's forward view remains aligned with the window position within ±45°. When the window shifts beyond this range ($\theta \in [-90^{\circ}, -46^{\circ}]$ or $\theta \in [46^{\circ}, 90^{\circ}]$), the system switches the reference window side to maintain glare evaluation accuracy. This ensures that occupant's perspective remains anchored to the actual light entry direction.

This study employs spatial glare autonomy (sGA) index proposed by Jones, which extends GA by using dual glare thresholds ($0.4 \leq DGP \leq 0.45$) and time proportion (≥ 5 % of the annual working hours) to evaluate the uniformity of glare distribution [20]. sGA is well-suited for assessing glare in multiple areas, offering a comprehensive evaluation of glare gradients. Although similar glare metric, such as the spatial disturbing glare (sDG) in the ClimateStudio plugin [62], sDG uses a broader and less specific threshold ($DGP > 0.38$), making sGA's dual-threshold approach more precise for performance-driven design.

3.2.2. Sampling and dataset generation

Based on a comparative analysis (Appendix Fig. A.1) and findings from previous studies [5,59,65], Latin hypercube sampling (LHS) was selected for its superior ability to uniformly and randomly cover the parameter space, making it well-suited for capturing nonlinear relationships in medium- to high-dimensional datasets [66]. The minimum sample size should be ten times the number of input variables (i.e., 150 samples in this study) according to previous research [67,68]. However, 150 samples were insufficient to cover all building features adequately as ML model requires more data for training and testing. While increasing the sample size further would significantly extend dataset preparation time with slightly improvements in model accuracy. Using empirical and trial-and-error methods [69], this study designed 2000 samples for both south- and north-facing window scenarios, respectively. These two datasets were used for training the ML models. Considering the dimensional requirements of the DVs in this study, this sample size could fulfill the requirements for dataset construction.

3.3. Ensemble learning supported model prediction

This study was conducted on a personal computer equipped with an Intel Core i9-12900KF processor (3.2 GHz), NVIDIA GeForce RTX 3090, and 64GB of RAM. All ML models were implemented using the Scikit-learn library in Python [70].

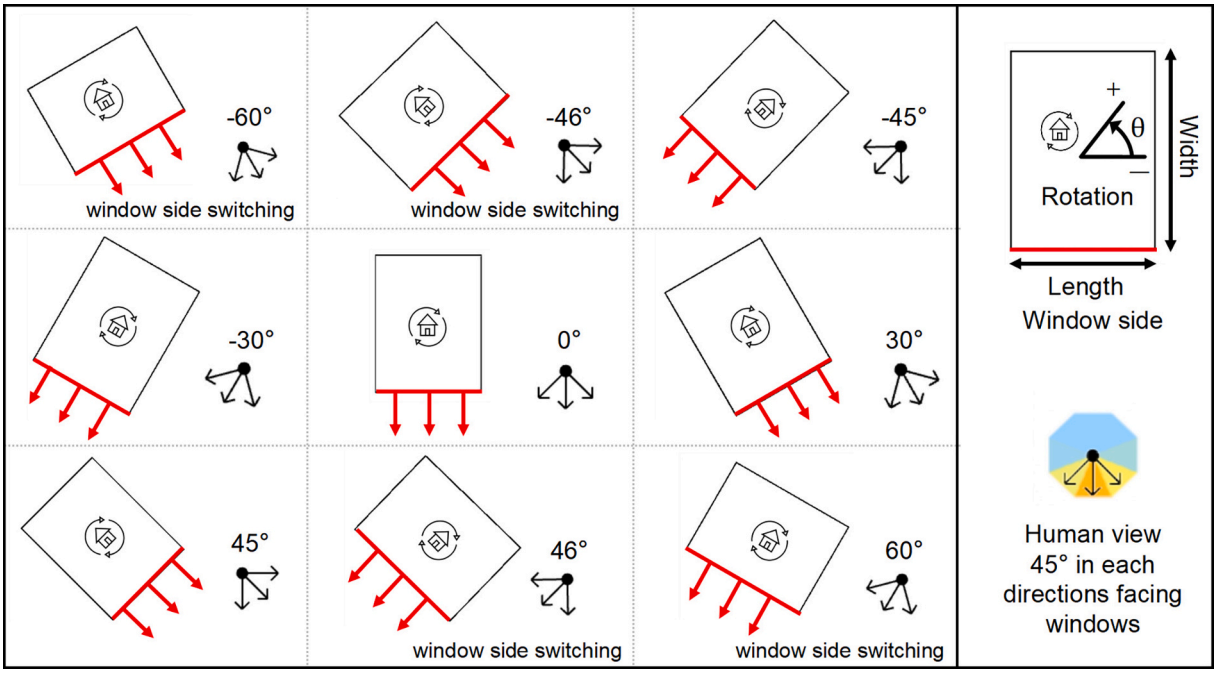


Fig. 4. Occupant-centered horizontal field of view toward the window direction, with the wall where the window is opened defined as "Length" and the opposite wall defined as "Width".

3.3.1. Data pre-processing

The dataset used in this study was generated from simulation scenarios involving both south-facing and north-facing window openings, with 2000 samples per direction, totaling 4000 samples. Each sample corresponds to a unique configuration of building features, window placement, and sun position. Due to the randomized nature of the sampling process, some combinations produced extreme or atypical glare outcomes. To ensure data quality, Interquartile range (IQR) method is employed [71] for data cleaning. This technique eliminates data points lying significantly outside the middle 50 % of the distribution, which could otherwise distort model training and evaluation (Appendix Fig. A.2). This approach guarantees that the data used for IVC evaluation is of both high quality and representative.

3.3.2. Ensemble learning approach

This study employs EL techniques to improve predictive accuracy and robustness. The stacking ensemble follows a two-layer architecture that combines the predictions of multiple base learners (Fig. 5). A meta-learner is then used to determine the optimal weighting of these predictions, enhancing overall model performance [72].

(1) Level 1 (base-learner): Each base model is trained using K-fold cross-validation to ensure full data utilization and prevent data leakage. In each fold, the model is trained on K-1 subsets and validated on the remaining subset. The out-of-fold predictions are assembled into a meta-feature matrix $M \in R^{n\times m}$, where n is the number of base models and m is the number of samples.
(2) Level 2 (meta-learner): A linear regression model serves as the meta-learner, which takes M as input and learns the optimal combination of base model outputs. The final stacked prediction is computed as:

$$\hat{y}_{stacking} = \sum_{i=1}^{m} \omega_i f_i(x) + b \tag{4}$$

where $f_i(x)$ is the prediction from the i-th base learner, ω_i is its corresponding weight, and b is the bias term. The meta-learner assigns higher weights to more accurate models for specific data patterns, improving both prediction accuracy and generalization.

3.3.3. Model selection and stacking

To build a robust and generalizable stacking ensemble for IVC prediction, this study follows a two-stage learning process. It involves selecting diverse base learners, training them independently, and combining their outputs with a meta-learner.

- Selection and training of base models

This study first evaluates six state-of-the-art models for constructing a robust stacking ensemble: XGBoost (XGB), LightGBM (LGBM), CatBoost (CB), AdaBoost (AB), Random Forest (RF), and Extra Trees (ET). The six base models are grouped into two categories: (1) Bagging-based models (RF, ET), which use bootstrap resampling to train multiple learners in parallel, reducing variance and mitigating overfitting through feature selection randomness. (2) Boosting-based models (XGB, LGBM, CB, AB), which are trained sequentially, with each model focusing on misclassified instances of its predecessors. All base models were trained using 80 % of the dataset, with each model trained independently under its respective ensemble strategy to capture various feature-space structures. K-fold cross-validation ($K = 5$) was applied in each fold to improve model generalization capabilities.

- Training and evaluation of meta-model

The meta-model was trained on out-of-fold predictions from the base models, ensuring that it combined model outputs without any data leakage from the training data. For final evaluation, 20 % of the dataset, untouched during base models and meta-model training, was reserved as an independent test set. This test set was solely used for performance assessment, ensuring an unbiased evaluation that reflects the model's ability to generalize to new data. By integrating diverse models and maintaining strict data separation throughout the training process, the stacking ensemble framework enhances prediction stability and generalization. In contrast to rule-based ensemble strategies, stacking allows the meta-model to learn optimal weights, improving adaptability to complex IVC scenarios.

3.3.4. Hyperparameter optimization

Hyperparameter optimization process can enhance model training efficiency, generalization ability, and prediction accuracy by identifying the best combination of hyperparameters [73]. Hyperparameter fine-tuning mainly focuses on learning rate, max depth, number of iterations, subsample, lambda, and alpha. Bayesian optimization was employed in this study to improve model performance, reduce

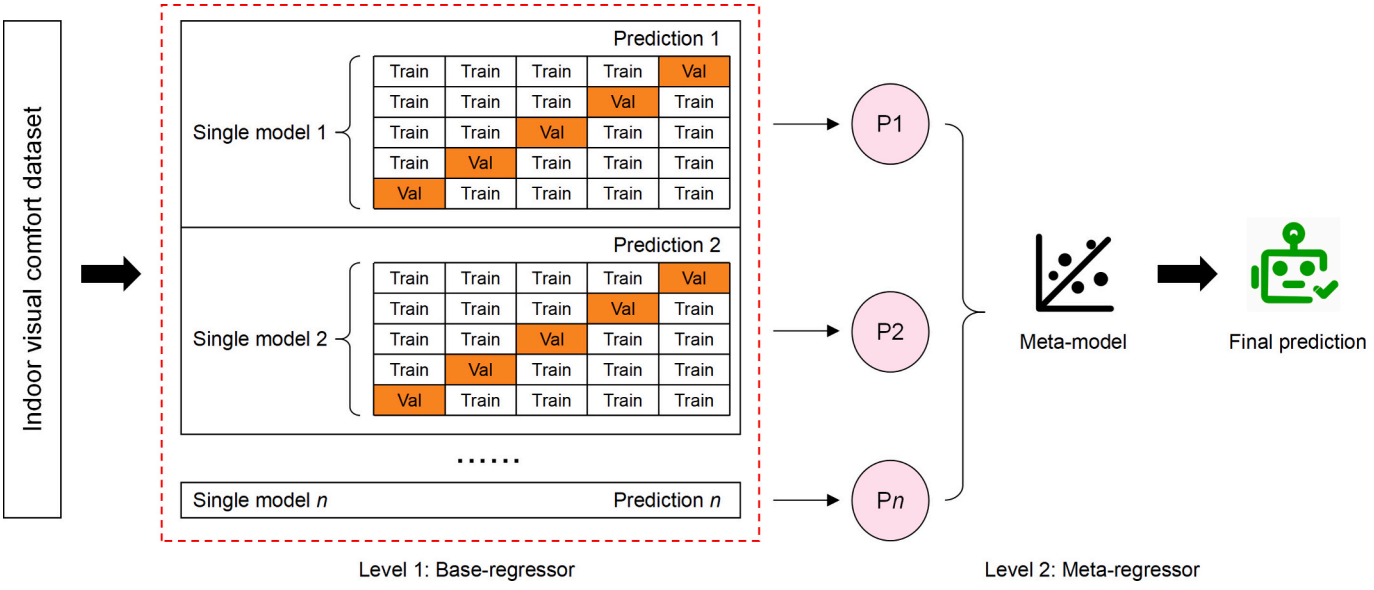


Fig. 5. The architecture of Stacking ensemble learning approach.

computational resources, and effectively prevent overfitting or underfitting. It offers the advantage of using a surrogate model (Gaussian process) to predict performance distribution, allowing for a more efficient search for the global optimum with fewer computations [74].

3.3.5. Model evaluation

Model performance is evaluated based on six EL models and the final stacking ensemble. All models were assessed on the held-out test set to provide an unbiased estimate of generalization performance. Three standard regression evaluation metrics were used to evaluate accuracy: the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) (Eqs. (5)–(7)) [75–77]. R^2 represents the proportion of variance in the dependent variable explained by the independent variables, with values approaching 1.0 indicating greater explanatory strength. RMSE and MAE provide insights into the prediction error, with RMSE penalizing larger errors more heavily and MAE reflecting average absolute errors. Lower values of RMSE and MAE indicate better model performance. The definitions of R^2, RMSE, and MAE are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{6}$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{7}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, and n is the sample size.

3.4. Model interpretation

The SHapley Additive exPlanations (SHAP) method enhances model transparency by quantifying the marginal contribution of each feature to the model's output [48]. Based on Shapley values from cooperative game theory [78], it quantifies the contribution of each feature to the prediction outcome, with the absolute Shapley values indicating the priority of DVs for adjustment. For IVC prediction, SHAP effectively captures nonlinear relationships between DVs and lighting parameters, clearly highlighting feature interactions. SHAP method provides both global and local interpretability, integrates the benefits of various model-agnostic methods, and is robust to perturbations, making it ideal for interpreting prediction models in this study. Two types of interpretation were conducted: (1) Global interpretation using SHAP beeswarm plots to visualize the overall importance of each DV across all samples, and (2) Local interpretation using dependence plots to explain individual predictions and explore how feature values influence model output in specific cases.

The calculation of Shapley value is as follows:

$$\phi_p = \sum_{S \subseteq \{x_1, \ldots, x_p\} \setminus \{x_p\}} \frac{|S|!(P - |S| - 1)!}{P!} \left(f(S \cup \{x_p\}) - f(S) \right) \tag{8}$$

where ϕ_p represents the Shapley value for feature P, S is a subset of features, x_p is the value vector for feature p, P is the total number of features, and $f(S)$ is the model prediction using only the features in subset S.

$f(x)$ is a linear function of Shapley value of feature P:

$$f(x) = \phi_0 + \sum_{p=1}^{P} \phi_p z_i' \tag{9}$$

where $f(x)$ is the predictive output and ϕ_0 is the expected value of prediction. z_i' is a binary feature: $z_i' = 1$ represents a present feature and $z_i' = 0$ represents absent features. Features with larger absolute Shapley values contribute more to the IVC factors.

4. Results

This section presents the sampled data on IVC performance, compares the predictive accuracy of different stacking architectures, analyzes the results of the optimal Stacking model, and demonstrates the application of proposed framework across various climate zones.

4.1. Dataset generation and pre-processing

The dataset contains 4000 IVC calculation samples for both south- and north-facing orientations. Fig. 6 shows the data distribution. The UDI index range for the sampled results is approximately 2.5 % higher in south-facing spaces than in north-facing ones, indicating a larger proportion of usable daylight throughout the year. North-facing indoor environments exhibit notable advantages in glare conditions. Median values of glare indicators—sGA and GA_{window}—increase by 20.7 % and 6.8 %, respectively, suggesting that most north-facing spaces are glare-free. In the ASE evaluation, the mean and median values for north-facing spaces are clustered closer to the coordinate origin and differ significantly from those of south-facing spaces.

4.2. Model performance and evaluation results

This subsection compares different stacking architectures to evaluate the trade-off between predictive performance and computational efficiency. It employs EL and Bayesian optimization for hyperparameter tuning, developing eight Stacking models across various room orientations and light environment indicators, and systematically validating their effectiveness in predicting IVC.

4.2.1. Model performance results

Six EL models were trained using Bayesian optimization for performance benchmarking (Fig. 7). The results showed that XGB, LGBM, and CB outperformed other models in terms of R^2, RMSE, and MAE. These three were therefore selected as base learners for stacking. Regarding computational efficiency, the XGB-based Stacking model required less than 1 min for training, significantly faster than LGBM-based (40 min) and CB-based (68 min). Given the need to balance accuracy and practical deployment, the XGB-based Stacking model was selected for further experiments due to its optimal balance of performance and speed.

Model combinations were validated using 5-fold cross-validation (Table 2). The three-model stack (XGB-LGBM-CB) yielded the best overall performance with $R^2 = 0.911$, RMSE = 4.084, and MAE = 2.986. This integration improved prediction accuracy by 0.9 % in R^2 and reducing error metrics by approximately 4.5 % in RMSE and MAE compared to two-model stacks (e.g., XGB-CB). Adding more base learners (e.g., four- to six-model stacks) increased training set fit (R^2 up to 0.988) but did not improve test performance. For example, the five-model stacking (XGB-LGBM-CB-RF-AB) showed higher RMSE (4.135) and MAE (3.016), indicating that increasing the number of base learners can lead to overfitting. Moreover, integrating Bagging-based models (e.g., RF) into Boosting-based stacks resulted in performance reduction. The XGB-LGBM-RF architecture showed RMSE and MAE increases of 8.5 % and 13.5 %, respectively, compared to the optimal three-model stack. These results underscore the critical influence of base learner selection on stacking performance. Therefore, this study adopts the XGB-LGBM-CB stacking architecture, with a linear regression model as the meta-learner to assign weights to base learner outputs. This architecture combines the efficiency and generalization capability of linear models with the complementary strengths of the base learners, enhancing

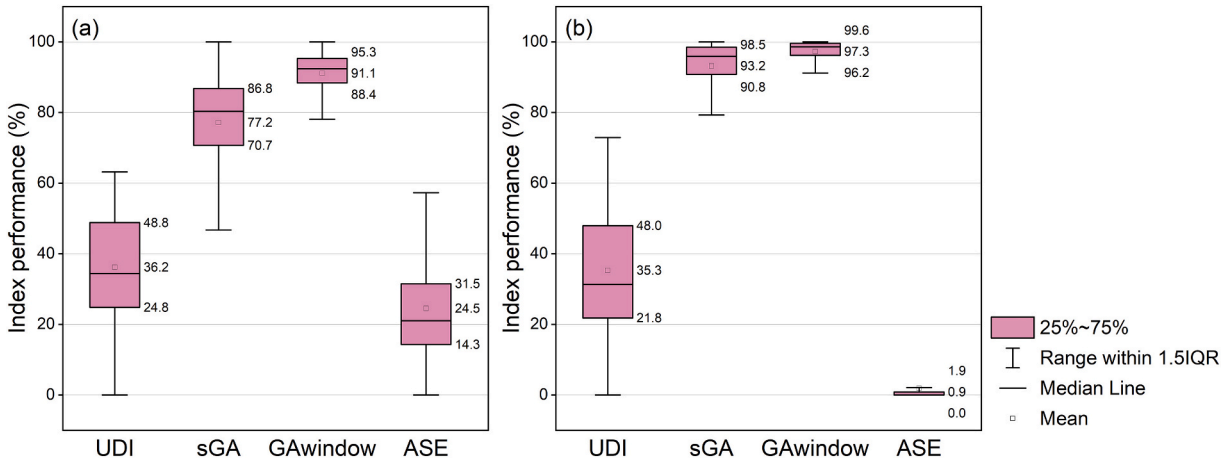


Fig. 6. Data distribution of UDI, sGA, GA_{window}, and ASE indicators: (a) south-facing window openings; (b) north-facing window openings.

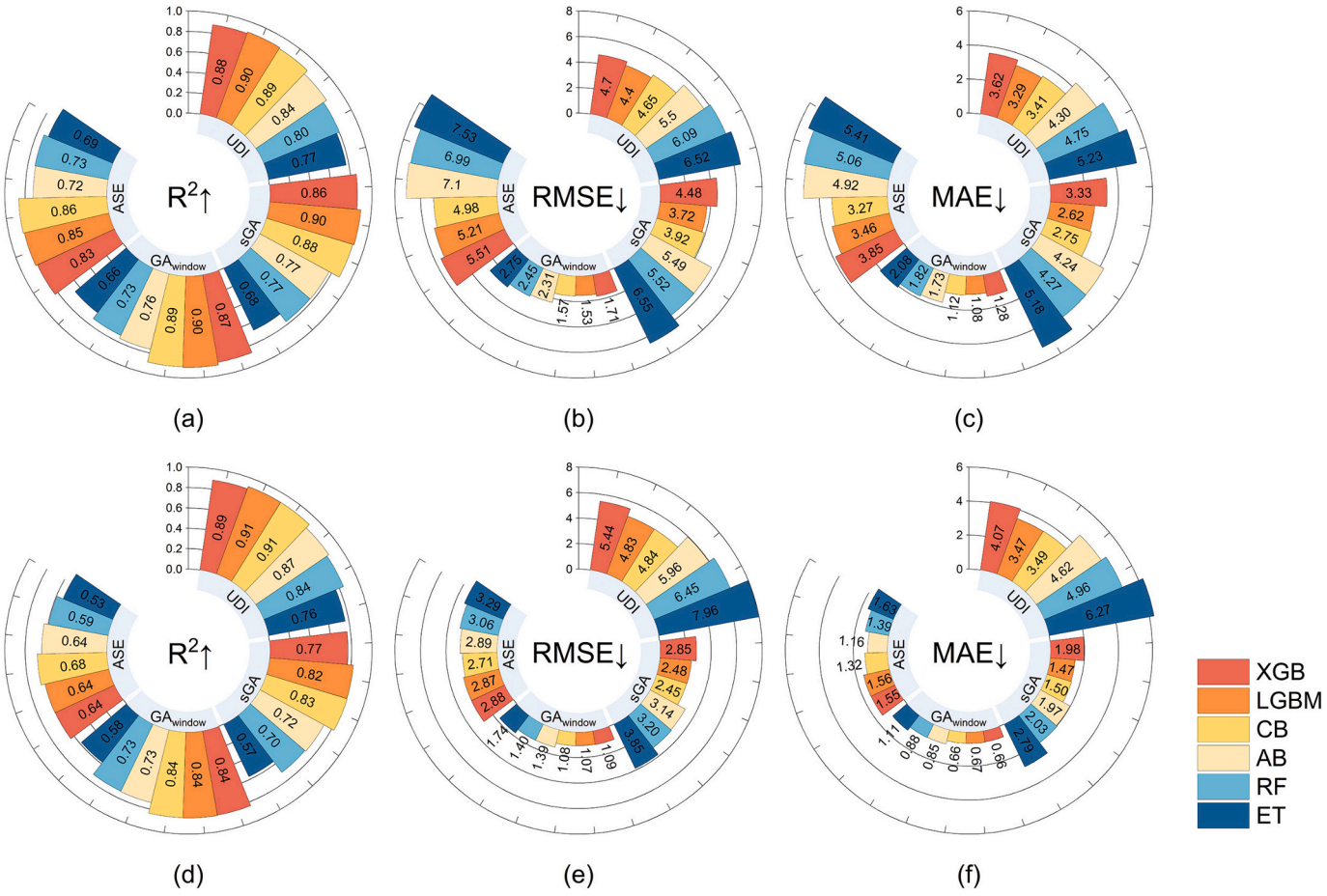


Fig. 7. Performance of six EL models based on R^2, RMSE, and MAE indicators: (a), (b), (c) represent south-facing window openings; (d), (e), (f) represent north-facing window openings.

prediction robustness under complex IVC scenarios.

4.2.2. *Evaluation results of stacking model*

As shown in Tables 3 and 4, the Stacking models outperform the individual base learners across four indicators. Among them, the best Stacking model achieved $R^2 = 0.922$, compared to 0.913 for the top-performing base model. Similarly, the Stacking model also demonstrated lowest RMSE (1.011) and MAE (0.663) in north-facing GA_{window} index compared to other models. This may indicate its superior ability to capture complex interactions among DVs.

Regression results for both training and test sets are shown in Fig. 8. For UDI prediction, the north-facing test set achieved a slightly higher $R^2 = 0.922$ than the south-facing set ($R^2 = 0.911$), with both scatter plots closely aligned along the y = x regression line, indicating high predictive accuracy. However, in glare-related indicators, south-facing models performed better due to greater variability in direct light. Specifically, sGA prediction yielded $R^2 = 0.914$ for the south-facing model versus $R^2 = 0.846$ for the north-facing model; for GA_{window}, the values of

Table 2
Average results of different Stacking model selections.

Stacking model architecture	Time cost (s)	Training set			Test set		
		R^2↑	RMSE↓	MAE↓	R^2↑	RMSE↓	MAE↓
XGB, RF	83.7	0.979	1.936	1.520	0.821	5.788	4.317
XGB, CB	18.1	0.922	1.231	0.976	0.903	4.262	3.125
XGB, LGBM	15.4	0.982	1.801	1.377	0.889	4.570	3.411
XGB, LGBM, CB	**21.2**	**0.987**	**1.547**	**1.205**	**0.911**	**4.084**	**2.986**
XGB, LGBM, RF	26.1	0.982	1.796	1.410	0.895	4.431	3.389
XGB, CB, RF	33.5	0.986	1.566	1.242	0.907	4.174	3.077
XGB, LGBM, CB, RF	44.7	0.988	1.447	1.230	0.910	4.106	2.994
XGB, LGBM, CB, AB	23.4	0.986	1.578	1.246	0.910	4.098	2.978
XGB, LGBM, CB, ET	25.8	0.985	1.625	1.287	0.910	4.104	3.014
XGB, LGBM, CB, RF, AB	28.6	0.986	4.599	1.280	0.909	4.135	3.016
XGB, LGBM, CB, RF, ET	36.2	0.988	1.449	1.100	0.911	4.097	2.986
XGB, LGBM, CB, ET, AB	28.6	0.987	1.512	1.213	0.911	4.090	2.975
XGB, LGBM, CB, RF, AB, ET	38.5	0.988	1.447	1.143	0.910	4.106	2.994

Abbreviation: XGB: XGBoost; LGBM: LightGBM; CB: CatBoost; AB: AdaBoost; RF: random forest; ET: extra tree.

Table 3
Indoor visual comfort performance results for the south-facing window openings.

Models	Training set			Test set		
	R^2↑	RMSE↓	MAE↓	R^2↑	RMSE↓	MAE↓
UDI index						
XGB	0.997	0.674	0.480	0.882	4.699	3.617
LGBM	0.986	1.583	1.123	0.897	4.398	3.291
CB	0.990	1.354	0.853	0.885	4.651	3.408
AB	0.995	0.946	0.583	0.839	5.499	4.298
RF	0.960	2.706	1.959	0.802	6.088	4.753
ET	0.990	1.340	0.934	0.773	6.523	5.232
Stacking	0.987	1.547	1.205	0.911	4.084	2.986
sGA index						
XGB	0.962	2.370	1.735	0.858	4.480	3.328
LGBM	0.992	1.119	0.782	0.896	3.715	2.621
CB	0.987	1.377	0.897	0.884	3.924	2.750
AB	0.982	1.625	1.163	0.773	5.485	4.242
RF	0.948	3.796	2.001	0.770	5.517	4.266
ET	0.985	1.507	1.173	0.676	6.547	5.180
Stacking	0.979	1.760	1.375	0.914	3.382	2.307
GA_{window} index						
XGB	0.984	0.672	0.496	0.869	1.713	1.278
LGBM	0.985	0.648	0.453	0.896	1.526	1.080
CB	0.989	0.564	0.369	0.890	1.569	1.117
AB	0.985	0.645	0.478	0.762	2.308	1.726
RF	0.935	1.356	0.923	0.731	2.453	1.821
ET	0.990	0.525	0.368	0.661	2.753	2.079
Stacking	0.984	0.682	0.552	0.913	1.394	1.064
ASE index						
XGB	0.962	2.485	1.793	0.832	5.509	3.851
LGBM	0.987	1.462	1.041	0.850	5.208	3.459
CB	0.986	1.512	1.048	0.863	4.979	3.265
AB	0.990	1.311	0.836	0.721	7.099	4.916
RF	0.938	3.201	2.201	0.730	6.992	5.060
ET	0.969	2.251	1.639	0.686	7.532	5.413
Stacking	0.977	1.944	1.518	0.883	4.600	3.089

Table 4
Indoor visual comfort performance results for the north-facing window openings.

Models	Training set			Test set		
	R^2↑	RMSE↓	MAE↓	R^2↑	RMSE↓	MAE↓
UDI index						
XGB	0.981	2.275	1.737	0.889	5.435	4.068
LGBM	0.983	2.201	1.504	0.913	4.830	3.465
CB	0.988	1.832	1.328	0.912	4.837	3.486
AB	0.997	0.904	0.521	0.867	5.957	4.619
RF	0.964	3.169	2.267	0.844	6.452	4.959
ET	0.978	2.476	1.881	0.763	7.962	6.274
Stacking	0.992	1.589	1.276	0.922	4.576	3.083
sGA index						
XGB	0.944	1.537	1.081	0.765	2.847	1.981
LGBM	0.986	0.778	0.514	0.822	2.480	1.468
CB	0.994	0.499	0.316	0.826	2.450	1.500
AB	0.982	0.872	0.615	0.715	3.138	1.969
RF	0.952	1.418	0.832	0.704	3.195	2.033
ET	0.910	1.949	1.311	0.570	3.852	2.786
Stacking	0.983	0.834	0.660	0.846	2.305	1.410
GA_{window} index						
XGB	0.973	0.447	0.301	0.838	1.086	0.659
LGBM	0.968	0.494	0.313	0.842	1.071	0.671
CB	0.987	0.308	0.194	0.840	1.077	0.661
AB	0.993	0.230	0.159	0.734	1.390	0.854
RF	0.933	0.710	0.402	0.729	1.402	0.875
ET	0.972	0.456	0.238	0.582	1.743	1.110
Stacking	0.981	0.381	0.305	0.859	1.011	0.663
ASE index						
XGB	0.954	0.944	0.612	0.640	2.876	1.547
LGBM	0.956	0.925	0.602	0.642	2.869	1.558
CB	0.991	0.428	0.242	0.681	2.707	1.322
AB	0.997	0.240	0.131	0.637	2.890	1.155
RF	0.911	1.317	0.596	0.593	3.060	1.392
ET	0.990	0.435	0.212	0.529	3.291	1.626
Stacking	0.970	0.757	0.509	0.709	2.587	1.452

R^2 were 0.913 and 0.859, respectively. This suggests the models are more effective in environments with broader illuminance ranges, where nonlinear patterns are more pronounced. Conversely, the reduced variability in north-facing spaces limits model sensitivity, often leading to underestimated errors. ASE results further emphasize directional differences. The south-facing model achieved $R^2 = 0.883$, while the north-facing model dropped to $R^2 = 0.709$. This discrepancy can be attributed to the lower temporal variance of diffuse light in north-facing spaces, reducing the model's ability to detect subtle shifts in light exposure—particularly in low-ASE regions. These findings highlight the need for greater sensitivity to weak lighting signals to enhance model performance in diffuse-light conditions.

Overall, these Stacking models demonstrated strong accuracy and robustness across all indicators. In south-facing spaces, dynamic lighting conditions facilitated more effective learning of nonlinear relationships. In contrast, north-facing spaces yielded lower absolute errors but require improved sensitivity to minor illuminance fluctuations—especially for ASE prediction. The variations in R^2, RMSE, and MAE between orientations (Tables 3 and 4) underscore the influence of lighting conditions on model performance and error metric behavior.

4.3. Model interpretation results

This subsection employs the SHAP method for interpretability analysis to deconstruct the multidimensional impact mechanisms of IVC prediction based on the XGB-LGBM-CB architecture Stacking model. The

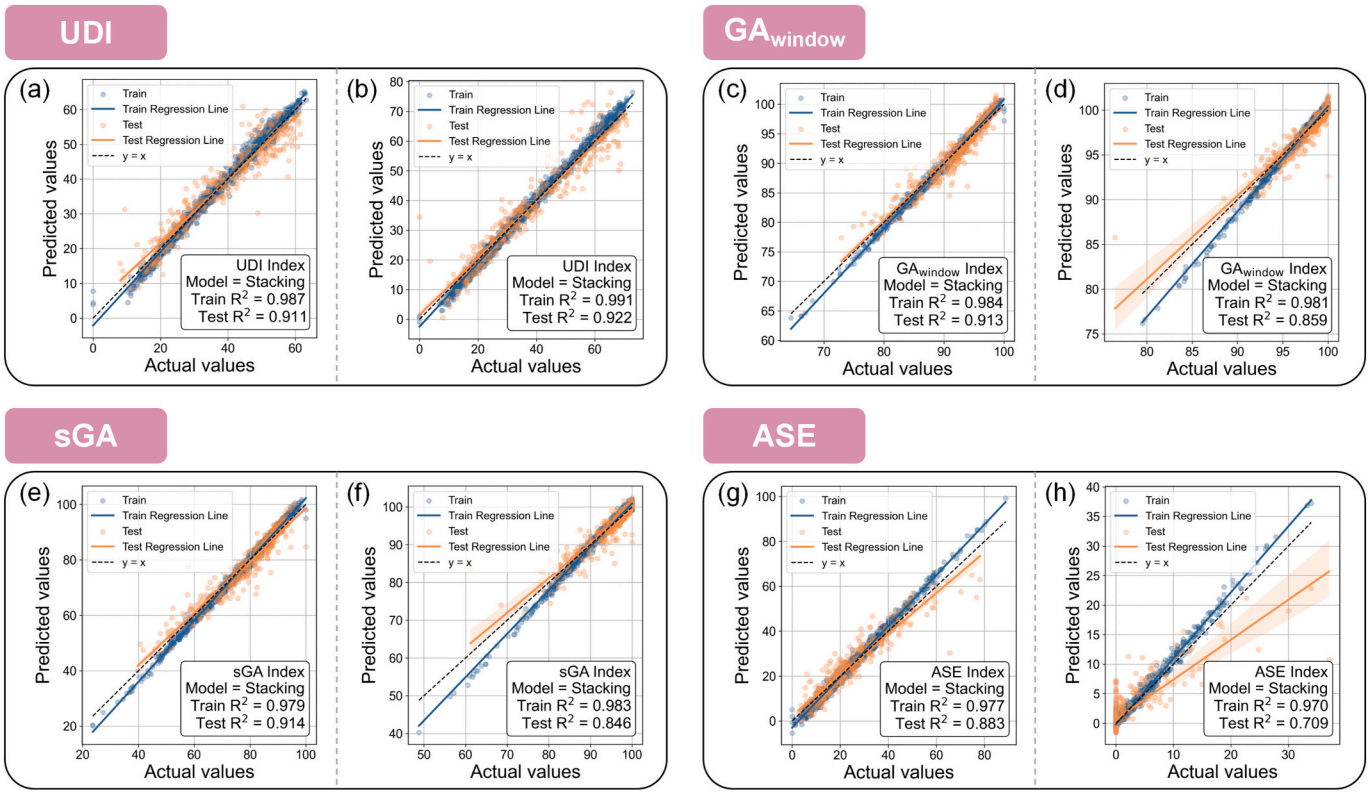


Fig. 8. Training and test performance of the Stacking model on the target index, with a 95 % confidence interval, (a), (c), (e), (g) represent the south-facing window openings; (b), (d), (f), (h) represent the north-facing window openings.

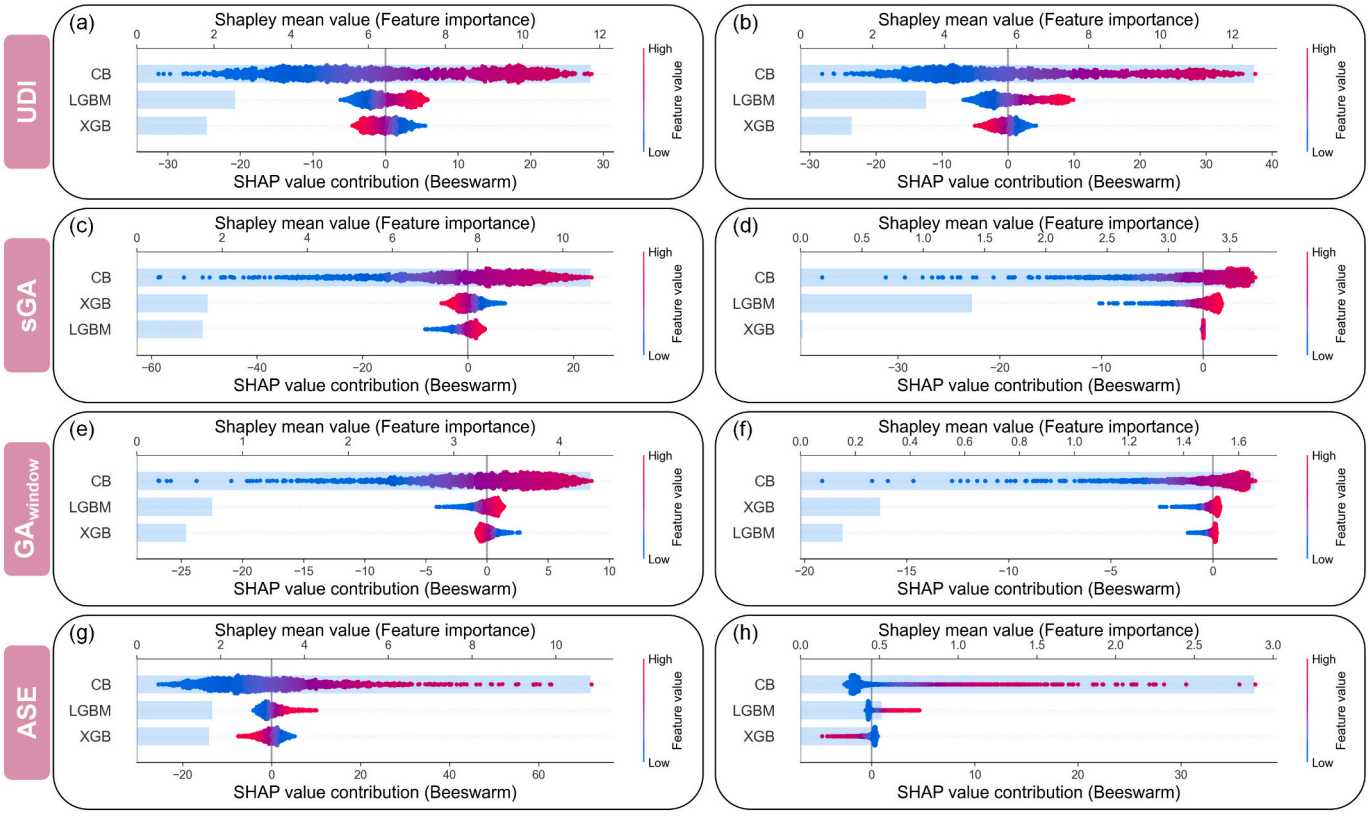


Fig. 9. Global interpretation of SHAP value contributions for the Stacking model based on the target index in the south and north directions: (a), (c), (e), (g) represent the south-facing window openings; (b), (d), (f), (h) represent the north-facing window openings.

global and local interpretation analyses reveal nonlinear interactions between DVs with SHAP values.

4.3.1. Global interpretation results

The SHAP method was used to deconstruct the Stacking model based on the XGB-LGBM-CB architecture. Eight subplots for the south-facing and north-facing windows reveal the explanatory advantages of the CB model in the automated evaluation system (Fig. 9). The average SHAP contribution for the CB model is about 75 %, significantly higher than LGBM (15 %) and XGB (10 %). These findings demonstrate that the CB model excels in capturing nonlinear interactions in the lighting environment. Given that the stacking model itself is not directly interpretable, the CB model was selected for this analysis.

Fig. 10 shows the quantified contribution weights of the DVs. Each subplot includes the SHAP value contribution on the left and a pie chart showing the contribution percentage on the right. The results reveal significant differences in feature importance between south- and north-facing spaces. Tables 5 and 6 rank the relative marginal contributions of the DVs, where higher percentages reflect greater impact on IVC. Among building features, building forms contribute approximately 46.6 %–52.7 % to prediction performance (south-facing average: 46.6 %, north-facing average: 52.7 %). Four features inside this category can consider as the key building factors in design decisions. Fenestration and shading device contribute between 22.6 %–24.9 % and 16 %–24.2 %, respectively, while surface material contributes only 6 %–7.2 %.

Further analysis of the four evaluation indicators reveals the following: (1) For the UDI index (Fig. 10a-b), room orientation, room length, and room width are the most influential factors. In south-facing spaces, orientation and room width contribute 18.9 % and 16.7 %, respectively, to the overall prediction. In north-facing spaces, room width, room length, and orientation together account for 53.6 %. Additionally, the shading angle and WWR contribute 11.9 % and 11.2 %, respectively, for south-facing rooms. (2) In glare evaluation (sGA and GA_{window}), the south-facing shading angle is the most significant factor, contributing 15.9 %. Combined with room length (15.4 %), these two factors form the dual control mechanism for light distribution (Fig. 10c). For north-facing spaces, WWR and orientation together contribute 48.8 % (Fig. 10d), emphasizing the importance of facade design. These results further highlight the role of room size and orientation in influencing comfort levels. (3) The ASE index confirms the central role of shading systems. For south-facing spaces, the shading angle contributes 17.5 % (Fig. 10g), with room length (15.8 %) and WWR (13.8 %) collectively accounting for 47.1 % of the explanatory dimension. For north-facing spaces, orientation contributes 33.9 % (Fig. 10h), with WWR contributing 18.9 %, requiring careful attention in design. The SHAP values for ASE and UDI highlight the relative contributions of different DVs to excessive-brightness risk—a necessary precursor to glare—although validating perceptual glare still requires occupant studies. In summary, six key DVs—room orientation, room length, room width, WWR, shading length, and shading angle—account for 80.6 % of the prediction performance and are critical to determining IVC. While the ranking of feature importance varies slightly, orientation, room length, and room width are consistently the most significant factors.

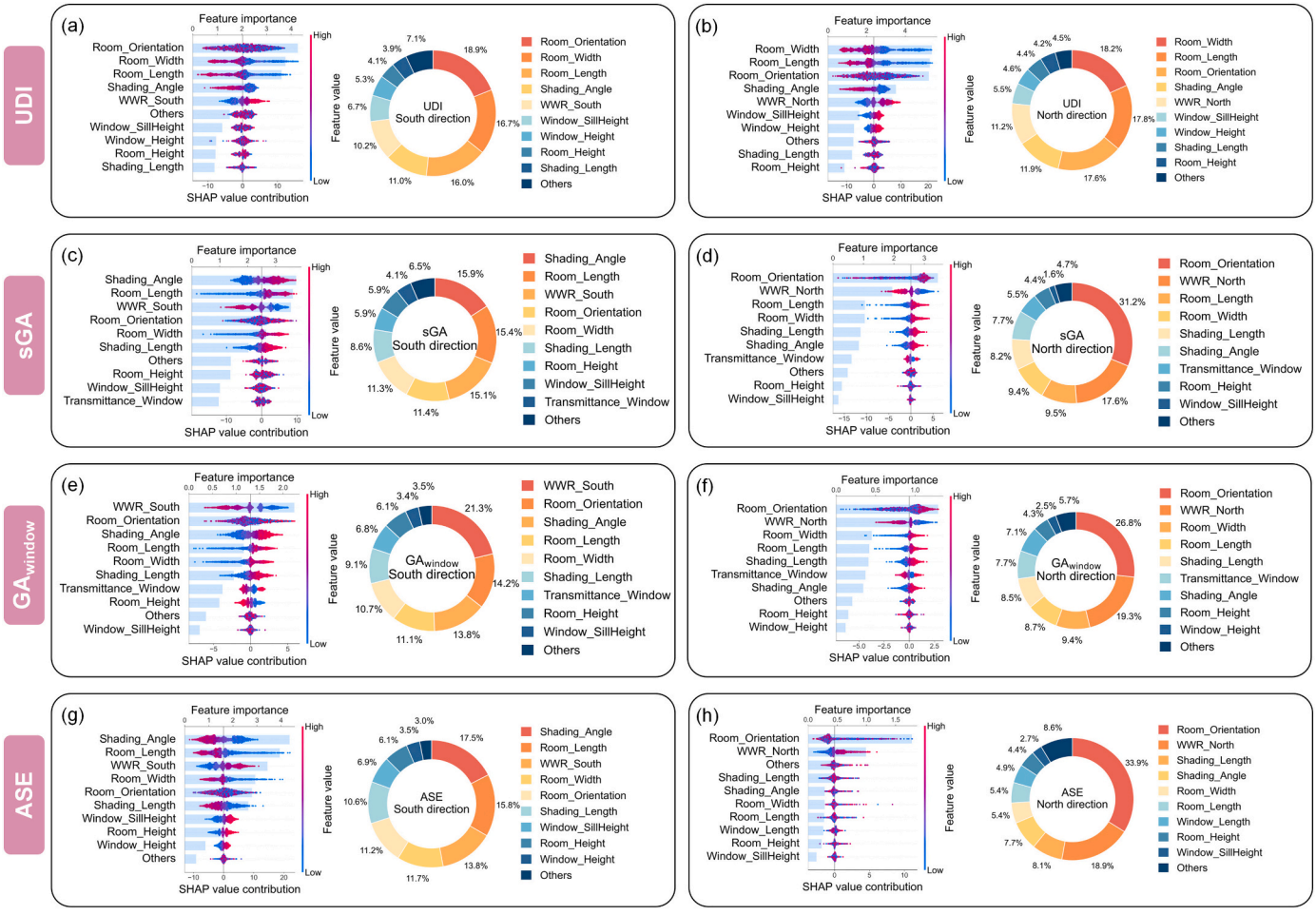


Fig. 10. Global interpretation of SHAP value contributions for the target index in the south and north directions, with corresponding feature importance percentages: (a), (c), (e), (g) represent the south-facing window openings; (b), (d), (f), (h) represent the north-facing window openings.

Table 5
Importance of DVs to relative marginal contributions toward south-facing window openings results.

DVs	UDI index		sGA index		GA_{window} index		ASE index	
	Total (%)	Ranking	Total (%)	Ranking	Total (%)	Ranking	Total (%)	Ranking
Building forms	**55.7**		**44.0**		**41.8**		**44.9**	
Room_Width	16.7	2	11.3	5	10.6	5	11.8	4
Room_Length	16.0	3	15.4	2	11.0	4	15.8	2
Room_Height	4.1	8	5.9	7	5.6	8	6.1	8
Room_Orientation	18.9	1	11.4	4	14.6	2	11.2	5
Fenestration	**22.8**		**25.1**		**26.9**		**24.7**	
WWR_South	10.2	5	15.1	3	22.1	1	13.8	3
Window_Height	5.3	7	3.6	10	0.9	10	3.4	9
Window_SillHeight	6.7	6	5.9	8	3.2	9	6.9	7
Window_Length	0.6	14	0.5	12	0.7	12	0.6	10
Shading device	**14.9**		**24.6**		**22.8**		**28.0**	
Shading_Length	3.9	9	8.7	6	8.9	6	10.6	6
Shading_Angle	11.0	4	15.9	1	13.9	3	17.4	1
Surface material	**6.6**		**6.3**		**8.5**		**2.4**	
Reflectance_Wall	1.6	11	1.2	11	0.9	11	0.6	10
Reflectance_Ceiling	0.5	15	0.3	15	0.2	15	0.6	12
Reflectance_Floor	0.8	13	0.3	14	0.3	14	0.5	13
Transmittance_Window	2.7	10	4.1	9	6.6	7	0.4	14
Reflectance_Shading	1.0	12	0.4	13	0.5	13	0.3	15

Table 6
Importance of DVs to relative marginal contributions toward north-facing window openings results.

DVs	UDI index		sGA index		GA_{window} index		ASE index	
	Total (%)	Ranking	Total (%)	Ranking	Total (%)	Ranking	Total (%)	Ranking
Building forms	**57.8**		**54.5**		**49.3**		**49.0**	
Room_Width	18.2	1	9.4	4	9.4	3	5.3	6
Room_Length	17.8	2	9.5	3	8.7	4	5.3	5
Room_Height	4.2	9	4.4	8	4.4	8	4.4	8
Room_Orientation	17.6	3	31.2	1	26.8	1	34.0	1
Fenestration	**22.0**		**21.5**		**24.2**		**28.9**	
WWR_North	11.2	5	17.7	2	19.3	2	18.9	2
Window_Height	4.6	7	1.1	10	2.5	9	2.3	10
Window_SillHeight	5.5	6	1.6	9	1.2	10	2.7	9
Window_Length	0.7	11	1.1	11	1.2	11	5.0	7
Shading device	**16.3**		**16.0**		**15.6**		**16.0**	
Shading_Length	4.4	8	8.3	5	8.5	5	8.2	3
Shading_Angle	11.9	4	7.7	6	7.1	7	7.8	4
Surface material	**3.9**		**8.0**		**10.9**		**6.1**	
Reflectance_Wall	0.6	12	0.6	13	0.7	14	1.4	11
Reflectance_Ceiling	0.3	15	0.5	14	1.0	12	1.2	13
Reflectance_Floor	0.4	14	0.9	12	0.9	13	1.1	14
Transmittance_Window	2.1	10	5.5	7	7.7	6	1.3	12
Reflectance_Shading	0.5	13	0.5	15	0.6	15	1.1	15

4.3.2. *Local interpretation results*

The local interpretation analysis in Fig. 11 reveals the nonlinear interaction effects between DVs. In the UDI analysis (Fig. 11a-b), WWR is positively correlated with room length and window sill height but negatively correlated with room width. Specifically, when room length exceeds 12 m and WWR $\geq$ 0.6, larger windows improve lighting efficiency by 23–28 %. This improvement is further enhanced when the window sill height is $\geq$0.4 m and dynamic shading is applied. The analysis of sGA and GA_{window} indicates significant differentiation (Fig. 11c-f). Room length and width are positively correlated with shading angle and orientation, suggesting that larger indoor spaces (length and width > 12 m) should prioritize light over glare control. To achieve this, a WWR $\geq$ 0.6 is necessary for both south- and north-facing spaces. For north-facing spaces, increasing room width by 1 m in orientations between −50° to 0° and beyond 50° improves IVC conditions (reducing luminance contrast) by approximately 10 %, even though direct solar glare is generally not a concern in these orientations. The ASE analysis (Fig. 11g-h) shows a significant negative correlation between shading angle and room width, suggesting that shading angles should be between −20° and 20° to reduce light exposure. This effect being particularly in south-facing spaces. The recommended WWR for south- and north-facing spaces is around 0.5 and 0.6, respectively, optimizing room dimensions (up to 10 m) while minimizing glare.

These findings demonstrate the model's ability to identify designs that meet or exceed the acceptable visual comfort threshold. For south-facing spaces, controlling WWR around 0.5 ± 0.1 balances room dimensions (up to 10 m) with daylight (Fig. 11a). Fig. 11c shows that shading angles between 25° and 35° effectively reduce glare and maintain IVC. In north-facing spaces, increasing WWR to 0.6–0.7 results in minimal ASE increases (Fig. 6b), suggesting that glare risk remains low despite larger window areas.

4.4. *Validation of the proposed framework*

This subsection evaluates the adaptability of the proposed framework across diverse climates. This study tested ten typical cities across five climate zones based on the Köppen-Geiger climate classification system [79]: Zone 1: Tropical climates (Singapore, Jakarta); Zone 2: Dry climates (Cairo, Riyadh); Zone 3: Temperate climates (Shanghai, Hamburg); Zone 4: Continental climates (Chicago, Helsinki); Zone 5: Polar climates (Iqaluit, Nuuk).

As shown in Table 7, the Stacking model demonstrates strong cross-

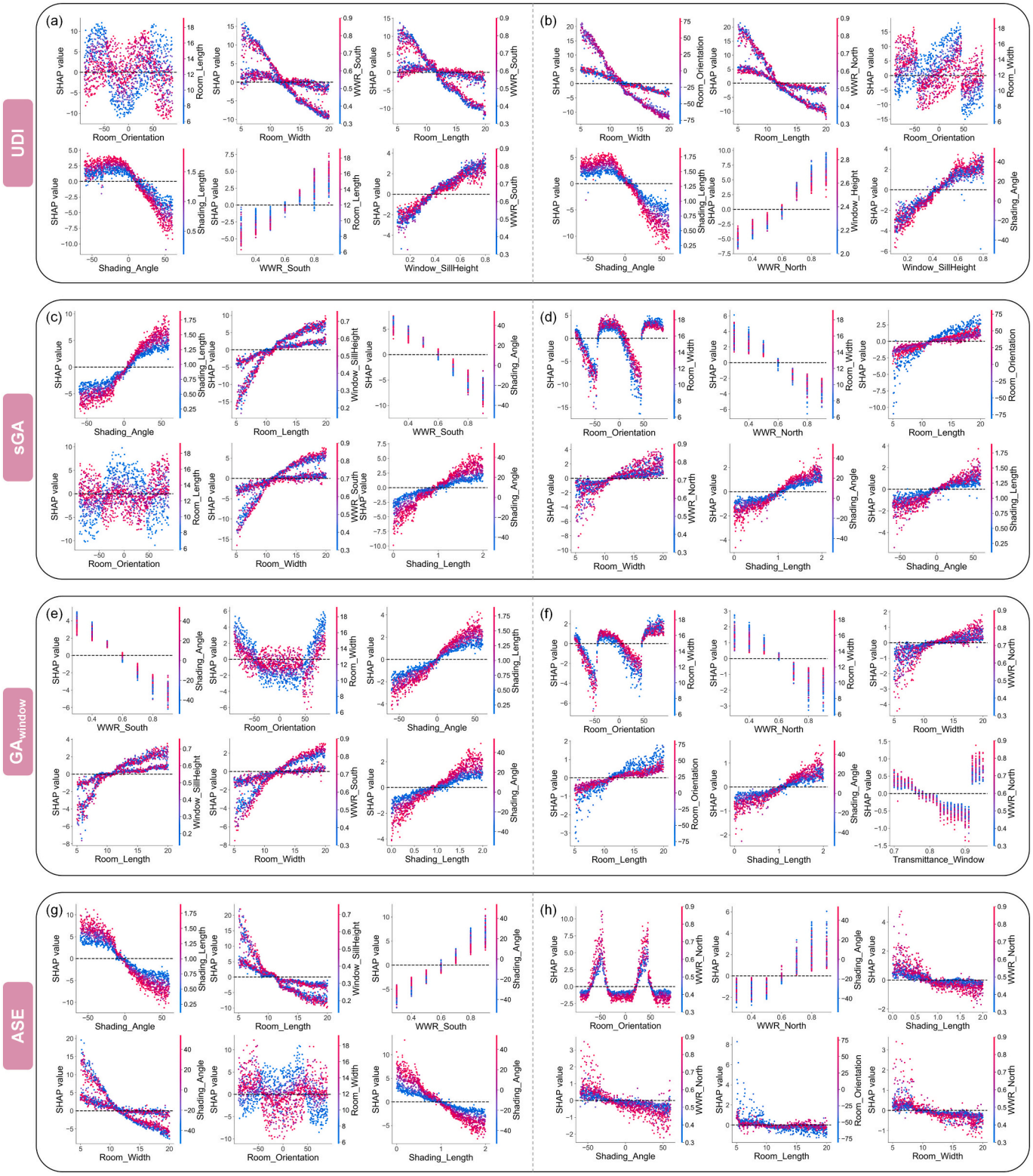


Fig. 11. Local interpretation of top 6 DVs for the target index in the south and north directions: (a), (c), (e), (g) represent the south-facing window openings; (b), (d), (f), (h) represent the north-facing window openings.

climate generalization capability, with an average R^2 of 0.893 across five climate zones. However, errors increased in tropical and highly variable climates, indicating a need for further optimization. Although retraining is required for each city due to local climatic and solar variations, the sampling and training process is efficient and performed only once per location. Once trained, the Stacking model functions as a fast and reusable surrogate, enabling rapid IVC predictions. This substantially reduces computational demands in later design stages, particularly when coupled with optimization algorithms for multi-objective optimization.

Table 7
Framework validation results for ten typical cities across different climate zones based on indoor visual comfort indicators using Stacking models.

Climate zones	Validated cities	Training set			Test set		
		R^2↑	RMSE↓	MAE↓	R^2↑	RMSE↓	MAE↓
Zone 1	Singapore	0.974	1.405	1.058	0.880	3.203	2.303
	Jakarta	0.973	1.217	0.876	0.840	3.031	2.002
Zone 2	Cairo	0.976	1.767	1.365	0.900	3.619	2.506
	Riyadh	0.983	1.490	1.154	0.906	3.631	2.562
Zone 3	Shanghai	0.976	1.616	1.230	0.892	3.462	2.320
	Hamburg	0.980	1.611	1.236	0.899	3.757	2.536
Zone 4	Chicago	0.979	1.601	1.237	0.897	3.592	2.418
	Helsinki	0.982	1.429	1.109	0.912	3.295	2.252
Zone 5	Iqaluit	0.980	1.520	1.160	0.886	3.485	2.337
	Nuuk	0.979	1.573	1.203	0.921	3.054	2.156

5. Discussion

The stacking ensemble architecture model in this study simultaneously predicts both daylight distribution and perceived glare, utilizing SHAP analysis to reveal the nonlinear interactions among DVs. By incorporating the occupant's viewing direction, the framework facilitates a dynamic assessment of IVC at the early stages of design, highlighting the multidimensional pathways through which daylight influences indoor environments. Its high predictive accuracy and interpretability not only enhance efficiency but also provide practical insights for optimizing IVC, thereby supporting the creation of healthier and more sustainable buildings.

5.1. Stacking architecture for indoor visual comfort prediction

The stacking architecture in this study marks several key advancements: (1) By leveraging generative design strategies in conjunction with parametric modeling, the framework effectively generates diverse indoor performance scenarios. A systematic sampling approach of building features creates comprehensive training datasets, addressing the limitations of traditional experience-based models and accelerating the development of predictive models. (2) While Boosting-based models outperformed Bagging-based approaches in this study, the performance differences were relatively modest in some cases. These discrepancies may be attributed to data variability and cross-validation randomness, highlighting the need for cautious interpretation of the results. Moreover, Boosting-based models (XGB, LGBM, CB, AB) consistently outperformed Bagging-based models (RF, ET), particularly in capturing nonlinear feature interactions. This trend aligns with trends observed in previous building performance studies [40,80]. (3) A comparison of performance between south- and north-facing spaces further emphasized distinct daylight dynamics. South-facing rooms, dominated by direct solar exposure, exhibited greater illuminance variability, which enabled the model to capture more nonlinear patterns. In contrast, north-facing spaces, characterized by predominantly diffuse daylight, offered more stable but less responsive lighting conditions. These orientation-dependent dynamics underscore the importance of tailoring modeling strategies to specific building orientations, reinforcing the applicability of Stacking-based ML techniques for a broader range of building performance metrics. In particular, dynamic environmental features should be prioritized for south-facing spaces, while temporal variables—such as time of day and seasonal changes—are crucial for improving model sensitivity in north-facing spaces [41,42]. (4) The simulation-based sampling and model training process are required for each new location. Once trained, the surrogate model provides rapid predictions with minimal computational cost. Seamlessly integrated into the Grasshopper parametric platform, the model enables designers to input basic building parameters—such as room dimensions and window configurations—and receive immediate IVC feedback. This substantially reduces the computational load in downstream design iterations and multi-objective optimization processes.

5.2. Interpretable models for indoor visual comfort evaluation

SHAP analysis was employed in this study to deconstruct the decision mechanisms within the Stacking model, enabling designers to focus on performance-driven solutions and explore design alternatives more effectively. The results show that the CB model contributed the largest weight (75 %) to overall predictions and highlighted six key DVs—room orientation, room length, room width, window-to-wall ratio (WWR), shading length, and shading angle. These key variables are consistent with findings from previous daylight and glare studies [6,27,81,82]. Furthermore, interaction analysis indicated that the combination of WWR and shading device parameters plays a decisive role in shaping IVC across different orientations. Such quantified insights provide a theoretical foundation for optimizing façade configurations and shading systems.

The feature-priority mapping derived from SHAP analysis provides actionable guidance for sustainable building design. The six key DVs collectively accounted for 80.6 % of the total model impact (south-facing: 79.6 %, north-facing: 81.6 %). This suggests that building forms and shading devices should be prioritized during early design stages, while refinements in material selection can follow at later stages. South-facing spaces, especially larger ones, contributed on average 29.5 % more to model predictions than north-facing counterparts, underscoring the importance of dynamic shading to mitigate glare and excessive solar gain. Conversely, WWR had a relatively greater influence in north-facing spaces, contributing 9.8 % more than in south-facing spaces. This finding suggests that larger windows in north-facing rooms can improve diffuse daylight without introducing excessive glare.

Importantly, dynamic shading systems emerged as a key enabler for reconciling daylight provision with IVC. While large windows enhance daylight penetration under overcast conditions, automated shading can effectively reduce glare and overexposure in sunny periods. This dual functionality underscores the potential of dynamic façade systems to balance visual comfort and energy efficiency. Moreover, room length and width were negatively correlated with comfort, whereas WWR showed a positive correlation, reinforcing the need to tailor window and shading strategies to local climatic conditions. Overall, these findings highlight that increasing WWR in north-facing spaces can compensate for limited diffuse daylight, while dynamic shading in south-facing spaces can alleviate excessive solar exposure. Together, these strategies provide designers with a framework for making orientation-sensitive decisions that enhance IVC and promote sustainable design outcomes.

5.3. Limitations

Despite its contributions, this study has several limitations:

One limitation of this study is the potential overlap among metrics that evaluate similar daylight aspects. For instance, both daylight autonomy (DA) and UDI-a reward moderate illuminance levels, while ASE and UDI-e penalize excessive daylight exposure. Although this overlap

may lead to redundancy, the metrics still capture complementary dimensions of daylight quality. The combined use of sDA and ASE, as defined by IES LM-83-12 [61], offers a standardized balance of sufficiency and overexposure, though it may overlook finer spatial variations. Subtle differences across standards, such as EN 17037, which translates sDA/DA thresholds (at least 50 % and 95 % of the time, respectively) into DF targets for different European locations, also need to be considered when interpreting results. Moreover, the glare assessment applied here is limited by its reliance on the GA metric. GA primarily captures glare from direct solar visibility but does not account for reflections or scattering from nearby buildings, floors, or shading devices, thus missing many common glare scenarios. This is a point to consider when interpreting glare-related findings.

A further shortcoming is the reliance on simulation-derived metrics without incorporating occupant-reported comfort data. The absence of field measurements restricts validation against user experience, and future work should integrate post-occupancy surveys and seasonal comfort variations. Extending optimization methods to multi-objective frameworks that balance daylight, thermal comfort, and energy use could further improve applicability [26,33]. This may help designers develop climate-responsive and comfortable indoor spaces [7,36,37,83].

Lastly, while the stacking ensemble improves predictive accuracy, it increases computational cost and architectural complexity. This may hinder adoption by design teams less familiar with ensemble methods. Training time, sensitivity to parameter space, and deployment challenges remain obstacles for design practice. Transfer learning may help reduce computational load and improve scalability across diverse contexts.

6. Conclusions

This paper presented a framework that integrates a Stacking ensemble learning model with SHapley Additive exPlanations (SHAP) to predict and analyze indoor visual comfort (IVC) factors both efficiently and accurately. By following a "parameter input → performance prediction → mechanism explanation" workflow, the proposed framework enables rapid identification of critical design variables (DVs). The framework uncovers nonlinear interactions among DVs, thereby supporting the creation of occupant-centered, visually comfortable, and sustainable indoor environments.

The key contributions are as follows: First, an occupant-centered IVC evaluation mechanism was developed, combining daylight distribution and glare control into a joint "daylight-glare" assessment. Localized glare discomfort near windows was further addressed through sGA and GA_{window} metrics, incorporating occupant viewing direction to refine the performance evaluation framework. Second, a high-performance prediction model was developed. This study proposed a novel stacking architecture integrating XGBoost, LightGBM, and CatBoost. This ensemble achieved an R^2 of 0.911 with a training time of 21.2 s, outperforming both individual learners and traditional simulation-based approaches. Third, an interpretable decision-support system was implemented via SHAP method. This enabled deconstruction of complex DVs interactions and identification of six key variables—room orientation, length, width, window-to-wall-ratio, shading length, and shading angle—that together explain 80.6 % of IVC variance. Notably, building form (46.6 %–52.7 %) and fenestration (22.6 %–24.9 %) were prioritized as dominant design drivers in sustainable building design.

In future work, the following issues still need to be further discussed. Although the framework demonstrated strong performance across five climate zones and ten cities (average $R^2 = 0.893$), its applicability is currently limited to standard office spaces. Future work should test its adaptability to various building types, incorporate occupant behaviors, and consider long-term climate change trends by integrating CMIP6 climate projections. Additionally, the selected combination of daylight and glare indicators can be adjusted based on specific project needs. Expanding the model to support multi-objective optimization—addressing energy consumption, lighting quality, and resilience—would enhance its sustainability assessment capabilities.

CRediT authorship contribution statement

Yuxin Zhou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tomohiro Fukuda:** Writing – review & editing, Supervision, Project administration. **Nobuyoshi Yabuki:** Writing – review & editing, Supervision, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT-4o and ChatGPT-5 tools in order to assist with language editing and refinement of the original manuscript. The primary goal was to improve the readability of the manuscript and to ensure that the language adhered to the conventions of academic English writing. After using these tools, the authors thoroughly reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

Yuxin Zhou reports financial support was provided by The University of Osaka. Tomohiro Fukuda reports a relationship with The University of Osaka that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research team would like to express our gratitude for the support of open-source technologies and the scholars who have contributed to this field of research. Moreover, the authors would like to thank Liang Yuan from the College of Architecture and Urban Planning, Tongji University for discussing this study.

Appendix A. Appendix

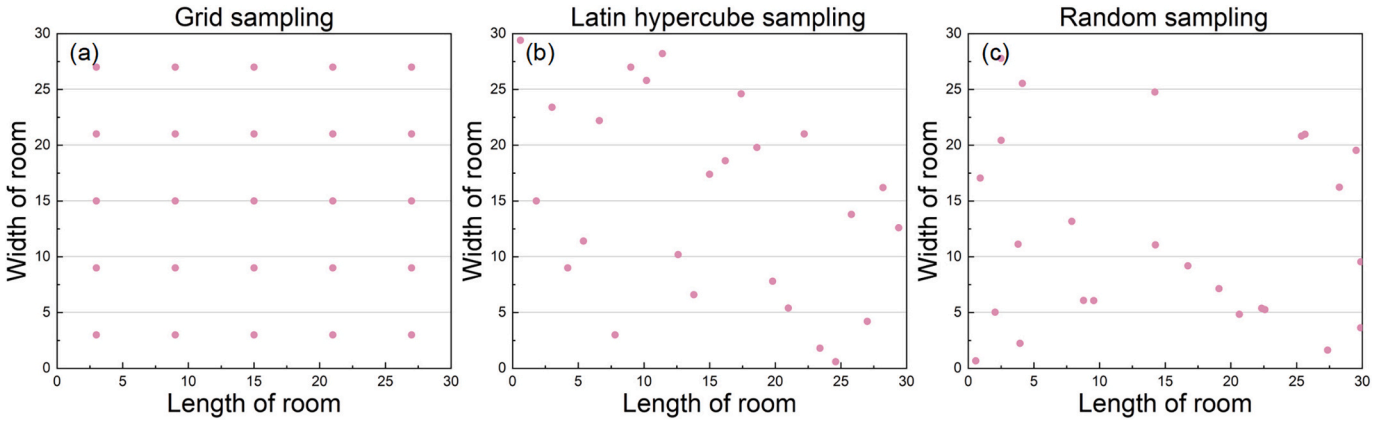


Fig. A.1. Comparison of the grid, Latin hypercube, and random sampling methods. Note: sampling tests were conducted using three common methods: grid sampling, LHS, and random sampling. The experiment used the length and width of room as examples for the sampling analysis. Considering sampling coverage, data representativeness, and computational efficiency, the LHS method was selected to generate the building performance dataset.

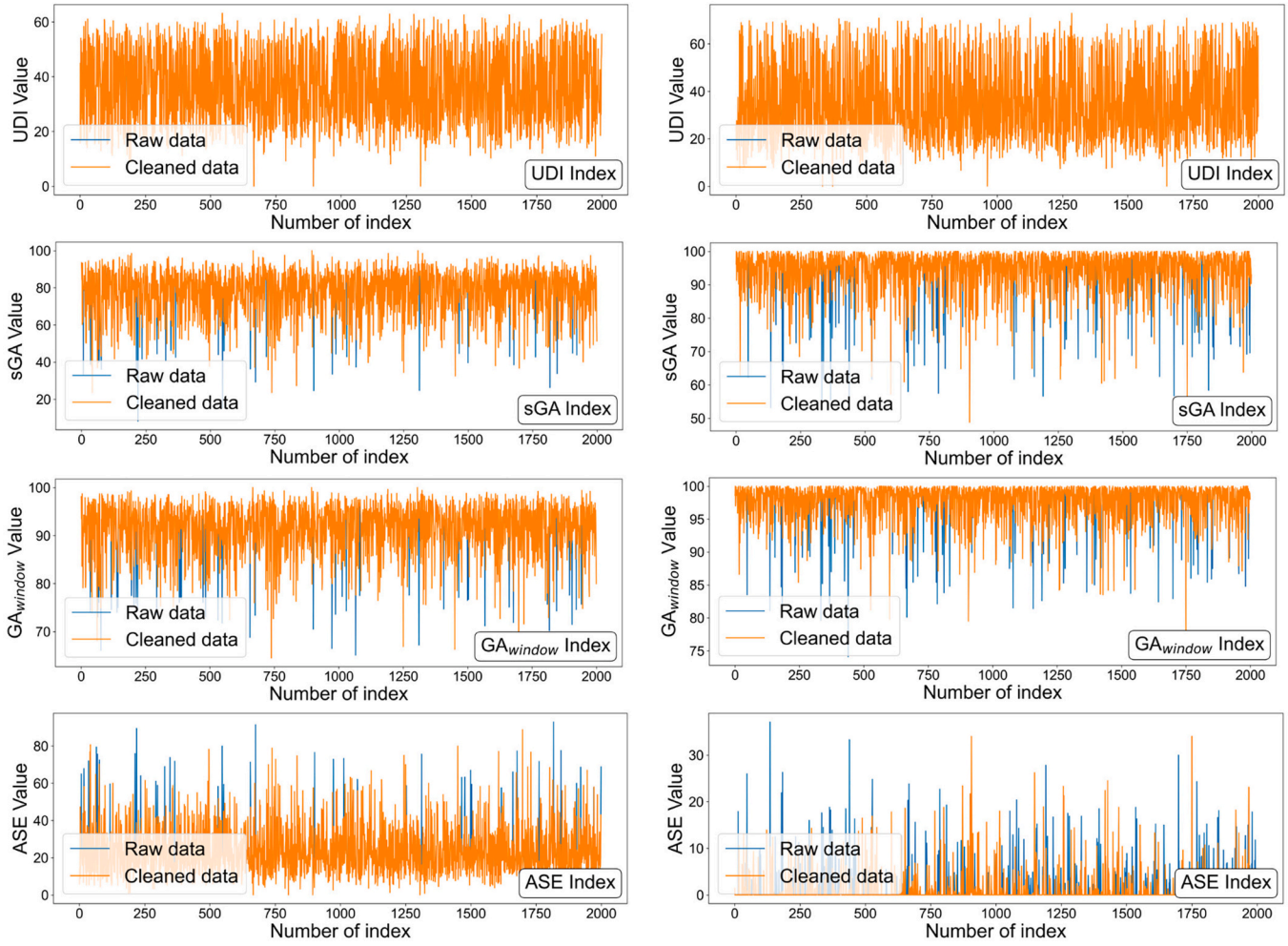


Fig. A.2. Data pre-processing for the IQR method of the target index: south-facing window openings (left) and north-facing window openings (right).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.autcon.2025.106582.

Data availability

To facilitate reproducibility, the complete source code and a sample dataset have been made publicly available at: https://github.com/sggesf/IndoorVisualComfort-Stacking-SHAP.git

References

[1] S. Akbari, M. Sheikhkhoshkar, F.P. Rahimian, H.B. El Haouzi, M. Najafi, S. Talebi, Sustainability and building information modelling: integration, research gaps, and future directions, Autom. Constr. 163 (2024) 105420, https://doi.org/10.1016/j.autcon.2024.105420.

[2] I. Konstantzos, S.A. Sadeghi, M. Kim, J. Xiong, A. Tzempelikos, The effect of lighting environment on task performance in buildings–a review, Energ. Build. 226 (2020) 110394, https://doi.org/10.1016/j.enbuild.2020.110394.

[3] T.M. Brown, G.C. Brainard, C. Cajochen, C.A. Czeisler, J.P. Hanifin, S.W. Lockley, R.J. Lucas, M. Münch, J.B. O'Hagan, S.N. Peirson, Recommendations for daytime, evening, and nighttime indoor light exposure to best support physiology, sleep, and wakefulness in healthy adults, PLoS Biol. 20 (2022) e3001571, https://doi.org/10.1371/journal.pbio.3001571.

[4] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J. V. Behar, S.C. Hern, W.H. Engelmann, The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants, J. Exposure Sci. Environ. Epidemiol. 11 (2001) 231–252, https://doi.org/10.1038/sj.jea.7500165.

[5] A. Razmi, M. Rahbar, M. Bemanian, PCA-ANN integrated NSGA-III framework for dormitory building design optimization: energy efficiency, daylight, and thermal comfort, Appl. Energy 305 (2022) 117828, https://doi.org/10.1016/j.apenergy.2021.117828.

[6] A. Ebrahimi-Moghadam, P. Ildarabadi, K. Aliakbari, F. Fadaee, Sensitivity analysis and multi-objective optimization of energy consumption and thermal comfort by using interior light shelves in residential buildings, Renew. Energy 159 (2020) 736–755, https://doi.org/10.1016/j.renene.2020.05.127.

[7] Y. Pan, Y. Shen, J. Qin, L. Zhang, Deep reinforcement learning for multi-objective optimization in BIM-based green building design, Autom. Constr. 166 (2024) 105598, https://doi.org/10.1016/j.autcon.2024.105598.

[8] M.A. Shameri, M.A. Alghoul, K. Sopian, M.F.M. Zain, O. Elayeb, Perspectives of double skin façade systems in buildings and energy saving, Renew. Sust. Energ. Rev. 15 (2011) 1468–1475, https://doi.org/10.1016/j.rser.2010.10.016.

[9] M. Marzouk, M. ElSharkawy, A. Mahmoud, Optimizing daylight utilization of flat skylights in heritage buildings, J. Adv. Res. 37 (2022) 133–145, https://doi.org/10.1016/j.jare.2021.06.005.

[10] Z. Shirzadnia, A. Goharian, M. Mahdavinejad, Designerly approach to skylight configuration based on daylight performance; Toward a novel optimization process, Energy Build. 286 (2023) 112970, https://doi.org/10.1016/j.enbuild.2023.112970.

[11] K. Konis, A. Gamas, K. Kensek, Passive performance and building form: an optimization framework for early-stage design support, Sol. Energy 125 (2016) 161–179, https://doi.org/10.1016/j.solener.2015.12.020.

[12] H. Yan, K. Yan, G. Ji, Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms, Build. Environ. 218 (2022) 109081, https://doi.org/10.1016/j.buildenv.2022.109081.

[13] M. Beccali, M. Bonomolo, G. Ciulla, V. Lo Brano, Assessment of indoor illuminance and study on best photosensors' position for design and commissioning of daylight linked control systems. A new method based on artificial neural networks, Energy 154 (2018) 466–476, https://doi.org/10.1016/j.energy.2018.04.106.

[14] L. Gradišar, M. Dolenc, R. Klinc, Towards machine learned generative design, Autom. Constr. 159 (2024) 105284, https://doi.org/10.1016/j.autcon.2024.105284.

[15] N.S. Shafavi, Z.S. Zomorodian, M. Tahsildoost, M. Javadi, Occupants visual comfort assessments: a review of field studies and lab experiments, Sol. Energy 208 (2020) 249–274, https://doi.org/10.1016/j.solener.2020.07.058.

[16] S. Carlucci, F. Causone, F. De Rosa, L. Pagliano, A review of indices for assessing visual comfort with a view to their use in optimization processes to support building integrated design, Renew. Sust. Energ. Rev. 47 (2015) 1016–1033, https://doi.org/10.1016/j.rser.2015.03.062.

[17] M. Andersen, J.M. Gagne, S. Kleindienst, Interactive expert support for early stage full-year daylighting design: a user's perspective on Lightsolve, Autom. Constr. 35 (2013) 338–352, https://doi.org/10.1016/j.autcon.2013.05.014.

[18] W.H. Ko, S. Schiavon, G. Brager, B. Levitt, Ventilation, thermal and luminous autonomy metrics for an integrated design process, Build. Environ. 145 (2018) 153–165, https://doi.org/10.1016/j.buildenv.2018.08.038.

[19] J. Wienold, J. Christoffersen, Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras, Energ. Build. 38 (2006) 743–757, https://doi.org/10.1016/j.enbuild.2006.03.017.

[20] N. Jones, Fast Climate-Based Glare Analysis and Spatial Mapping, 2019, https://doi.org/10.26868/25222708.2019.210267.

[21] R. Viula, R. Bokel, M. Tenpierik, Prediction of discomfort from glare from daylight in classrooms, Light. Res. Technol. 55 (2023) 712–729, https://doi.org/10.1177/14771535231173291.

[22] S. Wasilewski, L.O. Grobe, J. Wienold, M. Andersen, Efficient simulation for visual comfort evaluations, Energ. Build. 267 (2022) 112141, https://doi.org/10.1016/j.enbuild.2022.112141.

[23] Y. Ji, W. Wang, Y. He, L. Li, H. Zhang, T. Zhang, Performance in generation: an automatic generalizable generative-design-based performance optimization framework for sustainable building design, Energ. Build. 298 (2023) 113512, https://doi.org/10.1016/j.enbuild.2023.113512.

[24] Y. Fang, S. Cho, Design optimization of building geometry and fenestration for daylighting and energy performance, Sol. Energy 191 (2019) 7–18, https://doi.org/10.1016/j.solener.2019.08.039.

[25] H. Mo, Y. Zhou, Y. Song, Parametric design and spatial optimization of east–west-oriented teaching spaces in Shanghai, Buildings 12 (2022) 1333, https://doi.org/10.3390/buildings12091333.

[26] D. Yang, S. Ren, M. Turrin, S. Sariyildiz, Y. Sun, Multi-disciplinary and multi-objective optimization problem re-formulation in computational design exploration: a case of conceptual sports building design, Autom. Constr. 92 (2018) 242–269, https://doi.org/10.1016/j.autcon.2018.03.023.

[27] M. Marzouk, A. Eissa, M. ElSharkawy, Influence of light redirecting control element on daylight performance: a case of Egyptian heritage palace skylight, J. Build. Eng. 31 (2020) 101309, https://doi.org/10.1016/j.jobe.2020.101309.

[28] S. Lee, K.S. Lee, A study on the improvement of the evaluation scale of discomfort glare in educational facilities, Energies 12 (2019) 3265, https://doi.org/10.3390/en12173265.

[29] P. Bakmohammadi, E. Noorzai, Optimization of the design of the primary school classrooms in terms of energy and daylight performance considering occupants' thermal and visual comfort, Energy Rep. 6 (2020) 1590–1607, https://doi.org/10.1016/j.egyr.2020.06.008.

[30] D.L.R. de Garcia, F.O.R. Pereira, Method application and analyses of visual and thermal-energy performance prediction in offices buildings with internal shading devices, Build. Environ. 198 (2021) 107912, https://doi.org/10.1016/j.buildenv.2021.107912.

[31] M. Valitabar, A. GhaffarianHoseini, A. GhaffarianHoseini, S. Attia, Advanced control strategy to maximize view and control discomforting glare: a complex adaptive façade, Archit. Eng. Des. Manag. 18 (2022) 829–849, https://doi.org/10.1080/17452007.2022.2032576.

[32] E.T. Bailey, L. Caldas, Operative generative design using non-dominated sorting genetic algorithm II (NSGA-II), Autom. Constr. 155 (2023) 105026, https://doi.org/10.1016/j.autcon.2023.105026.

[33] C. Wu, H. Pan, Z. Luo, C. Liu, H. Huang, Multi-objective optimization of residential building energy consumption, daylighting, and thermal comfort based on BO-XGBoost-NSGA-II, Build. Environ. 254 (2024) 111386, https://doi.org/10.1016/j.buildenv.2024.111386.

[34] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. 14 (2020) 241–258, https://doi.org/10.1007/s11704-019-8208-z.

[35] S. Stevanović, H. Dashti, M. Milošević, S. Al-Yakoob, D. Stevanović, Comparison of ANN and XGBoost surrogate models trained on small numbers of building energy simulations, PLoS ONE 19 (2024) e0312573, https://doi.org/10.1371/journal.pone.0312573.

[36] H. Kim, G. Lee, H. Ahn, B. Choi, Interpretable general thermal comfort model based on physiological data from wearable bio sensors: light gradient boosting machine (LightGBM) and SHapley additive exPlanations (SHAP), Build. Environ. 266 (2024) 112127, https://doi.org/10.1016/j.buildenv.2024.112127.

[37] M. Haghirad, S. Heidari, H. Hosseini, Advancing personal thermal comfort prediction: a data-driven framework integrating environmental and occupant dynamics using machine learning, Build. Environ. 262 (2024) 111799, https://doi.org/10.1016/j.buildenv.2024.111799.

[38] M. Deng, X. Wang, C.C. Menassa, Measurement and prediction of work engagement under different indoor lighting conditions using physiological sensing, Build. Environ. 203 (2021) 108098, https://doi.org/10.1016/j.buildenv.2021.108098.

[39] I.D. Mienye, Y. Sun, A survey of ensemble learning: concepts, algorithms, applications, and prospects, IEEE Access 10 (2022) 99129–99149, https://doi.org/10.1109/ACCESS.2022.3207287.

[40] J. Lee, W. Wang, F. Harrou, Y. Sun, Reliable solar irradiance prediction using ensemble learning-based models: a comparative study, Energy Convers. Manag. 208 (2020) 112582, https://doi.org/10.1016/j.enconman.2020.112582.

[41] I.K. Nti, A.F. Adekoya, B.A. Weyori, A comprehensive evaluation of ensemble learning for stock-market prediction, J. Big Data 7 (2020) 20, https://doi.org/10.1186/s40537-020-00299-5.

[42] F. Li, H. Zheng, X. Li, F. Yang, Day-ahead city natural gas load forecasting based on decomposition-fusion technique and diversified ensemble learning model, Appl. Energy 303 (2021) 117623, https://doi.org/10.1016/j.apenergy.2021.117623.

[43] H. Park, D.Y. Park, B. Noh, S. Chang, Stacking deep transfer learning for short-term cross building energy prediction with different seasonality and occupant schedule, Build. Environ. 218 (2022) 109060, https://doi.org/10.1016/j.buildenv.2022.109060.

[44] S.R. Hong, J. Hullman, E. Bertini, Human factors in model interpretability: industry practices, challenges, and needs, Proc. ACM Hum.-Comput. Interact. 4 (2020) 1–26, https://doi.org/10.1145/3392878.

[45] E. Taveres-Cachat, G. Lobaccaro, F. Goia, G. Chaudhary, A methodology to improve the performance of PV integrated shading devices using multi-objective optimization, Appl. Energy 247 (2019) 731–744, https://doi.org/10.1016/j.apenergy.2019.04.033.

[46] D. Gaspar, P. Silva, C. Silva, Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron, IEEE Access 12 (2024) 30164–30175, https://doi.org/10.1109/ACCESS.2024.3368377.

[47] Y. Zhou, S. Booth, M.T. Ribeiro, J. Shah, Do feature attribution methods correctly attribute features? Proc. AAAI Conf. Artif. Intell. 36 (2022) 9623–9633, https://doi.org/10.1609/aaai.v36i9.21196.

[48] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc, 2017, https://doi.org/10.48550/arXiv.1705.07874.
[49] H. Chen, S.M. Lundberg, S.-I. Lee, Explaining a series of models by propagating Shapley values, Nat. Commun. 13 (2022) 4512, https://doi.org/10.1038/s41467-022-31384-3.
[50] A. Wojtuch, R. Jankowski, S. Podlewska, How can SHAP values help to shape metabolic stability of chemical compounds? J. Chemother. 13 (2021) 74, https://doi.org/10.1186/s13321-021-00542-y.
[51] S. Saremi, S. Mirjalili, A. Lewis, Grasshopper optimisation algorithm: theory and application, Adv. Eng. Softw. 105 (2017) 30–47, https://doi.org/10.1016/j.advengsoft.2017.01.004.
[52] T.G. Farr, P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf, The shuttle radar topography mission, Rev. Geophys. 45 (2007) 2005RG000183, https://doi.org/10.1029/2005RG000183.
[53] OpenStreetMap, OpenStreetMap. https://www.openstreetmap.org/, 2025 accessed December 8, 2024.
[54] L.K. Lawrie, D.B. Crawley, Development of Global Typical Meteorological Years (TMYx). https://climate.onebuilding.org, 2022.
[55] M. Sadeghipour Roudsari, M. Pak, A. Viola, Ladybug: A Parametric Environmental Plugin for Grasshopper to Help Designers Create an Environmentally-Conscious Design, 2013, https://doi.org/10.26868/25222708.2013.2499.
[56] G.J. Ward, The RADIANCE Lighting Simulation and Rendering System, in: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '94, ACM Press, 1994, pp. 459–472, https://doi.org/10.1145/192161.192286.
[57] S. Murakami, K. Iwamura, R.J. Cole, A Decade of Development and Application of an Environmental Assessment System for the Built Environment. https://www.ibecs.or.jp/CASBEE/english/download/CASBEE%20-%20A%20Decade%20of%20Development%20and%20Application%20of%20an%20Environmental%20Assessment%20System%20for%20the%20Built%20Environment%20-.pdf, 2025 accessed August 19, 2025.
[58] LEED, v4.1 | U.S. Green Building Council. https://www.usgbc.org/leed/v41, 2025 accessed May 11, 2025.
[59] H. Yan, G. Ji, K. Yan, Data-driven prediction and optimization of residential building performance in Singapore considering the impact of climate change, Build. Environ. 226 (2022) 109735, https://doi.org/10.1016/j.buildenv.2022.109735.
[60] A. Nabil, J. Mardaljevic, Useful daylight illuminances: a replacement for daylight factors, Energ. Build. 38 (2006) 905–913, https://doi.org/10.1016/j.enbuild.2006.03.013.
[61] I. Lm, Approved Method: IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE), Illum Eng Soc Https//Www IES Org/Product/Ies-Spatial-Daylight-Autonomy-Sda-and-Annual-Sunlight-Exposure-ASE. https://webstore.ansi.org/preview-pages/IESNA/preview_IES+LM-83-12.pdf, 2013.
[62] Annual Glare — ClimateStudio Latest Documentation. https://climatestudiodocs.com/docs/annualGlare.html, 2025 accessed May 9, 2025.
[63] M. Lesinski, A. Schmelcher, M. Herz, C. Puta, H. Gabriel, A. Arampatzis, G. Laube, D. Büsch, U. Granacher, Maturation-, age-, and sex-specific anthropometric and physical fitness percentiles of German elite young athletes, PLoS ONE 15 (2020) e0237423, https://doi.org/10.1371/journal.pone.0237423.
[64] Z. Skomina, M. Verdenik, N.I. Hren, Effect of aging and body characteristics on facial sexual dimorphism in the Caucasian population, PLoS ONE 15 (2020) e0231983, https://doi.org/10.1371/journal.pone.0231983.
[65] T. Østergård, R.L. Jensen, F.S. Mikkelsen, The best way to perform building simulations? One-at-a-time optimization vs. Monte Carlo sampling, Energy Build. 208 (2020) 109628, https://doi.org/10.1016/j.enbuild.2019.109628.
[66] I. Iordanis, C. Koukouvinos, I. Silou, Regression analysis in machine learning using conditioned latin hypercube sampling, in: 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), 2023, pp. 1–6, https://doi.org/10.1109/DESSERT61349.2023.10416433.
[67] D.D. Saurette, A. Biswas, A.W. Gillespie, Determining minimum sample size for the conditioned Latin hypercube sampling algorithm, Pedosphere 34 (2024) 530–539, https://doi.org/10.1016/j.pedsph.2022.09.001.
[68] M. Vořechovský, Hierarchical refinement of Latin hypercube samples, Comput. Aided Civil Eng. 30 (2015) 394–411, https://doi.org/10.1111/mice.12088.
[69] M. Rahbar, M. Mahdavinejad, M. Bemanian, A.H. Davaie Markazi, L. Hovestadt, Generating synthetic space allocation probability layouts based on trained conditional-GANs, Appl. Artif. Intell. 33 (2019) 689–705, https://doi.org/10.1080/08839514.2019.1592919.
[70] scikit-Learn: Machine Learning in Python — Scikit-Learn 1.7.1 Documentation. https://scikit-learn.org/stable/, 2025 accessed August 19, 2025.
[71] X. Wan, W. Wang, J. Liu, T. Tong, Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range, BMC Med. Res. Methodol. 14 (2014) 135, https://doi.org/10.1186/1471-2288-14-135.
[72] B. Pavlyshenko, Using stacking approaches for machine learning models, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), IEEE, Lviv, 2018, pp. 255–258, https://doi.org/10.1109/DSMP.2018.8478522.
[73] B.J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J.I.M. Alvarado, J.M. Janey, R.P. Adams, A.G. Doyle, Bayesian reaction optimization as a tool for chemical synthesis, Nature 590 (2021) 89–96, https://doi.org/10.1038/s41586-021-03213-y.
[74] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, S.-H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (2019) 26–40, https://doi.org/10.11989/JEST.1674-862X.80904120.
[75] T.O. Hodson, Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not, Geosci. Model Dev. Discuss. 2022 (2022) 1–10, https://doi.org/10.5194/gmd-15-5481-2022.
[76] L. Plonsky, H. Ghanbar, Multiple regression in L2 research: a methodological synthesis and guide to interpreting R^2 values, Mod. Lang. J. 102 (2018) 713–731, https://doi.org/10.1111/modl.12509.
[77] O. Rainio, J. Teuho, R. Klén, Evaluation metrics and statistical tests for machine learning, Sci. Rep. 14 (2024) 6086, https://doi.org/10.1038/s41598-024-56706-x.
[78] L.S. Shapley, 17. A value for n-person games, in: H.W. Kuhn, A.W. Tucker (Eds.), Contributions to the Theory of Games, Volume II, Princeton University Press, 2016: pp. 307–318. doi:https://doi.org/10.1515/9781400881970-018.
[79] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. 11 (2007) 1633–1644, https://doi.org/10.5194/hess-11-1633-2007.
[80] A. Rysanek, R. Nuttall, J. McCarty, Forecasting the impact of climate change on thermal comfort using a weighted ensemble of supervised learning models, Build. Environ. 190 (2021) 107522, https://doi.org/10.1016/j.buildenv.2020.107522.
[81] R. Singh, I.J. Lazarus, V.V.N. Kishore, Uncertainty and sensitivity analyses of energy and visual performances of office building with external venetian blind shading in hot-dry climate, Appl. Energy 184 (2016) 155–170, https://doi.org/10.1016/j.apenergy.2016.10.007.
[82] R.A. Mangkuto, M. Rohmah, A.D. Asri, Design optimisation for window size, orientation, and wall reflectance with regard to various daylight metrics and lighting energy demand: a case study of buildings in the tropics, Appl. Energy 164 (2016) 211–219, https://doi.org/10.1016/j.apenergy.2015.11.046.
[83] M. Kim, B. Kim, J. Koh, H. Yi, Flexural biomimetic responsive building façade using a hybrid soft robot actuator and fabric membrane, Autom. Constr. 145 (2023) 104660, https://doi.org/10.1016/j.autcon.2022.104660.