MTL-Light: An Explainable Chained Multi-Task Learning Framework for Rapid Daylighting Performance Prediction in Office Units

Gaoyang Liu , Yuting Chen , Yue Zeng *

Posted Date: 23 April 2026

doi: 10.20944/preprints202604.1668.v1

Keywords: daylighting performance; multi-task learning; explainable AI; SHAP; office units; surrogate modeling

 doi:10.20944/preprints202604.1668.v1

Article

MTL-Light: An Explainable Chained Multi-Task Learning Framework for Rapid Daylighting Performance Prediction in Office Units

Gaoyang Liu [1], Yuting Chen [2] and Yue Zeng [1,*]

1 Institute of Artificial Intelligence, Shaoxing University, Shaoxing 312000, China

2 School of Civil Engineering, Shaoxing University, Shaoxing 312000, China

* Correspondence: carsontseng@usx.edu.cn

Abstract

Accurate evaluation of indoor daylighting performance is essential for improving visual comfort and reducing lighting energy use in office buildings. However, simulation-based daylighting analysis is often too time-consuming to support rapid comparison of multiple design options in early-stage design. To address this issue, this study proposes MTL-Light, an explainable chained multi-task learning framework for fast daylighting performance prediction in typical office units. A parametric simulation dataset was constructed, and multiple representative daylighting indicators were extracted from the spatial distribution of daylight factors on the work plane. MTL-Light was then developed to jointly predict these indicators by modeling their interdependencies within a lightweight multi-task learning architecture. In addition, SHAP was employed to interpret the prediction results by quantifying the marginal contributions of geometric design variables. The results show that, compared with single-task models, MTL-Light achieves higher accuracy and more stable performance across multiple indicators, particularly for metrics sensitive to spatial distribution. Moreover, it reduces daylighting evaluation from minute-level simulation to millisecond-level inference. The interpretability analysis further indicates that room depth and window geometry dominate daylighting performance, while different indicators exhibit different sensitivities to geometric variables.

Keywords: daylighting performance; multi-task learning; explainable AI; SHAP; office units; surrogate modeling

1. Introduction

Office buildings are a major category of modern public buildings, with persistently high operational energy consumption—lighting typically accounts for a substantial proportion[1]. Enhancing natural light utilization reduces artificial lighting demand, cutting operational energy use while improving visual comfort [2]: natural light offers higher color rendering, a softer illumination environment, reduced visual fatigue, and improved work efficiency, aligning better with human physiological rhythms and health needs [3]. Reliable evaluation and rapid feedback of indoor daylighting performance are therefore of significant engineering value in office building design and optimization.

Indoor daylighting performance is currently evaluated via two primary approaches: on-site measurement and computational simulation. On-site measurement yields highly reliable data in real or physical model environments but requires experimental space construction or access to existing buildings, coupled with long-term illuminance monitoring—this is costly, time-consuming, and ill-suited for systematic comparison of multiple schemes and parameter combinations [4]. In contrast, computational daylighting simulation, valued for its flexibility and lower direct economic cost, has become the mainstream method in academic research and engineering practice [5], enabling

 doi:10.20944/preprints202604.1668.v1

quantitative prediction of indoor lighting environments under defined climate and sky model conditions.

Yet, amid the emphasis on rapid iteration and multi-scheme exploration in early design stages, traditional simulation methods face notable limitations [4–6]. First, setting up geometric models, material parameters, and boundary conditions is labor-intensive; high-resolution or year-round simulations incur substantial computational costs and long runtimes [7], leaving conventional workflows unable to provide rapid feedback during conceptual design [8,9]. Second, daylighting simulation demands expertise in building physics and proficiency with specialized software, making it challenging for non-experts to apply in early design—operational complexity can divert attention from daylighting design decisions to software operations [10,11]. Additionally, indoor daylighting performance depends on room geometry, window size and position, surface reflectance, and external obstructions, introducing high-dimensional nonlinear interactions among variables [12]. Consequently, exhaustive parametric simulations thus become intractable due to combinatorial growth in cases, hindering sensitivity analysis and efficient scheme exploration in practice [13,14]. These constraints relegate natural daylighting analysis to post-project verification in many cases, limiting its role in design-phase decision support.

With the widespread adoption of machine learning (ML) [15] in architectural design, ML-based prediction of indoor daylighting in office buildings offers distinct advantages over simulation-only methods[16]. Data-driven ML surrogates learn the relationship between design parameters and daylighting performance from simulated or measured datasets, enabling near-real-time evaluation of new schemes at minimal computational cost and supporting rapid alternative comparisons. Kumar et al. [17] constructed an ML model using laboratory-measured data, fusing multi-source information from sensors, digital cameras, and EVALGLARE software to improve the robustness of daylighting and glare prediction—focusing on validating ML applicability to real-time measurement data under specific spatial conditions. Ayoub [18] reviewed ML applications for predicting indoor daylighting performance, examining how building typology and climate regions influence predictive performance and highlighting the importance of data quality and feature selection for accuracy and practical relevance in design decision-making. Wagdy et al. [19] proposed an ML method integrating high dynamic range (HDR) images and user subjective feedback to improve glare prediction accuracy in open-plan offices, overcoming limitations of traditional glare assessment. However, this model was trained on only 80 survey data points and is primarily applicable to open-plan offices with low overall illuminance. Radziszewski and Waczyński [20] developed an artificial neural network-based tool, comparing 2763 traditional simulation results with ML predictions—finding that ML models drastically reduce computation time while maintaining small discrepancies with traditional simulations, validating their reliability for early building daylighting assessment.

Despite these advances, existing ML-based daylighting models still suffer from two critical limitations. First, most approaches treat different daylighting indicators as independent prediction tasks, ignoring their inherent physical coupling. Second, the majority of models remain black-box systems, limiting their ability to provide interpretable and actionable design insights [21]. To address these gaps, this study proposes an explainable chained multi-task learning framework that explicitly models inter-task dependencies while enhancing interpretability. SHapley Additive exPlanations(SHAP) [22], a post-hoc interpretability method quantifying the contribution of each input variable to predictions, has been widely applied in architecture and structural engineering to improve model transparency and engineering interpretability of results. Mashaly et al. [23] developed a neural network-based tool for predicting daylight performance (SDA) in early design stages, using large-scale climate-based daylight simulation data to enable rapid, globally applicable SDA and ASE predictions. Ngarambe et al. [24] systematically compared various ML algorithms for indoor illuminance prediction and found that, although high prediction accuracy can be achieved, the lack of interpretability remains a key limitation for practical engineering applications. By integrating SHAP into the proposed framework, the influence of geometric design parameters on

 doi:10.20944/preprints202604.1668.v1

different daylighting tasks can be explicitly quantified, thereby providing a transparent and reliable basis for design optimization.

Against this background, an interpretable multi-task learning framework is proposed for daylighting performance prediction in office buildings. First, a parametric simulation dataset is constructed under unified climatic and boundary conditions using Ecotect. Room depth and window dimensions are selected as key variables, and four tasks, namely Mean, Min, Max, and DF_{depth}, are defined to characterize daylighting performance. Second, a lightweight chained multi-task learning framework is developed, in which predictions from preceding tasks are incorporated as additional features to capture inter-task dependencies. Finally, SHAP is employed for global and local interpretability analysis, revealing the influence of geometric parameters and intermediate predictions and enabling actionable design insights.

2. Methodology

The methodology consists of three core components: parametric daylighting simulation, multi-task prediction, and interpretability analysis. This work aims to achieve efficient modeling and pattern identification of the natural daylighting performance of office buildings while ensuring the consistency of physical simulation. First, a parametric daylighting simulation dataset was constructed based on Autodesk Ecotect Analysis to characterize the daylighting response features of typical office units under different geometric configurations. Second, a lightweight chained multi-task learning framework (MTL-Light) was constructed to jointly predict multiple daylighting tasks, taking into account the multi-task output characteristics of the spatial distribution of daylighting coefficients. Finally, the SHAP interpretability analysis method was introduced to systematically interpret the model prediction results, revealing the influence mechanism of geometric design parameters on different daylighting tasks. The overall framework is shown in Figure 1.

 doi:10.20944/preprints202604.1668.v1

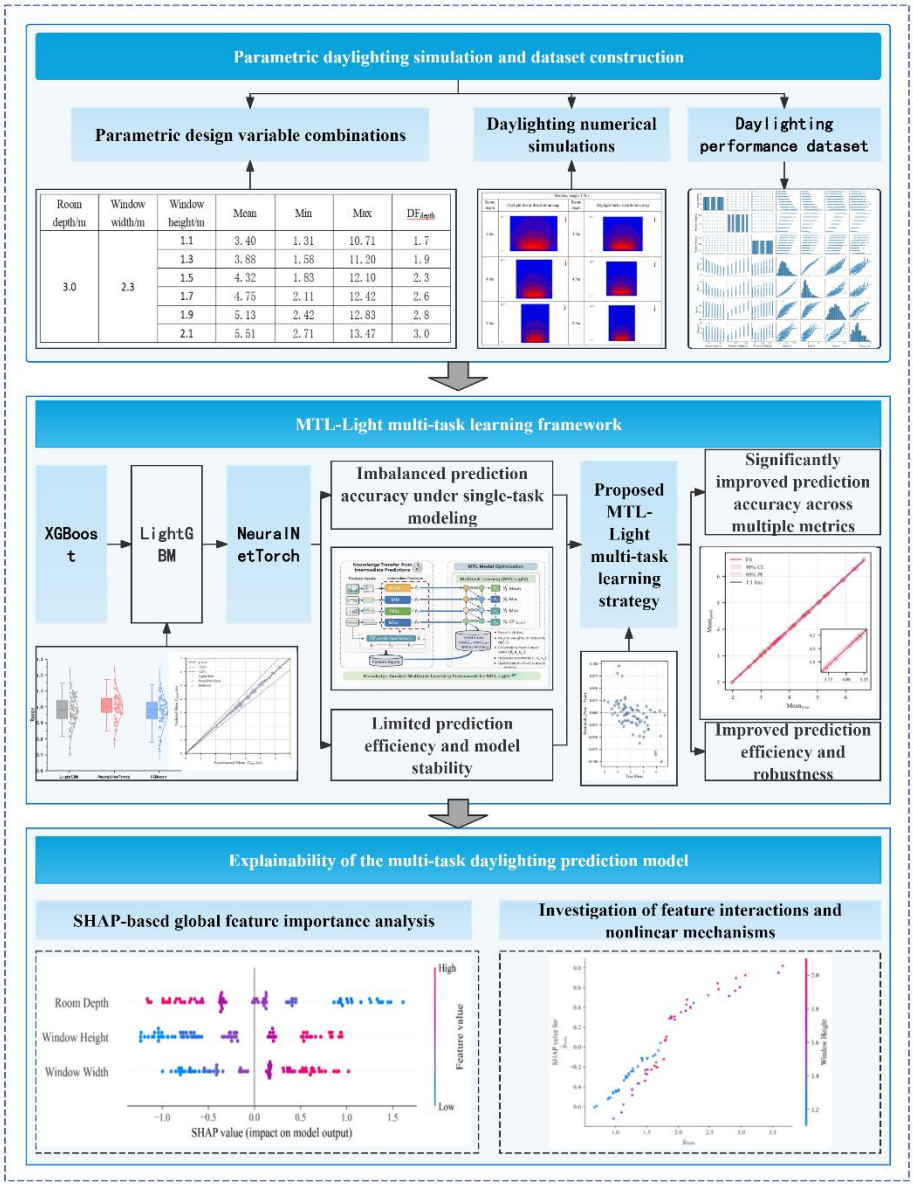


Figure 1. The framework of the proposed explainable Multi-Task Learning for daylighting

2.1 Simulation of Indoor Natural Daylighting in Office Buildings

Natural daylighting performance is an important consideration in office building design because it directly affects indoor visual comfort and lighting energy demand[25]. To quantitatively characterize daylighting behavior under different geometric configurations, the solar radiation and daylighting simulation module of Autodesk Ecotect Analysis was employed to perform parametric simulations and obtain the spatial distribution of daylight factor (DF) in typical office units[26].

All simulation cases assumed single-sided daylighting. External obstructions, including trees, shading devices, and surrounding buildings, were neglected, and no interior curtains or other shading elements were considered. The height of the working plane was set to 0.75 m. To ensure consistency with the Chinese Standard for Daylighting Design of Buildings (GB50033-2013)[27], the indoor surface reflectance were specified using intermediate recommended values, namely 0.75 for the ceiling, 0.50 for the walls, and 0.30 for the floor, while the glass transmittance was set to 0.62. The simulated building was assumed to be located in Hangzhou, Zhejiang Province, which belongs to daylight climate zone IV according to the standard. The CIE overcast sky model was adopted for daylighting evaluation.

 doi:10.20944/preprints202604.1668.v1

According to the minimum daylighting requirement for office buildings in GB50033-2013[27], and considering the regional correction coefficient $K = 1.1$, a minimum DF threshold of 3.3% was used as the benchmark for compliance assessment. Under these unified settings, indoor daylighting performance was characterized by four output indicators: the average daylight factor (Mean), the minimum daylight factor (Min), the maximum daylight factor (Max), and the daylighting-compliance depth, denoted asDF_{depth}. Here, $DF_{depth}$was defined as the maximum depth position at which the daylight factor first satisfied the 3.3% minimum threshold along the indoor spatial distribution. These four indicators were used as the target variables for subsequent predictive modeling and interpretability analysis.

2.2 Machine Learning Model Construction and Training

Data-driven surrogate modeling enables rapid approximation of the relationship between design parameters and daylighting performance, significantly reducing the computational cost of traditional simulations. By learning from simulated or measured data, machine learning (ML) models can efficiently predict daylighting performance for unseen parameter combinations, while supporting feature contribution analysis and interaction exploration for parametric design. To this end, a parametric daylighting simulation dataset for typical office units is constructed, and an interpretable prediction model is developed to enable rapid performance evaluation and sensitivity analysis in early design stages.

After comparing various single-task regression methods, a unified modeling process is established for multi-task natural daylighting prediction, adopting two task strategies: a baseline approach that outputs four tasks (Mean, Min, Max, DF_{depth}) in parallel within a single model; and a chain-based multi-task strategy, defined herein as the MTL-Light framework. Both strategies use room geometry and window size parameters as input features (e.g., D, B, Hw, and their derivatives) and four daylighting evaluation tasks as supervision signals, differing only in task coupling. The Baseline method directly learns the mapping from input x to multi-dimensional output $\hat{y} = [\hat{y}_{Mean}, \hat{y}_{Min}, \hat{y}_{Max}, \hat{y}_{DF_{depth}}]$, featuring a simple structure and training process suitable as a unified comparison benchmark. In contrast, MTL-Light explicitly injects inter-task dependencies via "predictive output featureization": first, $\hat{y}_{Mean} = f_1(x)$; $\hat{y}_{Mean}$ is then concatenated with original features to form an extended input for predicting $\hat{y}_{Min} = f_2(x, \hat{y}_{Mean})$; further, $\hat{y}_{Mean}, \hat{y}_{Min}$ are combined with original features to predict $\hat{y}_{Max}$; finally, joint information from $\hat{y}_{Mean}, \hat{y}_{Min}, \hat{y}_{Max}$, and original features is used to predict $\hat{y}_{DF_{depth}}$during training. The term “Light” indicates that inter-task coupling is achieved through hierarchical recursion and feature augmentation, without introducing complex shared encoders or explicit cross-task interaction modules. This design improves the learnability and prediction stability of extreme-value and threshold-based tasks, while maintaining low implementation complexity.

Model performance is evaluated using the coefficient of determination (R^2), mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE)—capturing average bias levels and sensitivity to large error samples, respectively. For a test set with n samples, let y_I and $\hat{y}_I$ denote the true and predicted values of the k output. The metrics are defined as follows:

1) MAE（Mean Absolute Error）

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (1)$$

2) RMSE（Root Mean Squared Error）

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

3) R^2（Coefficient of Determination）

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (3)$$

4) MAPE（Mean Absolute Percentage Error）

 doi:10.20944/preprints202604.1668.v1

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (4)$$

The performance of Baseline and MTL-Light are compared across the four tasks under identical data partitioning and training configurations, evaluating the gains from chain dependency injection for multi-task lighting prediction accuracy and stability.

2.2.1 Baseline Models

To evaluate the predictive performance of different modeling strategies for daylighting prediction, three regression models were adopted as baseline methods: LightGBM [28], XGBoost [29], and NeuralNetTorch [30]. LightGBM and XGBoost, as tree-based ensemble algorithms, are effective in modeling nonlinear relationships and feature interactions in tabular data, whereas NeuralNetTorch provides a representative neural network-based approach.

All models were implemented using an identical dataset, feature set, preprocessing procedure, and an 8:2 train–test split. Hyperparameter tuning was conducted exclusively on the training set. Model performance was assessed on the test set using MAE, RMSE, R^2, and MAPE, enabling comprehensive evaluation of prediction accuracy and error characteristics. Each model is trained and evaluated separately for the four outputs (Mean, Min, Max,DF_{depth}).

2.2.2 MTL-Light

To enable collaborative prediction of Mean, Min, Max, and DF_{depth}, and reduce information fragmentation from independent single-task training, the lightweight multi-task framework MTL-Light is proposed (Figure 1). Its core is chained multi-task learning: conditional dependencies between tasks are explicitly injected into subsequent tasks as "intermediate prediction results as features" , allowing the model to leverage global representation information from preceding targets to improve the learnability of subsequent tasks while learning the geometric parameter-light response mapping. For input feature vector x, the recursive form of MTL-Light is:

$$\hat{y}_{Mean} = f_1(\mathbf{x}) \quad (5)$$

$$\hat{y}_{Min} = f_2(\mathbf{x}, \hat{y}_{Mean}) \quad (6)$$

$$\hat{y}_{Max} = f_3(\mathbf{x}, \hat{y}_{Mean}, \hat{y}_{Min}) \quad (7)$$

$$\hat{y}_{DF_{depth}} = f_4(\mathbf{x}, \hat{y}_{Mean}, \hat{y}_{Min}, \hat{y}_{Max}) \quad (8)$$

Here, f1~f4 can be instantiated by the same type of regressor to ensure consistency in modeling form across stages. Each stage only expands the input feature dimension, avoiding complex shared encoders or cross-task interaction modules—embodying the "Light" design. This design is motivated by the complementary nature of the four tasks: Mean reflects overall lighting levels, Min and Max characterize distribution extremes, and DF_{depth} denotes the position meeting the 3.3% minimum daylighting coefficient threshold. Preceding tasks provide stable global priors for subsequent tasks, enhancing the robustness of extreme value and coverage task predictions.

 doi:10.20944/preprints202604.1668.v1

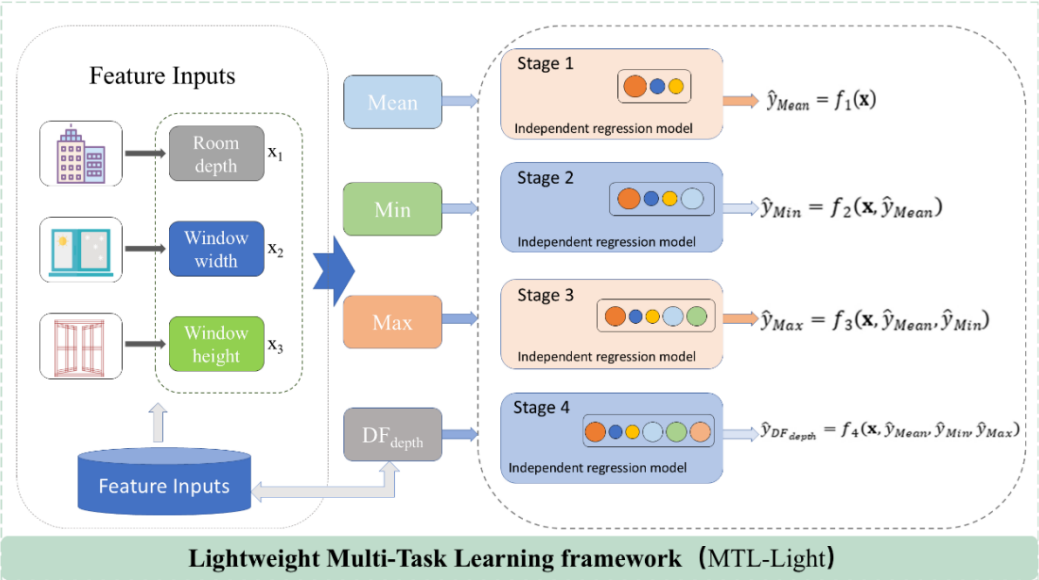


Figure 2. Stage-wise regression architecture of the MTL-Light framework

Table 1 presents representative optimal hyperparameter configurations for MTL-Light, with "valid_stacker = TRUE" indicating activation of the two-level stacking fusion mechanism. This training strategy provides a unified baseline for subsequent result comparison, model evaluation, and interpretability analysis.

Table 1. Hyperparameter settings

Hyperparameter	Optimal value
learning_rate	0.03
num_leaves	128
min_data_in_leaf	3
feature_fraction	0.9
valid_stacker	TRUE

MTL-Light training incorporates automated model selection, hyperparameter optimization, and ensemble learning to reduce subjective bias from manual tuning and improve method transferability. Under defined computational resource and time constraints, unified training and validation are performed on a candidate model set, with key hyperparameters optimized within a preset search space. The dataset is split into training (80%) and test (20%) sets; to enhance model selection robustness and mitigate single-partition randomness, 5-fold cross-validation is used for evaluation and screening within the training set, with a weighted ensemble strategy based on validation performance to fuse outputs of multiple sub-models. When stacking is enabled, the system trains a two-level fusion unit and automatically determines weights during validation, further improving generalization. All comparison methods use consistent data processing and evaluation workflows to ensure performance differences stem primarily from the modeling strategy itself. Section 4 compares the error performance of single-task baselines and MTL-Light across the four tasks, using SHAP analysis to examine differences in the contributions of key geometric variables and chain features.

2.3 SHAP Interpretability Analysis

Although MTL-Light achieves high predictive accuracy, its multi-component architecture retains a certain degree of black-box behavior, which limits its ability to provide directly interpretable

 doi:10.20944/preprints202604.1668.v1

and actionable design guidance. To improve the transparency of model predictions and quantify the marginal contribution of each input feature, the Shapley Additive exPlanations (SHAP) method was employed for interpretability analysis. Based on Shapley values from cooperative game theory, SHAP decomposes the prediction for an individual sample $x$into a baseline term and the additive contributions of all input features, enabling feature effects to be interpreted on a unified and comparable scale. For any sample x, the predicted value can be expressed as:

$$\hat{y} = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{9}$$

Here, $\hat{y} = f(x)$ is the model's predicted output for sample x; ϕ_0 is a baseline (typically the expected output of training sample predictions, $E[f(x)]$; ϕ_I is the marginal contribution of input feature i to the sample's prediction (i.e., the SHAP value), calculated as a weighted sum over all feature subsets excluding i; and M is the total number of input features (i=1,2,…,M).

The SHAP value ϕ_I is computed per the Shapley value definition: a weighted sum over all feature subsets excluding i, with the mathematical expression:

$$\phi_i = \sum_{S \subseteq F \backslash i} \frac{|S|!\,(M - |S| - 1)!}{M!} \left(v(S \cup i) - v(S)\right) \tag{10}$$

Where F is the set of all features（$|F| = M$）, S is any feature subset not containing i, $|S|$is the cardinality of S; $v(S)$ is a value function characterizing the expected model output performance when considering only subset S (defined herein as the conditional expectation $v(S) = E[f(x) \mid x_S]$; $v(S \cup \{i\}) - v(S)$ is the marginal gain from introducing feature i to subset S; and $\frac{|S|!(M - |S| - 1)!}{M!}$ is the corresponding Shapley weight. A positive ϕ_I indicates feature i increases the predicted value relative to the baseline, and vice versa. Using global feature importance (e.g., mean($|\phi_i|$)) and local dependency analysis, the impact of key process parameters on target predictions is characterized. Unlike individual sub-models being interpreted separately, MTL-Light is treated as a holistic predictor to ensure consistency between interpretation results and the final deployed and applied model. Specifically, SHAP values are calculated for all test set samples, and analysis is performed from both global and local perspectives.

2.4 Lighting Analysis Based on Interpretable Multi-Task Machine Learning

This study generates a dataset via parametric lighting simulation, establishing an integrated "multi-task prediction–interpretability analysis" workflow to enable rapid evaluation and mechanism identification of office unit daylighting performance.

During model construction and training, conventional single-task modeling and multi-task learning strategies are compared, and the chained multi-task method is defined as MTL-Light. Unlike training independent models for each task, MTL-Light explicitly leverages conditional correlations between tasks via "hierarchical recursion + feature enhancement."

To enhance the traceability and engineering applicability of model conclusions, SHAP for systematic model interpretation is employed. At the global level, statistical analysis of $|\phi_i|$ for all test set samples is conducted, adopting mean(|SHAP|) as a measure of feature importance to rank input variables—identifying key geometric and derived parameters influencing daylighting task predictions. SHAP summary plots are also employed to analyze the correspondence between feature values and contribution directions, revealing potential nonlinear trends and interaction effects.

At the local level, the SHAP contribution decomposition for representative scenario samples is analyzed, explaining the primary basis for model predictions under specific parameter combinations. Given the multi-task prediction output, the SHAP values for the four targets (Mean, Min, Max, DF_{depth}) are calculated, comparing sensitivity differences and dominant factors across tasks—transforming "high-precision prediction" into interpretable design insights to guide parameter adjustment and optimization in early design stages.

Based on these considerations, this study makes three key contributions:

 doi:10.20944/preprints202604.1668.v1

1. Establishing a parametric simulation dataset for office building daylighting under uniform climatic conditions, providing a consistent data foundation for multi-task prediction.
2. Proposing a lightweight chain-based multi-task learning method that effectively leverages correlations among daylighting tasks to improve multi-task prediction stability and accuracy.
3. Quantitatively revealing the influence of geometric design parameters on different daylighting tasks via SHAP interpretability analysis, enhancing the interpretability and practical value of model results for engineering design.

3. Dataset

3.1 Dataset Construction and Task Definition

Indoor natural daylighting is jointly influenced by spatial geometric factors (room depth, width, height, shape, etc.), window configuration (width, height, position, window-to-wall ratio, etc.), material optical properties, and external shading conditions. Including all these factors as independent variables leads to significant combinatorial explosion in the design space, making systematic sampling and simulation infeasible within engineering-realistic workloads. Based on existing daylighting theories and engineering experience, three geometric variables with significant influences on single-side window daylighting and direct controllability in the early design stage are focused on: room depth (D), window width (B), and window height (Hw). Parametric simulations are conducted under unified boundary conditions to construct a supervised learning dataset and the simulation sample set required for subsequent data-driven modeling.

The research object is a standardized office room, simplified to a rectangular space with a fixed floor height of 3.5 m and room width of 4.0 m. To systematically analyze the impact of geometric dimensions and window configuration on daylighting performance, D, B, and Hw are selected as key control variables: D (m) is the horizontal distance from the inner surface of the exterior window to the inner surface of the rear wall, with 8 discrete levels (3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5 m); B (m) is the net horizontal width of the side window, with 8 discrete levels (2.3, 2.5, 2.7, 2.9, 3.1, 3.3, 3.5, 3.7 m); Hw (m) is the net vertical height of the side window, with 6 discrete levels (1.1, 1.3, 1.5, 1.7, 1.9, 2.1 m). The window sill height is fixed at Hs=1.1 m, with a single rectangular side window used for all simulation cases to ensure consistent boundary conditions and structural configurations. Full factorial combination of D, B, and Hw yields 8×8×6=384 simulation cases. Variations in D, B, and Hw synchronously change the window area and derived indicators (window-to-wall ratio, window-to-floor ratio). All input conditions other than these geometric parameters and their derivatives are identical across cases, ensuring that differences in daylighting performance primarily stem from geometric parameter variations. Figure 3 presents the indoor daylight factor distributions for 64 simulation cases (window widths: 2.3–3.7 m; room depths: 3.0–6.5 m) using color-coded maps, where orange denotes high values and blue to indigo denotes low values.

For each simulation case, the daylight factor (DF) distribution on the work plane was obtained through numerical simulation. Due to the pronounced spatial heterogeneity of DF, with higher values near window openings and substantial attenuation with increasing distance, a single statistical indicator is insufficient to comprehensively describe indoor daylighting performance. To characterize indoor daylighting performance from multiple perspectives, three representative metrics were extracted from the DF field of each case, including Mean (overall daylighting level), Max (local peak value), and Min (worst-case condition). In addition, a compliance-related indicator, $\mathrm{DF}_{\mathrm{depth}}$, was defined to represent the maximum room depth that satisfies the minimum daylighting requirement.

Together, Mean, Min, Max, and $\mathrm{DF}_{\mathrm{depth}}$ form the output task system, revealing coupled variation laws between daylighting performance and parameters D, B, and Hw, and providing a unified data foundation for subsequent predictive modeling and interpretability analysis. To demonstrate response characteristics under different parameter combinations, the results section presents representative slices—e.g., comparing daylighting task variations across different window

 doi:10.20944/preprints202604.1668.v1

widths (B) and room depths (D) at fixed window height (Hw), and visualizing simulation results and distribution patterns of typical cases.

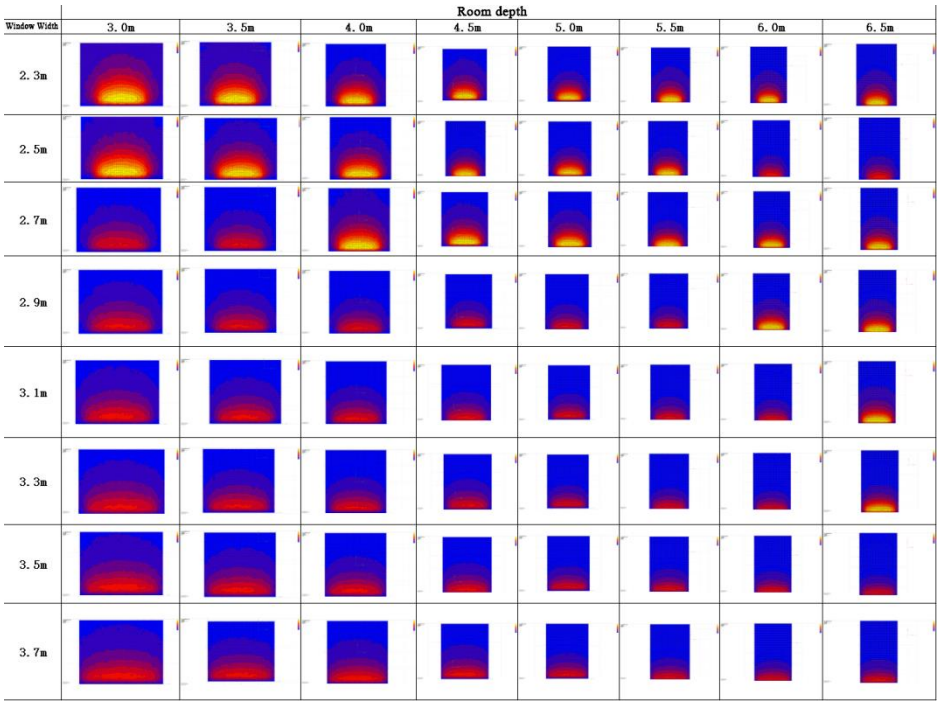


Figure 3. Simulated spatial distribution of daylight factor under different window width and room depth combinations

3.2 Dataset Characteristics

The statistical information of all variables is presented in Table 2, which details the characteristic parameters of the daylighting dataset and the Maximum, Minimum, Mean, Standard Deviation, Standard Error, and Sample Variance of each individual feature in the dataset.

Table 2 summarizes the statistical characteristics of all input and output features. For input variables, room depth ranges from 3.00 to 6.50 m with a standard deviation of 1.15 (significant variation), while window height and width are more concentrated—centered at mean values of 1.60 m and 3.00 m, respectively. Among output variables, Mean (%), Min (%), and Max (%) exhibit distinct distribution characteristics: Max remains consistently high and stable (mean: 12.62%, standard deviation: 1.09), Min shows large variation (mean: 1.74%, standard deviation: 0.70), and Mean falls in a moderate range (4.25%). DF_{depth} (m) spans 0.60–4.00 m with a standard deviation of 0.69, indicating significant differences under varying room geometric conditions. These results show that the four output variables differ in dimensionality and volatility but collectively characterize indoor daylighting performance and its correlation with spatial dimensions. Treating them as interrelated prediction tasks and adopting a multi-task learning framework for modeling thus fully exploits complementary information among tasks, enhancing overall prediction performance and exploring potential inter-variable influence mechanisms.

Table 2. Dataset Statistics

Feature	Unit	Maximum	Minimum	Mean	Standard Deviation	Standard Error	Sample Variance
Room Depth	m	6.50	3.00	4.75	1.15	0.06	1.32
Window Height	m	2.10	1.10	1.60	0.34	0.02	0.12
Window Width	m	3.70	2.30	3.00	0.46	0.02	0.21

 doi:10.20944/preprints202604.1668.v1

Mean	%	8.43	1.82	4.25	1.26	0.06	1.59
Min	%	4.64	0.63	1.74	0.70	0.04	0.49
Max	%	15.36	10.20	12.62	1.09	0.06	1.19
DF_{depth}	m	4.00	0.60	1.83	0.69	0.04	0.48

3.3 Feature Distribution Characteristics and Correlation Study

To systematically analyze the relationship between indoor daylighting performance tasks and design parameters, statistical analysis of the distribution characteristics and correlations of input parameters and output tasks is conducted firstly. Figure 4 presents pairwise relationship distributions between input variables (room depth, window height, window width) and daylighting performance tasks (Mean, Min, Max, DF_{depth}); Figure 5 quantifies linear correlations among all variables via a correlation coefficient matrix.

Figure 4 shows that input parameters are discretely and uniformly distributed within set intervals, ensuring sufficient parameter space coverage to support generalization learning of subsequent ML models for different design conditions. For output tasks, Mean, Min, Max, and DF_{depth} exhibit distinct continuity and clear variation trends under most variable combinations. Pairwise scatter plots indicate significant positive correlations among the daylighting performance tasks. Approximately linear relationships are observed between Mean and Min, as well as between Mean and Max, while DF_{depth} also exhibits strong correlations with all DF-related tasks. This indicates that different daylighting evaluation tasks are not independent but jointly influenced by the same geometric parameters.

Correlation analysis (Figure 5) further supports these observations. Room depth is strongly negatively correlated with Mean, Min, and DF_{depth}, indicating that daylighting performance decreases as room depth increases. By contrast, window height and width are moderately to strongly positively correlated with all daylighting tasks, and window height shows a particularly pronounced effect on Max and DF_{depth}. In addition, strong correlations are observed among the daylighting tasks themselves, especially between Mean and Min and between Mean and Max. This confirms the inherent coupling among the prediction targets and supports the adoption of a multi-task learning framework.

Room daylighting performance has a complex nonlinear relationship with building geometric parameters, and different output tasks exhibit strong correlations and information overlap. Single-task modeling for individual tasks neglects these inherent correlations, limiting overall model prediction stability. Introducing a multi-task learning method that explicitly characterizes inter-task correlation structures is thus necessary to fully exploit synergistic variation laws between input parameters and multiple daylighting tasks, providing a data and theoretical basis for constructing the MTL-Light framework.

 doi:10.20944/preprints202604.1668.v1

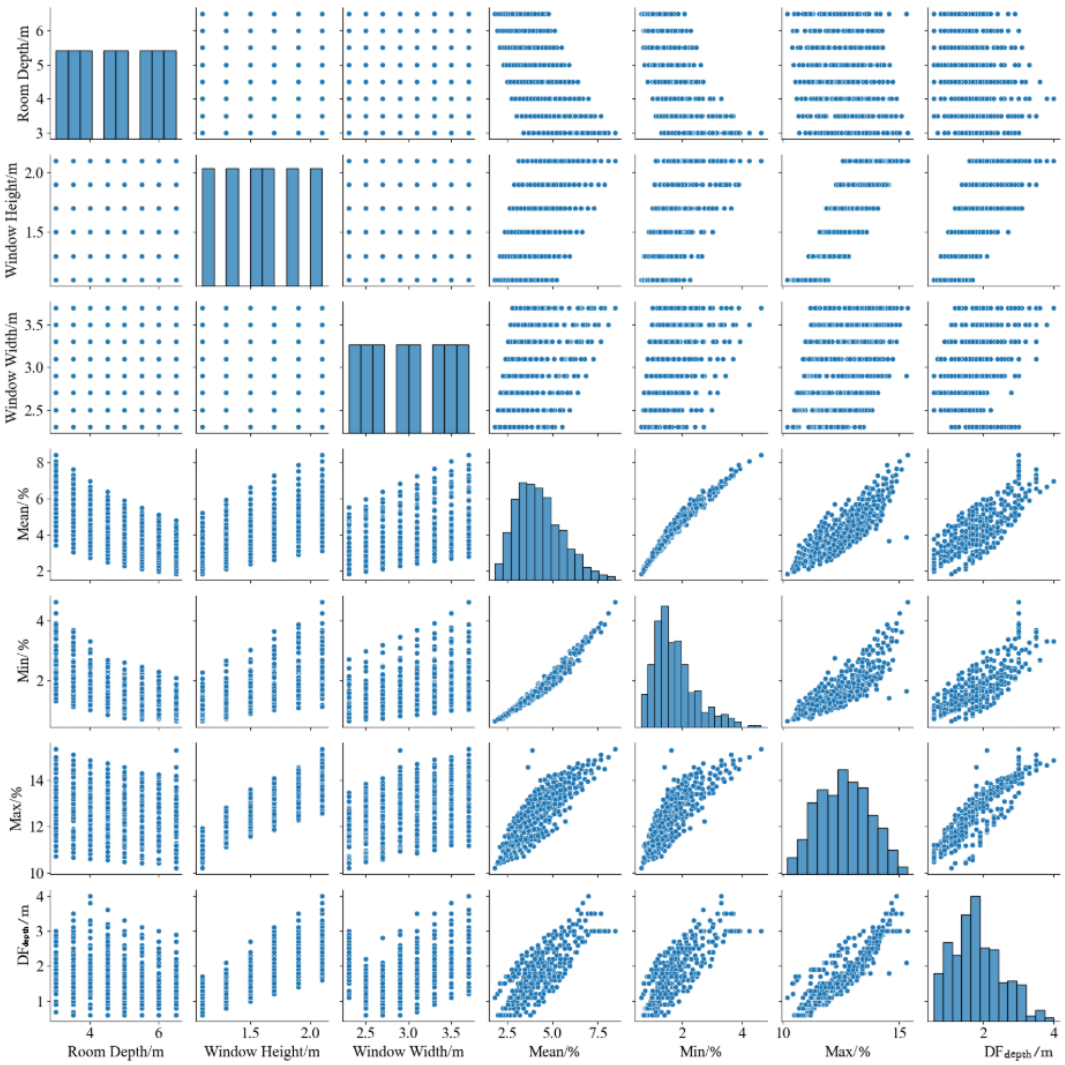


Figure 4. Pairwise relationships between input parameters and daylighting metrics

 doi:10.20944/preprints202604.1668.v1

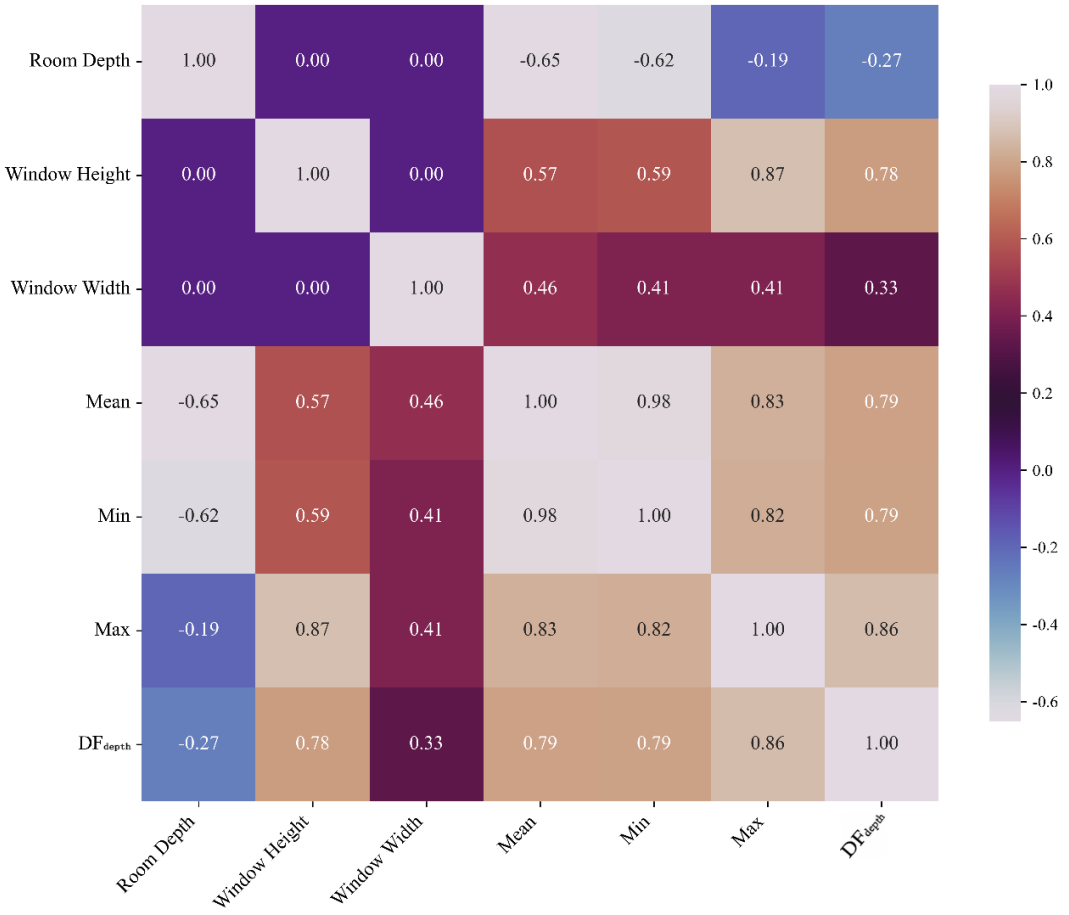


Figure 5. Correlation Matrix of Input Parameters and Daylighting Metrics

4. Prediction Performance Comparison and Error Analysis

4.1 Performance Evaluation of Single-Task Models

Prior to the analysis of multi-task prediction performance, single-task modeling methods are first investigated. Independent prediction of each output variable clarifies model performance under single-task settings, providing a baseline for subsequent multi-task comparisons and revealing potential limitations of single-task models for complex problems.

This section evaluates the prediction performance of different single-task models on four daylighting performance tasks (DF_{depth} (Fig. 6a), Max (Fig. 6b), Min (Fig. 6c), Mean (Fig. 6d)) to assess their accuracy and robustness under independent task settings. Figure 6 presents the distributions of model prediction errors, revealing the limitations of traditional single-task models in handling multiple prediction targets, particularly their inconsistent performance across different labels. LightGBM exhibits the most dispersed error range, particularly for DF_{depth} and Min tasks, indicating insufficient robustness to data fluctuations. NeuralNetTorch maintains relatively stable predictions overall but shows fluctuations in extreme value prediction for Max. XGBoost features the most concentrated error distribution, demonstrating higher reliability in capturing complex nonlinear relationships. Notably, while single-task models can adequately perform individual independent tasks, performance discrepancies across tasks highlight their limitations: they cannot ensure a single model performs well for all labels, and more importantly, they fail to account for potential inter-task correlations. Given the inherent connections among daylighting performance tasks, single-task modeling alone cannot fully exploit the dataset's overall value. Introducing multi-task learning thus mitigates single-task prediction instability and improves overall prediction accuracy and robustness via inter-task information sharing.

 doi:10.20944/preprints202604.1668.v1

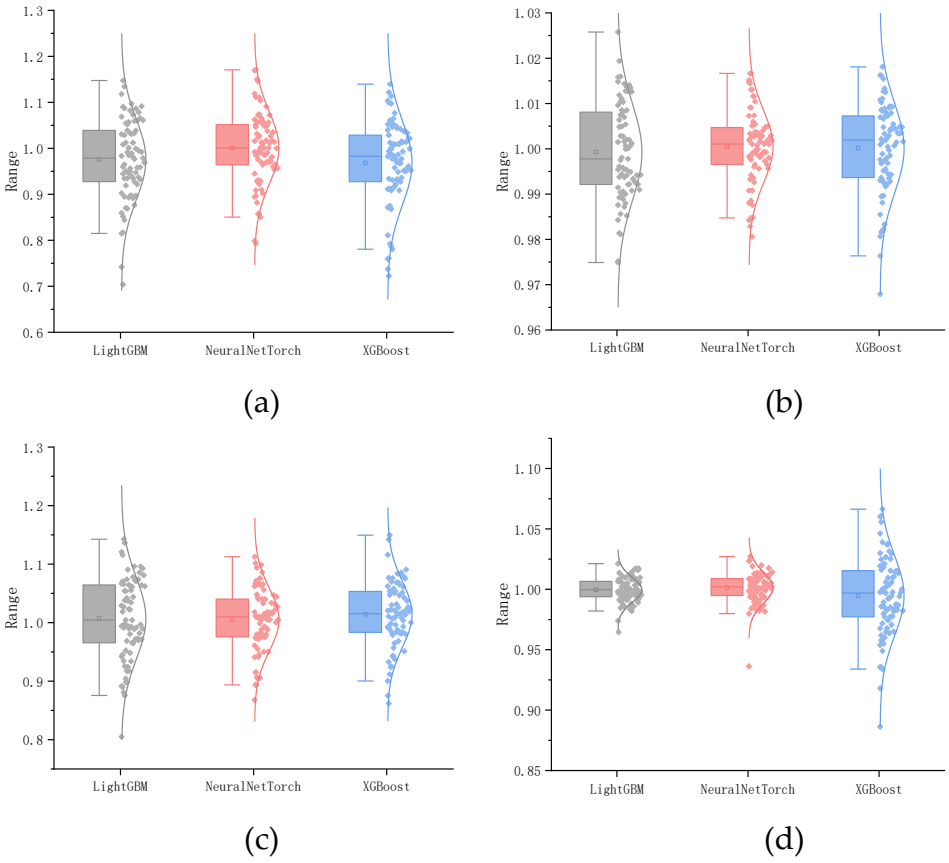


Figure 6. Error distributions of single-task models for different daylighting metrics. (a) DF_{depth} (b) MAX (c) MIN (d) MEAN

4.1.1 Comparative Prediction Performance Across Tasks

Figure 7 compares predicted results from different single-task models with experimental values for the four tasks (Mean (Fig. 7d), Min (Fig. 7c), Max (Fig. 7b), DF_{depth} (Fig. 7a)). In general, prediction points for all models cluster around the ideal line y=x, with most samples falling within the ±12% error band—indicating single-task methods can capture overall trends and maintain favorable prediction accuracy. However, a more detailed comparison reveals discrepancies in prediction stability across tasks. For Mean and Max prediction, XGBoost yields the most concentrated point cloud distribution, demonstrating superior reliability. In contrast, LightGBM and NeuralNetTorch exhibit systematic deviations in some results for Min and DF_{depth} prediction tasks, reflecting insufficient adaptability to extreme value and comprehensive tasks. These discrepancies embody the imbalance of single-task modeling when handling various output tasks: such models can achieve good fitting performance for certain tasks but incur significant errors in others. Given the intrinsic physical coupling among illumination-related tasks, single-task modeling often fails to fully exploit these inherent correlations, leading to inconsistent model performance across tasks. Multi-task modeling, by enabling information sharing among tasks, mitigates single-task prediction instability and enhances overall model accuracy and robustness. These results thus verify the limitations of single-task modeling and provide direct empirical support for multi-task prediction applications.

 doi:10.20944/preprints202604.1668.v1

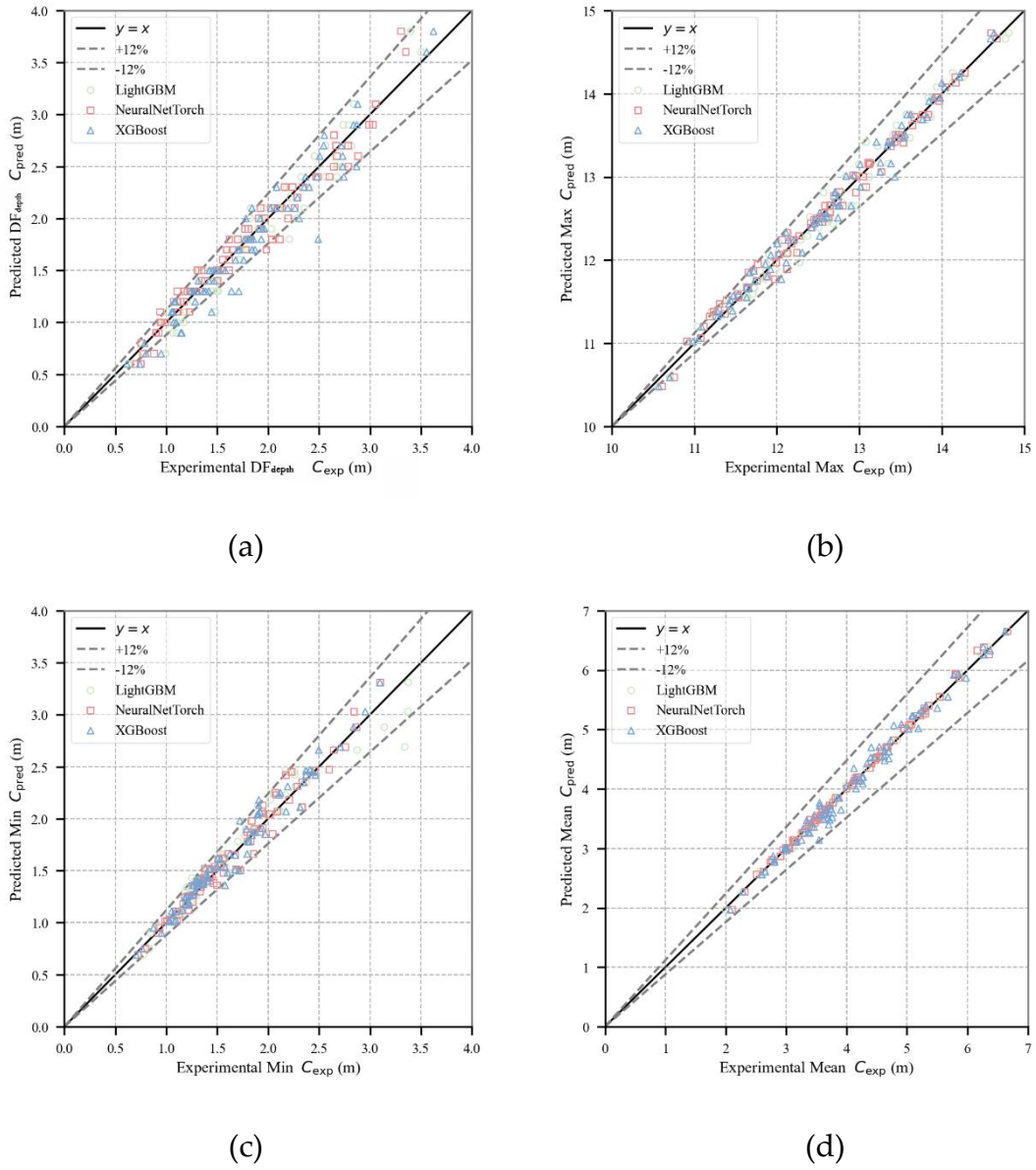
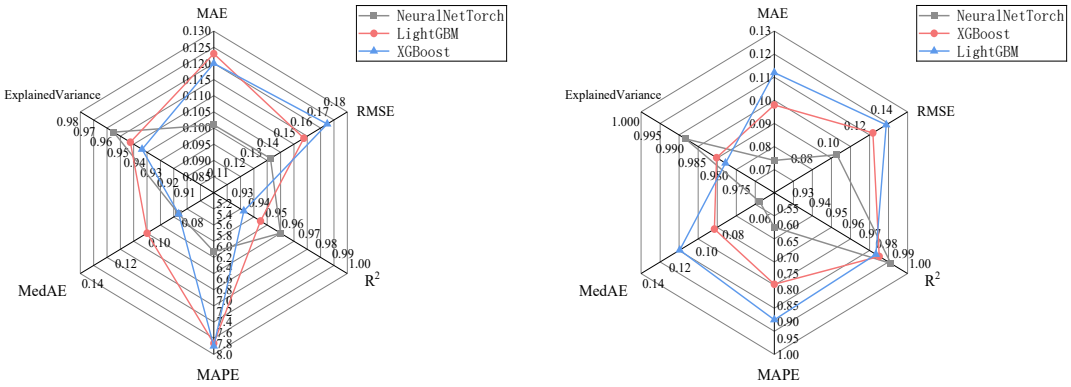


Figure 7. Predicted versus simulated values of single-task models. (a) DF_{depth} (b) MAX (c) MIN (d) MEAN

Analysis results show that single-task models can achieve reasonable accuracy on specific tasks but fail to fully capture inter-variable coupling relationships, with predictive performance susceptible to data complexity constraints. Multi-task methods, in contrast, enable mutual complementarity via information sharing, exhibiting superior overall efficacy.

4.1.2 Quantitative Comparison of Single-Task Models



 doi:10.20944/preprints202604.1668.v1

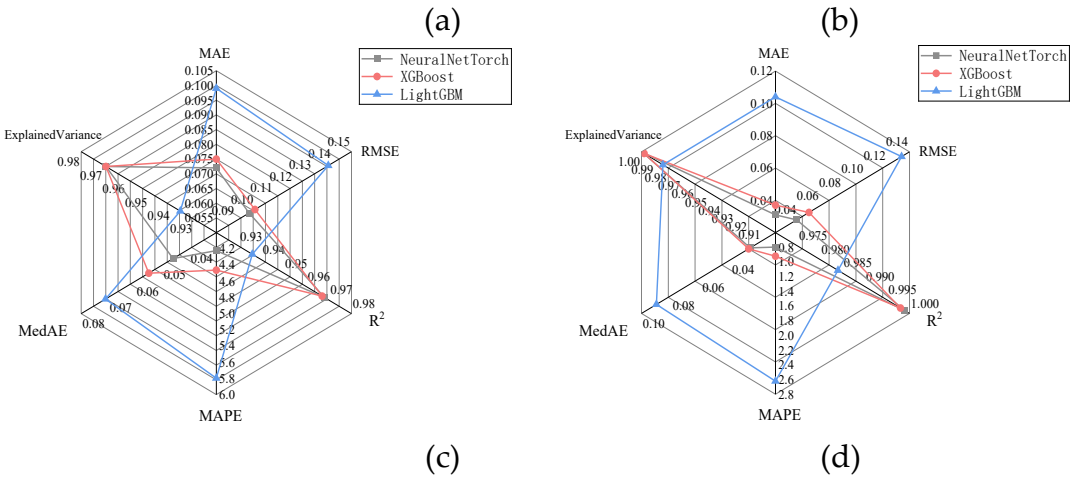


Figure 8. Radar comparison of single-task model performance metrics. (a) DF_{depth} (b) MAX (c) MIN (d) MEAN

Figure 8 presents evaluation results across the four daylighting tasks. Quantitative assessments reveal significant discrepancies in single-task model performance across tasks. For Mean (Fig. 8d) and Max (Fig. 8b) tasks, all models exhibit relatively low overall errors with R2 values generally close to 1.0, indicating strong fitting capability. In contrast, for sensitive tasks such as Min (Fig. 8c) and DF_{depth} (Fig. 8a) show markedly increased MAE and MAPE, revealing lower prediction stability and weaker generalization for low-illuminance and threshold-based tasks. A further comparison shows that different models exhibit varying performance advantages across individual tasks, demonstrating distinct task dependency: a single model can hardly maintain high prediction accuracy across all daylighting tasks simultaneously. This phenomenon indicates that independent modeling of multiple tasks tends to result in information isolation, making it difficult to fully characterize inherent inter-task correlations among daylighting tasks. Therefore, developing a multi-task learning framework that captures shared feature representations and inter-task dependencies is essential for improving the consistency and overall robustness of multi-task prediction.

4.2 Comparative Evaluation of MTL-Light and Baseline Models

Based on the preceding single-task prediction analysis, this section further compares the proposed MTL-Light model with baseline models and systematically evaluates its performance in terms of overall predictive accuracy, stability across tasks, and error distribution characteristics.

4.2.1 Overall Prediction Performance of MTL-Light

Figure 9 compares predicted results from the MTL-Light multi-task model with ground truth values for the four tasks (Mean (Fig. 9d), Max (Fig. 9b), Min (Fig. 9c), DF_{depth}(Fig. 9a)), including fitting curves with 95% confidence intervals (CI) and prediction intervals (PI). Results indicate: (1) For Mean and Max prediction, the point clouds cluster tightly around the ideal line $y = x$, with nearly all samples falling within the confidence interval, suggesting that the multi-task model achieves better fitting performance for overall trends and extreme values than the corresponding single-task models. (2) For Min and DF_{depth}prediction, while some points remain relatively scattered, the model exhibits smaller overall fluctuations and higher CI coverage than single-task models, indicating better adaptability to boundary condition and comprehensive task prediction. (3) Overall, multi-task learning effectively narrows the prediction intervals and reduces predictive uncertainty, in clear contrast to the unbalanced task-wise performance of single-task models.

This improvement is highly significant for office building indoor daylight simulation research. Traditional simulation methods rely on extensive computations and tedious modeling, while single-task prediction frameworks exhibit accuracy discrepancies when handling multi-task outputs, making them unsuitable as robust alternative tools. Multi-task learning, in contrast, not only improves overall prediction accuracy but also enhances model generalization via inter-task

 doi:10.20944/preprints202604.1668.v1

complementarity—enabling ML-based prediction to replace some complex simulation processes, providing a more efficient and reliable approach for rapid daylighting performance evaluation and optimization in office buildings.

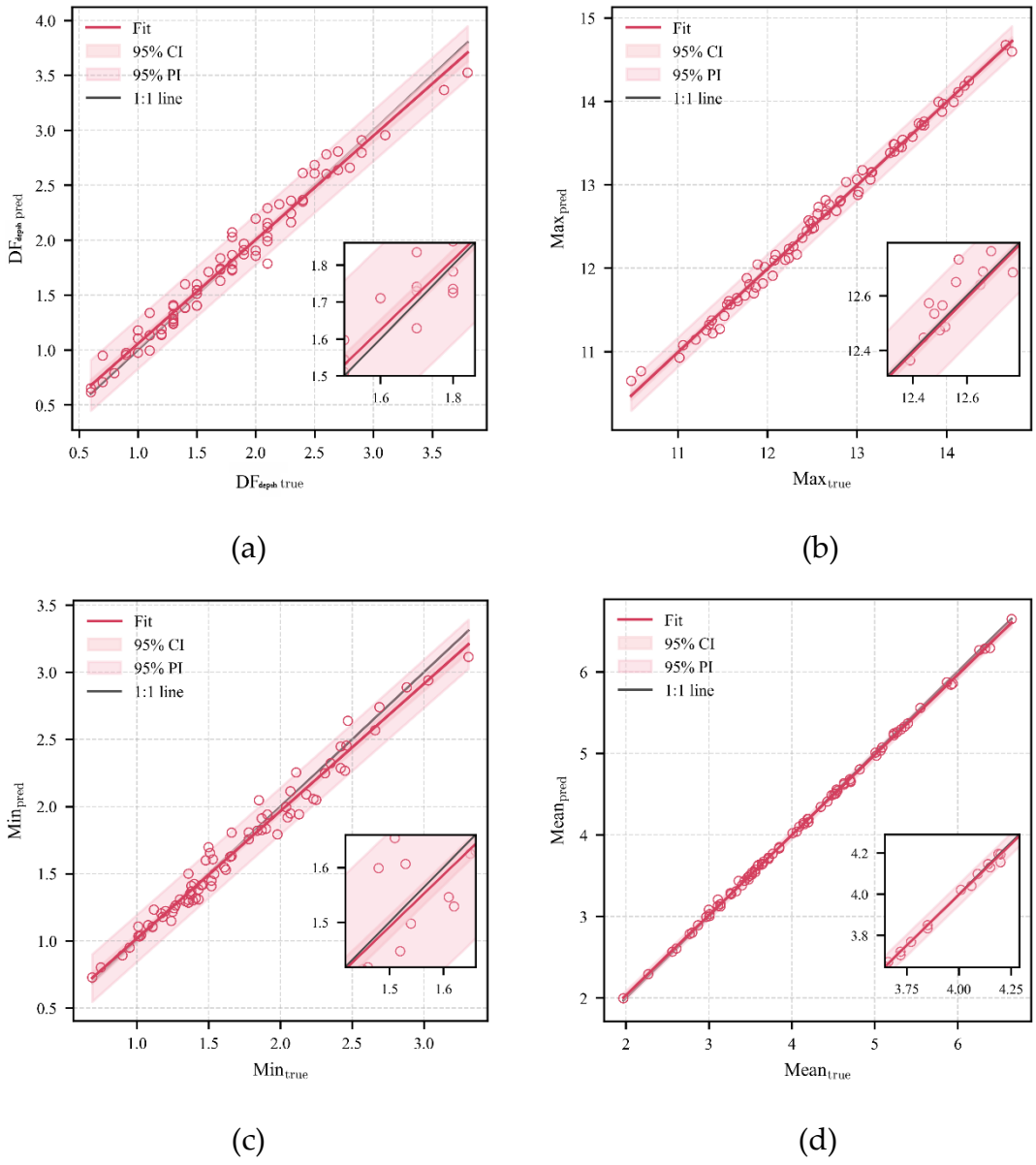
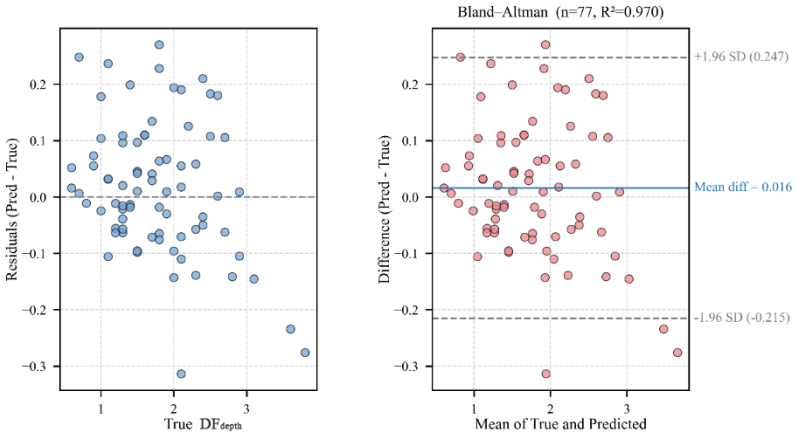


Figure 9. Prediction performance and uncertainty of the MTL-Light model across four daylighting metrics (a) DF_{depth} (b) MAX (c) MIN (d) MEAN

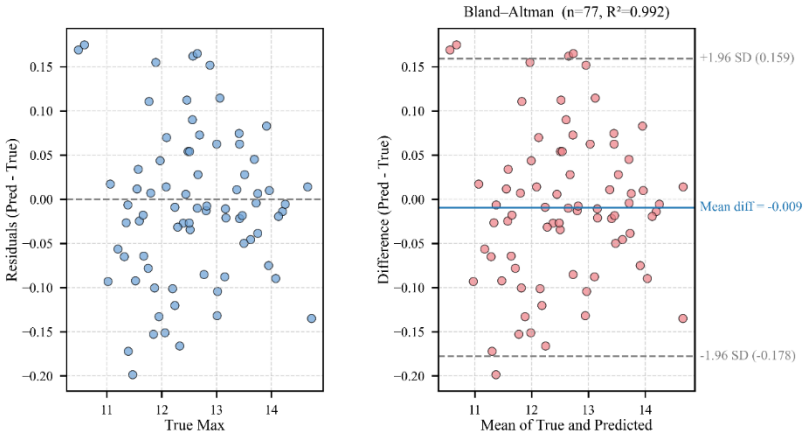
Beyond overall predictive performance comparison, it is necessary to examine error characteristics and consistency of the multi-task model across tasks. While fitting curves and CI reflect overall accuracy, they fail to reveal systematic prediction biases across numerical intervals. Residual analysis and the Bland–Altman method are thus introduced to compare the distribution of differences between predicted and actual values, enabling detailed evaluation of model stability and reliability across all task types. Figure 10 presents residual distribution and Bland–Altman analysis results for the multi-task model across the four tasks (Mean, Min, Max, DF_{depth}), verifying consistency and systematic biases between predicted and actual values. Most sample residuals cluster around zero; mean differences from Bland–Altman analysis are all close to zero, with over 95% of points falling within ±1.96 standard deviations (SD). These findings indicate excellent agreement between model predictions and actual values for primary tasks. In contrast, Min and DF_{depth} tasks

 doi:10.20944/preprints202604.1668.v1

exhibit relatively larger predictive fluctuations and more dispersed residual distributions, suggesting some uncertainty in model performance for low-value or comprehensive tasks. Nevertheless, compared with the uneven performance of single-task methods across tasks, multi-task learning effectively suppresses error propagation via inter-task information sharing, rendering overall predictions more stable. Importantly, this consistency verification not only validates the multi-task model's effectiveness but also provides methodological support for engineering practice. When integrated with simulations, the model further enhances daylighting performance evaluation reliability while maintaining high accuracy, offering a more robust technical basis for architectural design optimization and energy-saving strategy formulation. Residual and consistency analyses thus confirm the advantages of the multi-task framework in complex predictive tasks, laying a solid foundation for exploring its subsequent popularization and application.

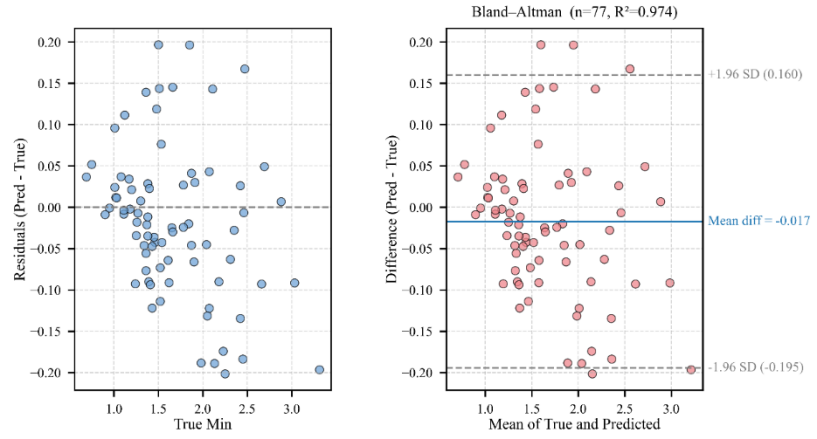


(a)

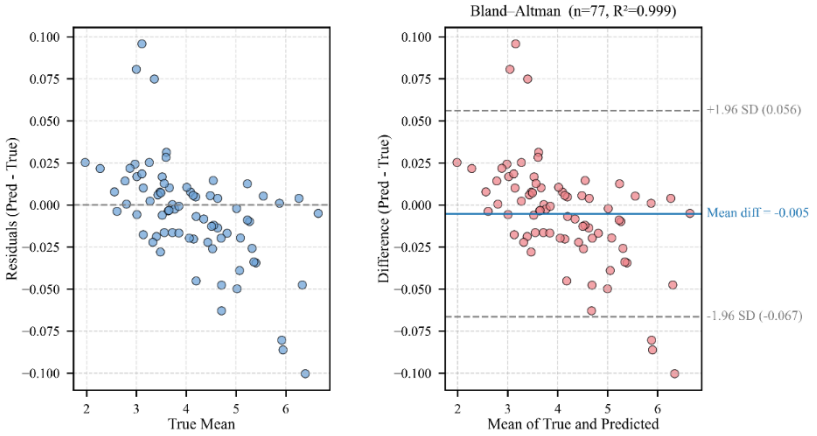


(b)

 doi:10.20944/preprints202604.1668.v1



(c)



(d)

Figure 10. Residual and Bland–Altman analysis of multi-task daylighting predictions. (a) DF_{depth} (b) MAX (c) MIN (d) MEAN

4.2.2 Improvement and Stability Analysis Across Tasks

Figure 11 presents multi-metric prediction performance of the MTL-Light model across tasks. Table 3 compares the predictive performance of three single-task models (NeuralNetTorch, LightGBM, XGBoost) with the improved multi-task model MTL-Light across the four output tasks (Mean, Min, Max, DF_{depth}). Results show significant performance variations of single-task models across tasks: for example, in the DF_{depth} task, LightGBM achieves a coefficient of determination (R2) of only 0.948 with a mean absolute percentage error (MAPE) as high as 7.797%; in Min task prediction, XGBoost yields a root mean square error (RMSE) of 0.100, significantly lower than in other tasks. This imbalance indicates that single-task modeling fails to effectively exploit inter-task correlations when processing each output independently, leading to large prediction errors in sensitive tasks (e.g., Min and DF_{depth}).

Table 3. Comparison of Predictive Performance between Single-Task Model and MTL-Light Model.

Label	Model	MAE	RMSE	R^2	MAPE
DF_{depth}	MTL-Light	0.092	0.118	0.969	5.598
	NeuralNetTorch	0.101	0.134	0.96	6.092

 doi:10.20944/preprints202604.1668.v1

Max	LightGBM	0.123	0.154	0.948	7.797
	XGBoost	0.12	0.168	0.938	7.847
	MTL-Light	0.067	0.086	0.992	0.545
	NeuralNetTorch	0.074	0.097	0.99	0.607
	XGBoost	0.098	0.124	0.983	0.783
	LightGBM	0.112	0.134	0.981	0.893
Mean	MTL-Light	0.022	0.032	0.999	0.541
	NeuralNetTorch	0.037	0.05	0.998	0.92
	LightGBM	0.031	0.039	0.997	0.797
	XGBoost	0.104	0.133	0.984	2.623
Min	MTL-Light	0.071	0.092	0.972	4.168
	NeuralNetTorch	0.072	0.097	0.968	4.222
	XGBoost	0.075	0.1	0.967	4.463
	LightGBM	0.099	0.138	0.936	5.797

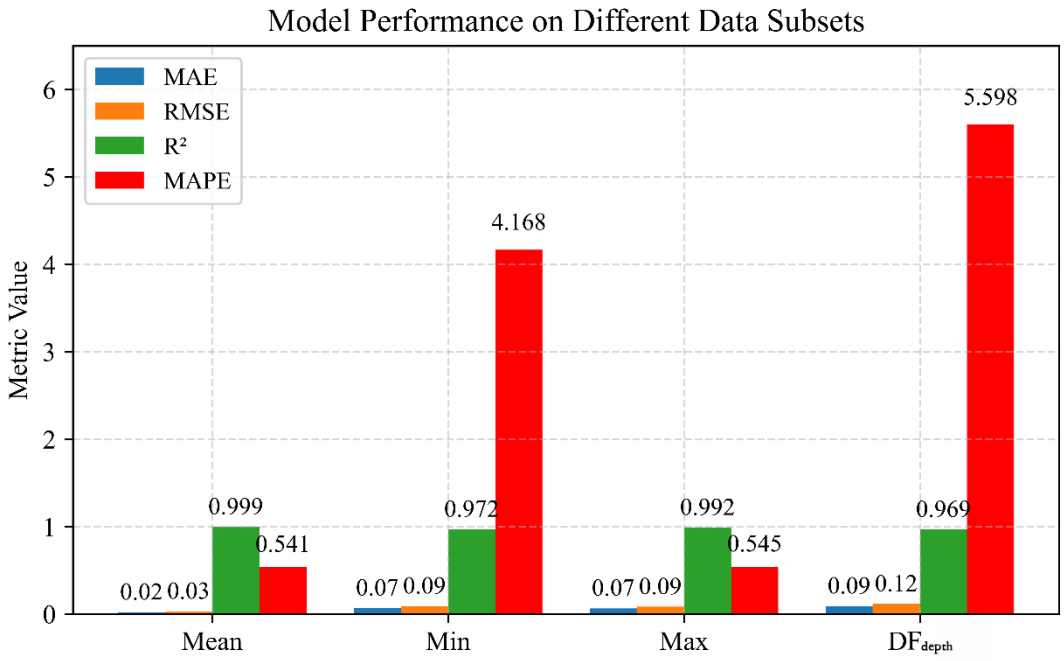


Figure 11. Multi-metric prediction performance of the MTL-Light model

In contrast, MTL-Light demonstrates remarkable advantages across all tasks (Figure 11): it achieves a mean absolute error (MAE) of merely 0.022 and an R2 of 0.999 in the Mean task; for more challenging prediction tasks such as Min and DF_{depth}, its MAPE remains below 4.168% and 5.092%, respectively, significantly lower than that of the corresponding single-task models.. Meanwhile, MTL-Light yields universally low RMSE (below 0.1) with a narrower prediction interval, exhibiting enhanced stability and robustness. These results verify that the multi-task learning framework more effectively captures coupling relationships among output variables via inter-task shared feature representation, improving overall model performance.

Such improvements are particularly valuable for engineering applications in simulation-based building daylighting studies. Traditional simulation methods rely on complex modeling and computation with low efficiency, while single-task prediction struggles to serve as a reliable

 doi:10.20944/preprints202604.1668.v1

alternative due to unbalanced accuracy. In contrast, MTL-Light not only improves overall predictive accuracy but also enhances generalization through inter-task complementarity, making it a more robust and broadly applicable modeling framework. Its core advantage lies in providing an efficient, stable, and interpretable technical approach for rapid daylighting performance evaluation and optimization in office buildings.

5. Interpretability Based on Multi-Task Prediction

Previous chapters verified the accuracy and stability of the MTL-Light model for illuminance prediction via a multi-task learning framework. However, while numerical performance evaluation demonstrates the model's overall prediction advantages, it fails to explain its internal decision-making logic and feature interaction mechanisms. To further clarify the contributions of different input variables and inter-task dependencies to prediction results, this chapter introduces the SHapley Additive exPlanations (SHAP) method for systematic interpretability analysis of the MTL-Light model's prediction process.

5.1 Global Feature Importance Analysis Based on SHAP

To deepen understanding of the MTL-Light model's decision-making mechanism across different prediction tasks, this section first conducts a systematic analysis of the feature importance of all input variables and prior prediction results. Analysis of SHAP value distributions reveals the dominant contributions of input variables and prior predictions to each output. It further identifies the direct effects of geometric variables on daylighting performance and uncovers the hierarchical coupling relationships among different illuminance prediction tasks, providing a basis for subsequent nonlinear mechanism interpretation and engineering application.

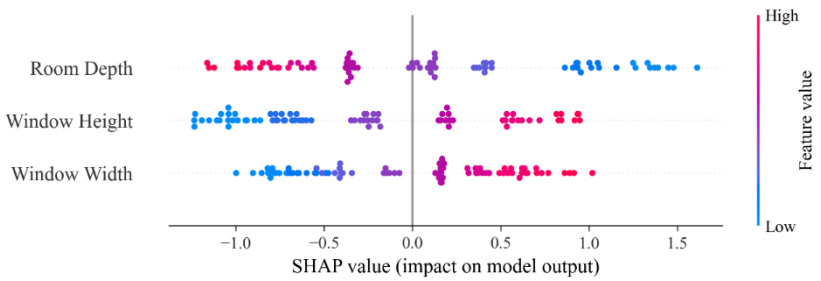
Figure 12 presents SHAP value distributions of the MTL-Light model under different prediction tasks, revealing the dominant roles and hierarchical dependencies of input variables and prior predictions on each output. For Mean illuminance prediction (Fig. 12a), Room Depth exhibits the most significant negative contribution, while Window Height and Window Width exert positive effects—with Window Height's impact particularly prominent. This indicates that mean illuminance is primarily governed by the dual effects of light attenuation from room depth and light enhancement from window geometric parameters. When $\hat{y}_{\mathrm{Mean}}$ is introduced as an input for Min illuminance prediction (Fig. 12b), its positive contribution far outweighs that of geometric variables, making it the primary explanatory factor. This demonstrates that overall illuminance levels effectively constrain minimum illuminance, implying the multi-task learning framework not only leverages raw feature information but also explicitly captures the Mean → Min dependency relationship.

When $\hat{y}_{\mathrm{Min}}$ is further incorporated for Max illuminance prediction (Fig. 12c), Window Height remains the strongest dominant factor, while $\hat{y}_{\mathrm{Min}}$ and Window Width exert clear positive effects, and the negative contribution of Room Depth is relatively weakened. This means maximum illuminance is not only dominated by geometric parameters but also indirectly constrained by minimum illuminance distribution, reflecting the hierarchical transfer of Mean/Min → Max. Finally, when $\hat{y}_{\mathrm{Max}}$ is introduced for $\mathrm{DF}_{\mathrm{depth}}$ prediction (Fig. 12d), its positive contribution is most prominent, followed by Window Width—with $\hat{y}_{\mathrm{Mean}}$ and $\hat{y}_{\mathrm{Min}}$ providing auxiliary positive support. The negative effect of Room Depth persists but is less pronounced. It can thus be concluded that $\mathrm{DF}_{\mathrm{depth}}$ prediction relies not only on peak illuminance but also integrates mean and minimum illuminance information, forming a key Max → $\mathrm{DF}_{\mathrm{depth}}$ link.

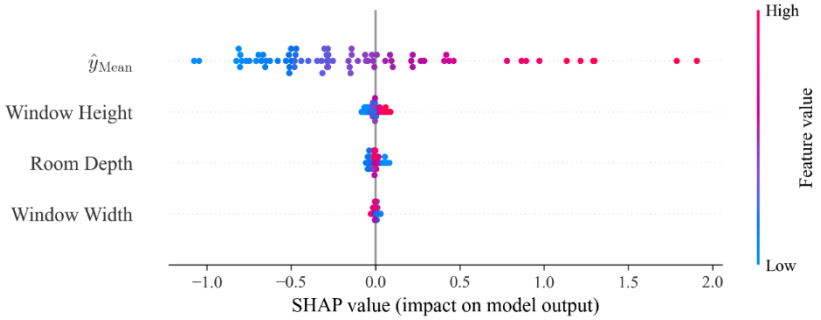
Overall, these four figures clearly delineate a hierarchical information flow extending from geometric inputs to output coupling: Room/Window → $\hat{y}_{\mathrm{Mean}}$ → $\hat{y}_{\mathrm{Min}}$ → $\hat{y}_{\mathrm{Max}}$ → $\mathrm{DF}_{\mathrm{depth}}$. By transferring prediction results across tasks, MTL-Light explicitly captures the physically inherent dependency relationships of daylight distribution, reducing uncertainty in sensitive tasks such as Min and $\mathrm{DF}_{\mathrm{depth}}$ and rendering overall predictions more stable and reliable. This mechanism highlights the unique advantages of multi-task learning in building daylighting simulation,

 doi:10.20944/preprints202604.1668.v1

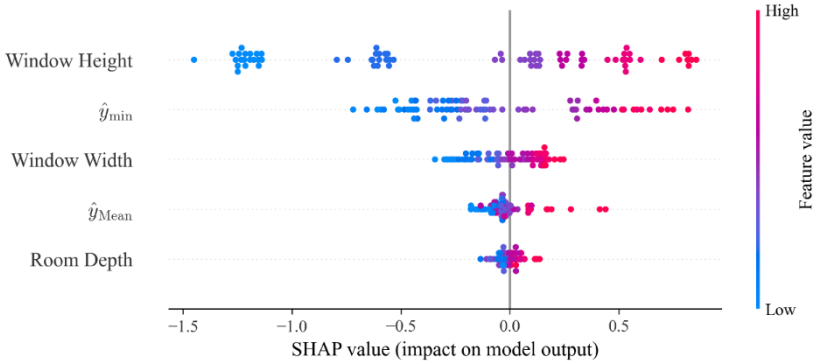
providing theoretical support and engineering implications for replacing complex simulation processes with data-driven approaches.



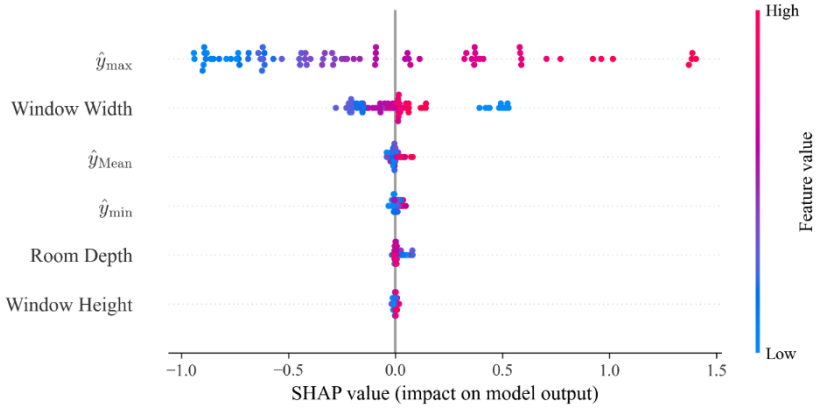
(a)



(b)



(c)



(d)

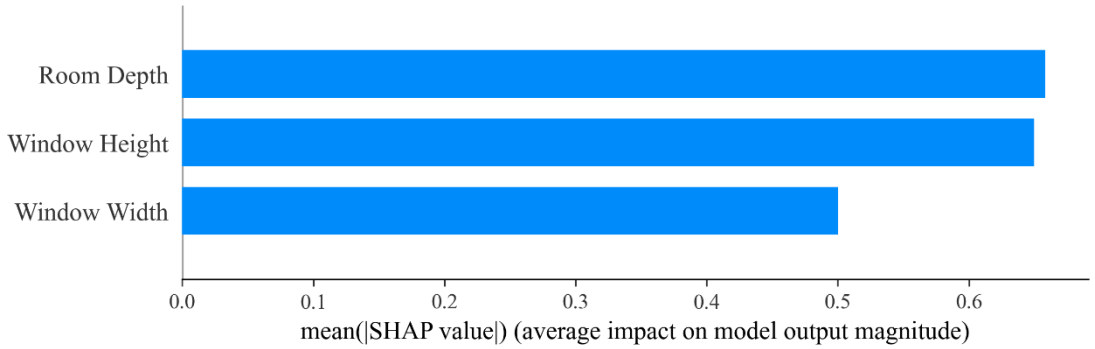
Figure 12. SHAP summary plots of the MTL-Light model across different prediction targets. (a) Base features only (b) $+\hat{y}_{\text{Mean}}$ (c) $+\hat{y}_{\text{Mean}}$ $+\hat{y}_{\text{Min}}$ (d) $+\hat{y}_{\text{Mean}} + \hat{y}_{\text{Min}} + \hat{y}_{\text{Max}}$

 doi:10.20944/preprints202604.1668.v1

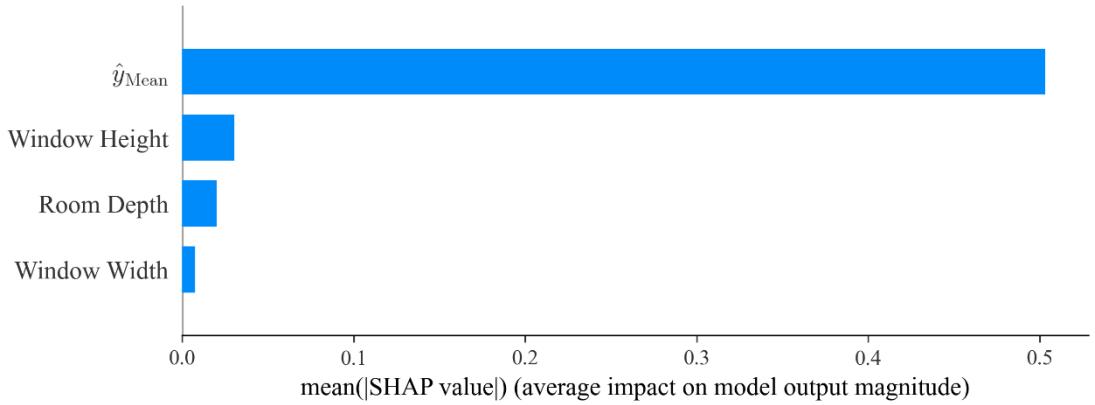
SHAP summary plots reveal the direction and magnitude of feature effects but do not allow direct comparison of overall feature importance. Therefore, SHAP bar plots (Figure 13) are used to quantify feature importance across the four output tasks based on mean absolute SHAP values.

For Mean prediction (Fig. 13a), room depth and window height show comparable importance and jointly dominate the model output, while window width plays a secondary role. For Min prediction (Fig. 13b), $\hat{y}_{Mean}$ contributes significantly more than geometric variables, indicating that overall daylighting levels strongly constrain minimum illuminance. For Max prediction (Fig. 13c), window height remains the most influential factor, followed by $\hat{y}_{Min}$, with window width also contributing noticeably. For $\mathrm{DF}_{\mathrm{depth}}$ prediction (Fig. 13d), $\hat{y}_{Max}$ shows the highest importance, while geometric variables become less dominant, suggesting that this task relies more on combined information from multiple sources.

Overall, these results demonstrate that the MTL-Light model not only captures the dominant effects of geometric variables but also effectively incorporates preceding task predictions as additional features, leading to improved stability and accuracy in multi-task daylighting prediction.

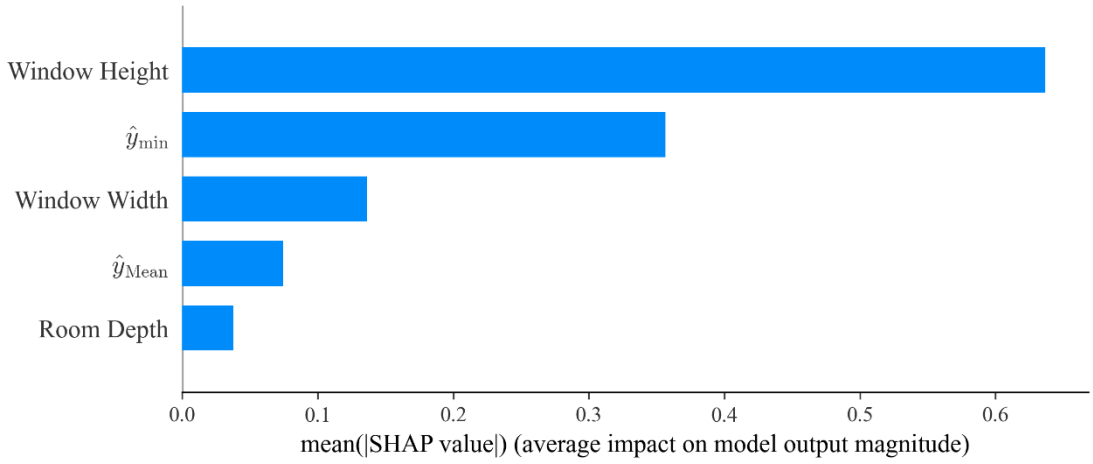


(a)

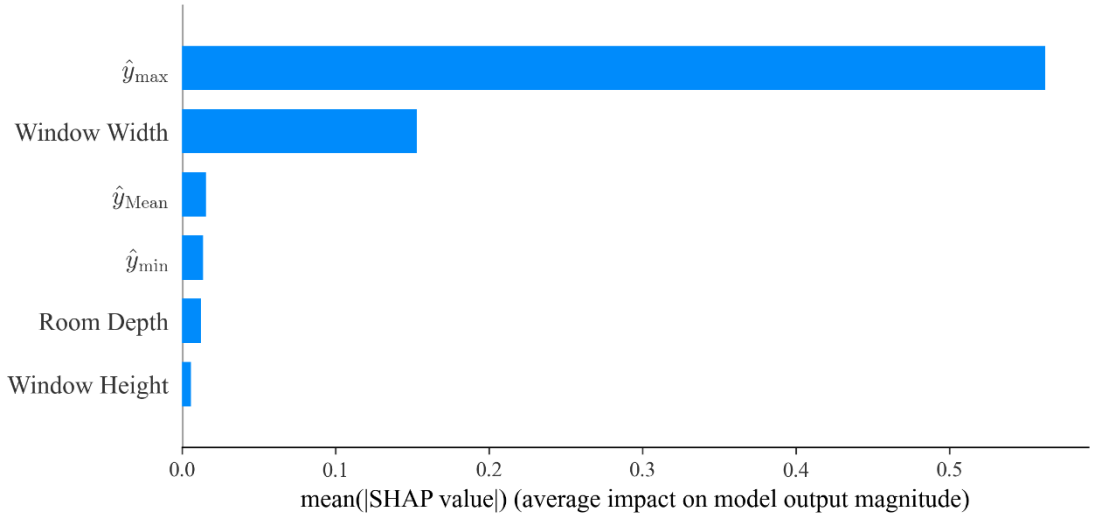


(b)

 doi:10.20944/preprints202604.1668.v1



(c)



(d)

Figure 13. Mean absolute SHAP value plots of the MTL-Light model across different prediction targets. (a) Base features only (b) $+\hat{y}_{Mean}$ (c) $+\hat{y}_{Mean}$ $+\hat{y}_{Min}$ (d) $+\hat{y}_{Mean}+\hat{y}_{Min}+\hat{y}_{Max}$

5.2 Feature Interaction Relationship and Nonlinear Mechanism Mining

This section presents representative SHAP interaction effect plots—including room depth effects for Mean, $\hat{y}_{Mean}$ effects for Min, window height effects for Max, and $\hat{y}_{Max}$ effects for DF_{depth}. These results intuitively reveal the mechanistic characteristics of the multi-task learning framework, wherein preceding task prediction results are incorporated as input features for subsequent task modeling, further highlighting differences in relative importance between geometric variables and each prediction task. Additional SHAP interaction effect plots are provided in the appendix as supplementary validation to ensure overall completeness and conciseness.

Figure 14 presents typical SHAP dependence plots of the MTL-Light model for different output tasks, uncovering interaction mechanisms of key variables and preceding task predictions on subsequent tasks. For the Mean task (Fig. 14a), room depth is negatively and approximately linearly correlated with SHAP values, suggesting that increasing room depth consistently reduces its contribution to average illuminance. Points are color-coded by window width, showing that, at the same room depth, larger window widths can partially offset this negative effect, reflecting the synergistic interaction among geometric variables (Fig. 14b), $\hat{y}_{Mean}$ and SHAP values exhibit a strong positive correlation with an approximate 1:1 ratio, demonstrating the decisive role of overall average

 doi:10.20944/preprints202604.1668.v1

illuminance in minimum illuminance. This underscores the multi-task framework's advantage: integrating preceding task predictions as input features enhances boundary condition constraint capability. For Max prediction (Fig. 14c), $\hat{y}_{\mathrm{Min}}$ and SHAP values also exhibit a strong linear relationship, indicating that maximum illuminance is directly governed by geometric parameters while being indirectly constrained by low-illuminance levels. Such cross-task dependency helps the model maintain stability in extreme value prediction. For $\mathrm{DF}_{\mathrm{depth}}$ prediction (Fig. 14d), $\hat{y}_{\mathrm{Max}}$ SHAP values increase monotonically with predicted values, reflecting peak illuminance's dominant influence on the comprehensive $\mathrm{DF}_{\mathrm{depth}}$ task. Meanwhile, the color distribution of points, representing window width, reveals the secondary influence of geometric conditions, indicating that $\mathrm{DF}_{\mathrm{depth}}$ is determined by the combined effects of peak illuminance levels and spatial configuration.

These interaction effects verify the MTL-Light model's modeling mechanism: different task outputs are driven not only independently by geometric features but also intercoupled via preceding task predictions, improving overall prediction consistency and interpretability. This characteristic holds particular engineering significance for architectural daylight simulation, uncovering inherent logical connections among multiple illuminance distribution labels and providing a solid theoretical foundation for ML-driven rapid evaluation and optimization of architectural daylight performance.

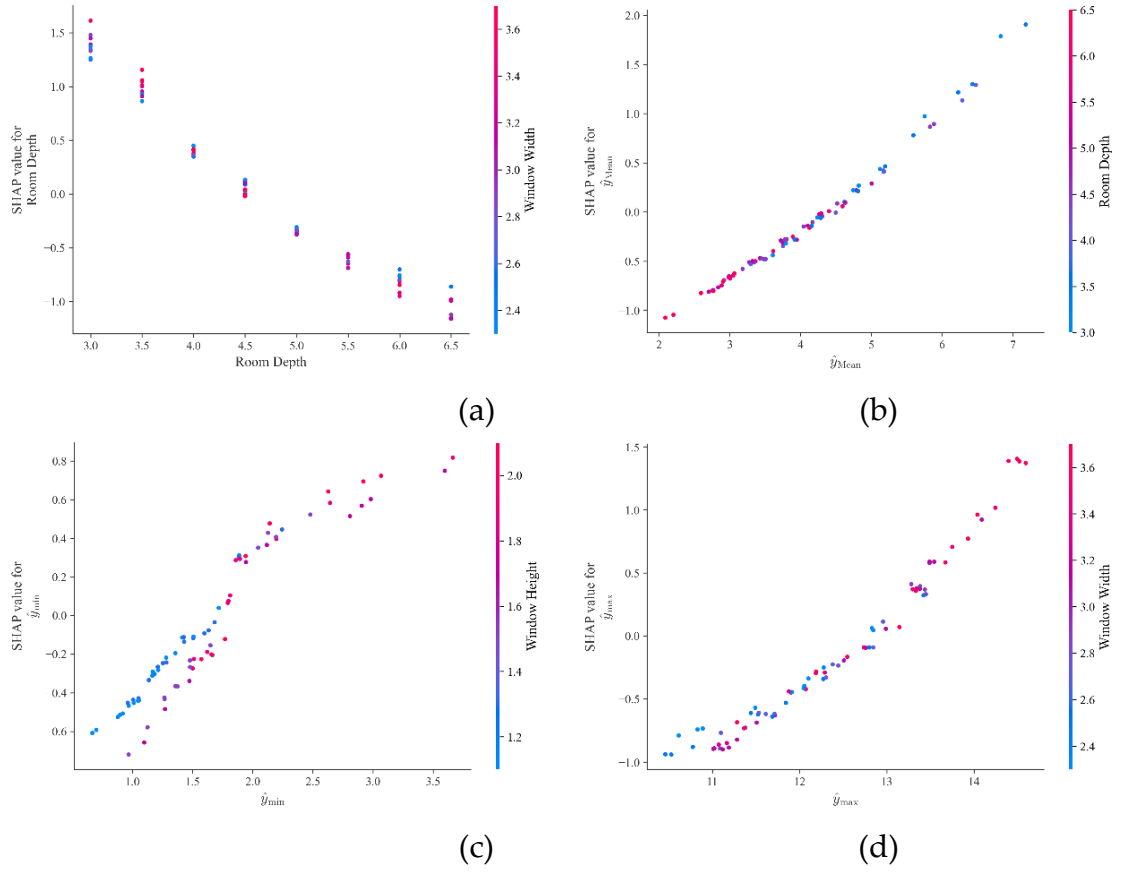


Figure 14. SHAP interaction effects in the MTL-Light model. (a) SHAP value for Room Depth with respect to Window Width (b) SHAP value for $\hat{y}_{\mathrm{Min}}$ with respect to Room Depth (c) SHAP value for $\hat{y}_{\mathrm{Min}}$ with respect to Window Height (d) SHAP value for $\hat{y}_{\mathrm{Max}}$ with respect to Window Width.

6. Conclusions

This study presents MTL-Light, a lightweight and explainable chained multi-task learning framework for rapid daylighting performance prediction in typical office units. Based on a unified parametric daylighting dataset generated from geometric modeling and simulation, the proposed framework establishes an effective surrogate modeling pipeline that links key design variables to

 doi:10.20944/preprints202604.1668.v1

multiple daylighting outputs, providing a reusable basis for model development, benchmarking, and early-stage design evaluation.

1. A unified parametric daylighting dataset and surrogate modeling pipeline were established. By integrating parametric geometric modeling with daylighting simulation, this study constructed a dataset from geometric design variables to multiple daylighting outputs, providing a reusable basis for model training, benchmarking, and rapid performance evaluation in early design stages.
2. MTL-Light achieved accurate and consistent prediction across all four daylighting tasks. Experimental results show that all output tasks reached high predictive accuracy ($R2 \geq 0.969$). In particular, the framework performed best on the Mean task and maintained similarly robust performance on the Max task, indicating the effectiveness of chained multi-task learning in improving information sharing and prediction consistency across correlated daylighting metrics.
3. MTL-Light remained reliable for more sensitive tasks, although MAPE should be interpreted cautiously. For the Min and DF_{depth} tasks, the relatively higher MAPE values are mainly attributable to output discreteness and the amplification effect associated with small target values. This suggests that model performance for such daylighting metrics should be assessed using multiple indicators, including MAE, RMSE, and R2, rather than MAPE alone.
4. The proposed framework provides both rapid evaluation and interpretable design support. MTL-Light can simultaneously predict Mean, Min, Max, and DF_{depth} in a single inference, enabling efficient parameter screening and multi-task comparison in early design stages. Combined with SHAP analysis, it also reveals the sensitivity and directional influence of key design variables, thereby offering interpretable support for room depth and window design decisions while reducing dependence on repeated numerical simulations.

Author Contributions: Conceptualization, Gaoyang Liu and Yue Zeng; methodology, Yuting Chen; software, Yuting Chen and Yue Zeng; validation, Gaoyang Liu and Yue Zeng; formal analysis, Yuting Chen; investigation, Yuting Chen; resources, Gaoyang Liu and Yue Zeng; data curation, Yuting Chen; writing—original draft preparation, Yuting Chen and Yue Zeng; writing—review and editing, Yuting Chen and Yue Zeng; visualization,Yuting Chen and Yue Zeng; supervision, Gaoyang Liu and Yue Zeng; project administration, Gaoyang Liu; funding acquisition, Gaoyang Liu. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This work was financially supported by the Shaoxing Science and Technology Planning Project (No. 2024B13004).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to acknowledge the financial support to this project by the Shaoxing Science and Technology Planning Project (No. 2024B13004).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. A.A. Akinbami, Integrating natural light for wellbeing, performance, and quality care delivery in healthcare environments, (2024). https://doi.org/10.24382/5191.
2. S. Idrissi Kaitouni, P. Sangkyu, M.O. Mghazli, F. El Mansouri, A. Jamil, M. Ahachad, J. Brigui, Design parameters influencing the energy performance and indoor comfort of net zero energy building "NZEB" designed for semi-arid urban areas: digital workflow methodology, sensitivity analysis and comparative assessment, Sol. Energy 268 (2024) 112264. https://doi.org/10.1016/j.solener.2023.112264.
3. L. Le-Thanh, H. Nguyen-Thi-Viet, J. Lee, H. Nguyen-Xuan, Machine learning-based real-time daylight analysis in buildings, J. Build. Eng. 52 (2022) 104374. https://doi.org/10.1016/j.jobe.2022.104374.

4. M.I. Ayoosu, A.L. Utsaha, K.E. Gabriel, A.M.M. Vishigh, M.E. Tuleun, I.G. Sen, M.I. Ayoosu, A.L. Utsaha, K.E. Gabriel, A.M.M. Vishigh, M.E. Tuleun, I.G. Sen, Daylighting performance assessment: a review of methodologies, Path Sci. 11 (2025) 8001–8011. https://doi.org/10.22178/pos.113-16.
5. L. Giovannini, V. Lo Verso, L. Valetti, J. Daltrozo, A. Pellegrino, Analysis of integrative lighting through field measurements and annual daylight simulations in offices, Lighting Research & Technology 57 (2025) 421–450. https://doi.org/10.1177/14771535241311606.
6. S. Jain, V. Garg, A review of open loop control strategies for shades, blinds and integrated lighting by use of real-time daylight prediction methods, Build. Environ. 135 (2018) 352–364. https://doi.org/10.1016/j.buildenv.2018.03.018.
7. A review of daylighting calculation metrics, variables and methods in evaluating the daylight performance of a residential building | AIP conference proceedings | AIP publishing, (n.d.). https://pubs.aip.org/aip/acp/article-abstract/2813/1/020016/2904941/A-review-of-daylighting-calculation-metrics (accessed January 10, 2026).
8. I.L. Wong, A review of daylighting design and implementation in buildings, Renew. Sustain. Energy Rev. 74 (2017) 959–968. https://doi.org/10.1016/j.rser.2017.03.061.
9. Q. Li, J.C. Baltazar, J. Haberl, Prediction of daylighting performance using machine learning algorithm based on radiance simulated data, in: IBPSA, 2023: pp. 1476–1482. https://doi.org/10.26868/25222708.2023.1212.
10. H. Nourkojouri, Z.S. Zomorodian, M. Tahsildoost, Z. Shaghaghian, A machine-learning framework for daylight and visual comfort assessment in early design stages, (2021). https://doi.org/10.48550/arXiv.2109.06450.
11. M.I. Ayoosu, Y.-W. Lim, P.C. Leng, T.T. Aule, A.B. Abdurrahman, S. Aminu, Tropical daylight availability and sky typologies for daylighting evaluation and design, J. Tour. Hosp. Environ. Manag. 7 (2022) 157–170. https://doi.org/10.35631/JTHEM.727013.
12. H. Nourkojouri, N.S. Shafavi, M. Tahsildoost, Z.S. Zomorodian, Development of a machine-learning framework for overall daylight and visual comfort assessment in early design stages, J. Daylighting 8 (2021) 270–283. https://doi.org/10.15627/jd.2021.21.
13. D.H.W. Li, S. Li, W. Chen, S. Lou, Simple correlations between point daylight factor, average daylight factor and vertical daylight factor under all sky conditions and building design implications, Indoor Built Environ. 31 (2022) 1700–1714. https://doi.org/10.1177/1420326X211061111.
14. Y.-W. Lim, M.H. Ahmad, D.R. Ossen, Internal shading for efficient tropical daylighting in malaysian contemporary high-rise open plan office, Indoor Built Environ. 22 (2013) 932–951. https://doi.org/10.1177/1420326X12463024.
15. A. Shehadeh, O. Alshboul, Enhancing occupational safety in construction: predictive analytics using advanced ensemble machine learning algorithms, Eng. Appl. Artif. Intell. 159 (2025) 111761. https://doi.org/10.1016/j.engappai.2025.111761.
16. J. Ngarambe, I. Adilkhanova, B. Uwiragiye, G.Y. Yun, A review on the current usage of machine learning tools for daylighting design and control, Build. Environ. 223 (2022) 109507. https://doi.org/10.1016/j.buildenv.2022.109507.
17. S.K.T. M, C.P. Kurian, S.G. Colaco, V. Mathew, Machine learning model for glare prediction in offices with simple architectural features, (2022). https://doi.org/10.3992/jgb.17.4.79.
18. M. Ayoub, A review on machine learning algorithms to predict daylighting inside buildings, Sol. Energy 202 (2020) 249–275. https://doi.org/10.1016/j.solener.2020.03.104.
19. A.W.M. Ibrahim, Predicting glare in open-plan offices using simplified data acquisitions and machine learning algorithms, QUT ePrints (n.d.). https://eprints.qut.edu.au/204266/ (accessed January 10, 2026).
20. K. Radziszewski, M. Radziszewska, Machine learning algorithm-based tool and digital framework for substituting daylight simulations in early-stage architectural design evaluation, 2018. https://doi.org/10.22360/simaud.2018.simaud.001.
21. K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095.

 doi:10.20944/preprints202604.1668.v1

22. D. Lowing, K. Techer, Toward a consensus on extended Shapley values for multi-choice games, Mathematical Social Sciences (2025) 102407. https://doi.org/10.1016/j.mathsocsci.2025.102407.
23. I. Mashaly, M. El-Hussainy, A. Sherif, K. Tarabieh, Daylighting performance prediction tool for early design stages using machine learning, J. Build. Eng. 111 (2025) 113496. https://doi.org/10.1016/j.jobe.2025.113496.
24. J. Ngarambe, A. Irakoze, G.Y. Yun, G. Kim, Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances, Sustainability 12 (2020) 4471. https://doi.org/10.3390/su12114471.
25. S. Vanage, H. Dong, K. Cetin, Visual comfort and energy use reduction comparison for different shading and lighting control strategies in a small office building, Sol. Energy 265 (2023) 112086. https://doi.org/10.1016/j.solener.2023.112086.
26. C.T. Do, N.T.T. Dung, Y.-C. Chan, Daylighting assessment of window layouts and architectural elements in early design stages, Sci. Rep. 15 (2025) 39507. https://doi.org/10.1038/s41598-025-23389-x.
27. Standard for Daylighting Design of Buildings, (n.d.).
28. H. Lian, Y. Ji, M. Niu, J. Gu, J. Xie, J. Liu, A hybrid load prediction method of office buildings based on physical simulation database and LightGBM algorithm, Appl. Energy 377 (2025) 124620. https://doi.org/10.1016/j.apenergy.2024.124620.
29. W. Tian, R. Qiu, Z. Tang, Y. Zhong, X. Liu, X. Wang, F. Zou, R. Li, XGBoost assisted in identifying the structural damage behavior of fiber-reinforced concrete during the soundless demolition process through electrical resistance changes, Construction and Building Materials 478 (2025) 141376. https://doi.org/10.1016/j.conbuildmat.2025.141376.
30. S.A. Emami, P. Castaldi, A. Banazadeh, Neural network-based flight control systems: present and future, Annu. Rev. Control 53 (2022) 97–137. https://doi.org/10.1016/j.arcontrol.2022.04.006.