

<https://doi.org/10.1038/s40494-025-01855-z>

Toward sustainable and differentiated protection of cultural heritage illustrated by a multisensory analysis of Suzhou and Kyoto using deep learning

Jiao Zhang^{1,2,3}✉, Yangyang Shi⁴, Liang Zhao⁵, Chenshu Cai⁶ & Katsunori Furuya³

Residents' perception is essential to cultural heritage (CH) and place identity, making its integration into sustainable conservation important. This study analyzes online reviews from Suzhou and Kyoto using deep learning to extract multisensory descriptions and physical elements. Sentiment analysis shows that overall perceptions are predominantly positive. In Suzhou, visual perception dominates, with a dynamic spatial experience, while other sensory inputs are limited. Kyoto offers richer multisensory engagement and greater openness, though with less spatial variation. Visitors differ significantly in their perceptions of sensory experiences and physical settings. Multiple linear regression indicates that multisensory engagement enhances overall perception, shaped by the environment. However, cost and accessibility are key negative factors influencing impressions. This study highlights the importance of incorporating multisensory public perceptions into CH conservation and supports differentiated, site-specific strategies. The findings contribute to a nuanced understanding of sustainable heritage space optimization based on multisensory experiences and contextual factors.

Cultural heritage (CH) has increasingly been recognized as a vital element of sustainability. Traditionally, sustainability debates have focused primarily on ecological and physical aspects—such as environmental conservation and urban infrastructure—while often overlooking the multifaceted contributions of CH. UNESCO has long underscored the significance of CH in sustainable development, as evidenced by the 2011 Recommendation on the Historic Urban Landscape (HUL)¹ and its incorporation into the UN 2030 Agenda for Sustainable Development². These frameworks demonstrate that CH not only bolsters environmental and economic sustainability but also plays a crucial role in promoting social equity and reinforcing cultural identity.

Many urban CH sites have been integrated into city systems, functioning as essential components that provide Cultural Ecosystem Services (CES) and enhance the well-being of urban residents^{3–6}. Through the 1972 World Heritage Convention, UNESCO has spearheaded global CH conservation by designating 897 sites worldwide, and numerous countries and regions have also implemented policies to protect both tangible and intangible CH⁷. Despite these efforts, urbanization, commercialization, and

tourism development continue to challenge CH preservation⁸. Overuse and rapid urban expansion threaten the integrity of many heritage sites, underscoring the need for a balanced approach between heritage protection and sustainable development⁹.

CH not only attracts higher tourist flows and augments local commercial value—thereby significantly influencing city investments^{10,11}, but its excessive development and commercialization can also yield adverse effects¹². For example, the protection of historical sites may constrain urban planning and contribute to traffic congestion, as observed in ancient urban areas like Beijing¹³. Moreover, CH can affect housing equity; in areas such as Hangzhou's West Lake, the presence of CH has led to a significant premium in surrounding property prices, as it is viewed both as an irreplaceable resource and a key selling point for developers¹⁴.

Both the positive and negative impacts of CH are poised to have long-term and far-reaching effects on urban and CH sustainability. Especially for cities that center their tourism and consumption industries around these resources, achieving a balance between heritage protection and sustainable development is imperative^{4,5,15,16}. Otherwise, CH may deviate from its

¹Fine Arts School, Hangzhou Normal University, Hangzhou, China. ²Institute of Industrial Science, The University of Tokyo, Tokyo, Japan. ³Faculty of Horticulture, Chiba University, Matsudo, Japan. ⁴Division Forest, Nature and Landscape, KU Leuven, Leuven, Belgium. ⁵School of Finance, Shandong Technology and Business University, Yantai, China. ⁶College of Software Technology, Zhejiang University, Hangzhou, China. ✉e-mail: 20240025@hznu.edu.cn

authentic historical form, resulting in negative experiences for stakeholders and diminishing benefits for the broader community^{17–20}. Subsequently, as a special form of urban green space, CH serves as the main body of the urban CES and promoting urban sustainable development. Exploring a win-win solution between CH and sustainable development is a concern for scholars, planners, and policymakers. Finally, adopting a human-centered sustainable development approach—one that is firmly based on residents' well-being—offers a robust framework for ensuring equitable and enduring urban progress^{5,21}.

With rapid urbanization, public engagement with urban environments has attracted widespread scholarly attention^{22,23}. Prior studies on urban green spaces have demonstrated that natural settings can significantly contribute to residents' physical and mental well-being, for instance, by reducing stress through activities like walking or enjoying scenic views^{24–27}. Beyond their ecological benefits, cultural and spiritual values within urban environment are equally important^{28,29}. Modern urban planning increasingly integrates cultural and natural landscapes, reflecting a shift toward holistic, people-centered development³⁰.

Urban perception refers to how residents interpret and interact with their surroundings, including streets, parks, and historic sites³¹. While previous studies in the field of urban research have laid a valuable foundation by demonstrating how environmental perception influences well-being, understanding public perceptions of cultural space requires an additional focus on historical, cultural, and social dimensions^{7,32–35}. In this context, modern human-centered urban planning increasingly emphasizes public engagement and participatory decision-making, specifically aimed at enhancing the management and conservation of CH^{36,37}.

CH itself, which includes both tangible elements (e.g., historic buildings, gardens) and intangible aspects (e.g., traditions, folklore), is inherently linked to subjective perception and place attachment^{38,39}. For example, while heritage sites in locations like West Lake, Hangzhou, contribute to both ecological and cultural experiences, their true value lies in the rich cultural narratives and emotional connections they foster among local communities¹⁵. Intangible CH, being harder to preserve, relies on social perception, education, and intergenerational transmission^{40,41}. Consequently, CH conservation extends beyond material protection to include cultural significance and experiential continuity⁴².

Compared to natural environments, cultural heritage or landscape perception is more dependent on human mediation, involving multi-sensory engagement that shapes visitors' experiences⁴³. This highlights the need to study CH perception beyond visual elements, incorporating auditory, olfactory, and tactile dimensions. The perception of the site is shaped by multi-sensory interactions, with vision being the most extensively studied sense^{44–46}. Factors such as color, shape, and spatial composition significantly influence visitors' experiences^{47–52}. However, other sensory modalities, such as soundscapes, scents, and textures, remain underexplored due to methodological challenges⁵³. Emerging studies suggest that multi-sensory engagement enhances visitor satisfaction, cultural appreciation, and place attachment, emphasizing the need for a holistic approach to cultural perception^{54–57}.

Despite its importance, multi-sensory CH perception remains understudied across different cultural contexts. Unlike purely natural settings, CH environments worldwide include both natural and human-made elements, such as architecture, cultural activities, and intangible heritage⁵⁸. Subsequently, the sensory experience involved is not only a single visual factor; other senses may also affect tourists' overall perception, evaluation results, willingness to pay, or other behavioral decisions⁵⁹. Understanding multi-sensory experiences in various CH contexts not only provides feedback on heritage conservation strategies but also fosters cross-cultural learning and enhances the global appreciation of CH³³. By analyzing the public perception across different geographical and historical settings, this research contributes to a broader discourse on CH perception, offering insights for the sustainable management of heritage sites worldwide.

Traditionally, studies on public perception of CH have relied on surveys, interviews, and controlled experiments^{60–62}. While these methods provide valuable insights, they are often time-consuming, resource-

intensive, and geographically limited^{63,64}. In contrast, online reviews offer a rich, user-generated dataset that captures real-time, large-scale public perceptions of CH sites^{65,66}. The increasing accessibility of CH tourism has led to an abundance of online narratives, reflecting visitors' experiences, preferences, and cultural interpretations^{67–70}. These online reviews—which include text, images, and ratings posted on platforms such as TripAdvisor, Google Reviews, and social media—serve as valuable data sources for analyzing public perceptions of CH. Compared to traditional survey methods, online reviews provide more spontaneous and diverse feedback, covering a broader range of visitor demographics and capturing multi-dimensional insights into CH experiences. Social media and review platforms have been widely used in urban studies, disaster response, and environmental planning, demonstrating their reliability as perception assessment tools⁶². For CH, integrating online user-generated content into heritage conservation can enhance public participation and decision-making, helping policymakers understand how visitors engage with cultural understanding and what improvements are needed^{71,72}.

Despite the growing recognition of multi-sensory experiences in CH studies, previous research has rarely utilized large-scale online reviews to analyze visitor perceptions. Most existing studies either rely on small-scale surveys or analyze textual and visual data separately, failing to capture the holistic, multi-sensory experience of CH sites. Additionally, comparative studies across culturally similar heritage sites remain limited, making it difficult to identify shared characteristics, contextual differences, and opportunities for mutual learning.

To address this gap, this study leverages large-scale online reviews to investigate visitor perceptions of CH sites in Suzhou, China, and Kyoto, Japan, both of which share an East Asian CH context. By applying a multi-sensory analytical framework, this research aims to link subjective perceptions to objective physical spaces, uncovering how different sensory elements shape visitors' experiences. Furthermore, this study explores how such insights can enhance CH management, allowing for targeted improvements based on visitor feedback and cross-cultural knowledge exchange.

This study examines how multi-sensory perceptions influence CH experiences by comparing CH sites in Suzhou and Kyoto through large-scale online reviews. The key research questions are: What are the unique multi-sensory characteristics of CH in Suzhou and Kyoto based on subjective perception? What are the key differences in sensory experiences between the CH of Suzhou and Kyoto? How do these multi-sensory experiences influence overall visitor perception in each city? How can multi-sensory elements be integrated into sustainable conservation strategies for CH?

By addressing these questions, this research provides a systematic approach to understanding visitor perceptions. It offers insights into the multi-sensory composition of public perception and the unique characteristics of different CH sites, thereby informing differentiated conservation measures and integrating subjective visitor experiences into management optimization and sustainable preservation strategies.

Methods

Study area

Suzhou and Kyoto are famous for their CH; both cities have many renowned UNESCO World Cultural Heritage (WCH) sites and many non-WCH sites protected by local governments. Although these non-WCH sites do not have the same high reputation as the WCH sites, they still attract many visitors and were also selected as our research sites. In total, 20 CH sites from Suzhou and 20 CH sites from Kyoto were selected for this study to investigate whether the perception of the two cities demonstrates differences in online reviews and how landscape and multi-sensory elements affect overall perceptions (Fig. 1 and Table 1).

Suzhou is in East China, southeast of Jiangsu Province. It is one of China's first national historical and cultural cities, with a history of nearly 2500 years⁷³. The Suzhou Classical Garden and the Suzhou section of the Beijing-Hangzhou Grand Canal have been listed as UNESCO WCH sites⁷⁴.



Fig. 1 | Research Site. **a** The red dots indicate the locations of the two cities. **b** World Cultural Heritage sites in Suzhou. Blue dots represent World Cultural Heritage sites, red dots represent non-World Cultural Heritage sites, and darker colors indicate a

greater number of reviews. **c** World Cultural Heritage sites in Kyoto. Blue dots represent World Cultural Heritage sites, red dots represent non-World Cultural Heritage sites, and darker colors indicate a greater number of reviews.

Kyoto, located in western Japan, is an important city in the Osaka metropolitan area. Some of the city's historic buildings and gardens were listed as WCH sites in 1994 as part of the Cultural Wealth of the Ancient Capital of Kyoto⁷⁵.

Both Suzhou and Kyoto exhibit a rich historical and CH sites, boasting a long history with numerous gardens and historic buildings—several of which are recognized as WCH sites—thus making them prototypical CH cities. Moreover, both cities are economically developed and have a

Table 1 | Cultural Heritage Sites Included in This Study

Suzhou			Kyoto		
No.	Name	WCH (Y/N)	No.	Name	WCH (Y/N)
S1	Humble Administrator's Garden	Y	K1	Kiyomizu-dera Temple	Y
S2	Lingering Garden	Y	K2	Kinkakuji Temple	Y
S3	Master of the Nets Garden	Y	K3	Nijo Castle	Y
S4	Mountain Villa with Embracing Beauty	Y	K4	Tō-ji Temple	Y
S5	Canglang Pavilion	Y	K5	Ginkakuji Temple	Y
S6	Lion Grove Garden	Y	K6	Tenryū-ji Temple	Y
S7	Couple's Retreat Garden	Y	K7	Shimogamo Shrine	Y
S8	Garden of Cultivation	Y	K8	Ryōan-ji Temple	Y
S9	Tuisi Garden	Y	K9	Kamigamo Shrine	Y
S10	Grand Canal (Suzhou Section)	Y	K10	Ninnaji Temple	Y
S11	Tiger Hill	N	K11	Nishi Hongan-ji Temple	Y
S12	Tongli Historical Town	N	K12	Daigo-ji Temple	Y
S13	Mudu Historical Town	N	K13	Fushimi Inari-taisha	N
S14	Hanshan Temple	N	K14	Kyoto Imperial Palace	N
S15	Shantang Street	N	K15	Heian Shrine	N
S16	Jinxi Historical Town	N	K16	Sanjūsangen-dō	N
S17	Pingjiang Historical District	N	K17	Nanzen-ji Temple	N
S18	Luzhi Historical Town	N	K18	Koudai-ji Temple	N
S19	Fengqiao Scenic Area	N	K19	Kyoto Gyoen National Garden	N
S20	Qiandeng Historical Town (Kunshan)	N	K20	Katsura Imperial Villa	N
Total review number		37,649			17,485

Table 2 | Confusion Matrix of Translation Consistency between Human and Machine on Culturally Specific Terms

Human Translation	Machine Translation = Inaccurate (0)	Machine Translation = Accurate (1)	Total
Inaccurate (0)	16	3	19
Accurate (1)	11	170	181
Total	27	173	200

longstanding tradition of heritage protection and sustainable urban development. As prominent CH tourism destinations, they attract many visitors each year, generating ample online review data that supports comprehensive analysis of overall visitor perceptions. Collecting this big data further aids in understanding site perception and the distinctive urban characteristics of each locale. The similarities and differences between Suzhou and Kyoto provide a robust context for applying these methods to detect variations in multi-sensory experiences, ultimately helping to analyze the distinctive characteristics—especially the subtle multi-sensory differences—and to offer targeted recommendations for spatial environment management and heritage conservation.

Data collection—integrating photos and text for a holistic understanding of visitor perception

Previous studies have commonly used Visitor-Employed Photography (VEP) to analyze tourist perceptions, where visitors take photos to document their experiences and highlight visually significant elements^{76–78}. However, while photos capture explicit preferences in terms of landscape features and spatial composition, they do not provide insight into the emotions, reasoning, or sensory experiences that influence those choices.

In contrast, textual reviews, often generated in unstructured formats, reflect visitors' subjective evaluations, detailing their likes, dislikes, and overall impressions. However, these textual descriptions may lack clear

spatial references, making it difficult to determine which specific landscape elements or cultural features contributed to sentiments.

By combining photo and text analysis, this study provides a more comprehensive understanding of CH perception, offering several key advantages. First, it enhances interpretability by using photos to identify specific landscape elements that attract visitors, while text provides the reasoning behind their preferences. Additionally, it captures multi-sensory experiences, as text reveals subjective perceptions and sensory details such as sound and atmosphere, which cannot be directly inferred from images alone. Finally, this approach provides a holistic understanding by allowing us to explore correlations between landscape elements and textual sentiment, offering a more comprehensive analysis of CH perception.

The review data for the CH sites in Suzhou and Kyoto came from China's largest travel platform, Ctrip, and Japan's Google Maps. Their popularity attracts many visitors to upload reviews, generating more exposure and access voluntarily; subsequently, the two platforms are widely used in research^{79,80}. Moreover, because the visitors come from different language backgrounds, Google Automatic Translate was used to unify the languages. To address the requirement for formal agreements to harvest the surveys used in our research, we have created a Google Cloud account and are using a valid API key to access data through the official Google Maps Platform APIs, strictly adhering to the Google Maps Platform Terms of Service and Google Terms of Service; similarly, for Ctrip, we ensure adherence to Ctrip's terms of service and privacy policies, collecting only publicly available reviews without personal information, used anonymously for research, and not for commercial purposes or competition with Google services^{81,82}. Data collection was completed in October 2023, so the data collected is up to that time. After removing duplicate data and empty comments, Suzhou and Kyoto received 37,649 and 17,485 reviews, respectively (Table 1). To address the difference in review volume between Suzhou and Kyoto, we emphasize that such disparities are common in cross-site comparative studies and do not compromise the validity of perception analysis. Instead of relying on absolute review counts, we focus on content-based and proportional analysis, ensuring that key themes and sensory

Table 3 | Cohen’s Kappa statistics for translation consistency

Statistic	Value	Standard Error (ASE)	Approximate T	Degrees of Freedom (df)	Significance (p-value)
Cohen’s Kappa	0.657	0.049	13.40	—	< 0.001
Weighted Cohen’s Kappa	0.682	0.045	15.13	—	< 0.001

A kappa value of 0.657 indicates substantial agreement between human and machine translations on culturally specific terms.
p-value < 0.001 indicates the agreement is statistically significant.

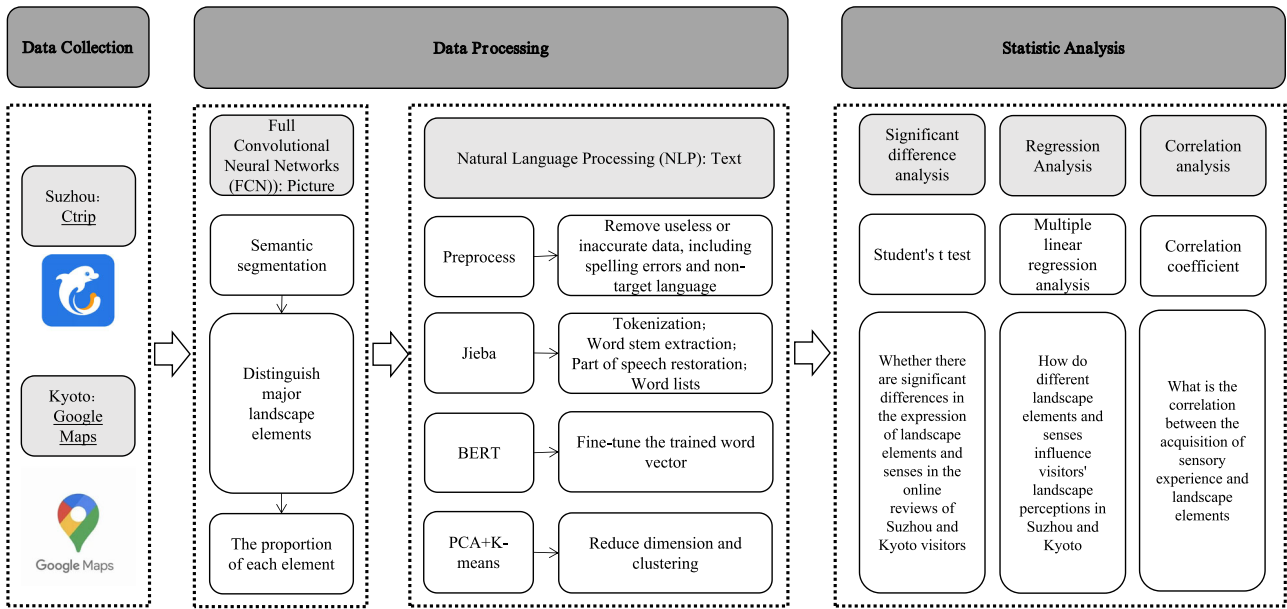


Fig. 2 | Picture and text processing workflow.

Table 4 | Classification of Semantic Segmentation Elements

Category	Elements
Architecture	wall building; edifice windowpane; window door; double door house column; pillar skyscraper grandstand; covered stand grandstand; covered stand stairway; staircase screen door; screen toilet; can; commode; crapper; pot; potty; stool; throne bar hovel; hut; hutch; shack; shanty tower stage step; stair
Landscape structure	fence; fencing lamp signboard; sign bench arcade machine bridge; span streetlight; streetlamp pole stool bannister; banister; balustrade; balusters; handrail sculpture ashcan; trash can; garbage can; wastebin; ash bin; ashbin; ashbin; dustbin; trash barrel; trash bin monitor; monitoring device bulletin board; notice board flag
Road	road; route sidewalk; pavement earth; ground railing; rail path traffic light; traffic signal; stoplight runway dirt track land; ground; soil
Transportation	car; auto; automobile; machine; motorcar boat bus; autobus; coach; char banc; double-decker; jitney; motorbus; motorcoach; omnibus; passenger vehicle truck; motortruck airplane; aero plane; plane van ship minibike; motorbike bicycle; bike; wheel; cycle
Sky	sky
Vegetation	tree grass plant; flora; plant life flower palm; palm tree pot; flowerpot
Mountain	mountain; mount rock; stone sand hill
Water	water sea river fountain swimming pool; swimming bath; natatorium waterfall; falls lake
People	person; individual; someone; somebody; mortal; soul
Animal	animal; animate being beast; brute; creature; fauna
Food	food; solid food
Interior	floor; flooring ceiling bed cabinet table curtain; drape; drapery; mantle; pall chair sofa; couch; lounge shelf mirror rug; carpet; carpeting armchair seat desk wardrobe; closet; press bathtub; bathing tub; bath; tub cushion base; pedestal; stand box chest of drawers; chest; bureau; dresser counter sink fireplace; hearth; open fireplace refrigerator; icebox pool table; billiard table; snooker table pillow bookcase blind; screen coffee table; cocktail table countertop stove; kitchen stove; range; kitchen range; cooking stove kitchen island computer; computing machine; computing device; data processor; electronic computer; information processing system swivel chair towel chandelier; pendant; pendent television receiver; television; television set; tv; tv set; idiot box; boob tube; telly; goggle box escalator; moving staircase; moving stairway ottoman; pouf; pouffe; puff; hassock buffet; counter; sideboard washer; automatic washer; washing machine plaything; toy barrel; cask basket; handbasket bag cradle oven ball tank; storage tank dishwasher; dish washer; dishwashing machine screen; silver screen; projection screen blanket; cover hood; exhaust hood sconce vase tray microwave; microwave oven fan crt screen shower plate radiator glass; drinking glass clock

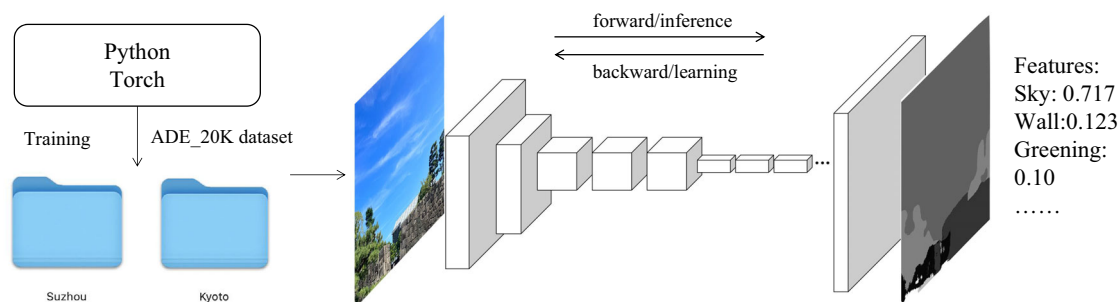


Fig. 3 | FCN flowchart of this study. FCN full convolutional neural network.

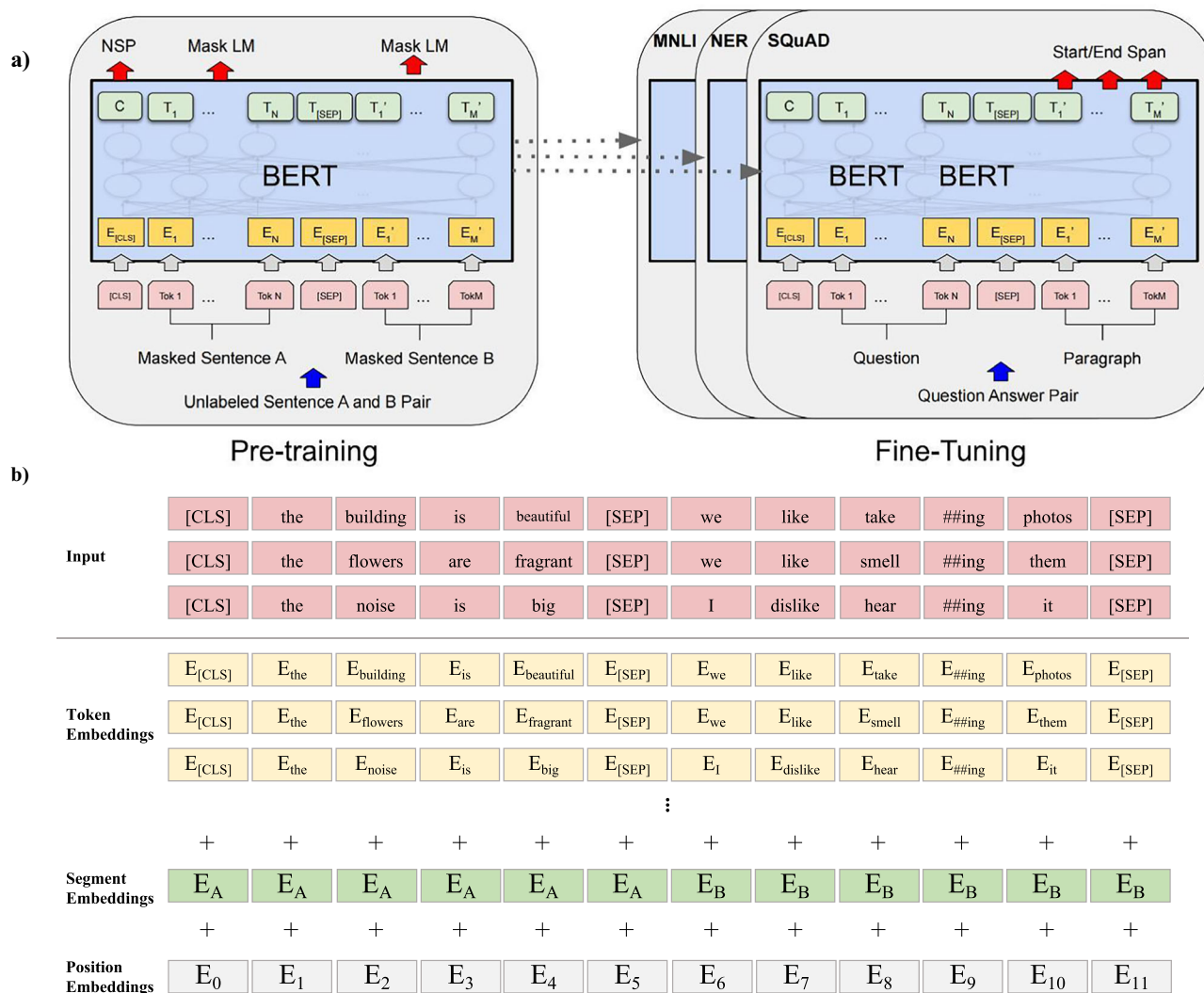


Fig. 4 | Overview of BERT pre-training, fine-tuning procedures, and input representation. **a** Overall pre-training and fine-tuning procedures for BERT. Apart from the output layers, the same architecture is used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different downstream tasks. During fine-tuning, all parameters are updated. [CLS] is

a special token added at the beginning of each input example, and [SEP] is a special separator token (e.g., between questions and answers). **b** BERT input representation. The input embeddings are the sum of token embeddings, segment embeddings, and position embeddings¹¹⁵. This study uses three short sentences as an example.

elements are assessed relative to the total dataset size. These methodological adjustments ensure that our results remain robust, comparable, and reflective of real visitor experiences, despite variations in sample size.

To ensure reliable cross-linguistic analysis, we adopted a translation strategy supported by recent advances in neural machine translation.

Transformer-based models like Google Translate⁸³ and large language models such as GPT-4⁸⁴ have shown strong performance in preserving semantic and cultural context across languages. To verify translation quality, we sampled 200 reviews from Suzhou and Kyoto containing culturally specific terms (proportional from Kyoto and Suzhou). Comparing machine

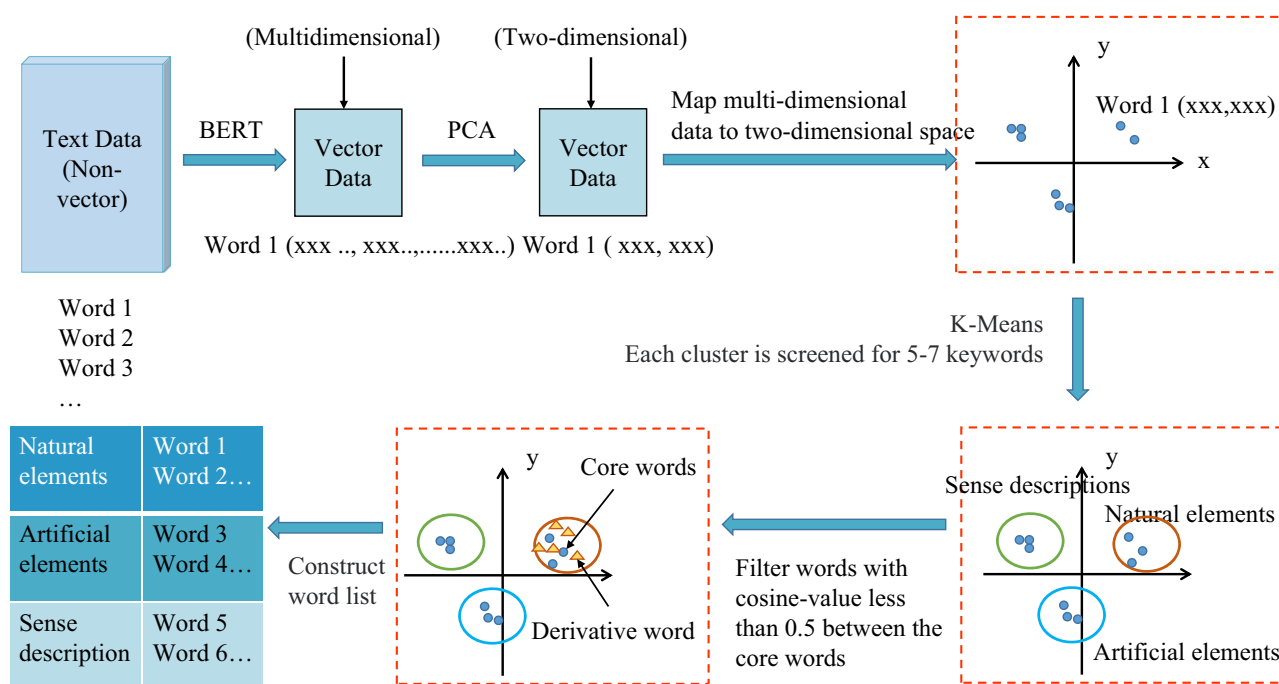


Fig. 5 | Technical framework of text processing.

translations to human-annotated versions, we observed 93.0% agreement and a Cohen's Kappa of 0.657 ($p < 0.001$), indicating substantial consistency (Table 2 and Table 3). To further reduce ambiguity, we applied synonym normalization and manually reviewed unclear terms during preprocessing.

Deep learning processing

In this study, two deep learning (DL) techniques are employed to comprehensively mine effective information from online reviews. We adopt a full convolutional neural network (FCN) for semantic segmentation of images to extract landscape element information, and natural language processing (NLP) for text analysis to capture sensory descriptions and identify CH features (Fig. 2).

The application of DL in multidisciplinary fields has greatly enhanced our ability to process various data types. For instance, DL-based analysis of large volumes of street view maps is widely used across many research domains^{37,85}. The application of DL in multidisciplinary fields has greatly enhanced our ability to process various data types. For instance, DL-based analysis of large volumes of street view maps is widely used across many research domains⁸⁰. Semantic analysis of photographs and emotional analysis of online text have become common practices—for example, in studies exploring the relationship between landscape facilities and residential rental prices⁸⁶. Although deep learning algorithms remain somewhat immature, they have already proven to be superior to manual methods, especially when processing and analyzing multimodal data such as text, images, and audio⁸⁷. For instance, some studies have demonstrated that the use of convolutional neural networks (CNNs) can improve data accuracy by more than 15%⁶⁹.

Geo-tagged photos from social media platforms are typically used to reveal visitors' landscape preferences, but most research has primarily focused on their geographic and temporal distribution rather than delving into the rich, underlying information. Manual extraction of elements from photos in online reviews is inefficient^{70,88}. Therefore, we employ FCN for automated image segmentation. Similarly, we use NLP methods to analyze the text in online reviews by clustering feature words with consistent word vectors, thereby facilitating both statistical and emotional analyses⁸⁰. Given the absence of a mature dataset for training models to identify characteristic

sensory vocabulary, we followed the approach of Koblet et al., who used a dictionary-based method to compile lists of sensitivity-related adjectives and nouns, and created a tailored vocabulary list for this study⁶. Based on this, we created the vocabulary list for this study. Python 3.7.4 was used for data acquisition, model construction, and processing.

Visitors tend to post photos that reflect their own preferences. Traditional methods such as VEP are widely used to understand visitor perceptions^{78,89}. However, manually classifying or counting landscape elements from a large volume of photos is both labor-intensive and prone to inaccuracies.

The FCN method addresses these challenges by replacing the fully connected layers of traditional semantic segmentation with convolutional layers. This allows FCN to accept input images of any size and to use deconvolution layers to up sample feature maps back to the original image size, enabling pixel-level classification, consequently, we adopted the FCN method for photo processing⁸⁸. For training the model, we used the ADE_20K dataset—annotated and published by the MIT CSAIL Computer Vision Group^{90–92}. This dataset comprises 25,000 multi-scene photos annotated with 150 object categories, making it suitable for analyzing the complex, multi-scene images generated by visitors⁹³.

In line with previous studies and the aims of our research, we removed irrelevant scene elements, retained 144 types of markers, and consolidated these into 12 major landscape element categories for subsequent model training (Table 4)^{6,47}. In total, 25,574 images from the dataset were used as the training set, and 2000 as the validation set. The FCN model was implemented and fine-tuned using PyTorch.

After deduplication and cleaning, the photos from online reviews of Suzhou and Kyoto were organized into two separate folders. Custom code was used to read the photo data, which was then input into the FCN to generate semantic segmentation results. These results allowed us to calculate the proportion of landscape elements favored by visitors, providing the basis for subsequent statistical analyses (Fig. 3).

For text data, we employed Bidirectional Encoder Representations from Transformers (BERT) as our primary tool for NLP. BERT is a transformer-based bidirectional encoder designed to pre-train deep

Table 5 | Environmental Feature Lexicon Used for Text Analysis

First layer	Second layer	Suzhou	Kyoto
Elements of natural experience	Greening	tree; grass; bamboo; pine; maple; willow; other plants; afforest	tree; grass; bamboo; pine; maple; willow; other plants; afforest
	Flower	flower; plum; orchid; chrysanthemum; lotus; peach; Osmanthus; flower related	flower; plum; chrysanthemum; lotus; peach flower; cherry; flower related
	Water	river; pool; pond; lake; waterfall; creek; shore; embankment; spring; water	river; pool; pond; lake; waterfall; creek; shore; embankment; spring; water
	Mountain	mountain	mountain
	Stone	stone	stone
Artificial environment experience elements	Animal	bird; insect; aquatic; others	bird; insect; aquatic; others
	Natural phenomenon	snow; wind; rain; smog; fog; cloud; thunder; sun; moon; star; sky; clear day; overcast; night	snow; wind; rain; smog; fog; cloud; thunder; sun; moon; star; sky; clear day; overcast; night
	Season	spring; summer; autumn; winter; season	spring; summer; autumn; winter; season
	Architecture	pavilion; tower; building; attic; pagoda; small room; terrace; boat room	pavilion; tower; building; attic; pagoda
	Cultural facilities	bell; picture; carved; statue	bell; picture; carved; statue
	Structure	seat; lamp; rockery; stone; bridge; gallery	seat; lamp; rockery; stone; bridge; gallery
	Space	room; venue; hall; lobby; shop; room; street; site; port	room; venue; hall; lobby; shop; room; street; site; port
	Serve	toilet; parking; management; commemorate; serve	toilet; parking; management; commemorate; serve
	Cost	fee; ticket; price; discount; expensive; value	fee; ticket; price; discount; expensive; value
	People	people type; number of people; age; student; friend; group; male; female	people type; number of people; age; student; friend; group; male; female
Senses description	Event or Festival	view and admire; opera; walk; take photos; relaxation; festival; fireworks	view and admire; opera; walk; take photos; relaxation; festival; fireworks
	Vision	eye; see; vision; red; yellow; green; blue; green; white; purple; pink; black; gold; silver; color; flash; light; shadow; multicolored	eye; see; vision; red; yellow; green; blue; green; white; purple; pink; black; gold; silver; color; flash; light; shadow; multicolored
	Hearing	ear, listen; voice; yell; sing; quiet; silence; tranquility; bustle; noisy; booming; sound; song; music; musical instrument; melody; tone; hear	ear, listen; voice; yell; sing; quiet; silence; tranquility; bustle; noisy; booming; sound; song; music; musical instrument; melody; tone; hear
	Olfaction	incense; smell; sniff; flavor; odor	incense; smell; sniff; flavor; odor
	Gustations	apor; appetizing; taste; sapor; sour; sweet; bitter; spicity; meal; eat; drink; edible; tea; liquor; coffee; dine	apor; appetizing; taste; sapor; sour; sweet; bitter; spicity; meal; eat; drink; edible; tea; liquor; coffee; dine
	Somatosensory	hot; heat; warm; temperature; cold; frosty; cool; wet; weather; air; touch	hot; heat; warm; temperature; cold; frosty; cool; wet; weather; air; touch

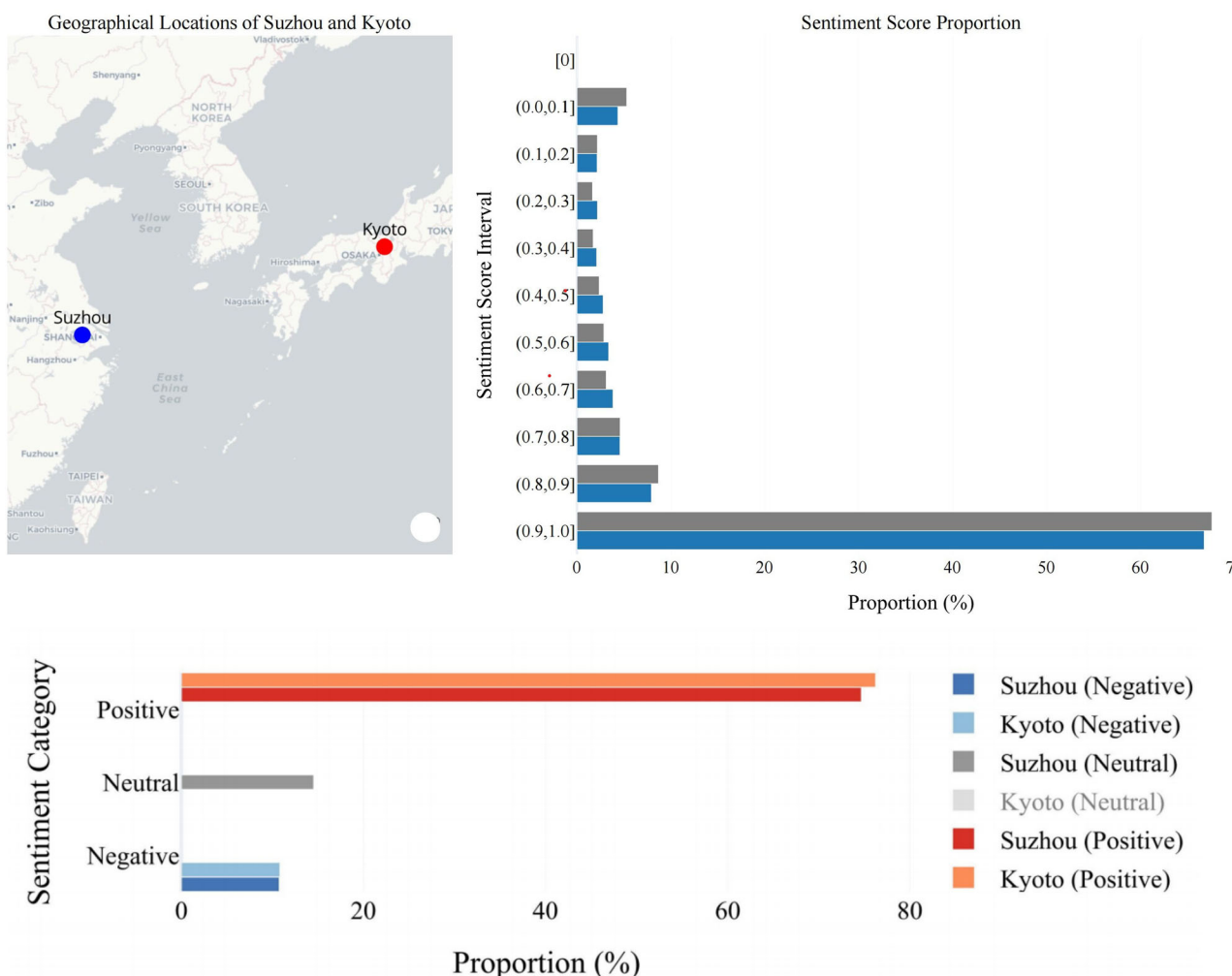


Fig. 6 | Overall Sentiment Score Distribution. Sentiment score distribution of online reviews for Suzhou ($N = 37,649$) and Kyoto ($N = 17,485$).

Table 6 | Distribution of Different Sentiment Scores

Classification	Interval	Suzhou		Kyoto	
		Number Of Reviews	Proportion	Number Of Reviews	Proportion
Positive	(0.9, 1.0]	25443	66.76%	11673	67.58%
Neutral	(0.8, 0.9]	3266	7.92%	1384	8.67%
	(0.7, 0.8]	1728	4.57%	799	4.59%
	(0.6, 0.7]	1171	3.83%	669	3.11%
	(0.5, 0.6]	1075	3.37%	590	2.86%
	(0.4, 0.5]	887	2.78%	486	2.36%
	(0.3, 0.4]	647	2.09%	365	1.72%
Negative	(0.2, 0.3]	620	2.18%	382	1.65%
	(0.1, 0.2]	816	2.14%	374	2.17%
	(0.0, 0.1]	1991	4.35%	761	5.29%
	[0]	5	0.01%	2	0.01%
Total		37649		17485	
Average score		0.8340		0.8293	

representations from unlabeled text by capturing contextual information⁷. In our approach, a pre-trained BERT model was fine-tuned on a large-scale corpus to generate word vectors that capture rich semantic information (Fig. 4). However, because the resulting word vectors include a

vast array of terms, we subsequently removed those deemed irrelevant to our study.

The text data from online reviews of Suzhou and Kyoto underwent an extensive preprocessing procedure. First, we cleaned the data by removing noise, such as spelling errors and non-target language content. We then used Jieba for tokenization, word stemming, and part-of-speech tagging, followed by stop word removal based on a standard stop word library. The resulting tokens were used to construct vocabulary lists and perform word frequency statistics. Given that the generated word vectors typically span several hundred dimensions, we applied Principal Component Analysis (PCA) to reduce these vectors to two dimensions, facilitating more intuitive clustering (Fig. 5).

Next, tokens were screened to form an environmental feature lexicon that aligns with the study's objectives, guided by K-means clustering outcomes. We created a list of terms to extract multi-sensory and landscape elements (Table 5). Recognizing that CH encompasses many artificial components, we explicitly separated natural elements from artificial ones. For the sensory descriptions—including aspects related to hearing, smell, and taste⁹⁴—we enriched our vocabulary using resources such as WordNet^{6,95,96}. This class includes nouns that make sounds or smells and adjectives that describe feelings. This sensory category includes nouns that denote sounds or smells as well as adjectives describing sensory experiences. Moreover, due to the indistinct clustering of terms associated with taste-smell and mountain-water-stone, we merged these into a single group. We also observed clusters that conveyed overall affective impressions; although these clusters encompass multiple sensory inputs and do not specify

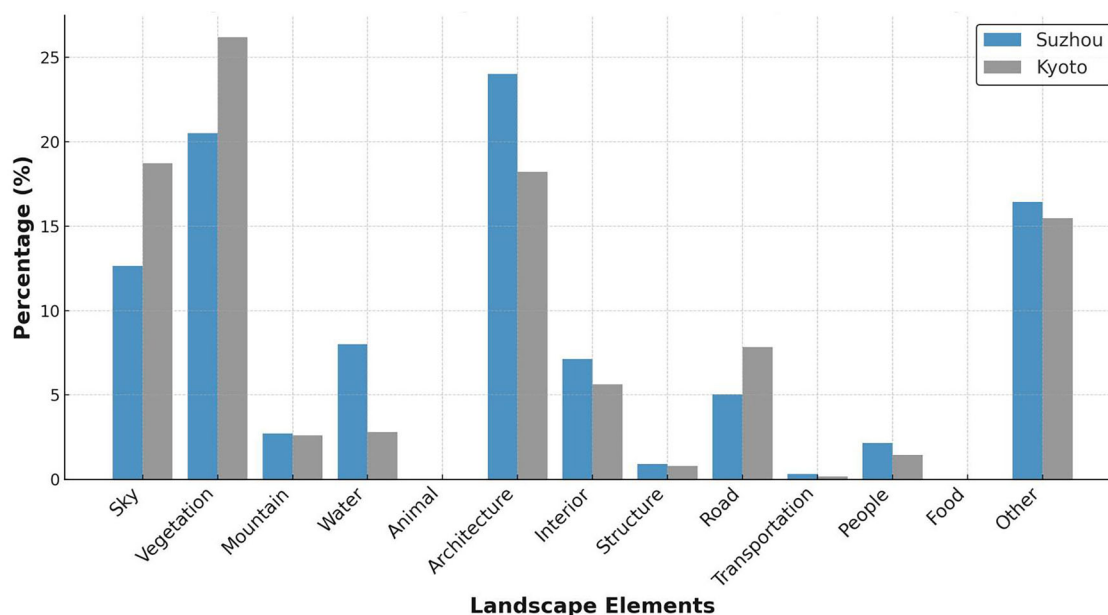


Fig. 7 | Proportion of landscape elements in photos.

individual senses, clear positive and negative groupings emerged, which were subsequently categorized as positive and negative feelings.

Statistical analysis

After the DL processing of photos and text, the resulting data were compiled and organized using Microsoft Excel. Subsequent statistical analyses were performed in Python using libraries such as Statsmodels and SciPy for quantitative assessments, while Plotly was employed for data visualization.

For significant difference analysis, independent sample t-tests were conducted to compare tourists' sensory experiences and the prevalence of landscape elements between the two cities, with a significance threshold set at $p < 0.05$.

Furthermore, multiple linear regression analyses were performed using various landscape elements and sensory factors (extracted from both text and photos) as predictor variables. This approach aimed to determine the extent to which these factors influence overall visitor perceptions.

To investigate the relationships between landscape elements and sensory experiences, Pearson correlation analysis was used. This method allowed us to quantify the degree of association between variables, again with statistical significance defined at $p < 0.05$.

Results

Public overall perception

In our study, we employed the Google Cloud Natural Language platform for text sentiment analysis⁹⁷. This platform processes input text through contextual and syntactic analysis to compute a sentiment score—which typically ranges from -1 (indicating strongly negative sentiment) to 1 (indicating strongly positive sentiment), with 0 representing neutrality—and then normalizes these scores to a 0 to 1 scale. Higher values thus correspond to more positive sentiment. In addition, the platform calculates a sentiment magnitude, reflecting the overall strength or intensity of the emotional content irrespective of its polarity.

Emotion recognition is a fundamental capability in modern natural language processing and human-computer interaction. General-purpose large models, such as ChatGPT, rely on understanding and processing emotional cues as part of their core functionality. These models are trained on vast amounts of data that encompass not only information but also the subtleties of human emotion, enabling them to interpret sentiment, tone, and context effectively. This capability underpins the generation of

contextually appropriate and empathetic responses, thereby enhancing human-AI interactions. In essence, emotion recognition is not only critical for specialized tasks like sentiment analysis but also forms a vital foundation for the broader understanding mechanisms in state-of-the-art large language models.

By integrating these advanced emotion recognition capabilities, the Google Cloud Natural Language platform provides a robust means to convert qualitative language descriptions from online reviews into quantifiable sentiment metrics. These metrics serve as reliable proxies for the public's overall perception of CH sites in Suzhou and Kyoto, thereby offering valuable insights for our study.

As illustrated in Fig. 6 and Table 6, the public's overall perception status of the CH sites in the two cities is as follows: both received a relatively positive overall perception level. Suzhou received a 66.90% positive score, 19.23% neutral score, and 13.19% negative score, with an average score of 0.829. Kyoto received a 66.76% positive score, 13.54% neutral score, and 11.45% negative score, with an average score of 0.834.

Photo data analysis

Using the semantic segmentation elements provided by the ADE_20K dataset, the photo data from online reviews was processed through training. Following FCN processing, as depicted in Fig. 3, the landscape element factor—defined as the proportion of a specific landscape element within an entire photo—was extracted for both Suzhou and Kyoto, with the results illustrated in Fig. 7.

Analysis of the photos uploaded by visitors reveals that the most frequently occurring landscape elements in Suzhou include architecture, vegetation, sky, water, and interior elements. In contrast, Kyoto's dominant landscape elements consist of vegetation, sky, buildings, roads, interior elements, and rock formations. This suggests that architecture plays a more prominent role in Suzhou's visual perception, whereas vegetation is more significant in Kyoto. Additionally, the proportion of the sky element in Kyoto is greater than that in Suzhou, indicating a higher degree of visual openness in Kyoto's landscape.

Furthermore, the proportion of water elements in Suzhou's images surpasses that in Kyoto's, highlighting the integral role of water in Suzhou's CH. This finding underscores the distinct characteristics of each city and the different ways in which visitors perceive and capture their surroundings.

Figure 8 presents the radar map depicting the landscape element factors of various CH in Suzhou and Kyoto. The visualization reveals that even

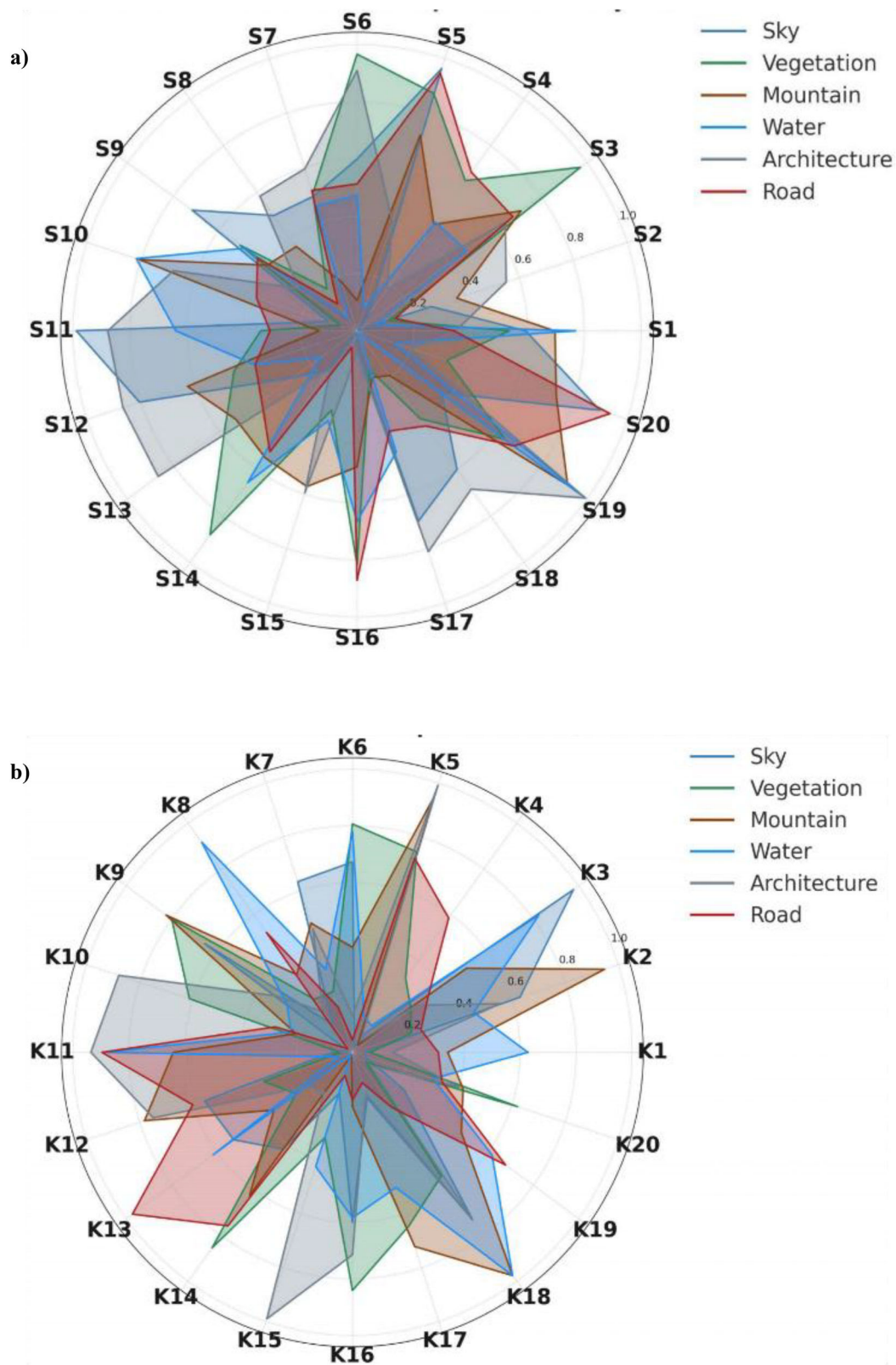


Fig. 8 | Radar chart of selected landscape elements. **a** Suzhou. **b** Kyoto.

within the same city, different CH sites exhibit distinct landscape characteristics. For instance, in Suzhou, S17 features the highest proportion of architectural elements, S8 is dominated by vegetation, S10 stands out for its water and sky elements, and S3 has the most significant interior

representation. Meanwhile, in Kyoto, K5 is characterized by abundant vegetation, K4 by a prominent sky element, K11, K13, and K16 by striking architectural features, and K8 by its well-defined road network, all of which are particularly appealing to tourists.

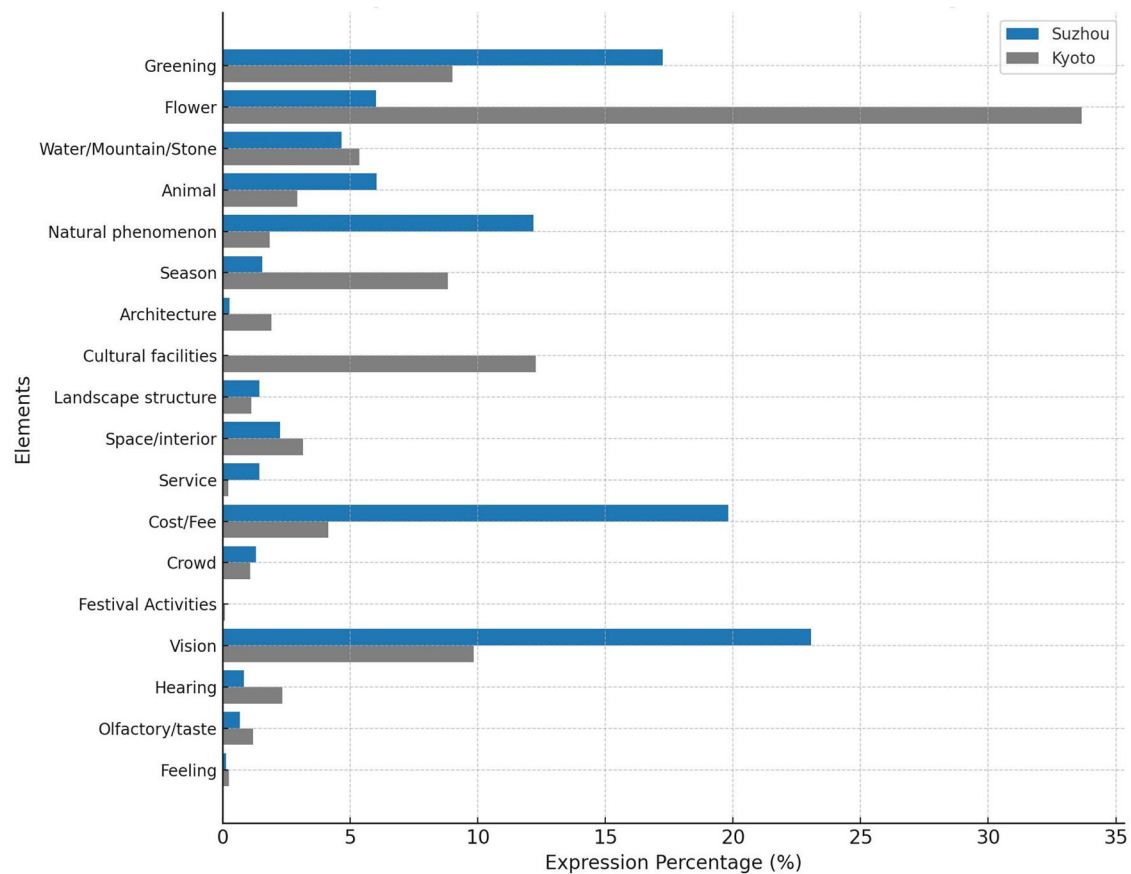


Fig. 9 | Expression of different elements in Suzhou and Kyoto.

Table 7 | Analysis of Significant Differences in Subjective Perception Between Suzhou and Kyoto

Photo data			Text data		
Elements	t	Sig.	Elements	t	Sig.
Sky	64.710	0.000	Greening	-2.250	0.024
Vegetation	45.879	0.000	Flower	43.838	0.000
Mountain	-.083	0.002	Water/ Mountain/Stone	-14.497	0.000
Water	-83.949	0.000	Animal	4.094	0.000
Animal	-3.422	0.001	Natural phenomenon	-11.988	0.000
Architecture	-52.834	0.000	Season	34.654	0.000
Interior	-23.026	0.000	Architecture	8.114	0.000
Structure	-9.067	0.000	Cultural facilities	3.809	0.000
Road	53.331	0.000	Landscape structure	-16.271	0.000
Transportation	-19.885	0.000	Space/interior	21.267	0.000
People	-23.527	0.000	Service	19.548	0.000
Food	-5.326	0.000	Cost/Fee	-47.963	0.000
Other	-11.572	0.000	Crowd	7.310	0.000
			Festival Activities	33.897	0.000
			Vision	23.514	0.000
			Hearing	-10.311	0.000
			Smell/taste	-10.397	0.000
			Feeling	-9.049	0.000
			Positive mood	-8.306	0.000
			Negative mood	18.192	0.000

Text data analysis

We extracted high-frequency words, including nouns, adjectives, and verbs, from the text data of online reviews. After processing these words using the NLP workflow illustrated in Fig. 5 and eliminating irrelevant clusters, we identified key feature word clusters for each city. These clusters represent potential characteristic factors influencing overall perceptions and encompass sensory elements, landscape features, and descriptions of experiential mood in Fig. 9.

To systematically analyze these perceptual clusters, we categorized them into three groups: natural elements, artificial elements, and sensory descriptions. As an intermediate step, we visualized these clusters within a two-dimensional coordinate system to refine the feature words and construct a structured word list. The feature list was developed using cosine similarity calculations, incorporating words with similarity values greater than 0.5 into the lexicon. A second layer of vocabulary classification was then conducted based on the initial classification framework, as illustrated in Table 3 and Fig. 5. Using this refined high-frequency vocabulary, we generated a chart map Fig. 8 to further illustrate the relationships among key terms.

Based on these results, we observed that in the sensory descriptions of Suzhou and Kyoto, landscape elements that evoke sensory responses are predominantly represented by nouns, whereas sensory perceptions and actions are mainly expressed through adjectives and verbs. While these terms are distributed across multiple clusters, their patterns remain discernible. Among the senses, vision is the most frequently described, particularly in relation to color. Consistent with prior photo analysis, both artificial and natural elements—such as buildings and vegetation—exhibit distinct clustering in terms of their visual impact. The second most frequently mentioned sense is auditory perception, which is often associated with sounds from birds, music, human voices, or an overall sense of tranquility.

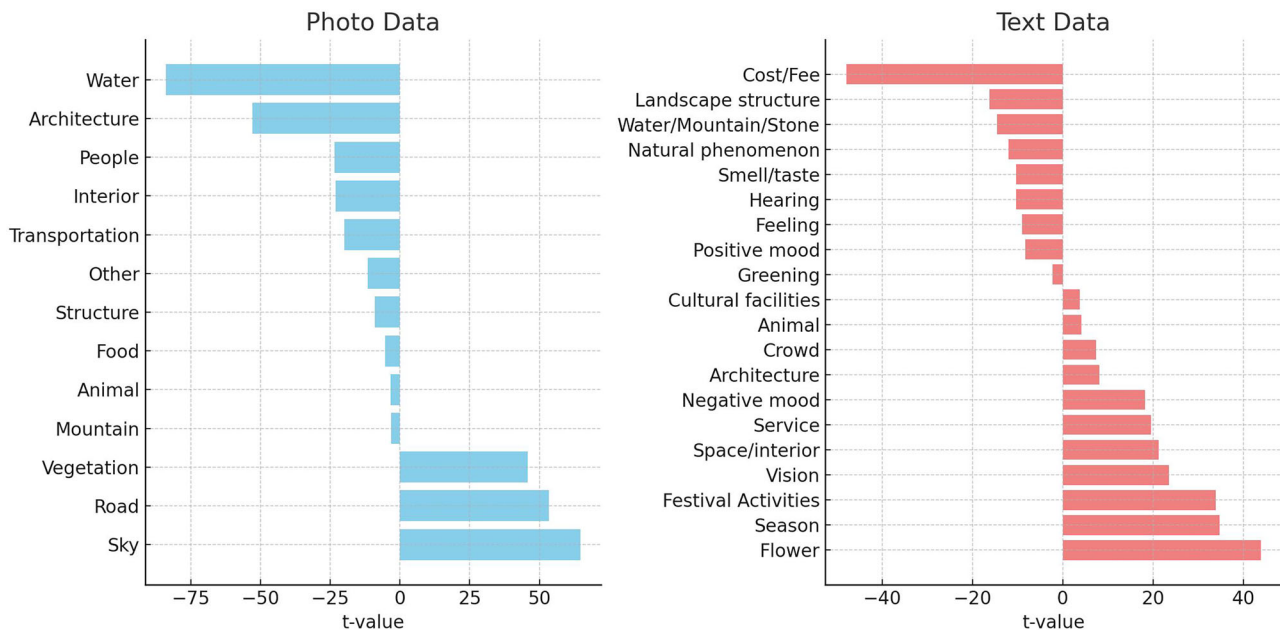


Fig. 10 | Differences in elements and perceptions in online reviews of the two cities.

In contrast, olfactory and gustatory descriptions frequently co-occur, leading us to analyze these two senses together. This clustering reveals strong associations with flowers and food, which may correspond to smell and taste, respectively. Descriptions of touch primarily relate to the perception of natural conditions, such as temperature. Additionally, the expressions of positive and negative experiential moods show distinct clustering patterns. Given that sensory experiences are inherently multisensory, we classified these mood-related descriptions within the broader framework of sensory perception.

Statistical analysis

Although Suzhou and Kyoto share an Asian cultural background, the architecture and landscape styles of their CH sites exhibit strong similarities. However, visitors' descriptions of landscape elements and sensory experiences reveal significant differences (Table 7 and Fig. 10), as reflected in both photographic and textual data.

In the photo data, the most pronounced differences were observed in water ($t = -83.94$, $p < 0.01$) and architecture ($t = -52.83$, $p < 0.01$), both of which were significantly more prominent in Suzhou. This result aligns with Suzhou's historical identity as a water town, where traditional gardens integrate pavilions, bridges, and canals as core elements of the landscape. In contrast, sky ($t = 64.70$, $p < 0.01$) and roads ($t = 53.33$, $p < 0.01$) were significantly more emphasized in Kyoto, likely reflecting the prevalence of open landscapes, Karesansui (dry rock gardens), and structured pathways in Japanese cultural sites. These findings indicate that Kyoto's landscapes emphasize expansive, open spaces, whereas Suzhou's CH landscape is more architecturally enclosed, and water integrated.

In the text data, significant differences were also observed in the frequency of descriptions of landscape elements and sensory experiences ($p < 0.01$), demonstrating distinct visitor perceptions of tangible landscape features and the sensory experiences they evoke. Among these, the most pronounced difference was found in cost perception ($t = -47.963$, $p < 0.01$), suggesting that discussions related to pricing and expenses were notably more frequent in Suzhou. Kyoto visitors, on the other hand, more frequently described flower-related elements ($t = 43.838$, $p < 0.01$) and seasonal aspects ($t = 34.654$, $p < 0.01$), likely due to the city's well-known seasonal aesthetics, such as cherry blossoms and autumn foliage. Additionally, Kyoto's CH sites descriptions emphasized festival activities ($t = 33.897$, $p < 0.01$) and vision-related experiences ($t = 23.514$, $p < 0.01$), whereas Suzhou's text data exhibited a stronger focus on landscape structure ($t = -16.271$, $p < 0.01$) and

natural elements such as water, mountains, and stones ($t = -14.497$, $p < 0.01$).

The most significant differences in textual descriptions relate to the cost and fees. While Suzhou exhibits a greater volume of sensory descriptions, Kyoto's depictions of seasonal flowers make it particularly appealing.

To understand the factors influencing overall perception in Suzhou and Kyoto, we conducted multiple linear regression analyses using landscape elements, artificial features, multi-sensory experiences, and photographic data as predictors (Table 8 and Fig. 11). To mitigate multicollinearity, we constructed separate regression models for each predictor category. The models demonstrated varying levels of explanatory power, with traditional R^2 values of 64.1% for Suzhou and 63.9% for Kyoto, while Pseudo R^2 values ranged from 0.22 to 0.36 in Suzhou and from 0.27 to 0.36 in Kyoto, depending on the predictor category. Notably, photo-based predictors explained 21.4% of the variance in Suzhou and 25.3% in Kyoto, indicating that photographic representations contribute moderately to visitor perceptions of CH sites.

To ensure the robustness of the regression models, a Kolmogorov-Smirnov (K-S) test was conducted to examine the normality of the residuals. The results confirmed that the residuals followed a normal distribution ($p > 0.05$), satisfying the assumption of normality. Additionally, a variance inflation factor (VIF) test was performed to assess multicollinearity, with all VIF values below 5, indicating the absence of severe collinearity among predictors.

Among text-based natural elements, greening had the strongest positive effect in Suzhou ($\beta = 0.3466$) and a smaller effect in Kyoto ($\beta = 0.1062$), suggesting that vegetation contributes more prominently to visitor perceptions in Suzhou. Seasonal descriptions positively influenced perception in both cities (Suzhou: $\beta = 0.0755$, Kyoto: $\beta = 0.0942$), reinforcing the importance of seasonal aesthetics in both cultural landscapes. However, descriptions of water, mountains, and stones negatively influenced perception in Suzhou ($\beta = -0.032$) but positively in Kyoto ($\beta = 0.0779$), indicating that while Kyoto's visitors associate these elements with a more favorable experience, Suzhou's visitors may perceive them as less influential or even unfavorable. The overall explanatory power of these natural elements was higher in Kyoto (Pseudo $R^2 = 0.36$) than in Suzhou (Pseudo $R^2 = 0.22$), suggesting that text-based natural features play a more central role in shaping Kyoto's visitor perceptions.

The influence of artificial elements varied across the two cities. Architecture had a positive effect in both Suzhou ($\beta = 0.1366$) and Kyoto

Table 8 | Summary of multiple linear regression models predicting overall perception in Suzhou ($N = 20$) and Kyoto ($N = 20$)

		Suzhou Estimate	Kyoto Estimate
Text-Natural element	Greening	0.3466	0.1062
	Flower	0.0673	0.1043
	Water/Mountain/Stone	-0.032	0.0779
	Animal	-0.0267	0.1388
	Natural phenomenon	-0.0415	-0.0184
	Season	0.0755	0.0942
		Pseudo R^2: 0.22	Pseudo R^2: 0.36
Artificial element	Architecture	0.1366	0.1162
	Cultural facilities	-0.2453	-0.0115
	Landscape structure	0.0104	0.1288
	Space/interior	-0.055	0.0271
	Service	-0.2424	-0.0886
	Cost/Fee	-0.0898	-0.0812
	Festival Activities	0.1221	0.0622
	Crowd	-0.0298	0.1027
		Pseudo R^2: 0.3	Pseudo R^2: 0.27
Multi-sensory	Vision	0.4069	0.107
	Hearing	-0.0635	0.1571
	Olfactory/taste	-0.1286	0.0448
	Feeling	-0.0875	0.0651
		Pseudo R^2: 0.29	Pseudo R^2: 0.253
Photos	Sky	0.0053	0.1123
	Vegetation	-0.02	0.1031
	Mountain	0.0278	-0.1747
	Water	0.0583	-0.2333
	Architecture	-0.0211	-0.0993
	Interior	-0.0292	-0.0939
	Structure	0.1164	-0.3388
	Road	-0.0785	-0.173
	Transportation	0.4476	-1.1192
	People	-0.0516	-0.2333
		Pseudo R^2: 0.214	Pseudo R^2: 0.253

Note. Traditional R^2 reflects the proportion of variance explained in linear models, while Pseudo R^2 is used in generalized linear models to indicate relative model fit. Pseudo R^2 values above 0.2 suggest acceptable explanatory power.

($\beta = 0.1162$), indicating its consistent role in shaping visitor perceptions. However, cultural facilities had a strong negative effect in Suzhou ($\beta = -0.2453$) but little influence in Kyoto ($\beta = -0.0115$), suggesting that built heritage is perceived differently between the two cities. Service and cost-related factors negatively influenced perception in both cities, with stronger effects in Suzhou (Service: $\beta = -0.2424$, Cost: $\beta = -0.0898$) than in Kyoto (Service: $\beta = -0.0886$, Cost: $\beta = -0.0812$). This result suggests that service quality and expenses play a more critical role in Suzhou's visitor satisfaction than in Kyoto. The model for artificial elements explained 30% of the variance in Suzhou (Pseudo $R^2 = 0.30$) and 27% in Kyoto (Pseudo $R^2 = 0.27$), indicating similar explanatory power across both cities.

Among sensory modalities, visual perception was the strongest predictor of overall perception in Suzhou ($\beta = 0.4069$), whereas its effect in Kyoto was much weaker ($\beta = 0.1070$). This suggests that Suzhou's visitors rely more on visual aesthetics in evaluating their experience. Auditory perception had a positive effect in Kyoto ($\beta = 0.1571$) but a negative effect in Suzhou ($\beta = -0.0635$), indicating that Kyoto's soundscape, potentially influenced by cultural elements like temple bells and ambient music, enhances visitor perception. In contrast, olfactory and taste-related experiences had a negative effect in Suzhou ($\beta = -0.1286$) but a minor positive effect in Kyoto ($\beta = 0.0448$), implying that Kyoto's food and scent-related experiences contribute more positively to overall perception. The explanatory power of the multi-sensory model was slightly higher in Suzhou (Pseudo $R^2 = 0.29$) than in Kyoto (Pseudo $R^2 = 0.253$), reinforcing the stronger role of visual factors in Suzhou's visitor experience.

In the photo-based regression model, transportation infrastructure had the most substantial effect in both cities, positively influencing perception in Suzhou ($\beta = 0.4476$) but negatively in Kyoto ($\beta = -1.1192$). This stark contrast suggests that transportation access enhances visitor perception in Suzhou but may be seen as disruptive or overwhelming in Kyoto. The presence of water in photos had a positive effect in Suzhou ($\beta = 0.0583$) but a significant negative effect in Kyoto ($\beta = -0.2333$), further supporting the idea that water is a defining and favorable feature in Suzhou's landscape but less valued in Kyoto's cultural setting. Similarly, structural elements positively influenced perception in Suzhou ($\beta = 0.1164$) but negatively in Kyoto ($\beta = -0.3388$), suggesting different visitor responses to the built environments in the two cities. The explanatory power of the photo-based model was similar in both cities (Suzhou: Pseudo $R^2 = 0.214$, Kyoto: Pseudo $R^2 = 0.253$), indicating that photographic elements play a comparable role in shaping visitor perceptions.

The analysis of landscape elements in photos not only provides the subjective preferences of visitors but also helps to objectively understand the situation of each cultural landscape entity. From the correlation heat map between landscape elements (Fig. 12), both the architectural elements of Suzhou and Kyoto have a negative correlation with vegetation elements, followed by the sky. The sky elements in Suzhou have a positive correlation with water, while Kyoto's water environment is less related to it.

The correlation analysis of the description of landscape elements and senses in the text demonstrates that the CH's visual perception in the two cities has a positive correlation with most of the natural and artificial landscape elements. The positive correlation between visual perception and architecture is stronger in Kyoto. Overall, the artificial (people, space) and natural (seasons, mountains and waters, greenery) elements are more relevant to visual perception in Kyoto, while in Suzhou, non-visual senses are more relevant to other landscape elements.

Regarding other sensory descriptions, different sensory experiences in Suzhou are positively correlated with various landscape elements. The description of hearing is positively correlated with landscape stone, natural phenomena, people, cultural facilities, landscape structures, space, and activity festivals in Kyoto. In Suzhou, artificial elements (structures, events-festivals, space) and natural elements (landscape stone, natural phenomena, flowers) demonstrate a positive correlation with smell-taste. Artificial (cultural facilities, structures) and natural (landscape stone, natural phenomena) elements correlate strongly with somatosensory descriptions.

However, in general, although the natural and artificial elements in the CH sites of Suzhou can cause more multi-sensory descriptions, the description of the five senses with positive experiences and the mood in Kyoto is more relevant.

Discussion

This study explores how multi-sensory perceptions influence cultural landscape experiences by analyzing large-scale online reviews of CH sites in Suzhou and Kyoto. To systematically address the key research questions, we interpret the findings from CH perception analysis, correlation analysis, and regression models in relation to multi-sensory experiences, visitor perceptions, and sustainable conservation strategies.

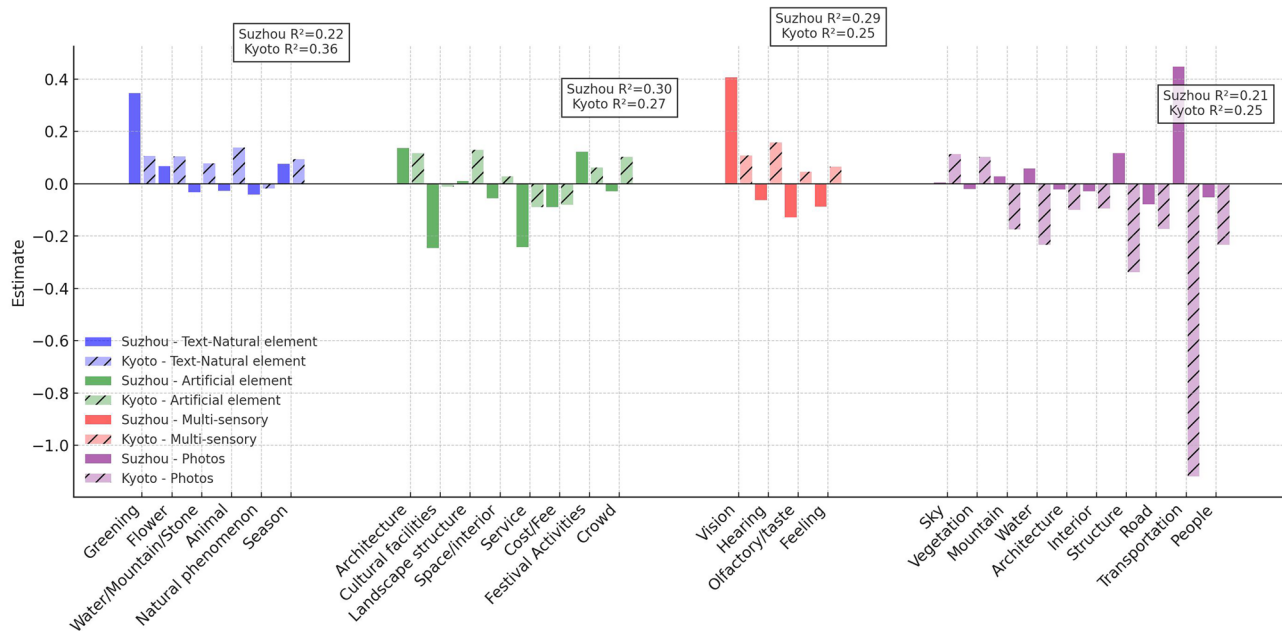


Fig. 11 | Influence of landscape elements and senses on overall perception.

This study highlights both the specificity and universality of CH sites in sustainable urban development research^{98,99}. Visitors' subjective perceptions are valuable for CH conservation policies, and with the rise of online reviews, local heritage values and landscape characteristics are increasingly reflected in digital discourse. These reviews contain implicit evaluations that are widely recognized for their policymaking significance in CH protection and sustainable development¹⁰⁰.

To analyze public perceptions, FCN and NLP offer reliable support¹⁰¹ and are widely applied in commercial strategy and data analysis¹⁰². However, few studies have integrated them into CH research, and sensory information in text analysis is often overlooked. This study addresses these gaps by proposing a new DL-based method for processing online reviews, demonstrating its potential for multi-sensory CH analysis and highlighting both its advantages and limitations.

To begin with, one particular aspect of cultural heritage (CH) protection lies in its unique contextual and historical characteristics. CH serves as a carrier of local culture, shaped by aesthetic, cultural, and environmental interactions⁷. While CH sites in Eastern cultural cities often share similar architectural and landscape styles, this study reveals significant differences in visitors' sensory experiences and perceptions of landscape elements, challenging the assumption of uniform subjective impressions.

Photo analysis reveals that while architecture, vegetation, and sky are the most perceived elements in both cities, their distribution differs. Kyoto's open landscapes emphasize visual spaciousness, whereas Suzhou's compact design follows the "heaven and earth in a pot" concept, creating diverse spatial changes in a confined area, leading to differences in landscape openness.

Suzhou's CH, originating from the Ming and Qing dynasties¹⁰³, developed under abundant rainfall and an intricate water system³, making it ideal for garden construction. Many banished officials built private gardens as a retreat, incorporating rockeries and vegetation to create dynamic spatial transformations that simulated nature in confined spaces^{104,105}. While symbolizing withdrawal from secular life, the architectural elements, plaques, and couplets still reflected their political aspirations. In contrast, public gardens—including temples and ancient streets along river systems—featured buildings, sky, and water as dominant landscape elements, offering a more open and accessible environment.

Kyoto's CH, though influenced by China's Sui and Tang Dynasties, evolved under Japanese geographical and cultural conditions, resulting in a distinct philosophy from Suzhou¹⁰⁶. It preserves buildings and gardens from

multiple historical periods, including temples and shogunate-era private gardens. Garden designs often simulate the sea using water or white sand and represent sacred mountains with stones, reflecting Zen and tea ceremony traditions. Unlike Suzhou, water is not an essential element in Kyoto's courtyards; instead, vegetation, sky, sand, and gravel paths dominate, creating a minimalist, open landscape that emulates nature.

The significance analysis of textual differences indicates that visitor descriptions of cultural heritage in Suzhou are more likely to focus on cost and expenses, which may reflect a sensitivity to economic considerations. Additionally, Suzhou exhibits a higher frequency of sensory descriptions. In contrast, Kyoto features significantly more references to seasonal events and floral activities, emphasizing the dynamic and aesthetically rich experiences shaped by seasonal changes.

Subjective perceptions of CH vary due to differences in landscape elements, management practices, and cultural experience activities. Therefore, CH conservation should adopt customized strategies that consider local history, natural environment, and existing built heritage, rather than applying uniform preservation methods. This approach aligns with findings from previous research^{13,107}.

In terms of overall perception, the multiple linear regression analysis highlights the significant role of multi-sensory experiences in shaping visitor satisfaction. Both multi-sensory engagement and visual aesthetics contribute positively to the overall perception of CH sites in Suzhou and Kyoto, reinforcing the importance of preserving architectural and natural elements.

The historical and cultural depth of CH in Suzhou and Kyoto has long been associated with multi-sensory engagement, as evidenced by ancient poems and artworks describing experiences such as listening to the rain under plantain trees, hearing temple bells at dusk, meditating in Karesansui landscapes, inhaling the fragrance of flowers, or touching natural elements like water and tree bark for restoration¹⁰⁸. However, in modern tourism-driven experiences, rapid visits may have weakened non-visual sensory engagement, leading to a diminished role of hearing, smell, and touch in visitor perception.

The role of visual elements is particularly significant, with vegetation playing a key role in shaping overall evaluations. This aligns with research on urban green spaces, which highlights the positive psychological effects of natural scenery^{47,109}. Some studies suggest that virtual reality (VR) or photographic representations of landscapes can replicate the relaxing effects of nature, but whether such technologies can fully substitute real-world CH landscapes remains an open question¹¹⁰. Beyond vision, non-visual sensory

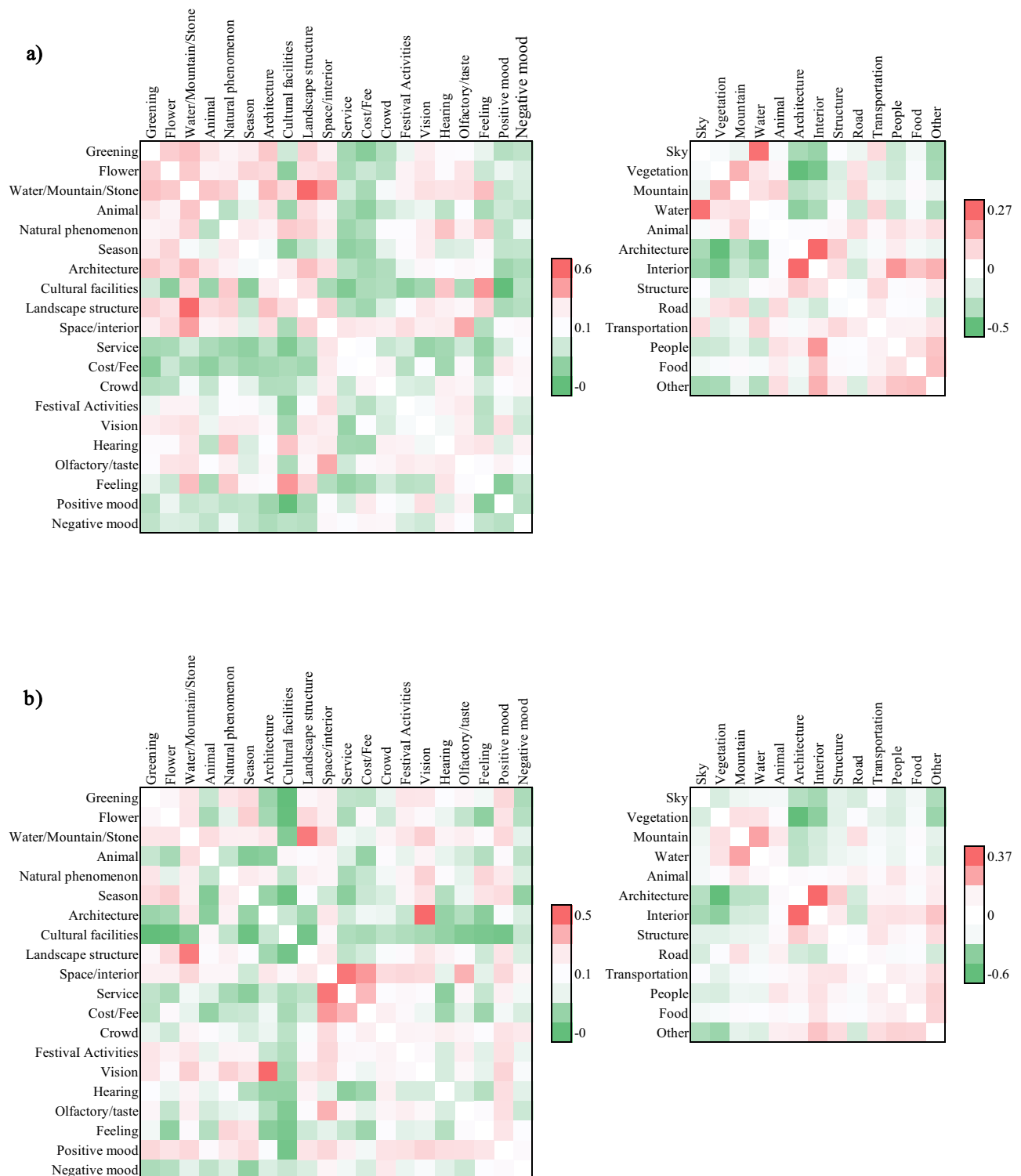


Fig. 12 | Correlation heat map. a Suzhou. b Kyoto.

factors also play a role in perception, albeit with mixed effects. While they cannot be captured in images, they are evident in visitor descriptions¹¹¹.

In Suzhou, hearing, smell-taste, and somatosensory factors negatively influence visitor perception, suggesting that issues such as noise, air quality, and thermal comfort require further improvement to enhance satisfaction. In Kyoto, the higher frequency of floral and seasonal terms suggests that flower-related experiences and seasonal aesthetics play a stronger role in shaping perceptions. The significance analysis of textual differences

indicates that visitor descriptions of cultural heritage in Suzhou are more likely to focus on cost and expenses, which may reflect a sensitivity to economic considerations. Additionally, Suzhou exhibits a higher frequency of sensory descriptions. In contrast, Kyoto features significantly more references to seasonal events and floral activities, emphasizing the dynamic and aesthetically rich experiences shaped by seasonal changes. Although Kyoto has fewer sensory descriptions than Suzhou, non-visual sensory elements contribute more positively to the overall perception. At the same

time, issues related to accessibility contribute the most to negative perceptions, it is essential to optimize the accessibility of various sites, reduce travel costs, and enhance accessibility equity.

In both cities, service quality and cost-related concerns negatively predict overall visitor perception, highlighting the importance of balancing accessibility, pricing, and visitor expectations. As CH sites function as public resources, facility management—including visitor capacity control, parking infrastructure, public rest areas, and maintenance—plays a crucial role in ensuring visitor satisfaction. High-quality multisensory experiences contribute to enhancing overall perception. Suzhou should focus on optimizing non-visual aspects to enrich sensory diversity and improve the visitor experience. In contrast, Kyoto has already gained stronger overall perception support through intangible sensory experiences. Future improvements could emphasize enhancing the richness of visual elements.

Based on the correlation analysis, different sensory experiences are closely linked to specific landscape elements. To preserve tangible CH, maintaining visual experiences is crucial, as it guides cultural perception and visitor engagement. Architectural landmarks and culturally significant vegetation enhance visitor interest, while the integration of natural landscapes and artificial structures strengthens visual sensory engagement.

Beyond vision, other senses also shape cultural perception, often influenced by landscape features or cultural activities. Regarding hearing, natural sounds such as birdsong and water flow are essential in shaping visitor experiences, reinforcing the perception of tranquility. However, urban noise pollution negatively impacts visitor evaluations, as confirmed by previous studies^{112,113}.

To counter this, it is important to curate soundscapes by incorporating traditional auditory elements such as temple bells or cultural music, enhancing the cultural experience and auditory perception.

Similarly, olfactory and taste experiences are strongly tied to flowers and traditional cuisine. Seasonal flower-viewing festivals, historically significant in both Suzhou and Kyoto, remain influential in enhancing sensory engagement. These experiences stimulate multiple senses simultaneously, particularly through floral scents and birdsong. The sense of touch, primarily related to temperature and material interaction, is also a frequently mentioned factor in visitor reviews. Intangible cultural activities, such as opera appreciation, offer a comprehensive sensory experience, combining visual, auditory, and tactile engagement with CH landscapes.

The excavation and preservation of sensory experiences play a key role in strengthening subjective conservation awareness among residents and visitors. Ultimately, both tangible landscape elements and intangible cultural activities contribute to multi-sensory engagement, positively impacting visitor perception and heritage conservation¹¹⁴.

This study highlights the importance of integrating multisensory elements into heritage preservation, with non-visual experiences potentially exerting an even greater impact on overall perception. Mapping tourist perceptions to specific locations enables targeted optimization, and findings confirm that high-quality sensory experiences consistently enhance overall satisfaction.

However, limitations remain. Current multimodal methods still struggle to effectively align emotional information between images and text in online reviews, due to variability in visual content and the complexity of cross-modal semantics. Additionally, image segmentation lacks consistency compared to structured data sources, and the incomplete sensory lexicon may lead to omissions in textual analysis.

With rapid advancements in technology, especially in multimodal deep learning and large-scale language-vision models, future research is expected to better address the alignment between images and text. These cutting-edge technologies can more accurately understand and integrate complex semantic relationships across visual and textual data, enhancing sensory simulation and emotion-aware analysis. For example, when analyzing online reviews of cultural heritage sites, it will be possible not only to detect emotions in the text but also to combine photos, videos, and other media uploaded by visitors to deeply capture their emotional responses and experiences related to intangible cultural elements. This will offer a more

holistic perspective for cultural heritage protection, fostering both sustainability and differentiated conservation strategies.

Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The code used during the current study is available from the corresponding author on reasonable request.

Received: 21 July 2024; Accepted: 4 June 2025;

Published online: 20 June 2025

References

1. UNESCO. Historic Urban Landscape Recommendation. Available online: <http://whc.unesco.org/en/activities/448> (accessed on Nations, U. Transforming our world: the 2030 Agenda for Sustainable Development. Available online: <https://sdgs.un.org/2030agenda>.
2. Battis-Schinker, E. et al. Towards quality of life indicators for historic urban landscapes – Insight into a German-Polish research project. *Environ. Sustain. Indic.* **10**, 100094 (2021).
3. Skrede, J. & Berg, S. K. Cultural heritage and sustainable development: the case of urban densification. *Historic Environ. Policy Pract.* **10**, 83–102 (2018).
4. Cucco, P., Maselli, G., Nesticò, A. & Ribera, F. An evaluation model for adaptive reuse of cultural heritage in accordance with 2030 SDGs and European Quality Principles. *J. Cult. Herit.* **59**, 202–216 (2023).
5. Koblet, O. & Purves, R. S. From online texts to landscape character assessment: collecting and analysing first-person landscape perception computationally. *Landsc. Urban Plan.* **197**, 103757 (2020).
6. Jenkins, V. Protecting the natural and cultural heritage of local landscapes: finding substance in law and legal decision making. *Land Use Policy* **73**, 73–83 (2018).
7. Marwick, B., Smith, P. World Heritage sites on Wikipedia: cultural heritage activism in a context of constrained agency. *Big Data Soc.* **8**, <https://doi.org/10.1177/20539517211017304>. 2021
8. Zhou, J. et al. Management effectiveness evaluation of world cultural landscape heritage: a case from China. *Herit. Sci.* **10**, 22 (2022).
9. Foster, G. & Saleh, R. The circular city and adaptive reuse of cultural heritage index: measuring the investment opportunity in Europe. *Resour. Conserv. Recycl.* **175**, 105880 (2021).
10. Medda, F. & Lipparini, F. Impact investment for urban cultural heritage. *City Cult. Soc.* **26**, 100413 (2021).
11. Li, J. et al. Promoting conservation or change? The UNESCO label of world heritage (re)shaping urban morphology in the Old Town of Lijiang, China. *Front. Archit. Res.* **11**, 1121–1133 (2022).
12. Su, W., Zhang, L. & Chang, Q. Nature-based solutions for urban heat mitigation in historical and cultural block: the case of Beijing Old City. *Build. Environ.* **225**, 109600 (2022).
13. Chen, Y. et al. Can smaller parks limit green gentrification? Insights from Hangzhou, China. *Urban For. Urban Green.* **59**, 19 (2021).
14. Feng, K., Song, S. & Zhou, W. The sustainability cycle of historic houses and cultural memory: Controversy between historic preservation and heritage conservation. *Front. Architectural Res.* **11**, 1030–1046 (2022).
15. Zhou, W., Song, S. & Feng, K. The sustainability cycle of historic houses and cultural memory: controversy between historic preservation and heritage conservation. *Front. Archit. Res.* **11**, 1030–1046 (2022).
16. Guzman, P. Assessing the sustainable development of the historic urban landscape through local indicators. Lessons from a Mexican World Heritage City. *J. Cult. Herit.* **46**, 320–327 (2020).

18. Peng, C. Protecting world heritage in China by enacting laws: Sustainable tourism development. *Chin. J. Popul., Resour. Environ.* **19**, 104–109 (2021).
19. Throsby, D. & Petetskaya, K. Heritage-led urban rehabilitation: evaluation methods and an application in Jeddah, Saudi Arabia. *City Cult. Soc.* **26**, 100397 (2021).
20. Wang, S. & Gu, K. Pingyao: the historic urban landscape and planning for heritage-led urban changes. *Cities* **97**, 102489 (2020).
21. Zhang, X., Zhou, L., Wu, Y., Skitmore, M. & Deng, Z. Resolving the conflicts of sustainable world heritage landscapes in cities: fully open or limited access for visitors?. *Habitat Int.* **46**, 91–100 (2015).
22. Macaulay, R., Lee, K., Johnson, K. & Williams, K. Mindful engagement, psychological restoration, and connection with nature in constrained nature experiences. *Landsc. Urban Plan.* **217**, 104263 (2022).
23. Navarrete-Hernandez, P. & Laffan, K. A greener urban environment: Designing green infrastructure interventions to promote citizens' subjective wellbeing. *Landsc. Urban Plan.* **191**, 103618 (2019).
24. Ha, J., Kim, H. J. & With, K. A. Urban green space alone is not enough: a landscape analysis linking the spatial distribution of urban green space to mental health in the city of Chicago. *Landsc. Urban Plan.* **218**, 104309 (2022).
25. Luo, S., Shi, J., Lu, T. & Furuya, K. Sit down and rest: Use of virtual reality to evaluate preferences and mental restoration in urban park pavilions. *Landsc. Urban Plan.* **220**, 104336 (2022).
26. Shanahan, D. F. et al. Health benefits from nature experiences depend on dose. *Sci. Rep.* **6**, 28551 (2016).
27. Yakınlar, N. & Akpınar, A. How perceived sensory dimensions of urban green spaces are associated with adults' perceived restoration, stress, and mental health?. *Urban For. Urban Green.* **72**, 127572 (2022).
28. Mascari, G. F., Mautone, M., Molledo, L. & Salonia, P. Landscapes, heritage and culture. *J. Cult. Herit.* **10**, 22–29 (2009).
29. Sen, S. & Guchhait, S. K. Urban green space in India: perception of cultural ecosystem services and psychology of situatedness and connectedness. *Ecol. Indic.* **123**, 107338 (2021).
30. Jelen, J., Šantrůčková, M. & Komárek, M. Typology of historical cultural landscapes based on their cultural elements. *Geografie* **126**, 243–261 (2021).
31. Porzi, L., Rota Bulò, S., Lepri, B., Ricci, E. Predicting and Understanding Urban Perception with Convolutional Neural Networks. In *Proc. of the 23rd ACM International Conference on Multimedia (MM '15)*. 139–148 (Association for Computing Machinery, 2015).
32. Castillo, A., Domínguez, M. & Yáñez, A. Citizen perception about world heritage and archaeology in three Spanish cities: first methodological case studies. *Complutum* **27**, 295–314 (1970).
33. Gkoltsiou, A. & Paraskevopoulou, A. Landscape character assessment, perception surveys of stakeholders and SWOT analysis: a holistic approach to historical public park management. *J. Outdoor Recreat. Tour.* **35**, 100418 (2021).
34. Özdemir, A. Cultural heritage readability: Children's perception of cultural landscape, Laodikeia Ancient City, Denizli. *Kastamonu Üniversitesi Orman Fakültesi Derg.* **18**, 32–45 (2018).
35. Riechers, M., Barkmann, J. & Tscharnkte, T. Diverging perceptions by social groups on cultural ecosystem services provided by urban green. *Landsc. Urban Plan.* **175**, 161–168 (2018).
36. Kaklauskas, A. et al. Emotions analysis in public spaces for urban planning. *Land Use Policy* **107**, 105458 (2021).
37. Wei, J., Yue, W., Li, M. & Gao, J. Mapping human perception of urban landscape from street-view images: a deep-learning approach. *Int. J. Appl. Earth Observ. Geoinf.* **112**, 102886 (2022).
38. Falanga, R. Understanding place attachment through the lens of urban regeneration. Insights from Lisbon. *Cities* **122**, 103590 (2022).
39. Zou, W., Wei, W., Ding, S. & Xue, J. The relationship between place attachment and tourist loyalty: a meta-analysis. *Tour. Manag. Perspect.* **43**, 100983 (2022).
40. Halder, S. & Sarda, R. Promoting intangible cultural heritage (ICH) tourism: Strategy for socioeconomic development of snake charmers (India) through geoeconomic, geotourism and geoconservation. *Int. J. Geoherit. Parks* **9**, 212–232 (2021).
41. De Paolis, L. T., Chiarello, S., Gatto, C., Liaci, S. & De Luca, V. Virtual reality for the enhancement of cultural tangible and intangible heritage: the case study of the Castle of Corsano. *Digit. Appl. Archaeol. Cult. Herit.* **27**, e00238 (2022).
42. Zeayter, H. & Mansour, A. M. H. Heritage conservation ideologies analysis—historic urban landscape approach for a Mediterranean historic city case study. *HBRC J.* **14**, 345–356 (2019).
43. Saviano, M., Di Nauta, P., Montella, M. M. & Sciarrelli, F. Managing protected areas as cultural landscapes: the case of the Alta Murgia National Park in Italy. *Land Use Policy* **76**, 290–299 (2018).
44. Cassidy, T. *Environmental psychology: Behaviour and experience in context*. (Psychology Press, 2013).
45. Ren, X. Combined effects of dominant sounds, conversational speech and multisensory perception on visitors' acoustic comfort in urban open spaces. *Landsc. Urban Plan.* **232**, 104674 (2023).
46. Shi, Z. & Muller, H. J. Multisensory perception and action: development, decision-making, and neural mechanisms. *Front. Integr. Neurosci.* **7**, 81 (2013).
47. Cai, K., Huang, W. & Lin, G. Bridging landscape preference and landscape design: a study on the preference and optimal combination of landscape elements based on conjoint analysis. *Urban For. Urban Green.* **73**, 127615 (2022).
48. Keleş, E., Atik, D. & Bayrak, G. Visual landscape quality assessment in historical cultural landscape areas. *Eur. J. Sustain. Dev.* **7**, 287–300 (2018).
49. Ma, B., Hauer, R. J., Xu, C. & Li, W. Visualizing evaluation model of human perceptions and characteristic indicators of landscape visual quality in urban green spaces by using nomograms. *Urban For. Urban Green.* **65**, 127314 (2021).
50. Elsadek, M., Sun, M., Sugiyama, R. & Fujii, E. Cross-cultural comparison of physiological and psychological responses to different garden styles. *Urban For. Urban Green.* **38**, 74–83 (2019).
51. Lengen, C. The effects of colours, shapes and boundaries of landscapes on perception, emotion and mentalising processes promoting health and well-being. *Health Place* **35**, 166–177 (2015).
52. Neale, C. et al. Color aesthetics: A transatlantic comparison of psychological and physiological impacts of warm and cool colors in garden landscapes. *Wellbeing, Space Soc.* **2**, 100038 (2021).
53. Dai, T. & Zheng, X. Understanding how multi-sensory spatial experience influences atmosphere, affective city image and behavioural intention. *Environ. Impact Assess. Rev.* **89**, 106595 (2021).
54. Liu, F. et al. Relationships between landscape characteristics and the restorative quality of soundscapes in urban blue spaces. *Appl. Acoust.* **189**, 108600 (2022).
55. Song, X. & Wu, Q. Study on smellscape perception and landscape application of fragrant plants. *Urban For. Urban Green.* **67**, 127429 (2022).
56. Zhang, T., Liu, J. & Li, H. Restorative effects of multi-sensory perception in urban green space: a case study of urban park in Guangzhou, China. *Int. J. Environ. Res Public Health* **16**, 4943 (2019).
57. He, M., Wang, Y., Wang, W. J. & Xie, Z. Therapeutic plant landscape design of urban forest parks based on the Five Senses Theory: a case study of Stanley Park in Canada. *Int. J. Geoherit. Parks* **10**, 97–112 (2022).
58. Chen, G., Shi, J. Y., Xia, Y. P. & Furuya, K. The sustainable development of urban cultural heritage gardens based on tourists'

- perception: a case study of Tokyo's Cultural Heritage Gardens. *Sustainability* **12**, 6315 (2020).
59. Kim, S., Park, E., Fu, Y. & Jiang, F. The cognitive development of food taste perception in a food tourism destination: a gastraphysics approach. *Appetite* **165**, 105310 (2021).
60. Pérez-Martínez, G., Torija, A. J. & Ruiz, D. P. Soundscape assessment of a monumental place: A methodology based on the perception of dominant sounds. *Landsc. Urban Plan.* **169**, 12–21 (2018).
61. Hoeven, A. V. D. Historic urban landscapes on social media: The contributions of online narrative practices to urban heritage conservation. *City Cult. Soc.* **17**, 61–68 (2019).
62. Huang, J., Obracht-Prondzynska, H., Kamrowska-Zaluska, D., Sun, Y. & Li, L. The image of the City on social media: a comparative study using “Big Data” and “Small Data” methods in the Tri-City Region in Poland. *Landsc. Urban Plan.* **206**, 103977 (2021).
63. Jeon, J. Y., Jo, H. I. & Lee, K. Potential restorative effects of urban soundscapes: Personality traits, temperament, and perceptions of VR urban environments. *Landsc. Urban Plan.* **214**, 104188 (2021).
64. Jo, H. I. & Jeon, J. Y. Urban soundscape categorization based on individual recognition, perception, and assessment of sound environments. *Landsc. Urban Plan.* **216**, 104241 (2021).
65. Bertrand, K. Z.; Bialik, M.; Virdee, K.; Gros, A.; Bar-Yam, Y. Sentiment in New York City: A high resolution spatial and temporal view. *arXiv preprint arXiv:1308.5010* (2013). <https://doi.org/10.48550/arXiv.1308.5010>.
66. Gullino, P. & Larcher, F. Integrity in UNESCO World Heritage Sites. A comparative study for rural landscapes. *J. Cult. Herit.* **14**, 389–395 (2013).
67. Bubalo, M., van Zanten, B. T. & Verburg, P. H. Crowdsourcing geo-information on landscape perceptions and preferences: a review. *Landsc. Urban Plan.* **184**, 101–111 (2019).
68. Karimiziarani, M., Jafarzadegan, K., Abbaszadeh, P., Shao, W. & Moradkhani, H. Hazard risk awareness and disaster management: extracting the information content of Twitter data. *Sustain. Cities Soc.* **77**, 103577 (2022).
69. Quinn, C. A. et al. Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. *Ecol. Indic.* **138**, 108831 (2022).
70. Tieskens, K. F., Van Zanten, B. T., Schulp, C. J. E. & Verburg, P. H. Aesthetic appreciation of the cultural landscape through social media: an analysis of revealed preference in the Dutch river landscape. *Landsc. Urban Plan.* **177**, 128–137 (2018).
71. Chesnokova, O. & Purves, R. S. From image descriptions to perceived sounds and sources in landscape: analyzing aural experience through text. *Appl. Geogr.* **93**, 103–111 (2018).
72. Stahl Olafsson, A. et al. Comparing landscape value patterns between participatory mapping and geolocated social media content across Europe. *Landsc. Urban Plan.* **226**, 104511 (2022).
73. Wang, Y., Wu, W. & Boelens, L. City profile: Suzhou, China—the interaction of water and city. *Cities* **112**, 103119 (2021).
74. UNESCO. Our Shared Heritage. Available online: <https://www.unesco.org/en/world-heritage>.
75. UNESCO. Historic Monuments of Ancient Kyoto (Kyoto, Uji and Otsu Cities). Available online: <https://whc.unesco.org/zh/list/688>.
76. Bapiri, J., Esfandiari, K. & Seyfi, S. A photo-elicitation study of the meanings of a cultural heritage site experience: a means-end chain approach. *J. Herit. Tour.* **16**, 62–78 (2021).
77. Mizuuchi, Y. Landscape assessment of forest trail using geotagged visitor employed photography: the case of the inariyama trail in the Takao Quasi-National Park, Tokyo. *J. For. Res.* **28**, 1–10 (2023).
78. Heyman, E. Analysing recreational values and management effects in an urban forest with the visitor-employed photography method. *Urban For. Urban Green.* **11**, 267–277 (2012).
79. Li, H., Hu, M. & Li, G. Forecasting tourism demand with multisource big data. *Ann. Tour. Res.* **83**, 102912 (2020).
80. Huai, S. & Van de Voorde, T. Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. *Landsc. Urban Plan.* **218**, 104307 (2022).
81. Cloud, G. Google Maps Platform Terms of Service. Available online: <https://cloud.google.com/maps-platform/terms>.
82. Trip.com. TRIP.COM TERMS AND CONDITIONS. Available online: <https://www.trip.com/contents/service-guideline/terms.html?locale=en-XX>.
83. Wu, Y. S. et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. 2016.
84. OpenAI. GPT-4 Technical Report. 2023.
85. Hu, C.-B., Zhang, F., Gong, F.-Y., Ratti, C. & Li, X. Classification and mapping of urban canyon geometry using Google Street View images and deep multitask learning. *Build. Environ.* **167**, 106424 (2020).
86. Su, S. et al. Do landscape amenities impact private housing rental prices? A hierarchical hedonic modeling approach based on semantic and sentimental analysis of online housing advertisements across five Chinese megacities. *Urban For. Urban Green.* **58**, 126968 (2021).
87. Bai, Y., Zhao, Y., Shao, Y., Zhang, X. & Yuan, X. Deep learning in different remote sensing image categories and applications: status and prospects. *Int. J. Remote Sens.* **43**, 1800–1847 (2022).
88. Jonathan, L., Evan, S., Trevor, D. Fully Convolutional Networks for Semantic Segmentation. *Computer Vision and Pattern Recognition*, <https://doi.org/10.48550/arXiv.1411.4038>. 2015
89. Fung, C. K. W. & Jim, C. Y. Unraveling Hong Kong Geopark experience with visitor-employed photography method. *Appl. Geogr.* **62**, 301–313 (2015).
90. Bolei Zhou, H. Z. et al. Scene Parsing through ADE20K Dataset. *Computer Vision and Pattern Recognition (CVPR)* 2017.
91. Bolei Zhou, H. Z. et al. Semantic Understanding of Scenes through ADE20K Dataset. *International Journal on Computer Vision (IJCV)* 2018.
92. Group, M.C.C.V. ADE20K. Available online: <https://ade20k.csail.mit.edu/>.
93. Zhou, B., Zhao, H. & Plug, X. Semantic Understanding of Scenes through the ADE20K Dataset. <http://groups.csail.mit.edu/vision/datasets/ADE20K/>.
94. Levin, B. *English verb classes and alternations: A preliminary investigation*; (University of Chicago Press, 1993).
95. Fellbaum, C. WordNet: An electronic lexical database and some of its applications. 1998.
96. Lynott, D. & Connell, L. Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* **41**, 558–564 (2009).
97. Language, C. N. Natural Language AI. Available online: <https://cloud.google.com/natural-language> (accessed on
98. Radosavljevic, U., Dordevic, A., Lalovic, K., Zivkovic, J. & Dukanovic, Z. Nodes and Networks: The Generative Role of Cultural Heritage for Urban Revival in Kikinda. *SUSTAINABILITY* **11**, 2509 (2019).
99. Guzmán, P. C., Roders, A. R. P. & Colenbrander, B. J. F. Measuring links between cultural heritage management and sustainable urban development: An overview of global monitoring tools. *Cities* **60**, 192–201 (2017).
100. Quintana, D. C., Díaz-Puente, J. M. & Gallego-Moreno, F. Architectural and cultural heritage as a driver of social change in rural areas: 10 years (2009–2019) of management and recovery in Huete, a town of Cuenca, Spain. *Land Use Policy* **115**, 106017 (2022).
101. Zhang, C., Chen, X. & Ji, S. Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks. *Int. J. Appl. Earth Observ. Geoinf.* **112**, 102885 (2022).

102. Aldunate, Á, Maldonado, S., Vairetti, C. & Armelini, G. Understanding customer satisfaction via deep learning and natural language processing. *Expert Syst. Appl.* **209**, 118309 (2022).
103. Zhang, T. T. & Lian, Z. F. Research on the distribution and scale evolution of Suzhou Gardens under the urbanization process from the Tang to the Qing Dynasty. *Land* **10**, 18 (2021).
104. Liang, H., Li, W. & Zhang, Q. Semantic-based 3D information modelling and documentation of rockeries in Chinese classical gardens: a case study on the rockery at Huanxiu Shanzhuang, Suzhou, China. *J. Cult. Herit.* **37**, 247–258 (2019).
105. Lu, L. & Liu, M. Exploring a spatial-experiential structure within the Chinese literati garden: the Master of the Nets Garden as a case study. *Front. Archit. Res.* **12**, 923–946 (2023).
106. Long, W. Study on the Differences between Chinese and Japanese Gardens. *J. Anhui Agric. Sci.* **35**, 2 (2007).
107. Sardaro, R., La Sala, P., De Pascale, G. & Faccilongo, N. The conservation of cultural heritage in rural areas: stakeholder preferences regarding historical rural buildings in Apulia, southern Italy. *Land Use Policy* **109**, 105662 (2021).
108. Wei, Z. Poetic and Picturesque Imagination in the Art of the Chinese Classical Garden. In *Proc. of the 3rd International Conference on Landscape and Urban Horticulture, Nanjing, PEOPLES R CHINA*, 207–212 (2011).
109. Li, J. et al. An evaluation of urban green space in Shanghai, China, using eye tracking. *Urban For. Urban Green.* **56**, 126903 (2020).
110. Yuan, S. et al. A virtual reality investigation of factors influencing landscape preferences: natural elements, emotions, and media creation. *Landsc. Urban Plan.* **230**, 104616 (2023).
111. Xu, J. et al. Multi-sensory experience and preferences for children in an urban forest park: a case study of Maofeng Mountain Forest Park in Guangzhou, China. *FORESTS* **13**, 1435 (2022).
112. Watts, G. The effects of “greening” urban areas on the perceptions of tranquility. *Urban For. Urban Green.* **26**, 11–17 (2017).
113. Liu, Y., Hu, M. & Zhao, B. Audio-visual interactive evaluation of the forest landscape based on eye-tracking experiments. *Urban For. Urban Green.* **46**, 126476 (2019).
114. Xiao, J., Wu, Y., Wang, M. & Zhao, Y. Using choice experiments to assess tourist values for intangible cultural heritage—the case of Changdao fishermen’s work song in China. *J. Cult. Herit.* **60**, 50–62 (2023).
115. Jacob, D., Ming-Wei, C., Kenton, L. & Kristina, T. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Google AI Language*, <https://doi.org/10.48550/arXiv.1810.04805> (2019).

Acknowledgements

We would like to sincerely thank all co-authors for their valuable contributions to this study. This research was supported by the Soft Science Research Program of Zhejiang Province (Grant Number 2025C25021), the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (B) (Grant Number 24K03144), and the China Scholarship Council. Their support made this work possible.

Author contributions

Jiao Zhang and Yangyang Shi developed the study and wrote the main manuscript text, Chenshu Cai analyzed the data. Liang Zhao is involved in the visualization. All authors were involved in data engagement and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Jiao Zhang.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025