2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 5-8, 2025. Vienna, Austria

2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3315-3358-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/SMC58881.2025.11343090

A Multimodal Perception System for Predicting Restorative Effect in University Open Spaces

Jiazhen Huang[1†], Ruoling Qi[1†], Jiayi Liu[2†], Tengfei Han[2], Fansheng Zhang[3], Jieqian Sun[4], Wei Zhao[2*], Wei Ju[1*]

***Abstract*— Growing mental health problems among college students underscore the urgent need to enhance the restorative effect of university campus. However, current evaluation tools lack systematic, scalable, and interpretable methods to quantify the psychological impact of open space design. This study proposes a multimodal perception system to predict the restorative effect of university open spaces by integrating visual, structural, and semantic features. We construct a large-scale dataset comprising 600 campus images and 12,147 subjective ratings collected through standardized psychological scales. Semantic segmentation is used to extract spatial visual indices from images, while semantic impressions are obtained via the Semantic Differential (SD) scale and converted into natural language descriptions. A dual-encoder alignment framework maps both index and text representations into a shared latent space, enabling cross-modal prediction of restorative scores. A Random Forest regressor is trained on this space to support score inference from either image- or text-based inputs. In addition, we apply a Rule-based Representation Learner (RRL) to extract interpretable spatial patterns associated with restorative outcomes. Experiments show that our method significantly outperforms traditional regression models, achieving an R^2 of 0.85 in predicting perceived restorative effect. The learned rules reveal both explicit visual drivers (e.g., greenery, sky openness) and implicit spatial logics (e.g., element interaction). This framework offers a lightweight and interpretable evaluation tool for health-oriented campus design, applicable across design and planning stages even without image data.**

I. Introduction

With the increasingly fierce competition in higher education, today's university students are faced with unprecedented mental health challenges. A study covering 15 countries during the COVID-19 pandemic reports that 31.2% of university students have exhibited depression symptoms and 26.0% have shown significant signs of stress [1]. The prevalence of these mental health issues much heightens the risk of mental disorders along with related comorbidities [2]. Therefore, exploring potential influencing factors of university students' mental well-being and seeking intervention strategies has become a critical research topic.

University open spaces play a critical role in supporting students' mental health and well-being [3]. In recent years, a growing body of research has highlighted the restorative potential of green and open spaces in relieving stress, enhancing attention recovery, and fostering emotional stability among college populations. However, quantifying the psychological impact of such spaces remains a complex challenge, particularly in campus environments characterized by diverse spatial typologies and user perceptions [4]. To quantify this impact, existing work usually introduces the concept of *restorative effect* [5], [6] from the field of environmental psychology, and relies on traditional psychological experiments or small-scale questionnaires for further investigation. Although they have provided useful insights, they often lack scalability, automation, and semantic depth. Specifically, these methods are typically labor-intensive, heavily reliant on human interpretation, and unable to generalize across different spatial contexts. Moreover, most existing frameworks focus primarily on objective physical features (e.g., green coverage, sky view), neglecting the subjective and semantic dimensions of spatial experience that strongly influence psychological restoration. Recent advances in machine learning and multimodal learning offer promising opportunities to address these limitations. In particular, the integration of visual scene understanding, semantic analysis, and human-centered perception modeling can support more nuanced and scalable assessments of restorative environments. Yet, few studies have attempted to unify these modalities within a coherent and interpretable framework.

Therefore, in this study, we propose a multimodal perception system to predict the restorative effect of university open spaces. Our approach combines image-based spatial analysis, subjective psychological ratings, and semantic representation learning into a unified, data-driven pipeline. We construct a large-scale dataset containing 600 campus images and over 12,000 responses collected through three standardized instruments: the Perceived Restorativeness Scale (PRS), the Stress Recovery Rating Scale (SRRS), and the Semantic Differential (SD) scale. Visual environmental indices are extracted using semantic segmentation, while SD scores are mapped into natural language descriptions to capture spatial semantics. To bridge the gap between quantitative indices and subjective perception, we develop a dual-encoder alignment model that maps environmental indices and semantic texts into a shared latent space. A Random Forest regressor is then trained on this space to predict restorative effect scores from either visual or textual inputs. Furthermore, we apply the Rule-based Representation Learner (RRL) to extract interpretable decision rules linking specific spatial configurations to restorative outcomes. Our contributions are concluded as follows:

*Corresponding authors.
†Equal contribution.
[1]School of Computer Science, Sichuan University, Chengdu, China.
[2]College of Architecture and Environment, Sichuan University, Chengdu, China.
[3]College of Arts, Sichuan University, Chengdu, China.
[4]Department of Automation, Tsinghua University, Beijing, China.

- We introduce a scalable and interpretable framework for predicting the restorative effects of university open spaces using multimodal data.
- We construct a novel dataset that integrates visual content, semantic impressions, and psychological ratings, providing a rich foundation for environmental perception modeling.
- We design a cross-modal alignment model and a rule-based reasoning module, enabling both accurate prediction and interpretable pattern discovery across different input formats.

II. Related Works

A. Restorative Effects

The field of environmental psychology proposes that the desire for contact with nature serves an important adaptive function, namely psychological restoration [5]. Consequently, it is imperative to explore strategies that can augment the restorative effects of the campus for the college student population, who often find themselves grappling with attentional depletion and high-pressure environments. To quantify the restorative effects on the environment, academics have developed a variety of assessment toolswhich provide a reliable measurement framework for relevant empirical studies [7]. Among them, the Perceived Restorativeness Scale (*PRS*) measures restorative potential by evaluating environmental performance; the Stress Response Recovery Scale (*SRRS*) focuses on assessing environments' regulatory functions on individual stress levels. They both have been validated to effectively reflect individuals' subjective environmental perceptions [8]. Therefore, we combine the results of these two scales as the restorative effects score in this paper.

Many studies on restorative effect show that individuals have differential preferences for restorative effects generated by different environmental factors. For example, Wang [9] states that open green lawns have significant landscape restorative characteristics. Han et al. [10] argue that the number and comfort of resting seats in open spaces positively influence individuals' willingness to stay. Currently, university open spaces are the focus of studies exploring the restorative effect of the environment.

In recent years, the rise of computer technology has opened new possibilities for research on restorative effects. Xin Han et al. [11] extractes urban streetscape elements through pixel-level semantic segmentation to construct multi-level explanatory variables for perceived stress. Haoran Ma [12] uses city-level graphs as inputs to train a graph neural network (GNN)-based model for predicting urban restoration quality. Wen et al. [13] integrates a fully convolutional network (FCN) with the k-means clustering algorithm to evaluate the psychological restorative effects of university open spaces. However, although the restorative effects of university open spaces have been validated, no systematic framework currently exists for their quantitative assessment. In response, our framework fills this gap.

B. Spatial Element Recognition and Environmental Index Quantification

Image segmentation is a key technique for achieving spatial element recognition and environmental index quantification. In restorative environment research, semantic segmentation enables precise differentiation of various visual elements within an image, providing structured inputs for subsequent environmental perception and model construction. In recent years, machine learning techniques have significantly advanced the development of image segmentation. For example, Xiao Fu et al. [14] proposes a multi-scale feature fusion framework based on the Pyramid Scene Parsing Network (PSPNet), enabling accurate quantification of environmental indices such as the Green View Index (GVI). Jingxian Tang et al. [15] combines semantic segmentation with ArcGIS-based 2D analysis methods to calculate physical parameters of street spaces, thereby achieving precise modeling of street spatial structures. In addition, Ruoyu Wang et al. [16] developes a two-stage model combining FCN-8s-based semantic segmentation with a random forest classifier to jointly predict multiple attributes in complex scenes. These studies demonstrate that semantic segmentation has become a vital approach for quantifying urban and campus environmental features. However, despite the effectiveness of the aforementioned studies in applying machine learning to the identification and quantification of environmental features, there is still a lack of a systematic evaluation framework for the restorative effects of university open spaces.

III. Preliminary Works

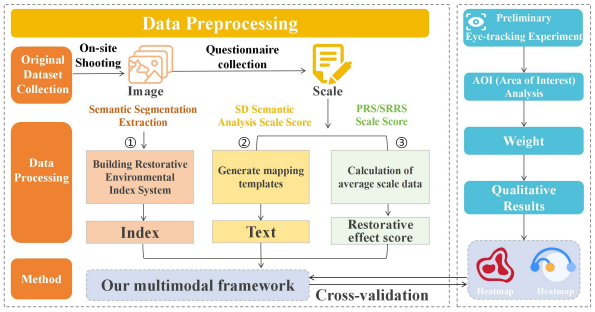

Fig. 1: **Data collection procedure of our work.**

A. Original Dataset Collection

1) Image dataset from photographs: Inspired by Wen et al. [13], we select five representative categories of university open spaces: vegetation areas, squares, waterfronts, courtyards, and playing fields. To ensure diversity and representativeness, we sample 12 campuses from 7 "Double First-Class" universities in Chengdu, a major higher education hub in Western China, as shown in Figure 2. Specifically, sampling campuses are chosen based on the following criteria: 1) Full-time research-oriented universities; 2) Inclusion of both traditional and newly developed campuses with planned areas ranging from 83 to 334 hectares; 3) Presence of all five targeted open space types; and 4) High density of student activity. We identify high-density gathering points

based on heatmaps of extracurricular activity distributions, and photograph these locations with standardized simulated views. All photographs are captured under clear weather conditions using a 28mm lens to approximate human field of visual perception, standardized in RGB format with a resolution of 150 PPI. A total of 600 images are collected as image dataset of university open spaces. This dataset serves as the foundation for both index calculation in Section III-B and scale collection in Section III-A.2.

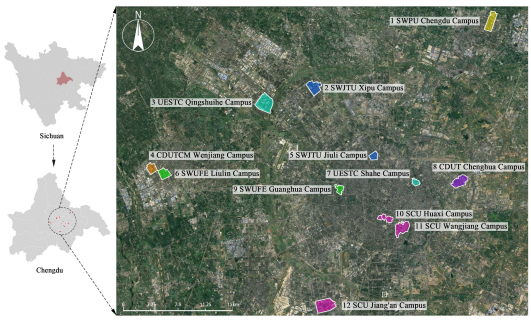

Fig. 2: **Final distribution map of our campuses selection.**

2) Scale dataset from questionaires: In addition, we conduct a large-scale survey on an online questionnaire platform to gather subjective perceptual scales from university students. Each participant evaluate at least six images of our image dataset using the Perceived Restorativeness Scale (PRS), Stress Recovery Rating Scale (SRRS), and Semantic Differential (SD) scale, along with basic demographic information. After screening for IP duplication and response validity, we obtain 2,000 valid questionnaires, yielding 12,147 image-level evaluations. In subsequent sections, the PRS and SRRS scales are aggregated to calculate the restorative effect score, while the SD scales are converted into natural language descriptions for multi-modal alignment.

B. Restorative Environmental Index System from Images

Considering the deeper perceptual pathways of visual elements and the combinatory relationships between them, we have constructed a multi-dimensional quantification system for restorative environmental indices in university open spaces, based on environmental psychology theory and taking into account the characteristics of campus spaces and student groups, as shown in Table I. Motivated by similar work [13], the final index system includes 4 main criteria layers and 12 specific indices, covering core logical dimensions such as visual openness, spatial enclosure, and functional support. To the best of our knowledge, our work is the first to construct a quantifiable index system specifically for campus environment restoratives. Based on this index system, the visual elements extracted from each image are further transformed into multi-dimensional perceptual indices.

C. Semantics and Scores from Scales

While the environmental index system provides a structured description of physical spatial elements, it overlooks the subjective, semantic qualities of space that also influence perceived restoration. For instance, spaces with vibrant, colorful vegetation may evoke stronger restorative effects than those with monotonous greenery, even if their objective indices are similar. To capture such perceptual nuances, we consider the Semantic Differential (SD) scale [17], a psychometric tool that quantifies individuals' affective responses to environments using bipolar adjective pairs (e.g., "clean–dirty", "open–enclosed") [18]. These semantic ratings reflect perceived spatial quality dimensions that are not directly observable from indices. Therefore, we incorporate this method into the process of quantifying environmental features in Section IV-C.

For the rest of scales, we adopt them to compute a composite restorative effect score for each image, serving as the prediction target in our framework. Specifically, each image is rated using two scales: the Perceived Restorativeness Scale (PRS) and the Stress Recovery Rating Scale (SRRS). Both scales have been validated in environmental psychology for measuring different dimensions of restorative potential. We normalize the ratings from both scales to the range $[0,1]$ and calculate the final effect score by averaging the two:

$$\text{Effect Score} = \frac{1}{2}\left(\text{PRS}_{\text{norm}} + \text{SRRS}_{\text{norm}}\right) \tag{1}$$

This composite score integrates attentional and emotional aspects of environmental restoration, providing a holistic target metric for model training and evaluation.

D. Eye Tracking Pre-study

To verify the interpretability of our perceptual system in the subsequent large-scale data analysis, we conduct a preliminary eye-tracking experiment to help identify and focus on potential spatial elements.

Eye-tracking technique reveals specific mechanisms of attention and cognition during image perception [19]. By collecting data related to Areas of Interest (AOIs), researchers record the number, duration, and sequence of fixations to determine the total dwell time on each AOI. Information provided by eye tracking helps to explain environmental preferences, since the technology can generate objective and automated visual attention maps. Building on this, we are able to assess the relative contribution of each spatial element by analyzing individuals' eye-movement distributions.

A total of 22 participants aged between 18 and 24 (M = 22.38, S.D. = 2.56), including both undergraduate and graduate students from various academic disciplines, are recruited for the experiment. The experimental device used is the aSee Studio wearable eye tracker developed by 7invensun. 120 images within our dataset from three campuses of Sichuan University, China, are selected for the this phase. we introduce task-based scenarios simulating high stress and attentional fatigue to emulate a sub-healthy mental state in participants, under which the influence of restorative effects becomes more pronounced. After completing the informed consent process and initial questionnaire, participants are equipped with the eye-tracking device and underwent a three-step calibration procedure. All images are randomized

Criterion Level	Index Level	Description	Calculation
Landscape environment	Naturalness (NI)	The inverse tangent of the ratio of natural components (e.g., greenery, water) to grey infrastructure (e.g., building, pavement).	$\arctan\left(\frac{\text{natural elements}}{\text{grey infrastructure}}\right)$
	Sky view index (SVI)	A prospect, vistas over the surroundings. Area proportion of sky in a view.	$\frac{S_{\text{sky}}}{S_{H\times W}}$
	Green view index (GC)	A characteristic with adequate greenery. Sum of the area proportions of all greenery components (e.g., tree, grass).	$\frac{S_{\text{greening}}}{S_{H\times W}}$
	Water Feature Coverage Rate (WR)	Percentage of Water Feature Area in Open Spaces. Proportion of Visible Water Feature Area.	$\frac{S_{\text{water\_feature}}}{S_{H\times W}}$
The built environment	Overhead shelter (OS)	A feeling of safety and shelter. Proportion of Ceiling and Canopy Area (with the canopy covering two-thirds of the tree's total area).	$\frac{S_{\text{ceiling}}+S_{\text{tree\_canopy}}}{S_{H\times W}}$
	Building Viewability Coefficient (BVC)	Percentage of Built-up Area in Open Spaces. Proportion of visible built-up area.	$\frac{S_{\text{visible\_building\_area}}}{S_{H\times W}}$
Event venues	Spatial division (SS)	A characteristic of not disturbing the sense of spatial unity.	$\frac{S_{\text{division\_coherent}}+S_{\text{complete\_activity\_space}}}{S_{H\times W}}$
	Free space (AS)	Sum of the area proportions of elements dividing a coherent and integral space for activity.	$\frac{S_{\text{free\_activity\_area}}}{S_{H\times W}}$
	Pavement Visibility (PV)	Pavement areas with a larger proportion or significant positioning in the environment. Proportion of pavement area.	$\frac{S_{\text{paving}}}{S_{H\times W}}$
	Walkability (WEVI)	The visual impact of perceived road conditions on the walking experience. The overall support of the outdoor environment for walking.	$\frac{\frac{1}{n}\sum_{i=1}^{n} Pl_i+\frac{1}{n}\sum_{i=1}^{n} F_i}{\frac{1}{n}\sum_{i=1}^{n} R_i}$
Rest facilities	Service facility (SF)	Total Proportion of Area Occupied by Functional and Decorative Service Facilities (e.g., benches, streetlights, signs, awnings, flower pots, outdoor flooring, sculptures, etc.).	$\frac{S_{\text{use}}+S_{\text{decorate\_service\_facilities}}}{S_{H\times W}}$
	Disturbance (SO)	Undisturbed environment; silent, calm. Sum of the area proportions of the disturbing components (e.g., truck, traffic lights).	$\frac{S_{\text{disturbing\_component}}}{S_{H\times W}}$

TABLE I: **Details of our Restorative Environmental Index System, with 4 criterion layers and 12 specific indices.**

and each participant is assigned to one of six independent scenario groups for the experimental experience. The experiment consists of four stages (Figure. 3):

1) Preparation: Before each scenario began, participants are instructed to close their eyes and rest for 3 minutes to adjust their mental state. **2) Stress Induction:** Participants are required to complete a timed 2-minute mental arithmetic task involving random 4-digit calculations, serving as a stress and attention manipulation. **3) Restorative Experience:** Participants, wearing the eye-tracking device, are guided through a scenario in which they imagined themselves resting after prolonged study. They then view landscape images in a natural observation manner for 3 minutes. **4) Assessment:** After each scenario, participants complete *PRS*, *SRRS*, and *SD* scales based on their subjective experience. Each full experiment session laste approximately 60 minutes. To ensure scientific rigor, a 15-second neutral gray background is inserted between images to allow visual resetting. Additionally, participants are allowed to begin the experience from any numbered image set to minimize random bias.

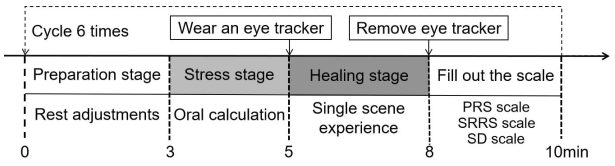

Fig. 3: **The four-stage scenario for preliminary study.**

After the experiment, we obtain three types of scale data [1] and extracted 20 eye-tracking indices. Through statistical analysis, images demonstrating the highest restorative effect are selected for each typical space type. The corresponding eye-tracking heatmaps, gaze trajectories, and fixation maps during the first minute of viewing are demonstrated in Figure 4. From these results, we are able to draw following preliminary conclusions qualitatively: university students tend to focus (either initially or most frequently) on: 1) natural elements (e.g., trees, lawns); 2) water features (e.g., lakes, rivers); 3) iconic architectural structures (e.g., pavilions, memorial halls); 4) man-made landscape features (e.g., statues of historical figures). In contrast, elements such as roads, sky, paved areas, and sports fields receive minimal visual attention. Notably, we observe some interesting findings: certain participants are more inclined to fixate on scenes that conveyed a strong sense of spatial depth or expansive vistas. This suggests that beyond explicit visual elements, underlying spatial organizational logics—such as hierarchical structures or implied accessibility—may influence restorative outcomes through indirect perceptual pathways.

[1]Note that scales collected here are different from those from questionaires, only used in this stage for preliminary analysis.

IV. Methods

A. Overview of our Framework

We develop a systematic and automated evaluation framework for assessing the restorative effect of university open spaces, grounded in a multimodal data system. Figure 5 presents the overview of our framework.

As shown in Figure 1, our framework integrates three primary data modalities: 1) visual data in the form of campus photographs; 2) quantitative environmental indices extracted via semantic segmentation based on a predefined restorative effect index system; 3) textual data representing perceived spatial qualities, collected through the SD method. The integration of these heterogeneous data sources enables a robust multi-method validation paradigm under a multimodal data architecture. Specifically, image data are first converted into a structured set of environmental indices. These indices are then aligned with corresponding semantic descriptors derived from SD analysis. This cross-modal alignment allows the system to automatically evaluate the restorative effect of a given campus scene by reading its visual indices and subsequently generating both a predicted restorative score and

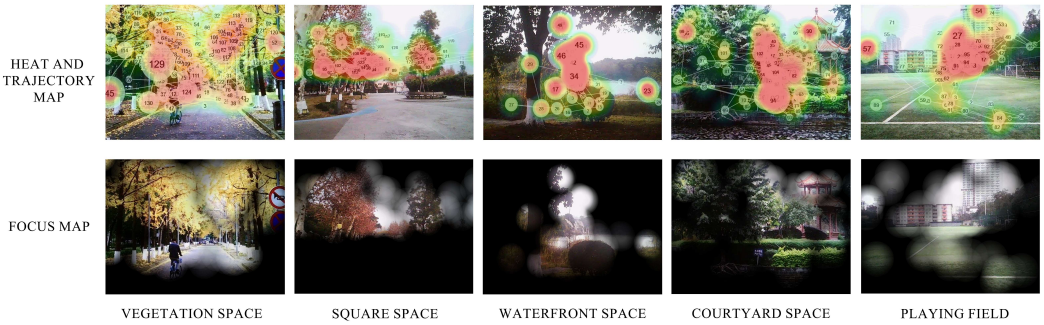

Fig. 4: **Eye-movement results from eye-tracking, including heat and trajectory maps (above) and focus maps (below).**

a natural language description of perceived spatial quality.

B. Environmental Index Calculation

We utilize image semantic segmentation techniques to extract visual perception elements of university open spaces. Specifically, we introduce Mask2Former [20], a general segmentation model that leverages attention mechanisms and multi-scale feature fusion to enhance its ability to model complex scenes with superior performance. The model is pre-trained on the COCO dataset [21], which comprises over 330,000 images (200,000 annotated), 1.5 million target instances, covering 80 objects and 91 scene categories. Considering the campus environment, we supplement a custom labeled dataset to enrich the element classes such as pillars. The model has achieved an accuracy of 92.3% and a mean Intersection over Union (mIoU) of 72.5% on the COCO test set, demonstrating a strong extraction capability for key elements. Finally, we perform segmentation on our university image dataset, and calculate environmental indices based on the restorative index system described in Table I. In detail, for each segmented image, we compute the area proportion of each labeled element (e.g., greenery, sky), relative to the total image area. These pixel-level ratios are then mapped to specific indices as feature vectors $\mathbf{x}_{\text{index}}$.

C. Multi-modal Semantic Alignment

While indices extracted from images offer a structured representation of environment, they only focus on objective visual perception based on certain spatial physical elements, failing to capture the nuanced, human-centered semantic perception of environmental quality. To bridge this gap, we propose a cross-modal alignment framework that maps low-level visual indices to high-level semantic descriptors derived from the SD scale. By integrating the above mechanisms, we have established a three-dimensional mapping relationship between "visual-semantic-effect" (see Figure 5 for details).

Our goal is to learn a shared latent representation that aligns these two modalities—visual perception and semantic understanding—enabling model to infer how quantified spatial elements translate into subjective perceptions of environmental quality. In implementation, we first generate mapping templates with GPT-4o [22] to enrich the formatted SD scales into natural language description texts $\mathbf{t}$, and then encode them with the pre-trained language backbone DeBERTaV3 [23]. To avoid high computational cost and the risk of semantic distortion, all parameters of backbone are kept frozen in the subsequent stages. The resulting high-dimensional embeddings $\mathbf{x}_{\text{text}}$ form the semantic representation space.

Next, we construct a shared probabilistic latent space $\mathcal{Z} \subset \mathbb{R}^{D_Z}$ by introducing a dual-encoder Variational Autoencoder (VAE) [24] architecture. Specifically, one encoder learns the distribution of the SD text embeddings $\mathbf{x}_{\text{text}}$ in the latent semantic space, while the second encoder maps environmental indices $\mathbf{x}_{\text{index}}$ into the same latent space. For text embeddings, a complete encoder-decoder was built on top, where the encoder $Enc_t(\cdot)$ maps $\mathbf{x}_{\text{text}}$ to a Gaussian distribution $\mathcal{N}(\mu_{\text{text}}, \sigma^2_{\text{text}})$, and the decoder $(\cdot)$ attempt to reconstruct the original embedding $\hat{\mathbf{x}}_{\text{text}}$ that sampled from latent vector $\mathbf{z} \sim \mathcal{N}(\mu_{\text{text}}, \sigma^2_{\text{text}})$ using the reparameterization trick. For environmental indices, we introduce an independent encoder $Enc_e(\cdot)$, which follows an analogous process to map $\mathbf{x}_{\text{index}}$ to $\mathcal{N}(\mu_{\text{text}}, \sigma^2_{\text{text}})$ in another latent space.

Among them, the $Enc_t(\cdot)$ employs a two-layer MLP for both encoder ($768 \to 512 \to 64$) and decoder ($64 \to 512 \to 768$), with the latent dimension set to 64. The $Enc_e(\cdot)$ uses a single 128-unit hidden layer ($12 \to 128 \to 128$) followed by two 64-unit heads for μ_{text} and $log(\sigma^2_{\text{text}})$. ReLU activation function [25] is applied to all layers. The training process consists of two stages:

1) Semantic Space Learning: Only the parameters of $Enc_e(\cdot)$ and $Dec(\cdot)$ are optimized. The goal is to learn a latent space $q_\theta(\mathbf{z}|\mathbf{x}_{\text{text}})$ that can effectively compress and reconstruct semantic information from the SD embeddings:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{\mathbf{z}\sim q_\theta(\mathbf{z}|\mathbf{x}_{\text{text}})}[\|\mathbf{x}_{\text{text}} - \hat{\mathbf{x}}_{\text{text}}\|^2] + \beta \cdot D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}_{\text{text}}) \parallel p(\mathbf{z})), \tag{2}$$

where β is a hyper-parameter, and $\|.\|^2$ denotes the L2 norm.

2) Latent Space Alignment: We then freeze $Enc_e(\cdot)$ and $Dec_e(\cdot)$, and only optimize $Enc_e(\cdot)$. To align the distributions of $\mathbf{x}_{\text{text}}$ and $\mathbf{x}_{\text{index}}$ in the latent space, we minimize the KL divergence between the two latent distributions:

$$\mathcal{L}_{\text{Align}} = D_{\text{KL}}(\mathcal{N}(\mu_{\text{index}}, \sigma^2_{\text{index}}) \parallel \mathcal{N}(\mu_{\text{text}}, \sigma^2_{\text{text}})). \tag{3}$$

Through this strategy, we guide the encoders to learn to map the visual environmental indices into a probabilistic region in the latent space that is consistent with their corresponding semantic representations. During inference, the model can

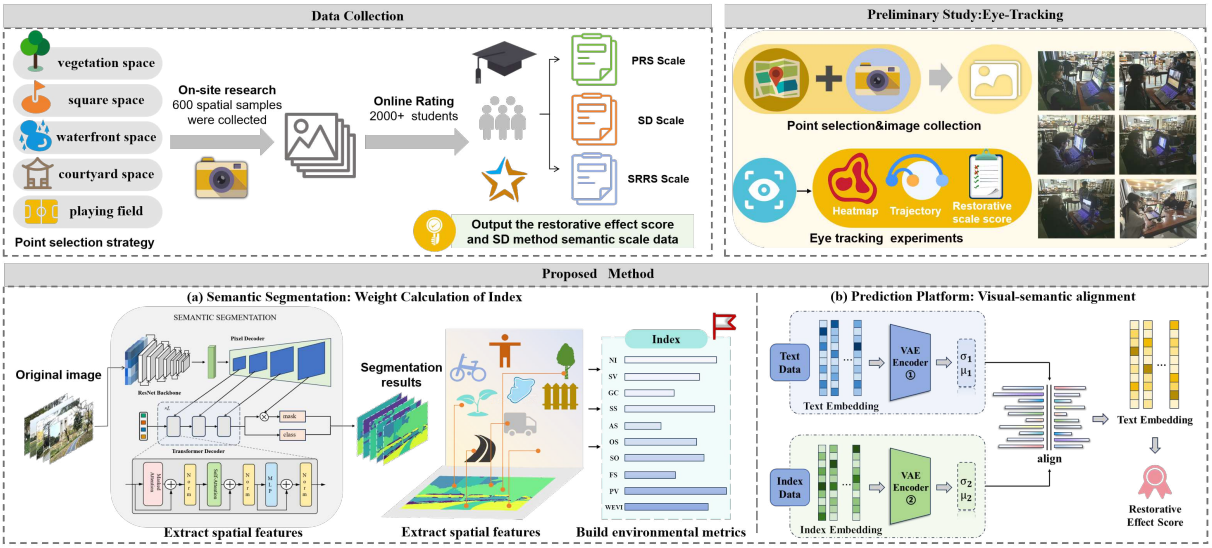

Fig. 5: **The overview of our framework.**

take a embedding vector as input (whether text or indices), encode it into a latent representation with the corresponding encoder, and decode it via the shared decoder $Dec(\cdot)$ to generate a prediction.

D. Restorative Effect Score Prediction

Our framework predicts restorative effect scores by first mapping diverse inputs into a unified semantic embedding space, and then applying a regression model trained on these embeddings. The system supports three input pathways:

- **Image-to-Score**: Images are segmented and converted into environmental indices. These indices are then encoded and mapped via the alignment mechanism into the semantic embedding space for prediction.
- **Text-to-Score**: Natural language descriptions are encoded directly into semantic embeddings.
- **SD-to-Score**: Structured SD ratings are mapped into descriptive texts, then processed as above.

In all cases, restorative effect scores are predicted using a Random Forest [26] regression model trained on semantic embeddings, ensuring a unified and modality-agnostic prediction process. Importantly, it also decouples the perception modality from regression model, allowing consistent inference even when only one type of data is available.

V. Findings and Analyses

A. Performance Evaluation

To comprehensively evaluate the performance of our model, we adopt Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2) as evaluation metrics. Four mainstream machine learning regression models are adopted as baselines: Multi-Layer Perceptron (MLP) [27], Support Vector Machine (SVM) [28], Random Forest (RF) [26], and Decision Tree (DT) [29]. These baselines are trained directly on formatted tabular SD scales. The results are shown in Table II.

Metric	MLP	SVM	RF	DT	Ours
$MSE\downarrow$	1.5708±0.04	1.6338±0.05	1.4874±0.03	1.4881±0.03	0.3108±0.02
$MAE\downarrow$	0.9910±0.01	0.9836±0.01	0.9688±0.01	0.9681±0.01	0.3357±0.01
$R^2\uparrow$	0.2126±0.02	0.1692±0.02	0.2410±0.02	0.2417±0.02	0.8478±0.01

TABLE II: **Performance comparison with baselines.**

As shown, our approach significantly outperforms traditional models across all metrics, with particularly strong gains in R^2 and error reduction. This demonstrates the advantage of aligning low-level spatial features with high-level semantic representations. The final Random Forest model operating on semantic embeddings achieves an R^2 of 0.8478 and MAE of 0.3357, reflecting both high accuracy and generalization ability.

To further assess the effect of alignment between indices and semantic texts in the shared latent space, we randomly select 50 validation samples within the latent space and employ the Procrustes analysis [30]. As shown in Figure 6, the two representations substantially overlap after dimensionality reduction, with the connecting lines indicating minimal Procrustes distances. This suggests that our alignment mechanism effectively transforms the environmental indices into representations closely aligned with the SD text embeddings.

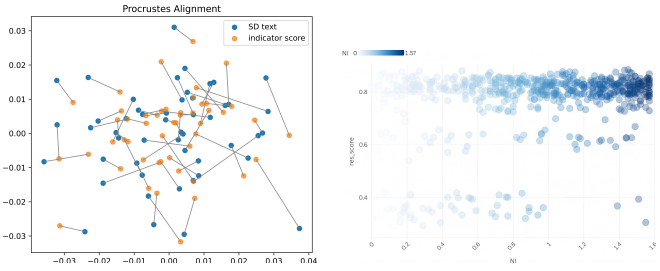

Fig. 6: (a) **Procrustes analysis of the distribution of environmental indices and SD text within VAE's shared latent space.** (b) **The relationship between Naturalness (NI) index and restorative effect scores.** The horizontal axis represents the index values, and the vertical axis represents the corresponding restoration scores.

B. Restorative Patterns in Rules

Figure 6(b) illustrates the non-linear and intertwined relationship between environmental indices and restorative effects. These relationships are often influenced by both intra-index correlations and spatial co-occurrence effects, consistent with Tobler's first law of geography [31].

To capture these complex interactions and uncover interpretable restorative patterns from indices, we adopt the Rule-based Representation Learner (RRL) [32] as a rule extraction tool. Since RRL is inherently a classification model, we discretize the continuous restorative effect scores into five ordinal classes: *very_low* ([0–0.3)), *low* ([0.3–0.5)), *medium* ([0.5–0.8)), *high* ([0.8–0.95)), and *very_high* ([0.95–1.0]). These classes serve as prediction targets for RRL, which is trained on environmental index features x_{index}. The model outputs a set of weighted decision rules that link specific spatial patterns to different levels of restorative effect. We perform 5-fold cross-validation to ensure generalizability. Unlike black-box regressors, RRL provides human-readable rules with interpretable scores for each class. This allows us to analyze which spatial configurations—such as high sky ratio or dense greenery—are most associated with strong restorative outcomes in university open spaces.

Representative rules are visualized in Figure 7, where several insightful patterns emerge from the learned rules:

- **Interaction effects:** When green coverage (*GC*) exceeds 0.247, the likelihood of above-average restoration is 72.5%. However, if pavement ratio (*PV*) also exceeds 0.170, the probability of a very low restorative level increases to 40%, and high-level restoration drops below 20%. This indicates that excessive pavement can suppress the positive effects of greenery.
- **Contextual dependency:** A low pavement ratio alone ($PV \leq 0.170$) does not guarantee higher restoration; 41% of such spaces are still rated at low levels. This suggests that single-element optimization is insufficient, and spatial configuration must be assessed holistically.

Beyond indices, we also analyze semantic spatial quality variables derived from the SD scales (Figure 7 below), where following rules are found:

- **Vegetation and maintenance:** When leisure facility maintenance exceeds 4.008, the probability of a very high restorative score is 86.53%. Similarly, when plant species diversity is greater than 3.156, 78.28% of spaces reach high-level restoration.
- **Aesthetic and safety factors:** High scores in pavement color (> 6.244), space accessibility (> 6.397), and spatial safety (> 5.092) all significantly improve restorative outcomes. These findings emphasize the importance of visual appeal, access, and a sense of security during movement.

Together, these rules reflect the multi-dimensional nature of restorative effects and validate the use of rule-based learning to uncover interpretable patterns linking spatial features to human-centered outcomes.

C. Cross-method Verification

Our analytical results demonstrate mutual reinforcement with our preliminary experimental findings. Eye-tracking data reveal that natural elements (e.g., trees, water bodies) and landmark architectures serve as visual foci, while low-attention elements such as roads and skies are identified as potential suppressive factors in RRL (e.g., high PV weakens GC's restorative effects). Furthermore, the observed "spatial depth preference" in eye-tracking aligns with the moderating effect of SS (spatially separated index) in RRL, suggesting that spatial organization logic (e.g., hierarchical separation) may influence restorative perception through nonexplicit pathways. Its directional consistency with RRL's large-scale rules validates the dual mechanism of environmental elements: explicit visual foci (natural elements) directly drive restorative effects, while implicit spatial structures (e.g., index interactions) indirectly regulate outcomes through complex nonlinear pathways.

In one failure case (Figure 7), our model overestimate the restorative effect of a courtyard. While the scene was rich in greenery, it received low ground-truth ratings from humans, likely due to a cluttered composition with overlapping sculptures and visible scaffolding. This suggests our model struggles when positive features (e.g., vegetation) are combined with negative, atypical spatial arrangements. Future improvements should focus on: (1) training on a more diverse dataset that includes such complex scenes, (2) incorporating metrics for spatial clutter, and (3) improving the model's ability to identify and penalize visually disruptive elements.

VI. Conclusion and Limitation

This study presents a multimodal perception framework that combines environmental psychology and machine learning to evaluate the restorative potential of university open spaces. By integrating visual indices, semantic impressions, and psychological scales, we construct a cross-modal system capable of predicting restorative effects from either image or text inputs. The Semantic Differential (SD) descriptions and a Rule-based Representation Learner (RRL) enhances both the interpretability and flexibility of the model. Our framework requires only a single image or textual description to generate a reliable restorative score, eliminating the need for field surveys or expert evaluation. This makes it applicable not only for post-occupancy assessment but also during early design stages, offering a lightweight and scalable tool for evidence-based, health-oriented campus planning.

Nonetheless, our framework's reliance on static 2D images from a single region restricts its ability to capture dynamic experiences and generalize to diverse environments. Furthermore, its visual-semantic alignment is brittle for atypical scenes. Future work could address this by using richer data modalities (e.g., video, 3D models) and exploring more robust alignment techniques.

References

[1] K. Batra, M. Sharma, R. Batra, T. P. Singh, and N. Schvaneveldt, "Assessing the psychological impact of covid-19 among college students:

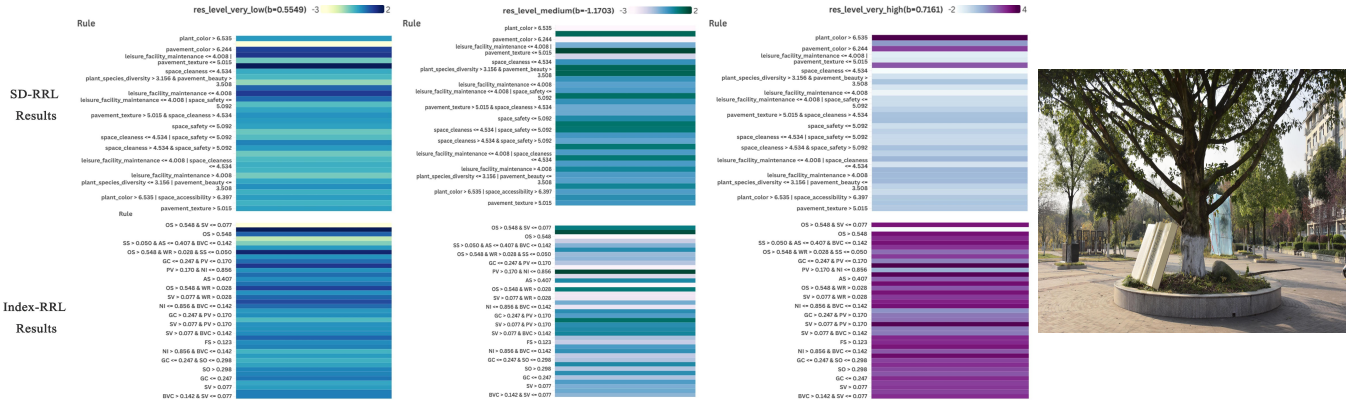

Fig. 7: (a) **Decision rules extracted by RRL, illustrating how environmental indices and spatial quality features from SD scales are associated with different levels of restorative effect.** Three representative classes (*very_low*, *medium*, and *very_high*) are selected for interpretability. (b) A failure case from predictions of our framework.

an evidence of 15 countries," in *Healthcare*, vol. 9, p. 222, MDPI, 2021.

[2] J. M. Twenge, B. Gentile, C. N. DeWall, D. Ma, K. Lacefield, and D. R. Schurtz, "Birth cohort increases in psychopathology among young americans, 1938–2007: A cross-temporal meta-analysis of the mmpi," *Clinical psychology review*, vol. 30, no. 2, pp. 145–154, 2010.

[3] J. Li, S. Chen, H. Xu, and J. Kang, "Effects of implanted wood components on environmental restorative quality of indoor informal learning spaces in college," *Building and Environment*, vol. 245, p. 110890, 2023.

[4] A. M. Salama, "When good design intentions do not meet users expectations: Exploring qatar university campus outdoor spaces," *ArchNet-IJAR: International Journal of Architectural Research*, vol. 2, no. 2, pp. 57–77, 2008.

[5] R. S. Ulrich, R. F. Simons, B. D. Losito, E. Fiorito, M. A. Miles, and M. Zelson, "Stress recovery during exposure to natural and urban environments," *Journal of environmental psychology*, vol. 11, no. 3, pp. 201–230, 1991.

[6] D. Li and W. C. Sullivan, "Impact of views to school landscapes on recovery from stress and mental fatigue," *Landscape and urban planning*, vol. 148, pp. 149–158, 2016.

[7] T. Hartig, "Issues in restorative environments research: Matters of measurement," *Psicología ambiental*, vol. 2011, pp. 41–66, 2011.

[8] T. R. Herzog, A. M. Black, K. A. Fountaine, and D. J. Knotts, "Reflection and attentional recovery as distinctive benefits of restorative environments," *Journal of environmental psychology*, vol. 17, no. 2, pp. 165–170, 1997.

[9] X. Wang, S. Rodiek, C. Wu, Y. Chen, and Y. Li, "Stress recovery and restorative effects of viewing different urban park scenes in shanghai, china," *Urban forestry & urban greening*, vol. 15, pp. 112–122, 2016.

[10] S. Han, D. Song, L. Xu, Y. Ye, S. Yan, F. Shi, Y. Zhang, X. Liu, and H. Du, "Behaviour in public open spaces: A systematic review of studies with quantitative research methods," *Building and Environment*, vol. 223, p. 109444, 2022.

[11] X. Han, L. Wang, S. H. Seo, J. He, and T. Jung, "Measuring perceived psychological stress in urban built environments using google street view and deep learning," *Frontiers in public health*, vol. 10, p. 891736, 2022.

[12] H. Ma, Y. Zhang, P. Liu, F. Zhang, and P. Zhu, "How does spatial structure affect psychological restoration? a method based on graph neural networks and street view imagery," *Landscape and Urban Planning*, vol. 251, p. 105171, 2024.

[13] H. Wen, H. Lin, X. Liu, W. Guo, J. Yao, and B.-J. He, "An assessment of the psychologically restorative effects of the environmental characteristics of university common spaces," *Environmental Impact Assessment Review*, vol. 110, p. 107645, 2025.

[14] X. Fu, "Do street-level scene perceptions affect housing prices in chinese megacities? an analysis using open access datasets and deep learning," *Environment and Planning B: Urban Analytics and City Science*, 2021. Accessed via open data and deep learning approaches.

[15] J. Tang and Y. Long, "Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in beijing," *Landscape and Urban Planning*, vol. 191, p. 103436, 2019.

[16] R. Wang, Y. Liu, Y. Lu, J. Zhang, P. Liu, Y. Yao, and G. Grekousis, "Perceptions of built environment and health outcomes for older chinese in beijing: A big data approach with street view images and deep learning technique," *Computers, Environment and Urban Systems*, vol. 78, p. 101386, 2019.

[17] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The measurement of meaning*. No. 47, University of Illinois press, 1957.

[18] S. Kaplan, "The restorative benefits of nature: Toward an integrative framework," *Journal of environmental psychology*, vol. 15, no. 3, pp. 169–182, 1995.

[19] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. oup Oxford, 2011.

[20] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1290–1299, 2022.

[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014.

[22] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[23] P. He, J. Gao, and W. Chen, "Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing," *arXiv preprint arXiv:2111.09543*, 2021.

[24] D. P. Kingma, M. Welling, *et al.*, "Auto-encoding variational bayes," 2013.

[25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[27] "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5, pp. 183–197, 1991.

[28] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[29] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, pp. 81–106, 1986.

[30] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[31] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[32] Z. Wang, W. Zhang, N. Liu, and J. Wang, "Learning interpretable rules for scalable data representation and classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.