Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Perception of Cross-cultural Garden Tourism: A Case Study of the New York Chinese Scholar's Garden at Snug Harbor

Lin Zhou
Jiangnan University
Fan Yang
Jiangnan University
Hainan Yan
Jiangnan University
Peijun Jia
7260310021@stu.jiangnan.edu.cn

Jiangnan University

Article

Keywords: Chinese gardens abroad, intercultural tourism, visitor perception, user-generated content, multimodal analysis

Posted Date: October 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11064062/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Perception of Cross-cultural Garden Tourism: A Case Study of the New York Chinese Scholar's Garden at Snug Harbor

Lin Zhou[1], Fan Yang[1], Hainan Yan[1] & Peijun Jia[1*]

[1]School of Design, Jiangnan University, Wuxi 214122, China.

These authors contributed equally: Lin Zhou, Fan Yang.

*Correspondence to: : 7260310021@stu.jiangnan.edu.cn

Abstract

Chinese gardens built abroad are tourism destinations and places where visitors encounter Chinese garden culture within a local social environment. Evidence remains limited on how visitors combine text and photographs to describe the same garden. Using the Chinese Scholar's Garden at Snug Harbor, Staten Island, New York, we analyzed 1,374 original textual records from Yelp, Google Maps and Twitter and 1,395 tourist photographs from Google Maps. We extracted keywords from the reviews, classified sentiment and segmented the photographs into semantic landscape elements. In the matched subset of 1,159 photographs and 139 reviews linked to non-empty review records, we used four regression models and Pearson correlation to examine associations between image elements and review sentiment scores. Of the screened reviews, 57.64% were positive, 38.11% neutral and 4.25% negative. Tree, plant and sky were the most frequently detected elements in the photographs, and architecture and vegetation recurred in both textual and visual records. Across the four models, held-out R^2 did not exceed 0.061, and no model explained more than 7% of the variance in review sentiment, indicating that the visual composition of photographs and the verbal evaluation of a visit are largely independent records of the same garden. The study offers a case-based exploratory workflow for examining what visitors record in reviews and photographs of an overseas Chinese garden and how those records relate to sentiment expression.

Keywords: Chinese gardens abroad; intercultural tourism; visitor perception; user-generated content; multimodal analysis.

1. Introduction

Cultural tourism research has shifted from attractions and heritage themselves toward visitor experience, destination image and the circulation of both in digital media. Text and photographs posted on social media record visits and participate in the formation of destination images online (McMullen, 2024; Liu et al., 2024). For Chinese gardens abroad, this material is useful because the garden is a Chinese cultural form encountered in a local social environment. The Chinese Scholar's Garden retains recognizable elements of Chinese garden design, including architecture, plants, rocks and water, while operating within a U.S. tourism context shaped by local platforms and visitor expectations. Understanding its online representation therefore requires attention to both the spatial features of the garden and the setting in which visitors recorded them.

Studies of garden form (Yu et al., 2021; Ding et al., 2023) and UGC (McMullen, 2024; Liu et al., 2024) have developed mostly along separate lines in work on overseas Chinese gardens. The two literatures rarely meet in the same case study. It remains unclear how visitors represent the same garden in both text and photographs, and whether visual labels in matched records are associated with review sentiment. This paper places these records side by side by describing what visitors mention or photograph, how sentiment is distributed and how visual labels vary with review sentiment scores.

Three questions guided the analysis of the Chinese Scholar's Garden in the United States:

Q1: What garden elements recur in the reviews, and which of them also appear in the photographs?

Q2: How are sentiment classes distributed across the reviews?

Q3: To what extent, if at all, do the semantic labels of visitor photographs predict the sentiment of the reviews to which they are linked? Because photographs and reviews on Google Maps are contributed independently, and because the sentiment distribution is highly skewed, we treat this question as exploratory and expect at most a weak association.

The analysis follows the data streams in sequence. Term frequency-inverse document frequency (TF-IDF; Salton & Buckley, 1988) was used for review keywords; the Robustly Optimized BERT Pretraining Approach model (RoBERTa; Liu et al., 2019) was used for review sentiment; and Mask2Former (masked-attention mask transformer; Cheng et al., 2022) with the ADE20K scene-parsing dataset (Zhou et al., 2017) was used for the photographs. The matched

image-text records were then examined with four regression models and Pearson correlations to assess label-sentiment associations.

The 1,395 Google Maps photographs were used to calculate label detection rates, mean pixel proportions and within-frame co-occurrence. Reviews were collected from Yelp, Google Maps and Twitter (now X; Yelp, 2026; Google Maps, 2026; X, 2026). Google Maps alone was used for the image analysis because its photographs are attached to a named place.

The paper first describes the case and data, then presents the multimodal results and the limits that apply to their interpretation.

2. Literature review

2.1 Cultural tourism and destination image

Cultural tourism research examines how visitors interpret cultural symbols, ways of life and place in specific settings. Recent studies identify authenticity, emotional engagement and visitor identity as relevant dimensions of cultural tourism experiences and sharing behavior (Panigrahi & Shrivastava, 2024; Nguyen et al., 2025; Han & Bae, 2022).

Destination image theory provides a basis for explaining such composite judgments. Baloglu and McCleary (1999) defined destination image as the cognitive and affective representation visitors form of a destination, and Beerli and Martín (2004) linked image formation to visitor attributes, travel experiences and destination characteristics. Recent work connects destination image with UGC, social media, authenticity, emotional experience and visitor motivation in heritage and tourism settings (Alarcón-Urbistondo et al., 2023; Li & Yu, 2022; Zhao & Agyeiwaah, 2025; Zhang & Xiao, 2025; Nath et al., 2025; Wang & Jailani, 2025). Research on intangible cultural heritage tourism has modelled how ritualized forms of experience shape visitor response, indicating that such dimensions of experience can be assessed and not only described (Yue, 2025). Fewer studies examine how visitors describe landscape features and express emotion within the same naturally occurring records.

These questions are particularly relevant in cross-cultural garden tourism because overseas Chinese gardens retain Chinese spatial and aesthetic forms while operating within a local social and tourism environment. In this paper, cross-cultural refers to the meeting of Chinese garden form with a U.S. tourism setting; the dataset does not support comparisons between defined cultural groups. UGC and image analysis can describe records of that encounter, while destination

image theory supplies the interpretive frame. Computer-vision studies have begun to compare the tourist gaze of visitors from different world regions (Zhang et al., 2021); the present dataset does not support such group comparisons and is not used for that purpose.

2.2 **Overseas Chinese gardens as cultural enclaves**

Overseas Chinese gardens retain recognizable Chinese design forms, yet they are encountered within the social and tourism settings of another country. Their meanings depend on the garden itself, its management and the ways visitors use it.

Work on overseas Chinese gardens has mostly examined their history, construction and design. Earlier texts and images helped circulate the Chinese garden in the West (Rinaldi, 2012). Later studies measured garden organisation and visitor stay with space syntax, spatial attributes and fractal analysis (Yu et al., 2021; Ding et al., 2023). These approaches describe garden form. The way visitors record these gardens in reviews and photographs has been examined less directly.

Research on overseas Chinese gardens has gradually extended from design history and cultural transmission to spatial analysis, while the visitor perspective has received little sustained attention. Research on heritage visitors further indicates that expectations mediate cultural heritage cognition and visitor experience, so what visitors record reflects what they anticipated as well as the site itself (Wang & Jailani, 2025). Taking the Chinese Scholar's Garden in the United States as a concrete case, this paper examines how Chinese garden elements are recorded, described and evaluated in the text and images of cross-cultural tourism.

2.3 **Tourist perception in user-generated content**

Visitor perception is central to destination image and experience evaluation. Questionnaires and interviews can provide detailed interpretation, but they often cover small samples and short observation periods. UGC records text, photographs and ratings posted voluntarily on platforms, and it has been used to study destination image, experience quality and electronic word-of-mouth (Alarcón-Urbistondo et al., 2023; Li & Yu, 2022; McMullen, 2024; Zain et al., 2025). Platform differences in user composition, expression habits and content length mean that multi-platform data require source-specific interpretation.

Recent UGC studies have examined both text and images. Text analysis can identify recurring places, activities and emotion words in reviews; visual analysis can count the landscape elements that appear in photographs (Xiao et al., 2022; Huai et al., 2022; Koruyan & Aktaş, 2022;

Qian et al., 2023). Deep-learning classifiers have also been used to classify tourist photographs and to derive destination image directly from image content (Cho et al., 2022). Multi-platform visual studies also show that photographs selected on different platforms can present different destination images (McMullen, 2024; Hu & Geng, 2024). Visual analytics tools have also been developed to compare several destination images drawn from tourism user-generated content within a single view (Li et al., 2022). Crowdsourced photographs have also been used to assess the visual quality of urban landscapes at scale (Qi et al., 2023). These studies support paired textual and visual analysis, with one boundary: the presence of an element in a photograph shows that it was recorded, whereas preference requires supporting evidence from other materials.

As UGC datasets have grown, tourism research has increasingly used natural language processing (NLP) and computer vision. Recent studies have used text mining, sentiment analysis, social media photographs and deep learning models to identify themes and landscape features in visitor expression (Mehra, 2023; Huang & Chelliah, 2025; Drašković & Milanović, 2025; Tunca, 2026; Bejoc et al., 2025; Pang, 2026). On the management side, the same data are used to trace digital influence as part of visitor flow management at heritage sites (Shen, 2025). Model outputs still need interpretation within the specific cultural setting, and a causal link between landscape elements and particular emotions requires independent evidence.

The relationship between textual and visual records of the same garden remains underexamined. For overseas Chinese gardens, reviews and photographs can sit alongside one another in the online destination image. The present study examines that relationship through an exploratory multimodal analysis of a single case.

3. **Methodology**

This paper uses multi-source user-generated content to analyse how visitors record the Chinese Scholar's Garden in reviews and photographs. The workflow includes data collection, data screening, text analysis, image semantic segmentation and image-text association analysis. Textual data provide keywords, word frequencies and sentiment results; image data provide detection rates, pixel proportions and co-occurrence relations of semantic labels. A matched subsample is then used to connect image labels with review sentiment scores for exploratory association analysis. Figure 1 outlines the workflow.

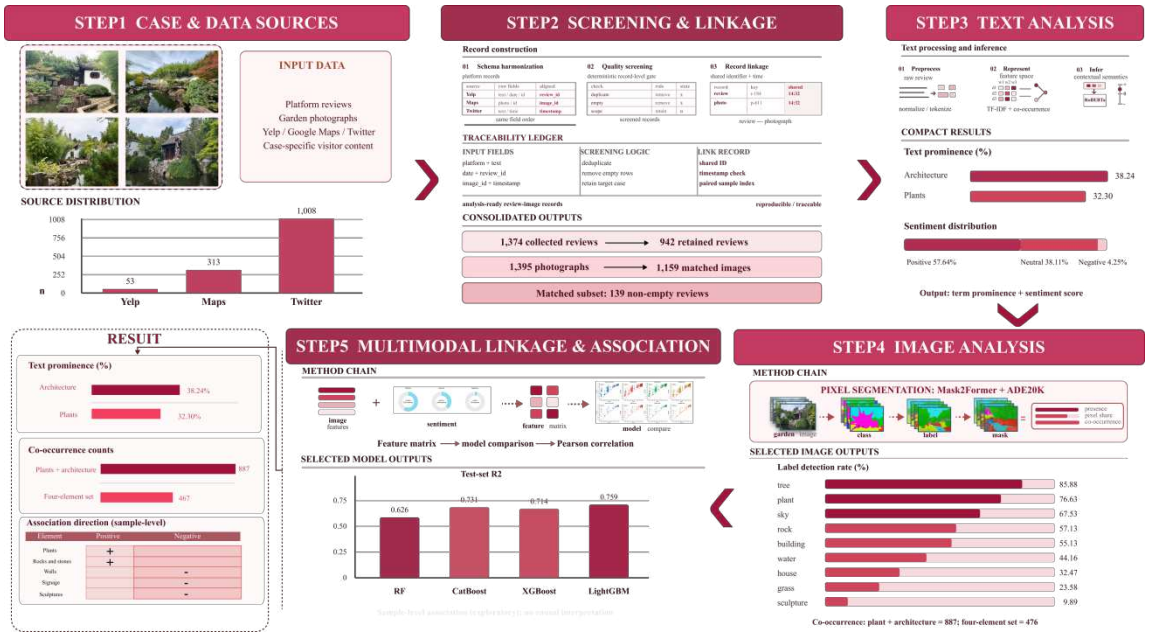


Figure 1. Framework of the proposed method.

3.1 Study area

The Chinese Scholar's Garden at Snug Harbor, Staten Island, New York (Figure 2), is located in a multicultural U.S. tourism context (Collins, 2019). It is the second Chinese garden built in the United States, after the "Ming Xuan" (Ming Room) at the Metropolitan Museum of Art. Modelled on the classical gardens of Suzhou, the garden contains architecture, plants, rocks and water features associated with Chinese garden design. Its travel-platform reviews and photographs provide a case for examining how visitors record an overseas Chinese garden in text and images. The garden was selected as the case site for this study of cross-cultural garden tourism.

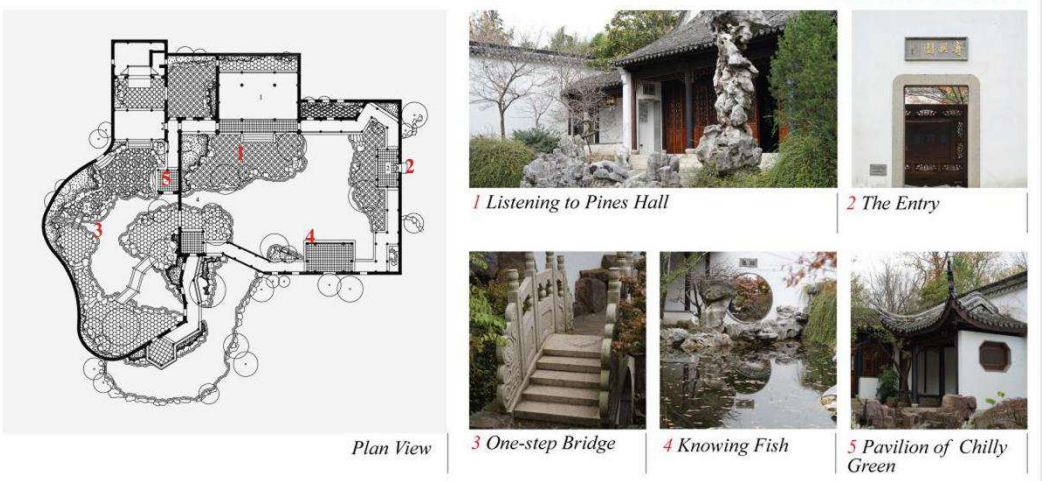


Figure 2. Study area. Note: Plans and photographs of the Chinese Scholar's Garden.

3.2 Data collection and preparation

Text data came from user-generated content on three platforms: Yelp, Google Maps and Twitter (now X). The three platforms differ in function and expression context: Yelp centres on place reviews, Google Maps links reviews and photographs to a specific place, and Twitter contains shorter public posts. We used the three text sources for the overall analysis of visitor perception and sentiment. Image analysis was limited to Google Maps photographs to preserve place-specific correspondence. Data collection covered 2008–2025 and yielded 1,374 raw textual reviews, of which Yelp, Google Maps and Twitter contributed 53, 313 and 1,008, respectively; the image data comprise 1,395 photographs, all from Google Maps.

During screening, we removed public or institutional posts unrelated to the garden, including non-visitor material about the Snug Harbor Cultural Center and CUNY TV. We translated a small number of non-English texts into English and removed links, special symbols and emojis. The final set contained 942 reviews: 53 from Yelp, 292 from Google Maps and 597 from Twitter. The 1,395 photographs are the complete set on the garden's Google Maps page. The image-text association analysis used only the 1,159 photographs and 139 reviews that could be matched to non-empty review records. The resulting association analysis describes this matched subsample rather than all photographs or reviews. Table 1 summarises the dataset before and after screening.

Table 1. Summary of the dataset.

Review type	Social media	Data collection	After data screening
Text data	Yelp	53	53
	Google Maps	313	292
	Twitter	1008	597
	Total data	1374	942
Image data	Google Maps	1395	1395
	Total data	1395	1395

3.3 Data analysis

Text analysis

The screened reviews were used for text analysis. After common stopwords were removed, term frequency-inverse document frequency (TF-IDF; Salton & Buckley, 1988) was used to weight the remaining words. We counted keyword frequencies and co-occurrences, with one co-occurrence recorded whenever two keywords appeared in the same review. Garden-element terms

were grouped into four categories: architecture, plants, water and rock. These counts and the co-occurrence network are used in Section 4.2 to describe the elements mentioned in reviews.

RoBERTa (Robustly Optimized BERT Pretraining Approach; Liu et al., 2019) was used for sentiment analysis through the cardiffnlp/twitter-roberta-base-sentiment-latest checkpoint. For each review, the model returned probabilities for positive, neutral and negative labels. The continuous sentiment score was calculated as P(positive) - P(negative), producing a scale from -1 to +1. Scores above 0.3 were classified as positive, scores below -0.3 as negative and scores between -0.3 and 0.3 as neutral. The class labels are used for the overall distribution in Section 4.3, and the continuous scores are paired with semantic labels from matched photographs in Section 4.5. Because the reviews are short texts, and because systematic comparisons show that the recovery of affective states from short online posts depends strongly on the text-mining method used (Macanovic & Przepiorka, 2024), these scores are treated as model-derived indicators rather than as direct measurements of visitor feeling.

Image segmentation

The image analysis used the 1,395 tourist photographs from Google Maps. Semantic segmentation was performed with Mask2Former (masked-attention mask transformer; Cheng et al., 2022) based on the ADE20K scene-parsing dataset (Zhou et al., 2017). The model returned pixel-level labels, mask maps and pixel distributions for each label. From these outputs, we calculated label detection rates, mean pixel proportions and within-image co-occurrence relations. Frequently retained raw labels included building, tree, grass, plant, water, house, rock, sculpture, sky, wall and earth; the merged categories used for the image-text association analysis, are described in Section 4.1. Figure 3 shows examples of the segmentation output.

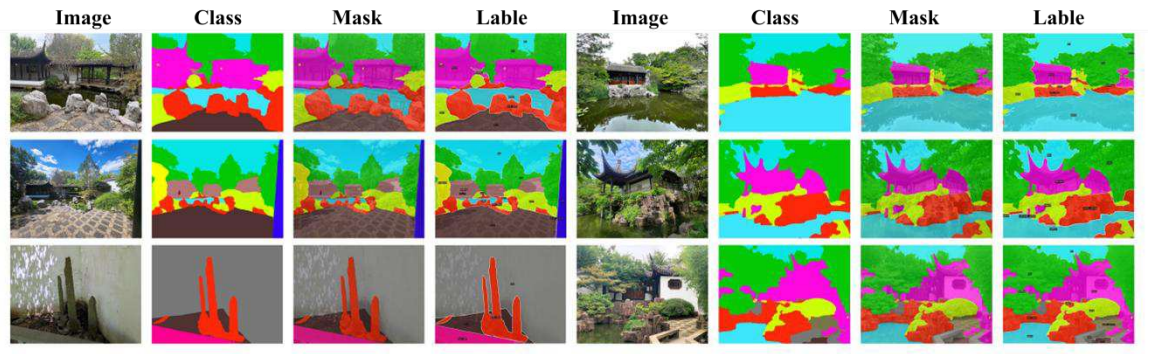


Figure 3. Semantic segmentation of garden images captured by tourists.

The image-text association analysis paired the semantic-label variables extracted from each matched photograph with the continuous RoBERTa sentiment score of the corresponding review. The resulting dataset contained 1,159 photograph records linked to 139 non-empty reviews. When several photographs were associated with the same review, the same review-level sentiment score was assigned to each linked photograph. The photograph records were randomly divided into a training set (80%) and a test set (20%) using random seed 42. The split was not grouped by review, so photographs linked to the same review could occur in both sets. A StandardScaler was fitted to the training predictors only and then applied unchanged to the test predictors.

Four tree-based regressors were fitted using fixed settings, with no additional hyperparameter tuning. Random Forest comprised 200 trees, with a maximum depth of 10, a minimum of 5 samples per split and 2 samples per leaf (Breiman, 2001). CatBoost used 200 iterations, a learning rate of 0.05, a depth of 5 and L2 leaf regularization of 3 (Prokhorenkova et al., 2018). XGBoost and LightGBM each used 200 trees, a learning rate of 0.05, a maximum depth of 5, a subsample ratio of 0.8 and colsample_bytree = 0.8. XGBoost additionally used min_child_weight = 2 (Chen & Guestrin, 2016), whereas LightGBM used num_leaves = 31 and min_child_samples = 5 (Ke et al., 2017). Random Forest, XGBoost and LightGBM used random_state = 42; CatBoost used random_seed = 42. Each model was fitted on the training set and assessed on the held-out test records using R^2 and RMSE, with RMSE retained on the original sentiment-score scale. Training-set R^2 was also recorded as a diagnostic of model fit. Five-fold shuffled cross-validation with random seed 42 was performed on the training set, using R^2 as the scoring metric. Built-in feature-importance scores were used to rank the semantic labels within each model. Pearson correlations were calculated between each label and the review sentiment score among photograph records in which that label was present; labels occurring in no more than five records were not analyzed. These procedures describe predictive performance and sample-level associations. Because the split was conducted at the photograph-record level and the same review-level sentiment score could be assigned to several photographs, the results do not constitute independent review-level validation or support causal claims about the effects of landscape elements on visitor sentiment.

Two further constraints shape the interpretation of the image-text analysis. First, photographs and reviews on Google Maps are contributed by users independently, so a photograph linked to a review record was not necessarily taken by the reviewer. Second, the outcome has limited variance:

95.75% of the screened reviews were neutral or positive, which caps the share of variance any predictor can explain; the review is also the independent unit, so the effective sample size for the association analysis is 139 rather than 1,159.

4. Results and discussion

4.1 Descriptive statistics of image semantic labels

To describe the basic composition of the visual data, this section reports descriptive statistics for the semantic segmentation results from the 1,395 tourist photographs from Google Maps. Table 2 reports selected garden-related labels, their detection rates and their mean pixel proportions. Detection rate indicates the percentage of photographs in which a label appears; mean pixel proportion gives the average image area occupied by that label across all photographs. Tree and plant had the highest detection rates, at 85.88% and 76.63%, respectively, and building and rock also appeared in about half of the photographs. These results show that vegetation and built garden elements were the most commonly detected components in posted tourist photographs. The frequencies describe image content rather than preference for any particular element.

Table 2. Descriptive statistics of semantic segmentation labels.

Semantic label	Images containing label (n)	Detection rate (%)	Mean pixel proportion (%)
tree	1,198	85.88	21.03
plant	1,069	76.63	10.66
water	616	44.16	8.86
building	769	55.13	8.82
house	453	32.47	4.35
rock	797	57.13	5.69
grass	329	23.58	1.99
sculpture	138	9.89	0.44

Note: Detection rate is the percentage of the 1,395 images in which the label was detected. Mean pixel proportion is calculated across all images, including images in which the label was absent.

Table 2 reports selected raw labels from semantic segmentation, whereas Table 4 uses the merged categories retained for image-text association analysis. In the coding scheme, tree, plant and grass were merged into Plant; building and house into Architecture; rock and sculpture into Rock; and water was retained as Water. Wall was kept as an independent label in the raw ADE20K output because it marks the garden boundary and is not included in the four categories above. The N values in Table 4 refer to records in which each merged category was detected, not to sums of the raw counts in Table 2. The two tables aggregate at different levels, so their counts are not directly comparable.

4.2 Visitor perception based on text analysis

This analysis covers the 942 screened textual reviews. After common stopwords were removed, TF-IDF weights were calculated for the remaining words. We then calculated keyword frequencies and co-occurrence counts, counting one co-occurrence whenever two keywords appeared in the same review. The resulting network, shown in Figure 4, comprises 49 nodes, including the garden name, landscape elements, descriptive terms and emotion words.

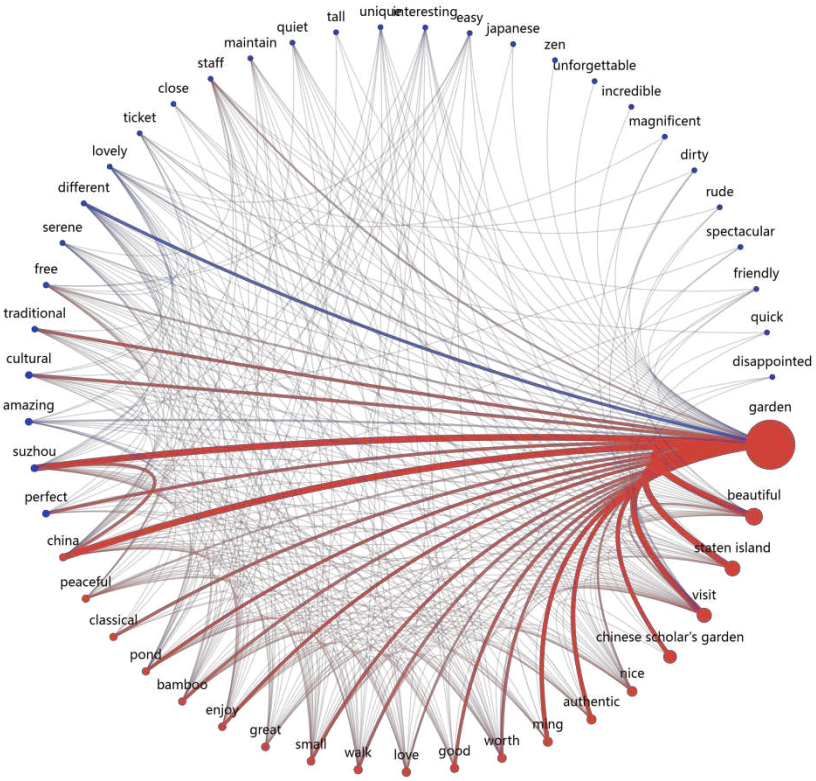


Figure 4. Co-occurrence network of social media reviews.
*The numbers next to the keywords indicate how many times these words appear in the dataset.

The largest nodes correspond to the most frequent terms. The terms "beautiful," "nice," "authentic" and "peaceful" occurred 139, 54, 49 and 33 times, respectively. These adjectives show that many reviews describe the garden positively, but they do not explain what cultural meanings reviewers attached to it. The 48 occurrences of "ming" point to references to Chinese history in some reviews; "worth" appears 44 times in comments about the value of the visit. "free" and "ticket" concern admission. At the smaller end of the network, "small" and "quick" refer to garden size and visit duration, whereas "japanese" and "zen" suggest mixed cultural references in some reviews.

The strongest links joined “garden” with “china” (169 co-occurrences) and “suzhou” (150). This points to repeated references connecting the garden with China and Suzhou. Pairs such as “dirty”–“maintain” and “rude”–“staff” may reflect comments about maintenance or staff, although the counts alone cannot show whether they came from a substantive complaint or repeated wording.

The network is dominated by terms about the garden environment, aesthetic impressions and cultural references. "beautiful" and "peaceful" often occur together, whereas "china" and "suzhou" link the garden to Chinese garden culture in the review corpus. Lower-frequency terms add information about size, visit duration and maintenance.

In content analysis, the coding scheme depends on both the conceptual framework and the material being coded (Neuendorf, 2017). We used the classical categories in Chinese garden scholarship (Peng, 1986; Tong, 1984) as a reference and adjusted the wording groups to the terms present in the dataset. Words referring to the same garden element were counted in one category. Repeated mentions were counted separately, so a review containing “building,” “water” and “flower” contributed one count to each category. Reviews with no garden-element term were excluded from this count.

Table 3. Classification of garden elements and frequencies.

Classification 1: Architecture		Classification 2: Plants		Classification 3: Water		Classification 4: Rock	
Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
architecture[a]	29	bamboo[a]	35	pond[a]	34	rock[c]	32
building[a]	25	plant[a]	29	water	22	stone[a]	4
pavilion[a]	19	tree[a]	26	waterfal[a]	13	hill	2
house[a]	13	flower[c]	15	river[a]	2	sculptures	2
bridge[a]	12	blossoms	7	waterscape	1	monoliths	1
door[a]	10	foliage	3			mountains	1
wall[a]	10	lotus	2				
hall	9	greenery	2				
roof[a]	8	branch[a]	2				
window[a]	8	magnolias	2				
structure[b]	5	jasmine	2				
Subtotal	148	Subtotal	125	Subtotal	72	Subtotal	42
Percentage	38.24%	Percentage	32.3%	Percentage	18.6%	Percentage	10.85%

Note: Related inflected or derived forms were grouped as follows.
N = 387. [a] The singular and plural forms of these words were applied. [b] Structure + structures + structured. [c] Flower + flowers + flowering, rock + rocks + rocking.

Table 3 contains 387 valid word frequencies. Architecture terms accounted for 148 (38.24%), followed by plant terms with 125 (32.30%), water terms with 72 (18.60%) and rock terms with 42

(10.85%). Architecture was the most frequent category. The architecture group included general terms such as "architecture" and "building" as well as roofs, doors and windows. Bamboo was the most frequent plant term. Its frequency is consistent with the bamboo grove near the entrance, although word counts do not show how reviewers understood bamboo culturally.

4.3 Sentiment distribution and textual interpretation

Figure 5 shows the sentiment distribution: 543 of the 942 screened reviews (57.64%) were positive, 359 (38.11%) neutral and 40 (4.25%) negative. These percentages describe only the reviews available on the three platforms; they should not be read as the satisfaction of all garden visitors. The same pattern appears in the frequent terms “beautiful” and “peaceful,” whereas “small” and “quick” point to mixed views of garden size and visit duration.

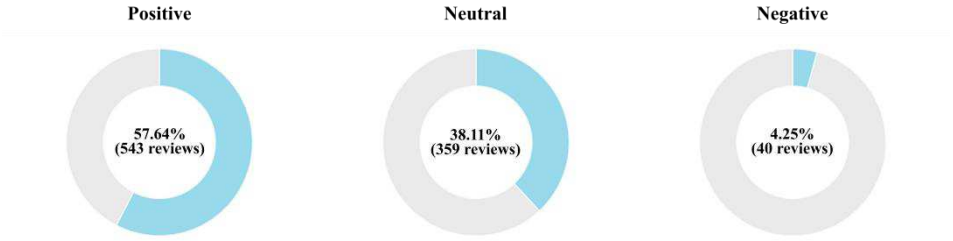


Figure 5. Sentiment analysis of tourist reviews about Chinese Scholar's Garden on social media.

The text results were predominantly positive, although comments on garden size and visit duration were mixed. Reviews frequently use "beautiful," "peaceful" and "authentic" and mention architecture and plants, while "small" and "quick" suggest a less uniformly positive view. Because nationality and cultural background were not recorded, these observations describe the sample rather than a defined cultural group.

4.4 Visual perception: image elements and composition

In the Google Maps photographs, tree, plant and sky were detected most often, each in more than 900 images (Figure 6). Wall, rock, building and earth also appeared frequently. These frequencies describe posted images rather than visitor preference. Earth refers mainly to ground surface, whereas wall marks the garden boundary; the labels were kept separate for that reason.

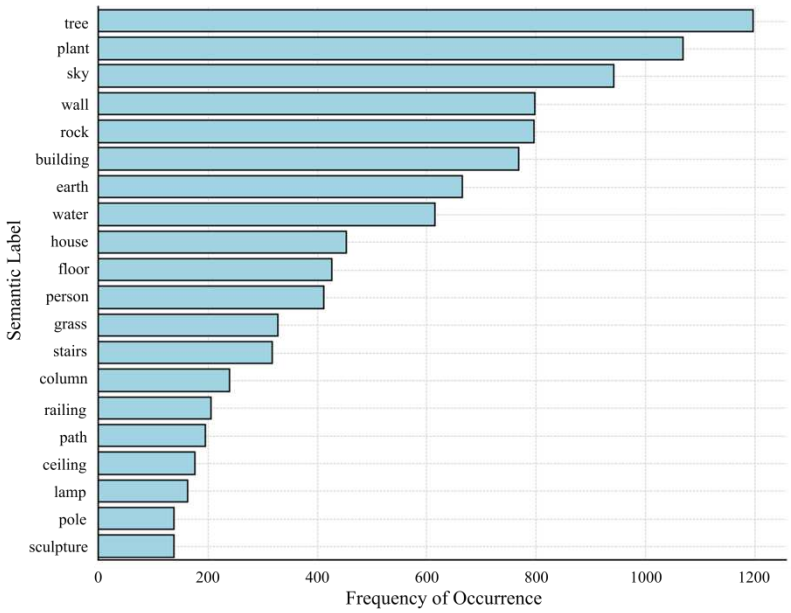


Figure 6. Top 20 most frequent semantic labels in tourist photos.

Tourist photographs often include several recognizable destination features in one frame (Albers & James, 1988; Day et al., 2002). For this analysis, architecture, rock, water and plants were treated as the principal garden elements. The scheme follows classical Chinese garden scholarship and was cross-checked against the NLP word-frequency results (Peng, 1986; Tong, 1984). Raw semantic labels were retained for co-occurrence analysis and grouped into Plant, Architecture, Rock and Water for the Sankey diagram. Figure 7 shows these combinations; nodes represent combinations at different levels and line width represents their frequency.

Figure 7 and Table 4 show that plants and architecture form the most frequent combination. Water and rock are also common, but their detection rates as single labels are somewhat lower. Plant + Architecture appears 887 times, making it the most frequent two-element combination. In this sample, architecture usually entered the frame together with plants.

Among the three-element combinations, Plant + Architecture + Rock appears 669 times, followed by Plant + Rock + Water (537) and Plant + Architecture + Water (501). The four-element combination, Plant + Architecture + Rock + Water, appears 476 times. These combinations describe co-occurrence within the frame; they do not reveal the order of visitors' attention or the emotions associated with these elements.

Water is detected as a single label less often than plants and architecture, but it accounts for a large share of the photographs in which elements co-occur. Rock + Water appears 552 times and Architecture + Water 501 times, and water is also part of the four-element combination. This

distribution may reflect the concentration of water features at a few representative viewpoints. The present data support a description of where water co-occurs, but they are not sufficient to interpret water as the "binding element" of the composition.

In the tourist photographs examined in this study, plants and architecture are the most common combination, while rock and water often appear as joint compositional elements. The text results also stress architecture, plants and bamboo. This correspondence suggests that visitors' records combine verbal descriptions of traditional garden motifs with image frames in which architecture and natural elements appear together.

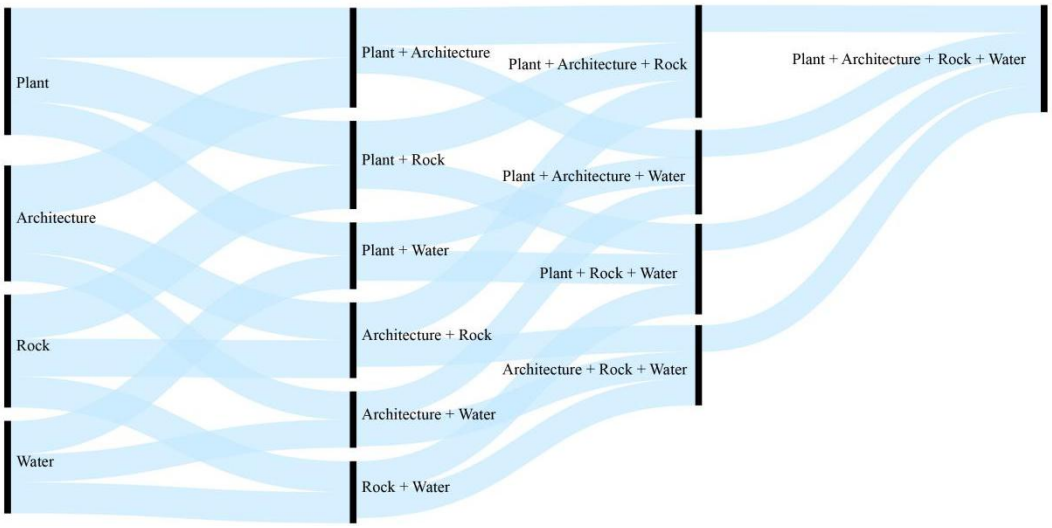


Figure 7. Multi-level Sankey diagram of co-occurrence of garden elements.

Table 4. Garden element combinations and frequencies.

Combination	Frequency	% of Plant (N=1276)	% of Architecture (N=907)	% of Rock (N=829)	% of Water (N=616)
Single-element combinations					
Plant	1276	100	-	-	-
Architecture	907	-	100	-	-
Rock	829	-	-	100	-
Water	616	-	-	-	100
Two-element combinations					
Plant + Architecture	887	69.5	97.8	-	-
Plant + Rock	784	61.4	-	94.6	-
Architecture + Rock	672	-	74.1	81.1	-
Plant + Water	593	46.5	-	-	96.3
Rock + Water	552	-	-	66.6	89.6
Architecture + Water	501	-	55.2	-	81.3
Three-element combinations					
Plant + Architecture + Rock	669	52.4	73.8	80.7	-
Plant + Rock + Water	537	42.1	-	64.8	87.2
Plant + Architecture	501	39.3	55.2	-	81.3

Combination	Frequency	% of Plant (N=1276)	% of Architecture (N=907)	% of Rock (N=829)	% of Water (N=616)
+ Water					
Architecture + Rock + Water	476	-	52.5	57.4	77.3
Four-element combination					
Plant + Architecture + Rock + Water	476	37.3	52.5	57.4	77.3

Note: Percentages are calculated based on the total number of photographs in which each individual element appears. A dash (–) indicates that the element is not included in the corresponding combination.



a.Plant + Architecture b.Plant + Architecture + Rock c.Plant + Rock + Water d.All Elements Combined

Figure 8. Representative photographs from tourist reviews.

Figure 8 presents representative photographs for four of the combinations reported above. Plants and architecture recur in both the review vocabulary and the image combinations. Their joint appearance indicates that traditional garden features were often recorded in the same frame. It does not show that visitors attached the same cultural meaning to them.

4.5 Image-text association analysis and interpretation

The Google Maps listing contained 1,395 photographs. We retained 1,159 photographs linked to 139 non-empty reviews after excluding images paired with empty review records. The platform records supplied the links. Each matched review was converted to a RoBERTa score from -1 to +1, and that score was assigned to its photographs. We then used the image labels and scores in the four regression models and calculated Pearson correlations to describe the direction of association.

After this photograph-record split, the model comparison (Figure 9) showed weak held-out prediction. CatBoost gave the highest test-set R^2 (0.061; RMSE = 0.267), followed by Random Forest (R^2 = 0.056; RMSE = 0.268), XGBoost (R^2 = 0.027; RMSE = 0.272) and LightGBM (R^2 < 0.001; RMSE = 0.276). Five-fold cross-validation produced similarly low mean R^2 values: CatBoost 0.083, Random Forest 0.080, XGBoost 0.045 and LightGBM 0.028. Training-set R^2 values were higher for XGBoost (0.710), LightGBM (0.641), Random Forest (0.475) and CatBoost (0.353), indicating that the models fit the training records far better than held-out records. No model explained more than 7% of the variance in review sentiment. Read against the restricted variance of the outcome (95.75% of reviews neutral or positive), this indicates that the visual

composition of the linked photographs and the verbal evaluation of a visit are largely independent records, rather than that the models failed to fit. Because the split was performed at photograph-record level while sentiment was assigned at review level, these values should be read as exploratory diagnostics rather than independent evidence of a stable predictive model.

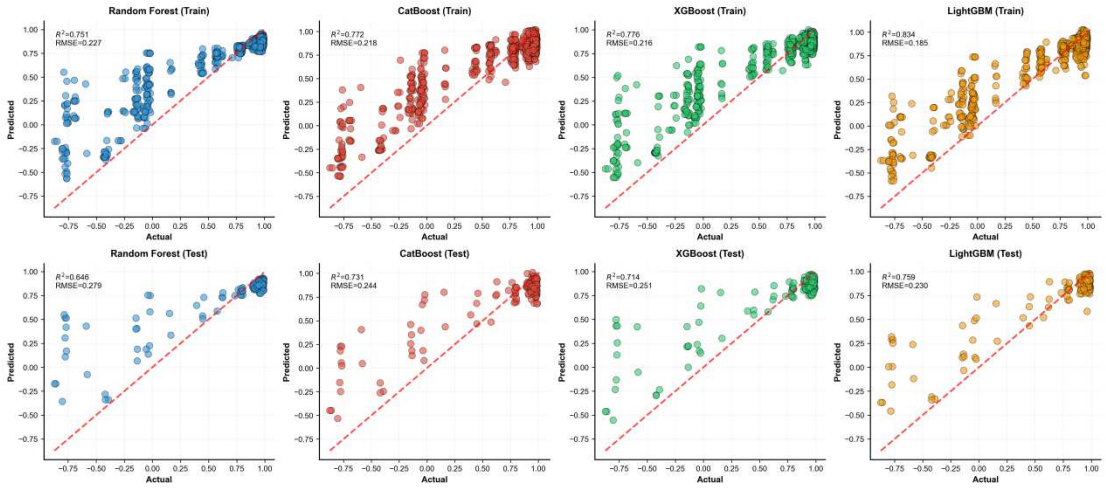


Figure 9. Predicted versus actual values of four regression models on training and test sets.

The residuals in Figure 10 changed sign across the predicted range. Negative values were more common near lower predictions and positive values near higher predictions, consistent with bias at both ends. The spread of residuals across the four models is also consistent with the weak held-out R^2 values reported above.

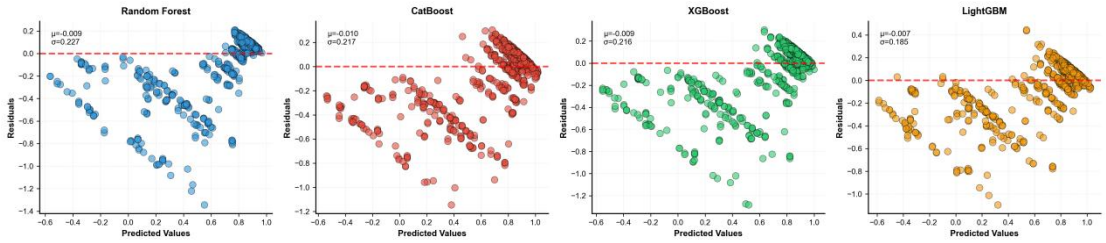


Figure 10. Scatter plots of residuals and predicted values for four regression models.

The four models did not rank the labels in the same order (Figure 11). Building, plant, tree, earth and wall appeared repeatedly among the higher-ranked variables, but their positions changed from one model to another. We treat these rankings as model-specific feature-importance diagnostics; they do not identify fixed drivers of visitor sentiment.

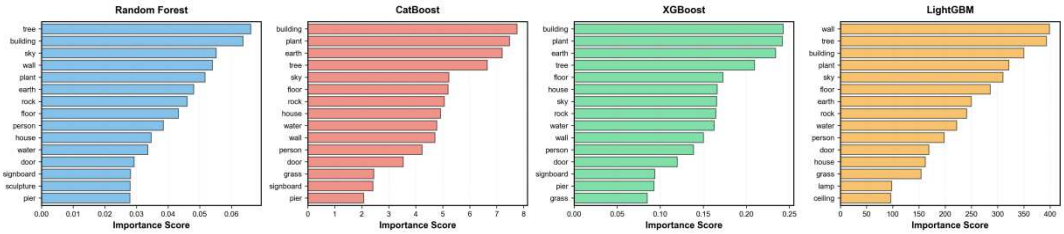


Figure 11. Feature importance scores of garden visual elements across four regression models.

Pearson correlations were calculated between visual labels and sentiment scores among records in which each label was detected. Figure 12 shows the direction of these associations. Natural elements generally had positive coefficients, whereas some artificial or functional elements had negative coefficients. In the exploratory analysis, coefficients with $p < 0.05$ were marked; these results are therefore interpreted as exploratory associations. Together with the coefficient magnitudes and model rankings, these results describe relationships within the matched sample. They do not show that any landscape element changed visitors' emotions.

Across the analyses, architecture and natural landscape were the most recurrent features of the garden's online representation, and reviews were predominantly positive. Maintenance, service and public hygiene appeared in some negative accounts. Word co-occurrence and image-sentiment associations describe patterns within the sample; neither provides evidence of a causal relationship. The descriptive results show a consistent garden image across text and images, while the predictive results show that this shared imagery does not translate into review sentiment. The two records are therefore complementary rather than substitutable.

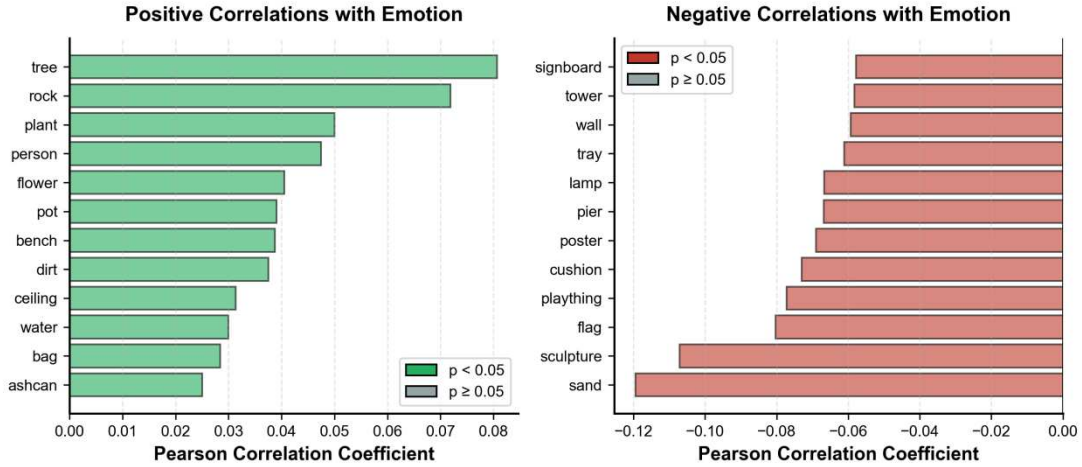


Figure 12. Pearson correlation analysis between garden visual elements and sentiment scores.

5. Conclusion

The Chinese Scholar's Garden at Snug Harbor provides a case for examining visitor records of an overseas Chinese garden through reviews and photographs. Positive sentiment accounted for 57.64% of the 942 retained reviews, and architecture, vegetation and related garden terms were the main topics in the text. The 1,395 photographs were dominated by trees, plants, sky, walls, rocks and buildings, and Plant + Architecture combinations appeared most often. In the matched image-text sample, natural labels tended to have positive correlations with review sentiment scores, whereas several artificial or auxiliary labels had negative correlations. The results are exploratory and remain bounded by the case, the selected platforms and the image-text matching procedure. The main contribution is a boundary condition: for overseas Chinese gardens, the visual and textual traces visitors leave on social platforms are not interchangeable evidence of experience.

The study covers one garden and a review set dominated by English-language material. The records do not include visitors' nationality, age, cultural background or motivation. The findings describe this case and this UGC sample, rather than a defined cultural group or a comparison between groups.

The three platforms differ in user base, posting motives and expression style, and only Google Maps supplied photographs. For that reason, the three sources are related but non-equivalent records and should not be treated as one common population. Voluntary posting also excludes visitors who left no review or photograph, which may bias the sample toward more positive responses. The 2008–2025 observation period may limit generalizability. The image-text analysis is exploratory because several photographs can share one review-level sentiment score and because the results depend on the matching procedure, sample size and available label set.

Future work could extend the time span, add reviews in other languages and examine additional overseas Chinese gardens. Multilingual models may reduce translation loss in German, French or Japanese data. Direct information on visitors' nationality, age, education and cultural background would allow more explicit group comparisons. Cross-case studies also require consistent platform definitions, review-level units and rules for matching reviews with images.

Acknowledgements

This research was funded by the 2024 National Social Science Fund Art Studies Project, "Research on the Design of Chinese Gardens Overseas" (Project No. 24BG134). The authors thank colleagues at the School of Design, Jiangnan University, for helpful discussions during the preparation of this manuscript.

Author contributions

L.Z. and F.Y. contributed equally to this work. L.Z. conceived the study and established the theoretical framework; L.Z. and F.Y. designed the research. F.Y. and P.J. collected and screened the user-generated content from Yelp, Google Maps and X (formerly Twitter). H.Y. and P.J. developed the analytical workflow and implemented the text, sentiment, semantic segmentation and regression analyses in software. H.Y. conducted validation and formal analysis and prepared the figures and tables. P.J. curated the data, conducted the image-text association analysis and wrote the original draft. L.Z., F.Y. and P.J. reviewed and edited the manuscript. L.Z. provided resources and supervision, administered the project and acquired funding. P.J. coordinated the submission and revision process as corresponding author. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The datasets analysed during the current study were derived from publicly available user-generated content (reviews and photographs) posted on Yelp, Google Maps and X (formerly Twitter), collected in accordance with the terms of service of each platform. The processed datasets supporting the findings of this study are available from the corresponding author on reasonable request. Raw third-party content is not redistributed by the authors, as it remains subject to the terms of the source platforms.

Ethical statements

This study used publicly available user-generated content (reviews and photographs) posted on Yelp, Google Maps and X (formerly Twitter). No human participants were recruited, no interventions were performed and no identifiable private information was collected; all records were publicly accessible and were anonymised before analysis. Ethical approval was therefore not required, and informed consent was not applicable.

References

1. Alarcón-Urbistondo, Pilar; Rojas-de-Gracia, María-Mercedes; Casado-Molina, Ana. Proposal for Employing User-Generated Content as a Data Source for Measuring Tourism Destination Image. Journal of Hospitality & Tourism Research 47(4), 643-664 (2023). https://doi.org/10.1177/10963480211012756
2. Albers, Patricia C.; James, William R. Travel photography. Annals of Tourism Research 15(1), 134-158 (1988). https://doi.org/10.1016/0160-7383(88)90076-x
3. Baloglu, Seyhmus; McCleary, Ken W. A model of destination image formation. Annals of Tourism Research 26(4), 868-897 (1999).
4. Beerli, Asunción; Martín, Josefa D. Tourists' characteristics and the perceived image of tourist destinations: a quantitative analysis—a case study of Lanzarote, Spain. Tourism Management 25(5), 623-636 (2004).
5. Bejoc, Marites B.; Gorro, Ken D.; Olivar, Gladies Mae C.; Campita, Rose Ann Casas. Beyond Instagram: Aspect-based sentiment analysis and cross-platform opinion mining in Philippine tourism user-generated content. Edelweiss Applied Science and Technology 9(9), 1178-1198 (2025). https://doi.org/10.55214/2576-8484.v9i9.10064
6. Breiman, Leo. Random Forests. Machine Learning 45(1), 5-32 (2001). https://doi.org/10.1023/A:1010933404324
7. Chen, Tianqi; Guestrin, Carlos. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794 (2016). https://doi.org/10.1145/2939672.2939785
8. Cheng, Bowen; Misra, Ishan; Schwing, Alexander G; Kirillov, Alexander; Girdhar, Rohit. Masked-attention mask transformer for universal image segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1290-1299 (2022).

9. Cho, Nahye; Kang, Youngok; Yoon, Jiyoung; Park, Soyeon; Kim, Jiyeon. Classifying tourists' photos and exploring tourism destination image using a deep learning model. Journal of Quality Assurance in Hospitality & Tourism 23(6), 1480-1508 (2022).
10. Collins, Brady. The view from the salad bowl: Community place attachment in multiethnic Los Angeles. Cities 94, 256-274 (2019). https://doi.org/10.1016/j.cities.2019.06.002
11. Day, Jonathon; Skidmore, Sarita; Koller, Theadora. Image selection in destination positioning: A new approach. Journal of Vacation Marketing 8(2), 177-186 (2002). https://doi.org/10.1177/135676670200800207
12. Ding, Mingjing; Zhang, Qingping; Li, Gang; Li, Weizheng; Chen, Fengyi; Wang, Yujie; Li, Qun; Qu, Ziya; Fu, Li. Fractal dimension-based analysis of rockery contour morphological characteristics for Chinese classical gardens south of the Yangtze River. Journal of Asian Architecture and Building Engineering 22(5), 2490-2505 (2023). https://doi.org/10.1080/13467581.2022.2160205
13. Drašković, Dražen; Milanović, Stefan. Aspect-based sentiment analysis of user-generated content from a microblogging platform. Journal of Big Data 12(1), 186 (2025). https://doi.org/10.1186/s40537-025-01244-0
14. Han, Ju Hyoung; Bae, So Young. Roles of authenticity and nostalgia in cultural heritage tourists' travel experience sharing behavior on social media. Asia Pacific Journal of Tourism Research 27(4), 411-427 (2022). https://doi.org/10.1080/10941665.2022.2075776
15. Hu, Tao; Geng, Juan. Research on the perception of the terrain image of the tourism destination based on multimodal user-generated content data. PeerJ Computer Science 10, e1801 (2024). https://doi.org/10.7717/peerj-cs.1801
16. Huai, Songyao; Chen, Fen; Liu, Song; Canters, Frank; Van de Voorde, Tim. Using social media photos and computer vision to assess cultural ecosystem services and landscape features in urban parks. Ecosystem Services 57, 101475 (2022). https://doi.org/10.1016/j.ecoser.2022.101475
17. Huang, Xiaoxu; Chelliah, Shankar. Attributes Influencing Tourist Satisfaction: Sentiment Analysis and Topic Modeling of Online Reviews. Journal of China Tourism Research 21(3), 819-838 (2025). https://doi.org/10.1080/19388160.2024.2440323
18. Ke, Guolin; Meng, Qi; Finley, Thomas; Wang, Taifeng; Chen, Wei; Ma, Weidong; Ye, Qiwei; Liu, Tie-Yan. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 30, 3146-3154 (2017).
19. Koruyan, Kutan; Aktaş, Erdem. Destination Image Analysis with User-Generated Content: A Computer Vision and Machine Learning Approach. Journal of Tourism and Gastronomy Studies 10(4), 2948-2964 (2022).
20. Li, Bingzhou; Yu, Yue. Network Interaction Utility of User-Generated Content and Destination Image Perception. Tourism Analysis 27(3), 343-362 (2022). https://doi.org/10.3727/108354222x16534530194813
21. Li, Changlin; Cao, Mengqi; Wen, Xiaolin; Zhu, Haotian; Liu, Shangsong; Zhang, Xinyi; Zhu, Min. MDIVis: Visual analytics of multiple destination images on tourism user generated content. Visual Informatics 6(3), 1-10 (2022). https://doi.org/10.1016/j.visinf.2022.06.001

22. Liu, Xinxin; Zhang, Ziqiong; Law, Rob; Zhang, Zili. Words meet photos: How visual content impact rating. International Journal of Hospitality Management 123, 103945 (2024). https://doi.org/10.1016/j.ijhm.2024.103945
23. Liu, Yinhan; Ott, Myle; Goyal, Naman; Du, Jingfei; Joshi, Mandar; Chen, Danqi; Lev y, Omer; Lewis, Mike; Zettlemoyer, Luke; Stoyanov, Veselin. RoBERTa: A robustly o ptimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019).
24. Macanovic, A.; Przepiorka, W. A systematic evaluation of text mining methods for sh ort texts: Mapping individuals' internal states from online posts. Behavior Research M ethods 56(4), 2782-2803 (2024). https://doi.org/10.3758/s13428-024-02381-9
25. McMullen, Melissa. Let’s see the whole picture: a multi-platform visual analysis of hi storic Mission San José on Instagram, Pinterest, Flickr, and TripAdvisor. Journal of H eritage Tourism 19(3), 365-379 (2024). https://doi.org/10.1080/1743873X.2024.2309215
26. Mehra, Payal. Unexpected surprise: Emotion analysis and aspect based sentiment analy sis (ABSA) of user generated comments to study behavioral intentions of tourists. To urism Management Perspectives 45, 101063 (2023). https://doi.org/10.1016/j.tmp.2022.1 01063
27. Nath, Atanu; Sit, Kokho (Jason); Saha, Parmita; Light, Duncan. Living the story: a d ual approach to exploring perceived authenticity and visitor motivation in heritage mu seums. Journal of Heritage Tourism 20(6), 848-871 (2025). https://doi.org/10.1080/1743 873x.2025.2585857
28. Neuendorf, Kimberly A. The content analysis guidebook. (2017).
29. Nguyen, Thi Kim Thanh; Le, Thi Hai Ha; Pham, Huy Hung. Emotional engagement as a mediator between perceived authenticity and tourist satisfaction at cultural festiva ls. Multidisciplinary Science Journal 7(11), 2025534 (2025). https://doi.org/10.31893/mu ltiscience.2025534
30. Pang, Jirui. Event image of rural football tourism in Guizhou: text mining and theme modeling of user-generated content. International Journal of Tourism Cities 12(1), 81-98 (2026). https://doi.org/10.1080/20565607.2025.2597820
31. Panigrahi, Rashmi Ranjan; Shrivastava, Avinash K. Intercultural competence, memorabl e experiences influence on Asian tourists' decisions. International Journal of Tourism Policy 14(5), 461-478 (2024).
32. Peng, Y. Zhongguo gudian yuanlin fenxi [Analysis of Chinese classical gardens]. Chin a Architecture & Building Press (1986).
33. Prokhorenkova, Liudmila; Gusev, Gleb; Vorobev, Aleksandr; Dorogush, Anna Veronika; Gulin, Andrey. CatBoost: unbiased boosting with categorical features. Advances in N eural Information Processing Systems 31 (2018).
34. Qi, Zhuoxu; Duan, Jin; Su, Hangying; Fan, Zhengxi; Lan, Wenlong. Using crowdsour cing images to assess visual quality of urban landscapes: A case study of Xiamen Isl and. Ecological Indicators 154, 110793 (2023). https://doi.org/10.1016/j.ecolind.2023.110 793
35. Qian, Lili; Guo, Juncheng; Qiu, Hanqin; Zheng, Chunhui; Ren, Lianping. Exploring d estination image of dark tourism via analyzing user generated photos: A deep learning approach. Tourism Management Perspectives 48, 101147 (2023). https://doi.org/10.101 6/j.tmp.2023.101147

36. Rinaldi, Bianca Maria. The Chinese garden: garden types for contemporary landscape architecture. (2012).
37. Salton, Gerard; Buckley, Christopher. Term-weighting approaches in automatic text retrieval. Information Processing & Management 24(5), 513-523 (1988). https://doi.org/10.1016/0306-4573(88)90021-0
38. Shen, Chenyi. Social media marketing and digital influence for visitor flow management in sustainable heritage tourism. Scientific Reports 15(1), 45767 (2025). https://doi.org/10.1038/s41598-025-28555-9
39. Tong, J. Jiangnan yuan lin zhi [A study of Jiangnan gardens] (2nd ed.). China Architecture & Building Press (1984).
40. Tunca, Sezai. Exploring visitor sentiment trends at Alanya Cleopatra Beach using natural language processing techniques: insights from online reviews. Tourism Recreation Research 51(4), 1292-1311 (2026). https://doi.org/10.1080/02508281.2025.2503992
41. Wang, Fang; Jailani, Siti Noormala Binti. Cultural Heritage Cognition and Visitor Experience Among Chinese Diaspora Visitors: Exploring the Mediating Role of Expectations. Tourism Review International 29(4), 275-290 (2025). https://doi.org/10.3727/194344225x17315216889069
42. Xiao, Xin; Fang, Chaoyang; Lin, Hui; Chen, Jingfu. A framework for quantitative analysis and differentiated marketing of tourism destination image based on visual content of photos. Tourism Management 93, 104585 (2022). https://doi.org/10.1016/j.tourman.2022.104585
43. Yu, Jia; Yokota, Takashi; Itami, Emiko; Yang, Jing. Correlations between Spatial Attributes and Visitor Stay in Chinese Gardens: A Case Study of the Ningbo Tianyige Museum Gardens. Urban Science 5(4), 74 (2021). https://doi.org/10.3390/urbansci5040074
44. Yue, Na. Assessing ritualized experiences in intangible cultural heritage tourism: a structural equation modelling methodology. GeoJournal 90(5) (2025). https://doi.org/10.1007/s10708-025-11505-3
45. Zain, Nur Adilah Md; Hanafiah, Mohd Hafiz; Asyraff, Muhammad Aliff; Ismail, Hazmal; Bafadhal, Aniesa Samira. User generated content versus mainstream media influence on hot spring tourism destination image formation. Tourism Review 80(6), 1261-1276 (2025). https://doi.org/10.1108/TR-04-2024-0250
46. Zhang, Aojia; Xiao, Honggen. Digitalising cultural heritage through online reviews: a sentiment analysis. Current Issues in Tourism (2025). https://doi.org/10.1080/13683500.2025.2566861
47. Zhang, Kun; Lin, Zhibin; Zhang, Jinyi. Tourist gaze through computer vision: Differences between Asian, North American, and European tourists. Annals of Tourism Research 88, 103039 (2021). https://doi.org/10.1016/j.annals.2020.103039
48. Zhao, Yuchen; Agyeiwaah, Elizabeth. How do Tourism Stakeholders Co-Create Destination Images with Photos on Social Media? Journal of Travel Research 64(7), 1519-1536 (2025). https://doi.org/10.1177/00472875241253006
49. Zhou, Bolei; Zhao, Hang; Puig, Xavier; Fidler, Sanja; Barriuso, Adela; Torralba, Antonio. Scene parsing through the ADE20K dataset. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 633-641 (2017).
50. Yelp. Official homepage (2026). https://www.yelp.com/ (accessed 2026-03-31).

51. Google Maps. Official homepage (2026). https://www.google.com/maps/ (accessed 2026-03-28).
52. X (formerly Twitter). Official homepage (2026). https://x.com/ (accessed 2026-03-31).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx