

Landscape Visual Perception Prediction and Spatial Optimization Using XGBoost Ensemble Learning

Yanhui Sui

Associate Professor, School of Architecture, Weihai Vocational College, Weihai, 264210, China, E-mail: rczhj0220@163.com

Project Management

Received February 28, 2026; revised June 11, 2026; accepted August 31, 2026
Available online September 10, 2026

Abstract: To overcome the difficulty of accurately analyzing complex nonlinear coupling mechanisms in Landscape Visual Perception (LVP) assessment and the lack of scientific decision-making support at the spatial optimization level, this research constructs a set of technical methods that integrate high-precision prediction and intelligent diagnosis to optimize space. This research proposes an integrated learning model based on eXtreme gradient boosting that deeply integrates three-dimensional spatial morphology and visual field perception indicators, and introduces 2nd-order Taylor expansion and regularization techniques to optimize the solution of the loss function. Furthermore, it uses the Shapley additive exPlanations algorithm to build a complete logical chain from feature attribution to spatial intervention. The results revealed that this method achieved a test-set coefficient of determination of 0.812 and a median absolute error of 0.25 in a heterogeneous multi-source data environment, significantly outperforming deep neural networks and traditional regression methods. In the spatial optimization simulation experiment, differentiated and precise intervention implemented based on the system attribution results increased the perceptual scores of typical low-quality block samples by an average of 37.2%, with a maximum improvement of 124.2%, verifying its decision-making efficiency under limited resources. This research forms a closed-loop path of "problem diagnosis-attribution interpretation-precise intervention", providing a scientific basis for urban stock renewal and refined management of human settlements.

Keywords: XGBoost; landscape visual perception; spatial optimization; shapley additive explanations; ensemble learning.

Copyright © Journal of Engineering, Project, and Production Management (EPPM-Journal).
DOI 10.32738/JEPPM-2026-270

1. Introduction

1.1. Research Motivation

Landscape Visual Perception (LVP) serves as the core link connecting people and the environment. Its quantitative prediction and spatial optimization have become key propositions in landscape planning, urban and rural design and other fields (Weber-Lewerenz and Traverso, 2023). LVP is influenced by the interplay of multi-dimensional factors, including natural elements, spatial form and humanistic cognition, and its internal mechanism exhibits complex nonlinear characteristics. This leads to significant challenges in accurately capturing perception patterns and scientifically deriving optimization paths (Williams, 2022). Traditional LVP assessment methods, which rely on subjective evaluation and single-dimensional quantitative analysis, are prone to individual bias and struggle to systematically integrate multi-source heterogeneous data or analyze complex relationships (Zhang et al., 2025). Furthermore, while basic data modeling technologies have been introduced, they often rely on single models that lack deep exploration of data correlations, resulting in poor generalization and a trade-off between prediction accuracy and spatial adaptability (Li et al., 2026). Therefore, constructing a technical framework that effectively integrates multi-source information, accurately characterizes nonlinear sensing mechanisms, and scientifically supports spatial optimization is a core issue in this field.

1.2. Related Work

Many scholars focus on LVP prediction modeling and spatial optimization. For example, Yue et al. (2023) used a Gaussian mixture model to predict soundscape evaluations and to provide visualization strategies for urban parks, while Freitas et al. (2023) integrated machine learning, computer vision, and interactive visualization technologies to analyze multi-sequential street-view images in studying urban LVP modeling. Additionally, Diao et al. (2025) proposed a new "nearest optimal network" method for analyzing the fitness landscape in evolutionary calculations.

Extreme Gradient Boosting (XGBoost), an optimized ensemble learning algorithm within the gradient boosting

framework, has been widely used in fields such as finance, healthcare, and e-commerce, and is especially popular for predictive modeling. By serially integrating weak learners and introducing regularization techniques, it offers significant advantages in processing complex data and mining deep correlations, demonstrating strong generalization capabilities and high prediction accuracy (Asselman et al., 2023). To demonstrate this, Mustapha et al. (2024) developed an XGBoost-based prediction model to address the high cost and time-consuming experimental measurement of the physical and mechanical properties of permeable concrete. Similarly, Schimohr et al. (2023) showed that XGBoost yielded lower errors and higher fit quality than traditional models in spatial distribution predictions for bike-sharing. Furthermore, Karimzadeh et al. (2024) used the stochastic finite-fault method to simulate and generate region-specific homogeneous datasets and introduced the XGBoost model to address the limitations of traditional linear regression.

1.3. Research Objective and Highlights

To sum up, existing related research has accumulated experience in the technical exploration of LVP prediction modeling and in the practice of high-precision prediction across multiple fields using XGBoost. However, there are still problems, including insufficient characterization of the multi-element nonlinear coupling mechanism in LVP, insufficient integration of multi-source 3D landscape information, and difficulty balancing model generalization and spatial adaptability. Therefore, this study proposes a solution that centers on XGBoost ensemble learning and integrates 3D landscape feature engineering with Shapley Additive Explanations (SHAP). It primarily consists of two levels: LVP prediction and spatial optimization. In terms of LVP prediction, this research uses XGBoost's serially integrated weak learner and regularization features to enhance the model's ability to process complex landscape data. Moreover, using its 2nd-order gradient approximation mechanism, it can solve the nonlinear modeling problem of LVP multi-dimensional element coupling. In terms of spatial optimization, this research integrates 3D terrain interactive features and SHAP interpretation strategies to build a causally transparent "prediction-optimization" framework, thereby providing a scientific technical path for intelligent urban landscape management and ecological civilization construction. From a theoretical perspective, the proposed framework treats LVP as a measurable interface between urban spatial form and human perceptual response. In this framework, Three-Dimensional (3D) landscape indicators represent the physical conditions of urban space, while predicted LVP scores reflect the perceptual consequences of these spatial conditions. The modeling process, therefore, not only estimates perception values but also translates complex spatial-perceptual relationships into interpretable evidence for planning decisions.

The decision-making logic is established through the connection between prediction accuracy and feature attribution. XGBoost captures nonlinear relationships among terrain, building morphology, visibility, and visual indicators, while SHAP identifies the positive or negative contribution of each factor to the predicted LVP outcome. Features with negative contributions indicate spatial constraints that should be prioritized for intervention, whereas features with positive contributions indicate spatial qualities that should be protected or strengthened. In this way, LVP modeling becomes a diagnostic basis for urban planning, supporting targeted decisions such as greening enhancement, visual corridor adjustment, enclosure control, and street interface optimization.

2. Methods and Materials

2.1. High-Dimensional Quantification and Engineering Construction of 3D Landscape Features

To address plane-indicator limits in vertical spaces, a 3D, form-centric index system is constructed. The technical path and key elements of this feature of engineering are shown in Fig. 1.

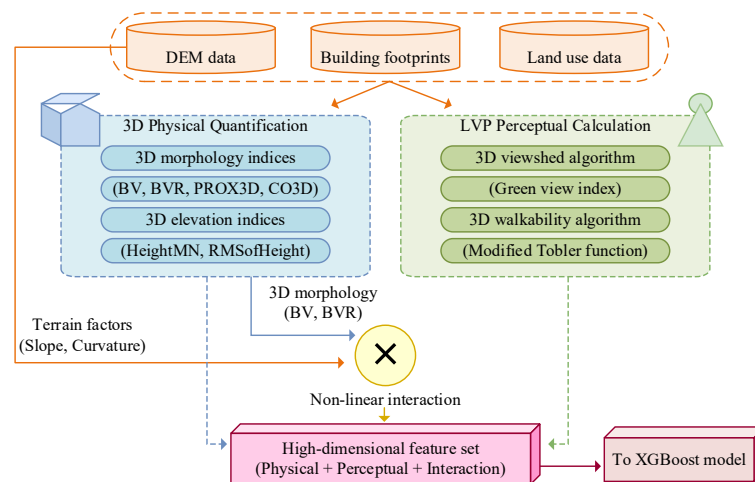


Fig.1. 3D urban landscape feature engineering flow chart

Fig. 1 illustrates the extraction of 3D physical and visual features from geographic data into a high-dimensional set. Integrating DEM, building bases, and land use yields 3D morphological and visual indicators (Xue et al., 2022). A terrain-architecture nonlinear interaction term quantifies cumulative landscape occlusion (Paláncz and Awange, 2013). Precise visual quantification employs an improved 3D visual field algorithm. The specific calculation mechanism is shown in Fig. 2.

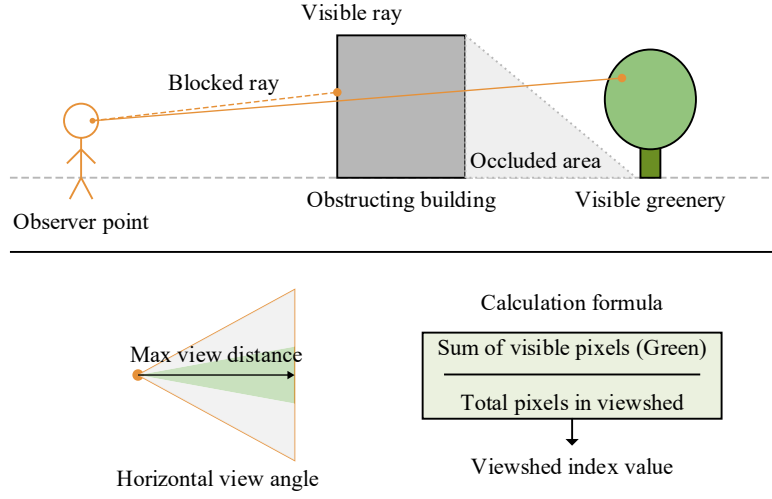


Fig. 2. 3D landscape visual domain computer graphics

Fig. 2 shows the 3D visual field algorithm that simulates sightlines to identify visible green spaces at building intersections. By generating dense line-of-sight beams from the observation point and calculating intersection intervals with 3D building bounding boxes, the algorithm effectively determines line-of-sight occlusion (Yuan et al., 2024). Based on these occlusion results, the visual perception characteristics are quantified. Among them, the landscape visibility rate (visual perception characteristics) is calculated as shown in Eq. (1).

$$V = \frac{N_{vis}}{N_{total}} \times 100\% \quad (1)$$

In Eq. (1), N_{total} is an unblocked visible pixel, and V is the percentage-based LVP. Combining V The modified Tobler function-based accessibility completes the 3D feature engineering for model training (Qin et al., 2023).

2.2. Construction and Optimization of XGBoost High-Precision Prediction Model

2.2.1. Integrated prediction architecture and objective function construction

To analyze high-dimensional nonlinear relationships (Grotti et al., 2023), an XGBoost model is built. The specific architectural design of this model is shown in Fig. 3.

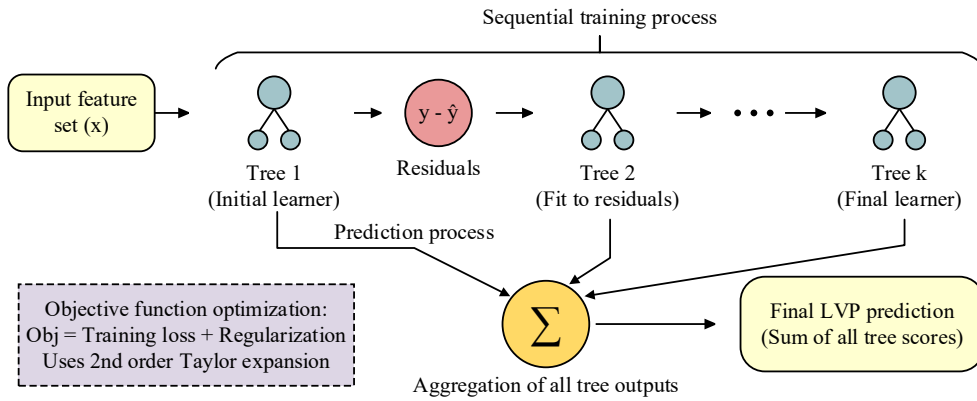


Fig. 3. XGBoost ensemble learning model architecture diagram

Fig. 3 shows XGBoost serially integrating decision trees to fit residuals for LVP predictions. Instead of traditional linear regression, this study uses XGBoost, which is based on a 2nd-order Taylor expansion and L2 regularization, to construct the objective function, thereby effectively capturing nonlinear coupling mechanisms while avoiding overfitting (Grotti et al., 2023; Zhang et al., 2023).

2.2.2. Predictive model optimization solution and training process

During optimization, the objective function is efficiently solved using 2nd-order gradient information. The specific optimization mechanism is shown in Fig. 4.

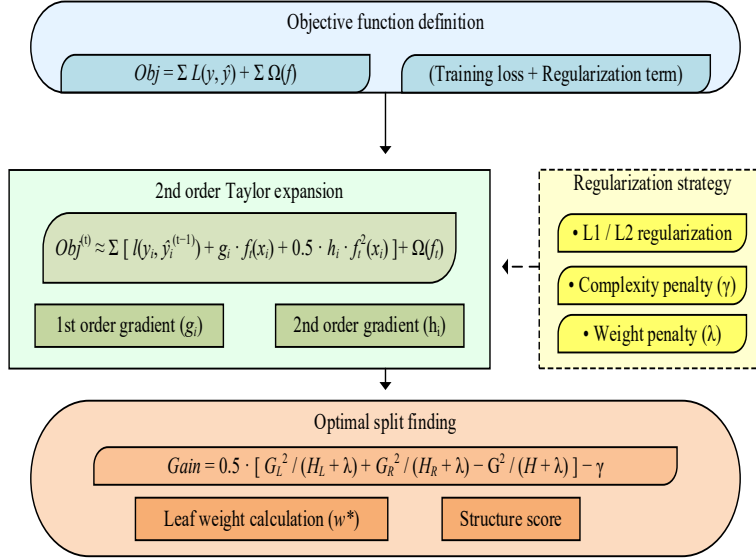


Fig. 4. XGBoost prediction model optimization mechanism diagram

Fig. 4 illustrates the use of Taylor expansion and regularization strategies for optimal splits and leaf weights. By retaining 1st- and 2nd-order derivative information, along with complexity and weight penalty terms, the model derives optimal split gains and leaf node weights to guide decision tree growth (Nguyen et al., 2023; Costa and Pedreira, 2022). The complete training and parameter iteration process after optimization is shown in Fig. 5.

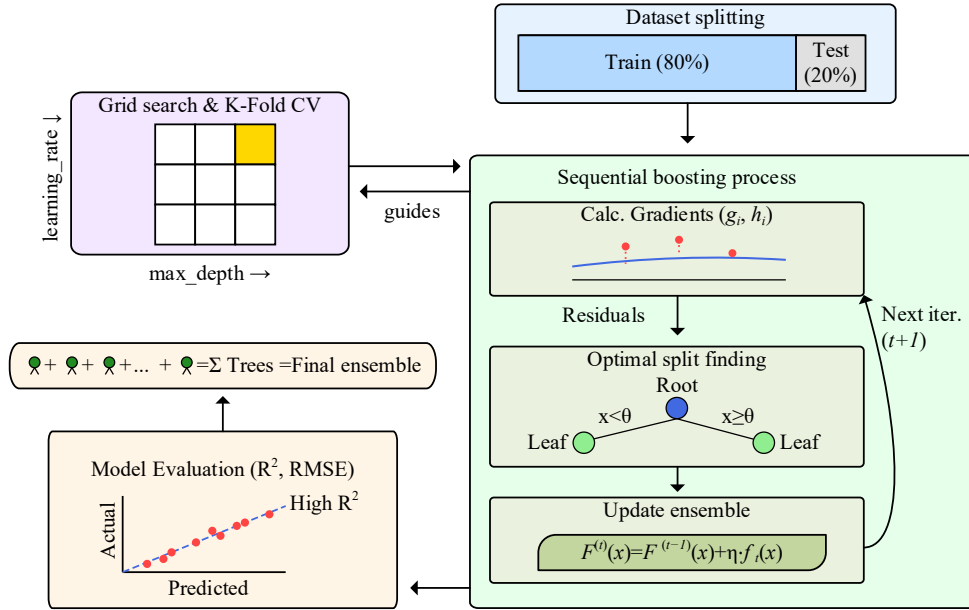


Fig. 5. XGBoost high-precision prediction training flow chart

Fig. 5 details data splitting, hyperparameter tuning, and sequential boosting. Data is split (8:2), hyperparameters tuned via grid search/cross-validation (Bates et al., 2023; Shams et al., 2024), and boosting fits residuals (Tan et al., 2023; Gunasekara et al., 2024). Model performance is comprehensively evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE) (Li et al., 2023; Liu et al., 2023).

2.3. SHAP Interpretation-Driven Spatial Optimization Framework Generation

To overcome XGBoost's black-box nature, SHAP decomposes forecasts into feature-level drivers, enabling causal optimization. The logical structure of this feature interpretation and optimization framework is shown in Fig. 6.

Fig. 6 outlines SHAP decomposing predictions into drivers to evaluate priority and value-added areas. Global ranking and local dependency graphs visualize nonlinear responses (Lee et al., 2023) and generate a spatial decision chain (Chen et al., 2023). Among them, the nonlinear interaction mechanism between terrain factors and 3D morphology is shown in Fig. 7.

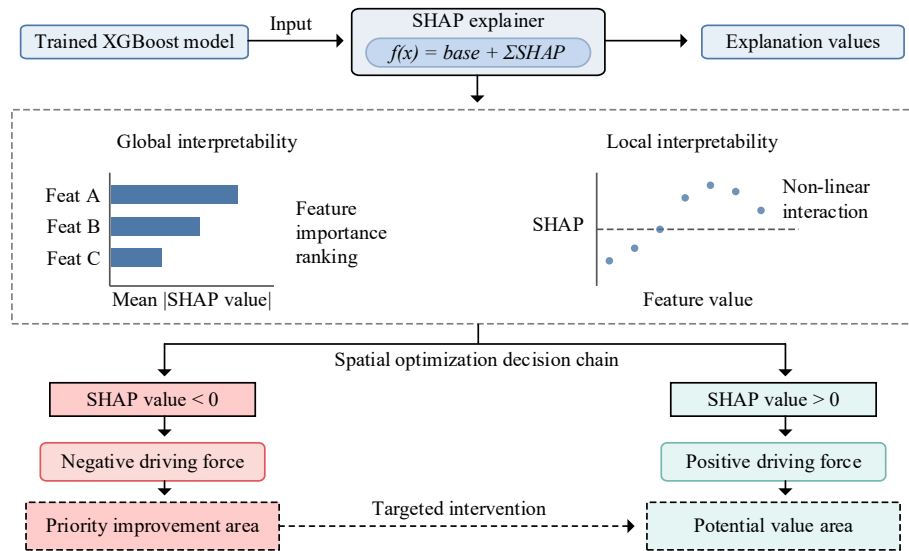


Fig. 6. SHAP feature contribution interpretation diagram

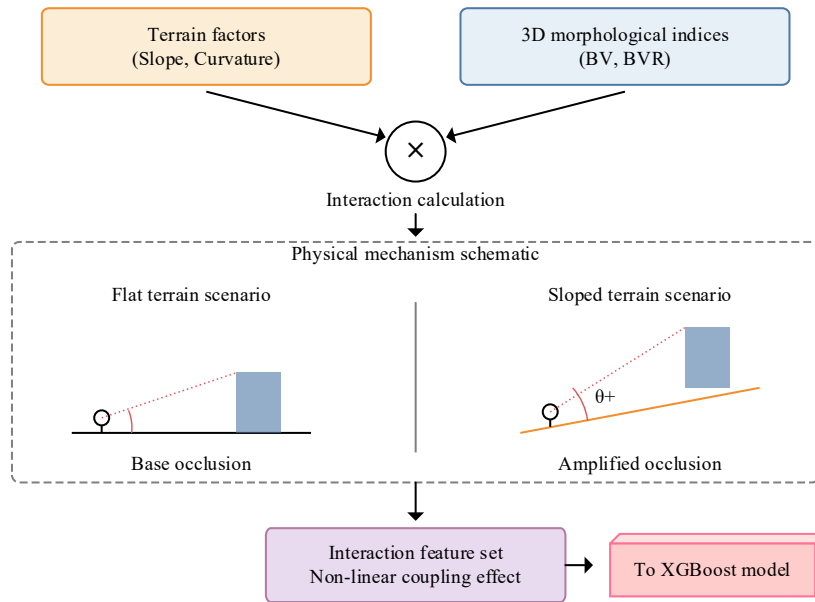


Fig. 7. LVP terrain interaction effect mechanism diagram

Fig. 7 shows rising slopes nonlinearly amplifying line-of-sight occlusion. This interactive mechanism captures terrain constraints on visual perception (Labetski et al., 2022). Locating abnormal SHAP values enables targeted morphological control, driving the LS-XG method. This study proposes an LVP prediction and spatial optimization method driven by XGBoost Ensemble Learning (LS-XG). Compared with existing XGBoost-SHAP applications that mainly use SHAP values to rank feature importance or explain prediction results, the methodological contribution of LS-XG lies in its spatially operational use of attribution information. First, the model input is not limited to conventional statistical or planar environmental variables, but integrates 3D landscape morphology, visual-field indicators, and terrain-building interaction features. Second, SHAP outputs are further converted into a spatial diagnosis mechanism, in which negative feature contributions identify priority intervention targets and positive contributions indicate spatial qualities to be maintained. Therefore, the framework extends XGBoost-SHAP from a post-hoc interpretation tool to a planning-oriented optimization method that links prediction, attribution, and targeted spatial intervention.

3. Results

3.1. Performance Comparison Verification of LS-XG

This study constructs a rigorous comparative experimental framework. The experiment selects Lasso Regression (Lasso), the Light Gradient Boosting Machine (LightGBM), and a Deep Neural Network (DNN) as benchmark models to evaluate improvements across linear regularization, ensemble algorithms, and deep learning. The data source uses the authoritative Place Pulse 2.0 dataset (official data archive: Figshare, DOI: 10.6084/m9.figshare.11859993).



Fig. 8. Example of Place Pulse 2.0 dataset (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

As shown in Fig. 8, an example subset of the Place Pulse 2.0 dataset comprises globally crowdsourced street-view image pairs that are evaluated across multiple perceptual dimensions, such as aesthetics and safety, to generate quantitative ratings for each spatial sample. After data cleaning, outlier removal, and matching with the 3D high-dimensional feature set, it is randomly divided into a training set and a test set in a ratio of 8:2. The experiments are uniformly deployed in the Python 3.8 compilation environment and equipped with NVIDIA RTX 3090 GPU for parallel computing acceleration.

3.1.1. Multi-model comparison of LVP prediction accuracy

Based on the unified test benchmark, a multi-model comparison is conducted to quantitatively evaluate the fitting performance of LVP prediction accuracy. The specific results are shown in Fig. 9.

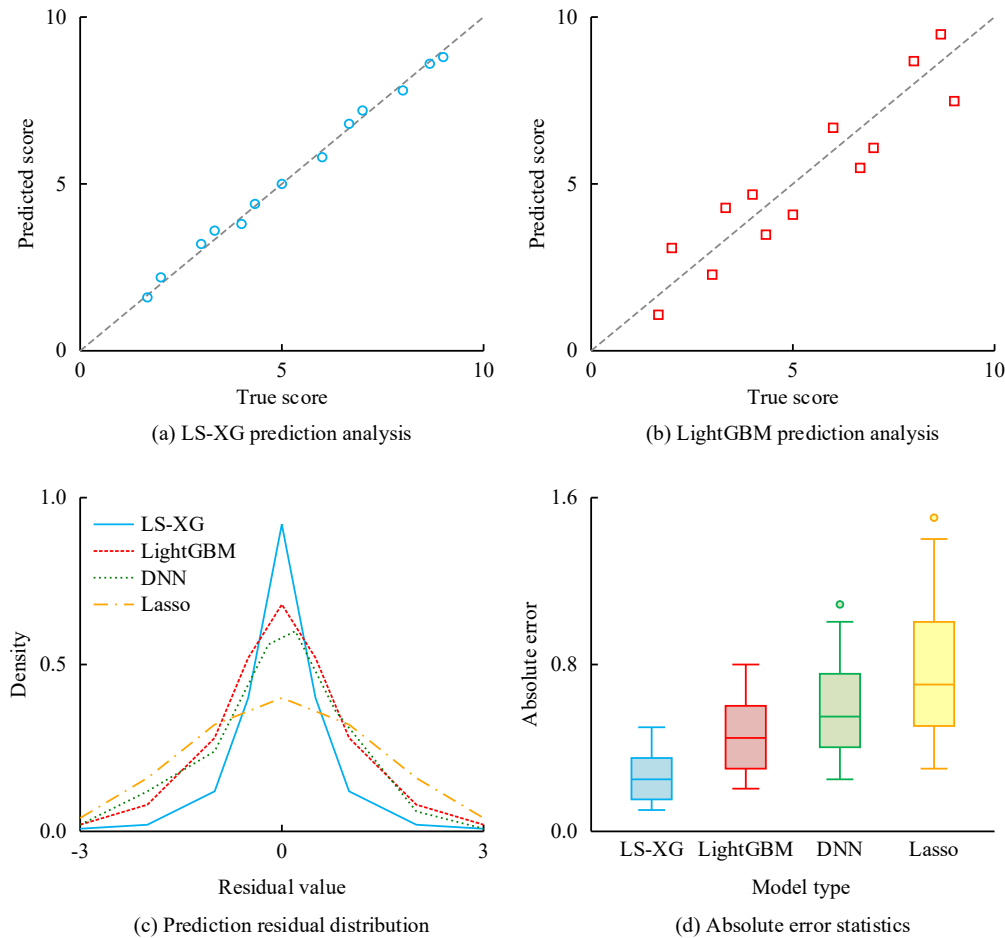


Fig. 9. Multi-model comparative analysis of LVP prediction accuracy

Fig. 9 provides a multi-model comparative analysis of LVP prediction accuracy, illustrating scatter plots, residual distributions, and error metrics. As shown in Fig. 9, the LS-XG model exhibits excellent nonlinear fitting aligned to the diagonal, significantly outperforming LightGBM's higher dispersion (Fig. 9(a), (b)). Its residual distribution peaks at 0.95 at the zero point with the narrowest convergence interval (Fig. 9(c)). Furthermore, LS-XG achieves a median absolute error of 0.25 (interquartile range: 0.20) without outliers (Fig. 9(d)), far surpassing DNN (0.55) and Lasso (0.70), strongly confirming its advantage in accuracy.

3.1.2. Validity verification of spatial form optimization decision-making mechanism

To evaluate the model's effectiveness in guiding spatial optimization, cumulative gain and Precision-Recall Curves (PRC) are used to assess its ability to capture high-perceived-value areas across different thresholds. The specific results are shown in Fig. 10.

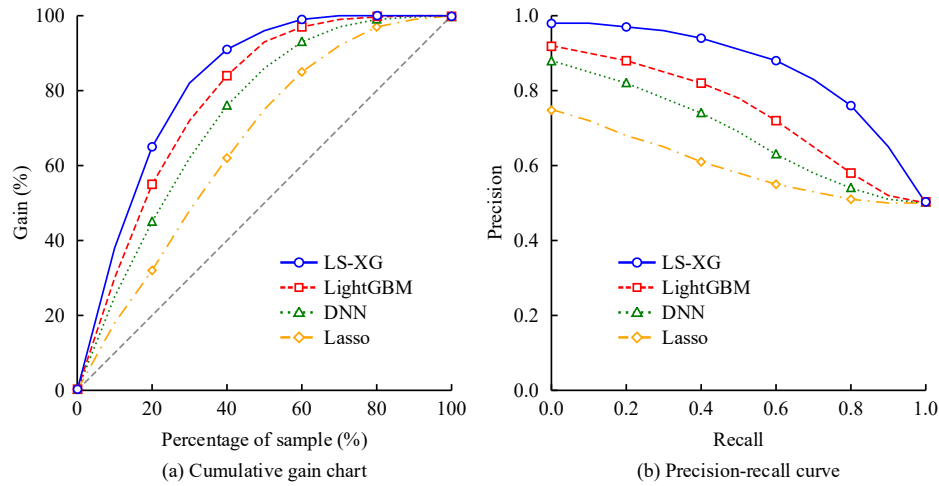


Fig. 10. Comparison of effectiveness for spatial form optimization

Fig. 10 compares the effectiveness of spatial form optimization across models using cumulative gain and PRC metrics. As shown in Fig. 10, LS-XG exhibits significant early recognition advantages, capturing 65% of high-perceived-value areas with only 20% of the sample size (Fig. 10(a)), thereby completely outperforming LightGBM (55%) and DNN (45%). Additionally, it covers the largest PRC area and maintains a precision of 0.76 with a high recall rate of 0.8 (Fig. 10(b)). This demonstrates its ability to effectively balance recall and precision during spatial interventions, minimizing the risk of misjudgment and confirming its decision-making reliability. To translate these abstract ranking curves into recognizable urban conditions, the high-, medium-, and low-ranked samples identified by the cumulative-gain analysis are illustrated with representative street-view scenes in Fig. 11.



Fig. 11. Representative practical spatial forms corresponding to spatial-optimization ranking (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

The optimum form (Fig. 11(a)) is characterized by a high green-view ratio, a continuous tree canopy, and a natural roadside enclosure, with balanced sky openness; such scenes are concentrated in the leading 20% of the gain curve, explaining why LS-XG captures 65% of high-perceived value areas at this early threshold. The sub-optimum form (Fig. 11(b)) shows a moderate green view supported by intermittent street trees and a pleasant residential interface, but with a more exposed sky and weaker vertical greenery; these samples occupy the middle segment of the curve, where the precision–recall trade-off becomes the decisive factor. The non-optimum form (Fig. 11(c)) is dominated by a hard, wire-fence- and road-oriented interface with a low green view and limited human-scale enclosure, corresponding to the tail of the ranking, where competing models such as LightGBM and DNN misclassify low-value scenes as high-value ones. By aligning each curve segment with a concrete spatial form, Fig. 11 demonstrates that the early recognition advantage of LS-XG is not merely a statistical artifact but reflects a physically interpretable gradient of spatial quality, thereby supporting reliable prioritization of intervention targets in real street environments.

3.1.3. Overall performance and robustness evaluation of LS-XG

To confirm stability against real-life urban sample unevenness and noise interference, the model's generalization boundary and robustness are evaluated across different training set proportions and characteristic noise intensities.

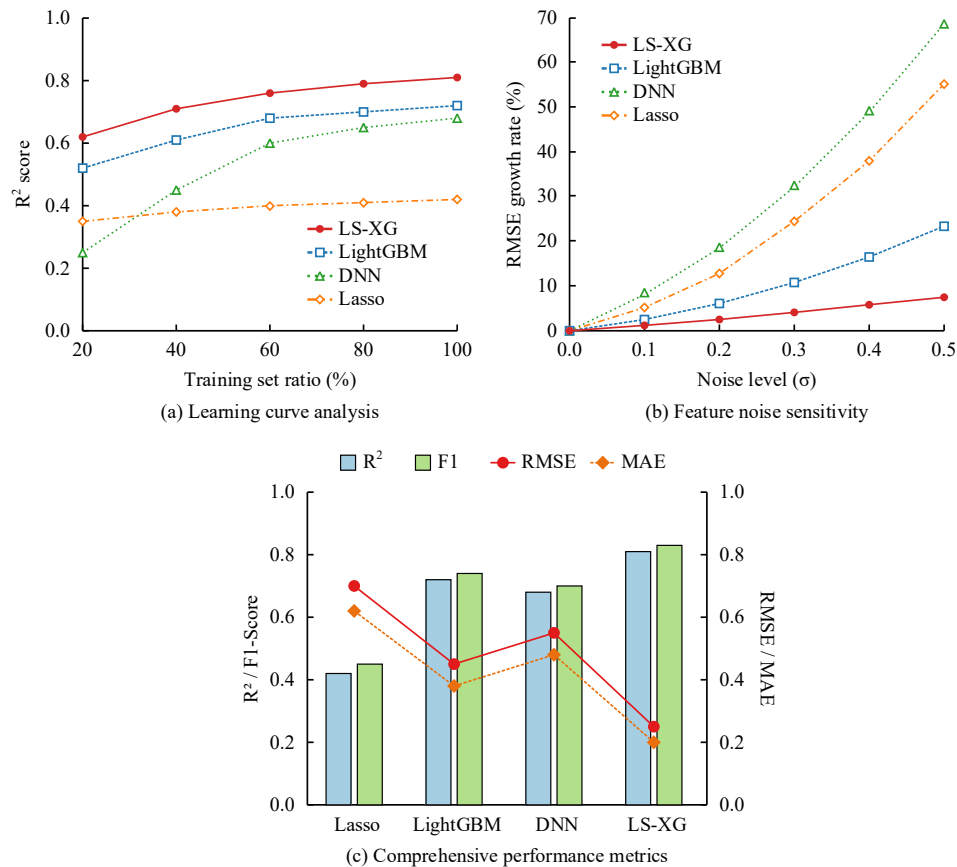


Fig. 12. Comprehensive performance index and robustness comparison

Fig. 12 presents a comprehensive performance index and robustness comparison across varying data scales and noise levels. As shown in Fig. 12, LS-XG achieves an R² score of 0.62 with just 20% of the training data, steadily rising to 0.81 as the sample size increases (Fig. 12(a)), demonstrating superior small-sample generalization over data-reliant DNNs. Furthermore, as noise intensity increases, LS-XG's RMSE growth rate increases slightly to 7.5% (Fig. 12(b)), far lower than DNN (>60%) and Lasso, reflecting extreme robustness in complex environments. Ultimately, LS-XG delivers the best comprehensive indicators, including an R² of 0.81 and an RMSE of 0.25 (Fig. 12(c)), thereby confirming its high predictive accuracy and decision-making stability. To make the accuracy curves in Fig. 12 spatially interpretable, the high-, medium-, and low-fitted samples are illustrated with representative street-view scenes in Fig. 13.

The R² is directly associated with these forms through the consistency of their visual features. For the optimum form (Fig. 13(a)), the continuous tree canopy and high, spatially stable green view produce low within-scene feature variance, so the model explains almost all perceptual variation, and the local fit approaches the upper R² of 0.81. The sub-optimum form (Fig. 13(b)) combines moderate street-tree greenery with a more exposed sky, introducing greater feature variance and a correspondingly moderate R². The non-optimum form (Fig. 13(c)), dominated by a hard road interface and a sparse green view, offers few high-value visual cues, thereby weakening the explainable signal and lowering the local R². Such scenes are also the most exposed to the noise that the 7.5% RMSE-growth test simulates. This form-by-form correspondence shows that the rising R² in Fig. 13a is not an abstract statistical gain but tracks a physically interpretable improvement in spatial quality from non-optimum to optimum environments.



Fig. 13. Representative practical spatial forms corresponding to prediction-accuracy results (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

3.2. Ablation Experimental Analysis of LS-XG

To verify the independent contributions of high-dimensional feature engineering and 2nd-order optimization strategies, a modular ablation experiment is constructed using the Place Pulse 2.0 dataset after outlier removal and high-dimensional spatial feature matching. All models are deployed in a Python 3.8 environment with an NVIDIA RTX 3090 GPU for accelerated parallel computing. The specific configuration and core variables of each experimental group are shown in Table 1.

Table 1. LS-XG core module ablation experiment settings and configuration table

Experimental group	Model variant	Feature set composition	Optimization strategy	Expected R^2
Baseline	LS-XG w/o-Vis	Physical morphology indices only	2nd-order Taylor expansion	0.652
Ablation 1	LS-XG w/o-Int	Physical+visual (no non-linear interaction)	2nd-order Taylor expansion	0.745
Ablation 2	LS-XG w/o-Opt	Full feature set (With interaction)	1st-order gradient descent	0.768
Proposed	LS-XG (Full)	Full feature set (With interaction)	2nd-order Taylor expansion	0.812

Table 1 details the gradient progressive ablation contrast scheme, outlining the feature set composition and optimization strategies across different model variants. Compared to the full-featured LS-XG, the models selected include LS-XG without visual features (w/o-Vis) to evaluate the limitations of a single physical form dimension, LS-XG without the nonlinear interaction term (w/o-Int) to verify the explanatory power of complex spatial coupling features, and LS-XG without 2nd-order optimization (w/o-Opt) replacing 2nd-order Taylor expansion with traditional 1st-order gradient descent to test optimization contributions to convergence accuracy (Table 1). Finally, the complete LS-XG integrates all features and the 2nd-order optimization mechanism, serving as a comparison top layer to assess overall performance.

3.2.1. Contribution analysis of feature engineering and optimization strategies

To quantify module contributions and verify the core roles of nonlinear interaction terms and 2nd-order optimization strategies, the study compares prediction accuracy and error convergence across ablation variants using the control variable method. The specific results are shown in Fig.14.

Regarding the error-density comparison in Fig. 15(a), the optimum form (Fig. 15(a)) exhibits a strong coupling between a high green view and balanced sky openness; the nonlinear interaction terms capture this coupling, which is why the full model attains a sharp error-density peak at 0.82, whereas removing the interaction terms flattens the peak to 0.45 for exactly these visually rich scenes. The suboptimal form (Fig. 15(b)) shows only partial green-sky coupling, yet the 2nd-order optimization still rapidly drives the training loss down and converges to 0.08, demonstrating stable behavior on intermediate scenes. The non-optimum form (Fig. 15(c)) is dominated by a hard wire-fence-road interface with weak feature coupling; without the interaction terms and 2nd-order information, the model struggles to fit these samples and is prone to local minima that the convergence curve in Fig. 15(b) avoids. Thus, the abstract density and loss curves map onto a clear gradient of real spatial forms, confirming that both the nonlinear coupling features and the 2nd-order strategy are most decisive precisely where the spatial form is more complex.

3.2.2. Spatial sorting capability and system comprehensive performance analysis

To confirm LS-XG's practical value in spatial optimization decision-making, this study compares spatial sorting capabilities, recognition accuracy for high-perceptual-value areas, and robustness to outlier samples in complex terrain. The specific results are shown in Fig.16.

Fig. 16 evaluates LS-XG's spatial sorting capability and overall performance against ablation variants across R2, NDCG, and absolute error. As shown in Fig. 16, the complete LS-XG exhibits the best overall performance with an R2 of 0.812 and an NDCG@20% of 0.854, while the gradient increase across variants confirms the key role of visual and interactive features in improving sorting accuracy (Fig. 16(a)). During complex terrain tests, LS-XG maintains a dense box distribution with a median absolute error as low as 0.15 and few outliers, whereas the w/o-Vis model's median error rises to 0.40 with maximum dispersion (Fig. 16(b)). This demonstrates the complete model's robustness and decision-making reliability on difficult spatial samples in concrete terms, as shown in Fig. 17.

The optimum sample (Fig. 17(a)) is an unambiguous high-value scene where the full model ranks at the top with a low absolute error of about 0.15, consistent with its dense box distribution. The sub-optimal sample (Fig. 17(b)) is a mid-ranked residential scene whose prediction remains stable, with few outliers, even under terrain variation. The non-optimum sample (Fig. 17(c)), a hard road-and-fence scene with a low green view, is the type of difficult sample where the ablated w/o-Vis

model degrades sharply to a median error of 0.40 with maximum dispersion, whereas the complete LS-XG keeps the error low and the ranking correct. By showing that the model maintains accurate predictions across the full range of real spatial forms, including the most challenging non-optimum scene, Fig. 17 provides practical evidence of the excellent robustness and decision-making reliability reported in Fig. 17(b).

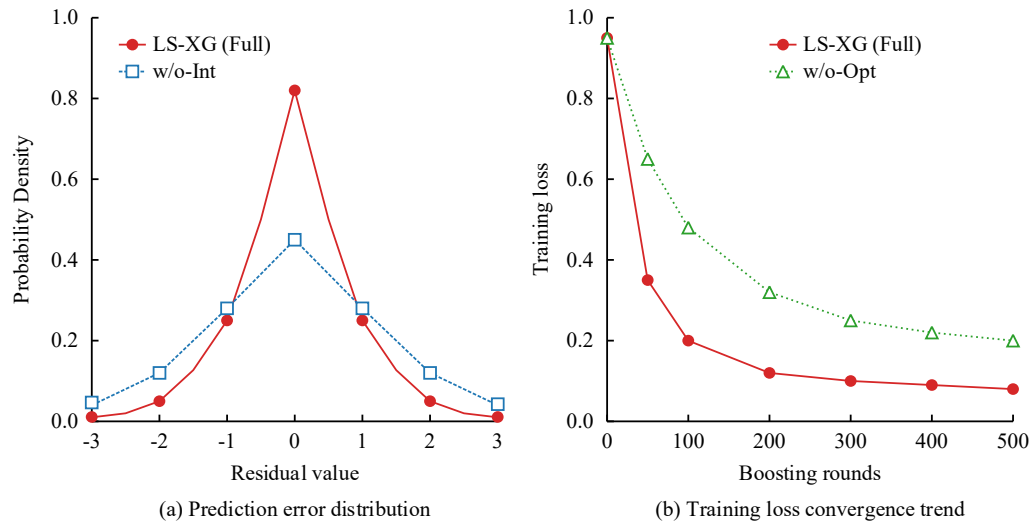


Fig.14. Contribution analysis of feature engineering and optimization strategies in LS-XG



Fig. 15. Representative practical spatial forms corresponding to the ablation results (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

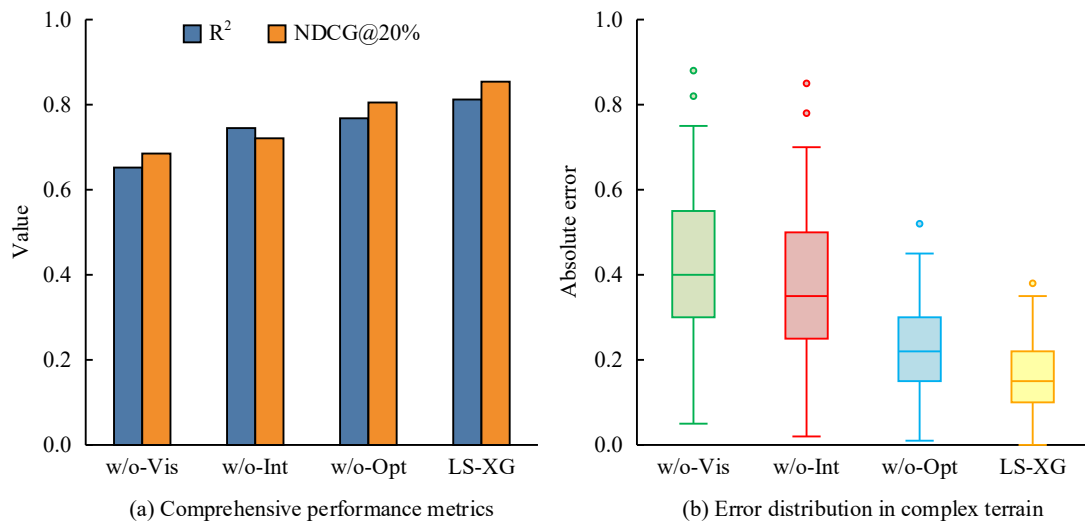


Fig. 16. LS-XG spatial sorting capability and comprehensive performance analysis



Fig. 17. Representative practical spatial forms used to verify robustness on difficult samples (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

3.3. Analysis of Simulation Application Effects of LS-XG

To verify the diagnostic and decision-making performance of LS-XG in urban renewal practice, simulation-based experiments are conducted using the Place Pulse 2.0 open-source dataset, with the "aesthetics" score as the perceived true value. The perceived true value is extracted from the aesthetics-related records in the Place Pulse 2.0 metadata. First, samples belonging to the aesthetics perception dimension are selected, and the corresponding image IDs, geographic information, and rating scores are retained. Then, these records are matched with the street-view samples and the constructed 3D spatial feature set based on the image ID and spatial location. Samples with missing aesthetics scores, invalid coordinates, or incomplete spatial features are removed. The retained aesthetics score is then used as the target variable for model training and simulation analysis, because it represents the aggregated public perception of visual aesthetic quality for each street-view scene. The data is spatially matched and cleaned using multidimensional street-view geographic metadata to construct a test sample set with complete physical and visual attributes, with Python 3.8 and an NVIDIA RTX 3090 GPU for acceleration.

3.3.1. Visual interpretation of perception driving factors based on SHAP

As an illustration, Fig. 18 presents four representative samples drawn from the Place Pulse 2.0 dataset used in this study. Each sample is a geo-located street-view image that spans the typical range of urban conditions, from tree-lined residential streets to bare elevated roads, providing the heterogeneous visual scenes on which the perceptual ratings are collected.



Fig. 18. Example samples from Place Pulse 2.0 dataset used in the simulation experiment (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

The "aesthetics" score used as the perceived true value is not rated directly on an absolute scale but is derived from the crowdsourced pairwise comparisons that constitute the Place Pulse 2.0 dataset. In the original data collection, participants are shown two street-view images side by side and asked to choose which one looked more beautiful, yielding a large set of binary win–loss votes for the aesthetics dimension. For each image, these pairwise outcomes are aggregated using a ranking algorithm that converts the relative number of wins and losses (while accounting for each opponent's strength) into a latent quality value, so that an image that frequently wins against strong competitors receives a higher value. The resulting latent values are then linearly rescaled to a continuous 0–10 aesthetics score, which serves as the regression target (perceived true value) in this study.

To quantify the actual improvement achieved by LS-XG, the core driving factors of perception are first analyzed using the SHAP algorithm, followed by the selection of representative low-scoring samples for a spatial morphology optimization simulation. The specific perceptual attribution diagnosis and optimization simulation results are shown in Fig. 19.

Fig. 19 illustrates the LS-XG perceptual attribution diagnosis and optimization simulation results, highlighting key driving factors and the nonlinear effects of the green view index. As shown in Fig. 19, the green view and the sky view exhibit the highest average SHAP absolute values (0.35 and 0.28, respectively), confirming the core contribution of visual elements to aesthetic ratings (Fig. 19(a)). Furthermore, the impact of the green view index on ratings shows a significant nonlinear, diminishing marginal effect, where SHAP values rise rapidly below 0.3 (a strong "weakening board" effect) and level off above 0.6 (Fig. 19(b)). At a 20% transformation intensity, the LS-XG-oriented precision intervention strategy increases the prediction score to 3.65, significantly outperforming the uniform optimization strategy's 2.78 (Fig. 19(c)),

thereby demonstrating decision-making efficiency under limited resources. Fig. 20 contrasts a real low-scoring scene with the spatial form targeted by the LS-XG-oriented precision intervention.

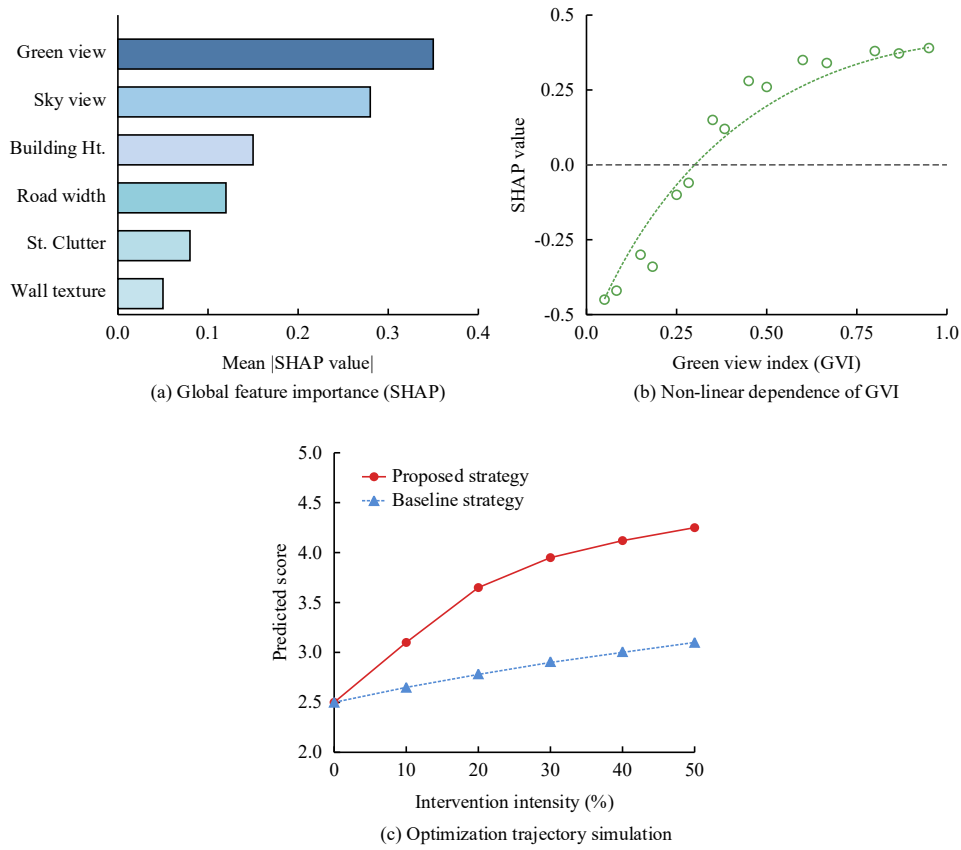


Fig. 19. LS-XG perceptual attribution diagnosis and optimization simulation results diagram



Fig. 20. Real-image illustration of optimization simulation (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

Note: Panel (a): Google Street View image from Place Pulse 2.0. Panel (b): author-generated optimization simulation based on panel (a).

The "before" scene is a bare elevated road with a low green view, corresponding to the uniform-strategy baseline score of about 2.78, whereas the "after" scene, in which street trees and vertical greening have been added under a 20% transformation intensity, illustrates the form that raises the predicted aesthetics score to about 3.65. This visual comparison demonstrates that the simulated score gain corresponds to a tangible and implementable improvement in the street-level spatial form.

3.3.2. Typical regional spatial optimization strategy generation and improvement simulation

To verify decision-making efficiency in actual spatial intervention, typical low-scoring samples are selected for differential strategy simulation to quantitatively evaluate the guiding value of LS-XG in improving urban stock space. The specific results are shown in Table 2.

Table 2 details the spatial diagnosis and optimization simulation data for typical low-scoring block samples, outlining diagnostic SHAP values, targeted strategies, and simulated improvement rates. As shown in Table 2, across the 10 lowest-

scoring block samples, LS-XG accurately identifies "insufficient green view" and "limited sky view" as the main constraints, with an average negative contribution of -0.24. Implementing targeted morphological control strategies (e.g., "adding vertical greening" or "reducing lane width") yields significant improvements in perception scores. Notably, repairing core shortcomings in sample Site_042 increases its score from 2.15 to 4.82, a maximum improvement of 124.2%. Overall, this precise attribution-oriented intervention strategy increases the average sample score by 37.2%, effectively verifying the framework's diagnostic accuracy and decision-making effectiveness for urban spatial updates. To make the two dominant constraints identified in Table 2 visually explicit, Fig. 21 illustrates the "insufficient green view" and "limited sky view" conditions together with their improved counterparts.

Table 2. Spatial diagnosis and optimization simulation data table of typical low-scoring block samples

Sample ID	Original score	Key deficit feature	Diagnostic SHAP	Optimization strategy	Simulated score	Improvement rate
Site_042	2.15	Green view index	-0.45	Add vertical greenery	4.82	124.20%
Site_108	2.33	Sky view factor	-0.38	Remove billboards	3.95	69.50%
Site_215	2.41	Building enclosure	-0.32	Increase setback	3.68	52.70%
Site_336	2.58	Road width	-0.28	Reduce lane width	3.42	32.60%
Site_411	2.65	Green view index	-0.25	Add street trees	4.15	56.60%
Site_509	2.84	Texture entropy	-0.21	Facade recoloring	3.55	25.00%
Site_622	3.01	Sky view factor	-0.18	Prune tree canopy	3.45	14.60%
Site_785	3.12	Building height	-0.15	Reduce height (Sim.)	3.48	11.50%
Site_891	3.25	Street clutter	-0.12	Clear street furniture	3.62	11.40%
Site_904	3.38	Green view index	-0.08	Add flower beds	3.92	16.00%
Avg.	2.77	/	-0.24	/	3.8	37.20%

Fig. 21(a) shows an insufficient green view, a bare bridge scene almost devoid of vegetation that yields a strong negative SHAP contribution, while Fig. 21(b) shows a limited sky view, where a dense overhead canopy encloses the street and reduces visual openness. In contrast, Fig. 21(c) presents a sufficient green view with continuous roadside vegetation, and Fig. 21 (d) presents an unobstructed, open-sky view with a wide band of sky. These paired examples make tangible why targeted strategies such as "adding vertical greening" and "reducing lane width" or canopy and corridor control move a sample from the deficient forms (a, b) toward the favorable forms (c, d), thereby improving the predicted aesthetics score.

Beyond accuracy improvement, the practical value of LS-XG can be demonstrated through managerial decision scenarios. In an urban renewal budgeting scenario, managers can rank low-scoring blocks according to predicted LVP deficits and negative SHAP values, so that limited funds are first allocated to locations with the highest expected perceptual gain. In a streetscape maintenance scenario, the model can determine whether a low score is primarily due to insufficient greenery, restricted sky view, excessive enclosure, or street clutter, thereby replacing uniform beautification with targeted maintenance actions. In a design review scenario, alternative interventions can be simulated before implementation, and the option producing the largest predicted improvement under cost, safety, and regulatory constraints can be selected. These scenarios show that LS-XG supports not only more accurate perception prediction, but also changes managerial decisions from experience-based ranking to evidence-based prioritization, intervention matching, and pre-implementation comparison.



Fig. 21. Real-image comparison of dominant constraints (source: Google Street View via Place Pulse 2.0; image capture period: 2007–2012)

4. Discussion

This study built LS-XG by integrating 3D landscape feature engineering with 2nd-order XGBoost optimization, demonstrating excellent fitting accuracy and decision-making efficiency. Empirically, achieving an R^2 of 0.812 and a lower RMSE than Lasso and DNN revealed that landscape perception is a 3D nonlinear coupling process rather than a simple linear superposition. Furthermore, introducing the "topography-building form" interaction term significantly improved performance, confirming that vertical heterogeneity (e.g., changes in slope) exerts a nonlinear regulatory effect on perception. Compared with the use of flat grid data by Yue et al. (2023), this study accurately captured vertical perceptual characteristics. Compared with the XGBoost model established by Mustapha et al. (2024) in material property prediction, this study further introduced 2nd-order Taylor expansion to approximate the objective function. Using the Hessian matrix for gradient descent retained curvature information, achieving faster convergence in complex terrain and effectively avoiding the overfitting flaws of deep learning models on small high-dimensional samples.

Second, this research broke the "black box" bottleneck by introducing SHAP attribution analysis to establish a logical bridge between prediction and spatial diagnosis. Results showed a significant diminishing marginal utility for key morphological indicators: specifically, when the green view was below 0.3, SHAP values increased exponentially but leveled off beyond that threshold, consistent with the Weber-Fechner law in psychophysics. Based on this, LS-XG generated a differentiated intervention strategy that prioritized "short board" areas, resulting in a 37.2% improvement over uniform optimization. This result echoed the fitness landscape optimization logic proposed by Diao et al. (2025) and further concretized the abstract optimization into operational urban design guidelines. Compared with the study by Schimohr et al. (2023), this study not only identified feature importance but also analyzed threshold intervals through SHAP dependency graphs to provide planners with quantitative intervention boundaries. This closed-loop framework effectively solved the interdisciplinary application problem proposed by Freitas et al. (2023), elevating AI from an evaluation tool to a generative design assistant. In practical planning use, SHAP results can support a stepwise decision process. Planners can first identify blocks with low predicted LVP scores, then examine the dominant negative SHAP factors for each block to determine the main spatial deficit. If the negative contribution comes from the green view index, priority can be given to street trees, vertical greenery, or small planting spaces. If the negative contribution comes from the sky view factor, visual openness can be improved through billboard control, canopy pruning, or visual corridor protection. If building enclosures or street clutter have strong negative effects, setback adjustments, facade regulations, or street furniture management can be considered. However, implementation is constrained by land ownership, construction cost, road safety requirements, infrastructure conditions, maintenance capacity, and planning regulations. Therefore, SHAP-based recommendations should be used as diagnostic and prioritization tools, while final planning decisions should combine model outputs with site surveys, budget constraints, engineering feasibility, and stakeholder negotiations.

Although LS-XG has significant advantages, it has certain limitations. The current validation primarily relies on the Place Pulse 2.0 dataset, which serves as a large-scale, widely used basis for street-level visual perception modeling. However, the reliance on a single perception dataset may limit the assessment of model transferability across different urban forms, cultural contexts, and planning systems. Therefore, the reported results should be interpreted as evidence of methodological feasibility in a representative, large-scale dataset, rather than as definitive proof of universal applicability. Future validation should incorporate additional datasets from different cities or locally collected street-view samples to test the robustness and adaptability of LS-XG under diverse spatial and social contexts. First, the single visual dimension of

the data source limits the overall capture of the experience. Tsai et al. (2023) pointed out that traditional visual methods struggle to fully capture the user's on-site experience, necessitating a multisensory framework (including smell and hearing) to better understand human-landscape interactions. Second, the research lacks verification of objective physiological effects. A systematic review by Iyendo et al. (2023) confirmed that the synergistic effect of soundscapes and natural landscapes could significantly regulate heart rate and parasympathetic nerve activity. Lacking these indicators, the evaluation reduces to mere "perception prediction" rather than physical and mental healing. Therefore, future research will focus on multimodal perception and physiological computing via wearable sensors. Meanwhile, in data simulation, drawing on the ideas of Karimzadeh et al. (2024), generative adversarial networks (GANs) can be used to expand extreme-scene samples, ultimately achieving a paradigm shift toward full-dimensional health diagnosis and treatment for human settlements.

5. Conclusion

To address the difficulty of analyzing nonlinear mechanisms in urban LVP and the insufficient spatial optimization support, LS-XG is constructed using XGBoost and SHAP. This method integrates 3D physical and visual features and uses a 2nd-order Taylor expansion algorithm for high-precision prediction. Thanks to feature interaction terms and the 2nd-order optimization strategy, the model achieves an R^2 of 0.812 and a median absolute error of 0.25 on multi-source heterogeneous data, significantly outperforming DNN and Lasso. Furthermore, it demonstrates strong robustness, with only a 7.5% increase in RMSE in the presence of noise. Through SHAP attribution, it accurately identifies "green view" and "sky view" as key constraints (average negative contribution of -0.24), guiding precise interventions that increase sample scores by an average of 37.2% (up to 124.2%) and effectively avoid resource waste from uniform strategies. For urban landscape managers, the framework changes the decision process from uniform improvement and experience-based prioritization to evidence-based spatial triage. Management attention should first be directed to blocks with low predicted LVP scores and strong negative SHAP contributions. Investment decisions can then be matched to the dominant deficit factor, such as green view, sky view, enclosure, or street clutter, rather than applying the same renewal measure across all sites. After intervention, predicted score changes can be used as a screening tool to compare alternatives and select actions with the highest perceptual gain under limited resources. Future work will overcome current limitations of a single visual modality by introducing multi-sensory emotional computing (e.g., hearing and physiological signals) and using GANs for sample expansion to ultimately enhance optimization of full-dimensional human settlement environments. More broadly, the value of LS-XG lies in shifting the evaluation of the urban landscape from static quality assessment to dynamic perceptual governance. Urban visual quality should not be understood only as an aesthetic outcome, but as a manageable spatial resource that can be diagnosed, prioritized, and improved through interpretable evidence. This perspective suggests that future urban renewal should move beyond generalized beautification and adopt a more precise logic of "where to intervene, what to adjust, and how much improvement can be expected." In this sense, AI-assisted LVP modeling should not replace professional planning judgment, but can provide a transparent evidence layer that makes perceptual equity, resource allocation, and spatial intervention more accountable.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author used Grammarly Premium solely for language editing and readability improvement. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

- Asselman, A., Khaldi, M., and Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, 31(6), 3360–3379. <https://doi.org/10.1080/10494820.2021.1928235>
- Bates, S., Hastie, T., and Tibshirani, R. (2023). Cross-validation: What does it estimate and how well does it do it? *Journal of the American Statistical Association*, 119(546), 1434–1445. <https://doi.org/10.1080/01621459.2023.2197686>
- Chen, H., Covert, I. C., Lundberg, S. M., and Lee, S. I. (2023). Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence*, 5(6), 590–601. <https://doi.org/10.1038/s42256-023-00657-X>
- Costa, V. G. and Pedreira, C. E. (2022). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765–4800. <https://doi.org/10.1007/s10462-022-10275-5>
- Diao, Y., Li, C., Zeng, S., Yang, S., and Coello, C. A. C. (2025). Nearest-better network for fitness landscape analysis of continuous optimization problems. *IEEE Transactions on Evolutionary Computation*, 29(5), 2089–2103. <https://doi.org/10.1109/TEVC.2024.3478825>
- Freitas, F., Berreth, T., Chen, Y. C., and Jhala, A. (2023). Characterizing the perception of urban spaces from visual analytics of street-level imagery. *AI and Society*, 38(4), 1361–1371. <https://doi.org/10.1007/s00146-022-01592-y>
- Grotti, E., Santana, P. B., Filho, J. G. P., and Gomes, H. M. (2023). Multi-objective robust optimization framework based on first and second order Taylor expansion applied to a vehicle suspension design. *Optimization and Engineering*, 25(2), 637–668. <https://doi.org/10.1007/s11081-023-09817-9>
- Gunasekara, N., Pfahringer, B., Gomes, H., and Bifet, A. (2024). Gradient boosted trees for evolving data streams. *Machine Learning*, 113(5), 3325–3352. <https://doi.org/10.1007/s10994-024-06517-y>

- Iyendo, T. O., Welch, D., and Uwajeh, P. C. (2023). Soundscape and natural landscape as a design construct for improving psycho-physiological health in cities: A semi-systematic literature review. *Cities and Health*, 8(3), 447–485. <https://doi.org/10.1080/23748834.2023.2280288>
- Karimzadeh, S., Mohammadi, A., Salahuddin, U., Carvalho, A., and Lourenço, P. B. (2024). Backbone ground-motion model using simulated records and the XGBoost machine learning algorithm: An application to the Azores Plateau (Portugal). *Earthquake Engineering and Structural Dynamics*, 53(2), 668–693. <https://doi.org/10.1002/eqe.4040>
- Labetski, A., Vitalis, S., Biljecki, F., Ohori, K. A., and Stoter, J. (2022). 3D building metrics for urban morphology. *International Journal of Geographical Information Science*, 37(1), 36–67. <https://doi.org/10.1080/13658816.2022.2103818>
- Lee, Y. G., Oh, J. Y., Kim, D., and Kim, G. (2023). Shap value-based feature importance analysis for short-term load forecasting. *Journal of Electrical Engineering and Technology*, 18(1), 579–588. <https://doi.org/10.1007/s42835-022-01161-9>
- Li, L. Z., Deng, C. N., Zhu, F., Song, X. C., Wang, E., and Xie, M. H. (2026). Beyond one-size-fits-all: Differentiated green development assessment integrating a hybrid approach in China's Yangtze River Economic Belt. *Environmental Impact Assessment Review*, 116. <https://doi.org/10.1016/j.eiar.2025.108076>
- Li, X. Q., Song, L. K., and Bai, G. C. (2023). Vectorial surrogate modeling approach for multi-failure correlated probabilistic evaluation of turbine rotor. *Engineering with Computers*, 39(3), 1885–1904. <https://doi.org/10.1007/s00366-021-01594-2>
- Liu, Z., Jiang, L., Shi, C., Zhang, T., Zhou, Z., Liao, J., and Hu, K. (2023). CRA-40/atmosphere—The first-generation Chinese atmospheric reanalysis (1979–2018): System description and performance evaluation. *Journal of Meteorological Research*, 37(1), 1–19. <https://doi.org/10.1007/s13351-023-2086-x>
- Mustapha, I. B., Abdulkareem, Z., Abdulkareem, M., and Ganiyu, A. (2024). Predictive modeling of physical and mechanical properties of pervious concrete using XGBoost. *Neural Computing and Applications*, 36(16), 9245–9261. <https://doi.org/10.1007/s00521-024-09553-w>
- Nguyen, H., Cao, M. T., Tran, X. L., Tran, T. H., and Hoang, N. D. (2023). A novel whale optimization algorithm optimized XGBoost regression for estimating bearing capacity of concrete piles. *Neural Computing and Applications*, 35(5), 3825–3852. <https://doi.org/10.1007/s00521-022-07896-w>
- Paláncz, B. and Awange, J. L. (2013). Pareto optimality solution of the multi-objective photogrammetric resection-intersection problem. *Earth Science Informatics*, 6(1), 1–20. <https://doi.org/10.1007/s12145-012-0107-x>
- Qin, Y., Chi, X., Sheng, B., and Lau, R. W. (2023). GuideRender: Large-scale scene navigation based on multi-modal view frustum movement prediction. *The Visual Computer*, 39(8), 3597–3607. <https://doi.org/10.1007/s00371-023-02922-x>
- Schimohr, K., Doeblér, P., and Scheiner, J. (2023). Prediction of bike-sharing trip counts: Comparing parametric spatial regression models to a geographically weighted XGBoost algorithm. *Geographical Analysis*, 55(4), 651–684. <https://doi.org/10.1111/gean.12354>
- Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M., and Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83(12), 35307–35334. <https://doi.org/10.1007/s11042-023-16737-4>
- Tan, S., Hooker, G., Koch, P., Gordo, A., and Caruana, R. (2023). Considerations when learning additive explanations for black-box models. *Machine Learning*, 112(9), 3333–3359. <https://doi.org/10.1007/s10994-023-06335-8>
- Tsai, M., Brough, M., and Cushing, D. F. (2023). Developing a multisensory methodology to explore older people's landscape experience in Australian aged-care facilities. *Ageing and Society*, 43(7), 1707–1729. <https://doi.org/10.1017/S0144686X21001471>
- Weber-Lewerenz, B. C., and Traverso, M. (2023). Navigating applied artificial intelligence (AI) in the digital era: How smart buildings and smart cities become the key to sustainability. *Artificial Intelligence Applications*, 1(4), 214–227. <https://doi.org/10.47852/bonviewAIA32021063>
- Williams, A. (2022). Human-centric functional computing as an approach to human-like computation. *Artificial Intelligence Applications*, 1(2), 112–121. <https://doi.org/10.47852/bonviewAIA2202331>
- Xue, B., Xiao, X., Li, J., Zhao, B., and Fu, B. (2022). Multi-source data-driven identification of urban functional areas: A case of Shenyang, China. *Chinese Geographical Science*, 33(1), 21–35. <https://doi.org/10.1007/s11769-022-1320-2>
- Yuan, R., Wang, S., Liu, G., and Peng, M. (2024). Non-line-of-sight ultraviolet positioning using linearly-arrayed photon-counting receivers. *IEEE Journal on Selected Areas in Communications*, 42(9), 2535–2549. <https://doi.org/10.1109/JSAC.2024.3413960>
- Yue, R., Meng, Q., Yang, D., Wu, Y., Liu, F., and Yan, W. (2023). A visualized soundscape prediction model for design processes in urban parks. *Building Simulation*, 16(3), 337–356. <https://doi.org/10.1007/s12273-022-0955-3>
- Zhang, H., Yang, Z., Sudjianto, A., and Zhang, A. (2023). A sequential Stein's method for faster training of additive index models. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10), 13941–13950. <https://doi.org/10.1109/TNNLS.2023.3273474>
- Zhang, K. Q., Jia, L. J., and Xu, S. (2025). Recognizing mixing patterns of urban agglomeration based on complex network assortativity coefficient: A case study in China. *Applied Sciences*, 15(4), 2024–2049. <https://doi.org/10.3390/app15042024>



Yanhui Sui obtained her master's degree in agriculture from Shandong Agricultural University in 2003. She currently serves as an Associate Professor at Weihai Vocational College. She has been invited as a consultant and has delivered numerous academic lectures on topics such as landscape architecture design, landscape architecture project management, and landscape plant maintenance. She has also served and continues to serve as a reviewer for international conferences and journals. She has published articles in multiple journals and conference proceedings. Her research areas include landscape architecture evaluation and management, as urban digital twins.