Understanding Public Experiences of Urban Greenspace: A Novel Data-driven Multimodal Method based on Online Review Data and Natural Language Processing

Zian Wang [1] *, Yifan Yang [2], Peter van Oosterom [3], Steffen Nijhuis [4], Stefan van der Spek [5]

[1] Faculty of Architecture and Built Environment, Delft University of Technology, Delft, 2628 BL, the Netherlands - Z.W.Wang@tudelft.nl
[2] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, 3584 CL, the Netherlands - y.yang6@uu.nl
[3] Faculty of Architecture and Built Environment, Delft University of Technology, Delft, 2628 BL, the Netherlands - P.J.M.vanOosterom@tudelft.nl
[4] Faculty of Architecture and Built Environment, Delft University of Technology, Delft, 2628 BL, the Netherlands - S.Nijhuis@tudelft.nl
[5] Faculty of Architecture and Built Environment, Delft University of Technology, Delft, 2628 BL, the Netherlands - S.C.vanderSpek@tudelft.nl

Keywords: Urban green space, Urban experience, Natural language processing, Vision-language model, Multimodal, Online review.

Abstract

Understanding public experiences in urban greenspace is essential for supporting more human-centric design and management. While traditional survey methods are often time- and labor-intensive, user-generated content (UGC) offers a rapid and scalable alternative for capturing public experiential insights. However, extracting detailed user experience information from this data remains methodologically challenging. This study proposes a novel multimodal analytical framework based on online review data and natural language processing techniques, combining LoRA fine-tuned RoBERTa language model with CLIP vision-language model to analyze multidimensional ecosystem service experience patterns in urban greenspace from user-generated text and image reviews. Results demonstrate that the proposed approach achieves more robust extraction and analysis of user experience insights compared to conventional deep learning and lexicon-based methods, exhibiting greater capacity to process contextually embedded experiential information. The multimodal framework enables more comprehensive capture of user experiences than either text or image data alone, with particular gains on dimensions that are difficult to represent through a single modality. Applying the analytical framework to Amsterdam and Rotterdam as case studies, statistical and spatial analysis reveals heterogeneity in user urban greenspace experiences and identifies key experiential bundles alongside their associated synergies and trade-offs. This study offers a novel approach to quantifying urban greenspace experiences from a user perspective, and provides insights for evidence-based urban greening practices.

1. Introduction

Urban greenspaces play an important role in cities, serving critical functions in regulating urban climate and improving environmental quality, as well as providing recreational and activity opportunities for urban residents and contributing to public health and well-being (Markevych et al., 2017). With the ongoing urbanization process and rapid urban population growth, there is an increasing public demand for urban greenery, especially high-quality urban greenspaces in cities. Understanding how people experience urban greenspace is therefore critical to assessing their performance and efficacy, thereby informing future urban greening practices that better align with public preferences and deliver well-being benefits.

Traditional assessments of urban greenspace experiences have predominantly relied on questionnaires and on-site observation, which are valuable but often time- and labor- intensive (Wang et al., 2025). With the advent of big data, every citizen increasingly acts as a “sensor”, and user-generated content (UGC) from online social media and review platforms, such as X (Twitter), Instagram, Google Maps, and TripAdvisor, has emerged as a promising alternative for capturing public insights (Ghermandi & Sinclair, 2019). This user-produced data offers rich, first-person accounts that reflect multidimensional experiences of individuals within urban environments. Such data’s broad spatial-temporal coverage, coupled with favorable accessibility, also makes it particularly well-suited for capturing large-scale user experiential information in and across cities.

However, leveraging UGC data to investigate user experience faces three key methodological challenges: the inherent limitations of UGC data, the effective extraction of meaningful user insights, and the integration of multimodal heterogeneous data:

(1) First, UGC data—particularly from commonly used social media sites—are often criticized for inherent limitations, including high noise levels, variable data quality, and potential biases in representing the target population (Goodchild & Li, 2012). Against this backdrop, online review data from platforms such as Google Maps and TripAdvisor may present several distinctive advantages. Unlike the often geographically dispersed and thematically scattered content of conventional social media, review platforms allow users to post evaluations and opinions about specific points of interest (POIs), thereby providing more spatially explicit insights. Review data have also been found to encompass more detailed descriptions of user experiences (e.g., recreation, activity) and emotional expression, rendering them better-aligned with urban and greenspace analysis tasks (Huai & Van de Voorde, 2022). Crucially, online review platforms may feature broader population representativeness than traditional social media: a recent study found that 97% of respondents used review platforms and 69% had written a review in the past year (Paget, 2025). Multiple analyses have also validated that self-

selected samples of online review data can yield assessments representative of wider populations (Greaves et al., 2012). While review data have seen growing application across diverse research domains, including understanding urban experiences, evaluating public services satisfaction, and assessing businesses' electronic word-of-mouth (E-WoM), their adoption in urban greenspace research remains limited (Xiang et al., 2017).

(2) Second, the unstructured nature of UGC data poses substantial challenges for extracting detailed and targeted user experience insights compared to conventional survey-based approaches. Earlier UGC-based urban greenspace research predominantly applied sentiment analysis to quantify the emotional polarity of textual data and associated it with the objective physical characteristics of greenspace. However, reducing human experience to a single affective metric oversimplifies human–environment interactions and overlooks actual behavioral and experiential dimensions in urban greenspace. In response, a growing body of research adopts advanced natural language processing (NLP) techniques to extract nuanced subjective insights, following two primary approaches. The Lexicon approach uses predefined lexicons to identify specific experiential elements by computing keyword frequencies, thereby classifying users' different experience patterns. This approach, however, may overlook contextual meaning, syntactic structure, and word relationships, and remains susceptible to data noise. The second and increasingly prominent approach employs deep learning methods, notably Transformer-based models, to develop task-specific models tailored to the analytical context. These models learn semantic representations directly from data, allowing them to capture contextual meaning and nuanced associations between terms beyond predefined lexicons. While deep learning approaches have demonstrated considerable flexibility and utility in extracting subjective insights, their adoption in urban greenspace research remains scarce, and effective models or datasets capable of evaluating the multiple dimensions of user experience in greenspace settings are still lacking.

(3) Lastly, UGC data typically encompasses both textual and visual content, but current research predominantly focuses on a single modality, particularly texts, overlooking the rich experiential insights embedded in user-generated images. Studies suggest that users tend to share images and text together to express emotions and perspectives (Zhu et al., 2023). The act of photographing reflects a subjective, perceptually grounded engagement with the environment, and visual records are therefore considered to hold distinct advantages in capturing actual, uninfluenced experiences of landscape, encompassing dimensions that text alone cannot adequately convey (Dunkel, 2015). Still, the extraction of meaningful information from images and its subsequent integration with textual data remains methodologically challenging. The typical approach has relied on computer vision models, such as scene recognition and semantic segmentation, to identify visual elements in images, but these models are constrained by predefined label sets and cannot flexibly adapt to diverse analytical tasks. The recent advent of vision-language models, exemplified by the CLIP (Contrastive Language-Image Pre-training) model, presents new possibilities: by mapping image and text encoders into a shared semantic space, these models support the analysis and transformation of visual content into semantic concepts without depending on fixed category systems. Their zero-shot classification capacity makes them especially suited to the recognition of complex, subjective experiential qualities, offering a technical foundation for more flexible and multimodal analytical configuration in research.

Addressing these methodological challenges, this study proposes a novel analytic framework combining online review data with multimodal NLP techniques to extract in-depth user greenspace experience insights from UGC. The research question is: *how can multimodal NLP be applied to extract user urban greenspace experiences from online review data*? The framework integrates Transformer-based model and vision-language model to jointly process textual and visual reviews and, by developing targeted datasets corresponding to our analytic tasks, extracts detailed user experience insights across multiple experiential dimensions. The framework is applied, validated, and interpreted using Amsterdam and Rotterdam, the Netherlands, as empirical cases.

2. Study Area and Data

As key cities of the Dutch metropolitan region, Amsterdam and Rotterdam are the most populous cities in the Netherlands, with approximately 942,000 and 674,000 residents, respectively. Both cities are characterized by rich urban green-blue infrastructure and have actively pursued green strategies aimed at fostering more sustainable and liveable urban environments, making them well-suited as case study sites.

Google Maps reviews were selected as the UGC data source for this study. Google Maps is among the most widely used online review platforms, reported to cover 83% of market share that substantially exceeded alternative platforms, with its user base continuing to grow (Paget, 2025). Google Maps reviews have been recognized as a valuable resource for capturing public insights and supporting urban analysis (Zhao et al., 2025).

This study adheres to the ethical principles and ensures that all data collection and use comply with Google's fair use principle for non-commercial purposes, the EU Directive on Copyright, and the Copyright Law of the Netherlands (Auteurswet). Data anonymization and aggregation were applied throughout to protect privacy.

Data collection was conducted between November and December 2025. To identify urban greenspace-relevant locations within the study cities, we performed POI searches using a set of urban greenspace-related category labels, as listed in Table 1. The resulting POI lists were manually reviewed and cross-referenced with urban land use data to retain only actual urban greenspace sites with valid reviews, yielding a final set of 248 POI sites. These sites were subsequently compared against the official greenspace inventories of both cities, revealing coverage of 45 out of 48 officially listed urban greenspaces. The absent sites were caused by a lack of corresponding POI tags on the review platform.

Table 1. POI tags applied to filter valid urban greenspace sites in the study cities

POI tag
“amusement_park”, “beach”, “botanical_garden”, “city_park”, “cultural_landmark”, “cycling_park”, “dog_park”, “farm”, “fountain ”, “garden”, “hiking_area”, “historical_landmark”, “historical_place”, “lake”, “monument”, “national_park”, “natural_feature”, “nature_preserve”, “off_roading_area”, “park”, “picnic_ground”, “planetarium”, “plaza”, “scenic_spot”, “skateboard_park”, “state_park”, “tourist_attraction”, “town_square”, “water_park”, “wildlife_park”, “woods”

Subsequently, Google Maps reviews published between 2016 and 2025 for these sites were collected with Python. Each review contains an anonymized ID, review text (if present), review image (if present), star rating, and upload timestamp (see Table 2 for examples). As Google Maps reports review timestamps in relative formats (e.g., "1 week ago"), exact dates were transformed by back-calculating from the collection date. Reviews from multiple POI entries corresponding to the same site were merged. In total, 78,491 reviews were collected.

Table 2. Example of collected urban greenspace review

Review ID	Vondelpark-13154
Rating	5/5
Text	Beautiful park to walk, to run, or to people watch. Spread out a blanket and have lunch, read a book, or stroll through the park and see sculpture works by Picasso. You can see a lot of birds, even green parrots.
Image(s)	 
Time stamp	2025-10-02

Content was transcribed for privacy protection.

During data filtering, only reviews with valid text and/or image content were retained as valid entries given our analytic tasks, and reviews consisting solely of star ratings were therefore excluded. To ensure data robustness, only urban greenspace sites with five or more valid reviews were included. As the collected reviews cover multiple languages, all reviews were translated into English using the Google Translate API; prior studies have confirmed that translated reviews retain validity comparable to original-language content (Gelashvili et al., 2024). After filtering, 75,081 reviews from 228 urban greenspace sites were retained.

3. Methods

3.1 Analytic Framework

The analytical framework is presented in Figure 1. An ecosystem services framework was employed to conceptualize human–environment interaction in urban greenspace and structure the experiential dimensions under investigation. Seven dimensions were incorporated: aesthetic appreciation, recreation activity, outdoor exercise, social interaction, culture and education, nature connection, and biodiversity observation. Table 3 presents the experiential framework and corresponding review examples.

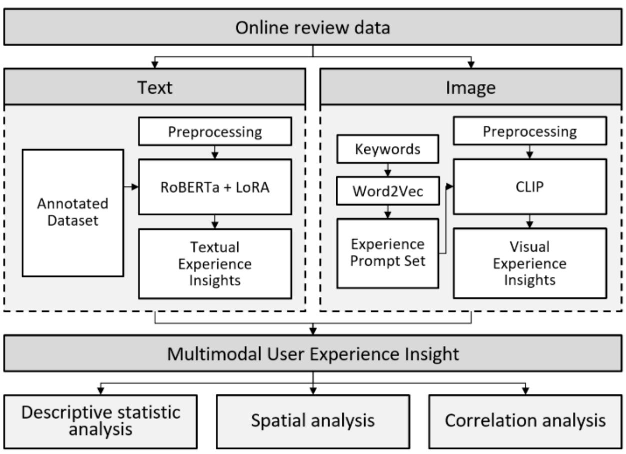


Figure 1. Analytical framework

3.2 Data Annotation and Labeling

To develop a labelled dataset, 2,500 reviews were randomly sampled and manually annotated with respect to the seven experience dimensions. Two researchers with backgrounds in environmental psychology and urban studies independently annotated each review against each dimension, with the review assigned 1 if it involved and expressed a positive experience relevant to the corresponding dimension, or 0 otherwise. Given that reviews mentioning ecosystem service experiences in negative terms were found to be very rare, a binary rather than three-way labelling structure was adopted to ensure training performance. Annotation followed a multi-label setup in which each dimension was evaluated independently. Disagreements were discussed until a consensus was reached. Annotated examples are provided in Table 3.

Table 3. Included experience dimensions, corresponding criteria, and examples

Dimension	Criteria	Example
Aesthetic appreciation	Appreciating the beauty or aesthetic value in various aspects of the environment	"It is a beautiful place, very peaceful."
Recreation activity	Spending leisure time engaging in static or dynamic outdoor activities	"It is relaxing and has many recreational facilities. It is also a nice place to picnic."
Outdoor exercise	Engaging in light, moderate, or vigorous physical activity	"It is good to either cycle or walk."
Social interaction	Engaging in mutual contact, influence, and interaction between individuals	"Great place to chill and meet friends or go with family."
Culture and education	Engaging in learning, researching, cultural expression, or heritage-related experiences	"A great place to learn a little about the country's history."
Nature connection	Experiencing emotional, cognitive, or experiential value with contact with nature	"Nice flora of flowers and trees. Great places to enjoy nature in peace."
Biodiversity observation	Experiencing biodiversity through e.g., observing the presence, diversity, or ecological value of living organisms	"Full of animals and wide paths. It is nice to watch ducks building their habitat."

Content was transcribed for privacy protection.

3.3 Text-based Experience Analysis Employing LoRA Fine-tuned RoBERTa Model

The annotated dataset was used to develop text-based urban greenspace experience classification models built upon the pretrained language model RoBERTa (Robustly Optimized BERT Approach). Because of its deep bidirectional encoder architecture, RoBERTa better captures semantic features from review text than conventional deep learning approaches (Liu et al., 2019). However, pretrained models may underperform on specific tasks such as urban experience recognition, and full fine-tuning of the complete model entails substantial computational cost. We therefore adopted LoRA (Low-Rank Adaptation), a parameter-efficient fine-tuning method that introduces low-rank decomposition modules into the weight matrices of the self-attention mechanism, enabling effective training and rapid transfer by updating only a small subset of parameters (Hu et al., 2021).

To accommodate the distinct semantic characteristics of each experience dimension, an independent binary classification model was trained for every dimension to improve recognition precision and sensitivity for specific experience types. The dataset was randomly split into training, validation, and test sets at a 7:2:1 ratio. All review texts underwent standardized preprocessing prior to model input, including special character removal, tokenization, and sequence normalization to a uniform length of 128 tokens via truncation or padding. Model training employed cross-entropy loss to measure the divergence between predicted probability distributions and ground-truth labels. Optimization was performed using the AdamW optimizer with a learning rate of 4e-5, combined with a linear warm-up followed by a linear decay scheduling strategy, whereby the learning rate was gradually increased during the initial training phase before being linearly reduced, thereby enhancing training stability.

In the implementation of LoRA fine-tuning, low-rank decomposition updates were applied exclusively to the Query and Value weight matrices within the self-attention layers of RoBERTa, while all remaining pretrained parameters were kept frozen. A dropout mechanism with a rate of 0.1 was additionally introduced in the classifier layer to mitigate overfitting and improve model generalizability. Training was conducted for a maximum of 20 epochs, with early stopping applied to terminate training if no improvement in validation loss was observed over three consecutive epochs to prevent overfitting.

Following training, model performance was evaluated with the independent test set. To contextualize the performance of the proposed approach, a series of comparative evaluations was conducted incorporating methods commonly employed in previous research: a lexicon-based method, a Long Short-Term Memory network (LSTM), and XGBoost. All baseline methods were applied to the same annotated dataset to ensure consistent and comparable evaluation.

3.4 Image-based Experience Analysis Employing CLIP Model

To extract the experiential information embedded in user-generated images, a multimodal experience recognition module based on CLIP was introduced into the analytical framework. CLIP is a contrastive vision-language model pretrained on large-scale image-text paired dataset, enabling the alignment of visual and textual representations within a shared semantic space. For the purposes of our task, a set of semantic concept prompts oriented towards urban greenspace experiences was first constructed. Drawing on the definitions of ecosystem service experience dimensions adopted in our study and the representative vocabulary identified in the annotated dataset, researchers developed a base prompt list of phrases for each experience dimension, encompassing experiential expressions with potential visual representation, including common activity types, landscape features, and use scenarios. Word2Vec was subsequently applied to the collected reviews to identify semantically related, high-frequency terms associated with the base list, thereby expanding the prompt sets. The expanded prompts were then subject to manual screening to retain only elements that were both conceptually distinct and visually relevant, with the aim of improving the accuracy and specificity of CLIP-based matching.

During the experience recognition stage, images and the textual prompt set associated with each experience dimension were jointly assessed by the CLIP model, enabling the computation of similarity between image features and each set of experience-related descriptions. An experience dimension was considered present in an image when the corresponding similarity exceeded a predefined threshold of 0.30. Consistent with the text-based approach, a multi-label setup was adopted in which each experience dimension was assessed independently. As a complementary measure, the BLIP model was employed to automatically generate natural language captions for each image, capturing the primary subjects and scene-level information to provide additional semantic support for experience recognition.

For reviews containing both text and images, data fusion was performed following a union-based decision rule, whereby an experience dimension was considered present if it was detected in either modality. This multimodal fusion strategy enables the complementary analysis of user-generated textual and visual content, and mitigates the risk of underrepresenting experiential information that may be difficult to convey through a single modality alone. Comparative analyses using text-only and image-only inputs were additionally conducted to benchmark the performance gains related to multimodal integration.

3.5 Statistical and Spatial Analysis

The performance of the developed analytical framework was first evaluated using four standard metrics: accuracy, precision, recall, and F1 score. These are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

$$Precision = \frac{TP}{TP+FP}, \quad (2)$$

$$Recall = \frac{TP}{TP+FN}, \quad (3)$$

$$F1 = \frac{2\times Precision \times Recall}{Precision+Recall}, \quad (4)$$

where TP = true positives
TN = true negatives
FP = false positives
TN = false negatives

Second, to compare the coverage of the multimodal approach against unimodal alternatives, an Identification Coverage Rate was defined as the proportion of reviews in which the selected experience dimension was identified under a given modality:

$$C_{m,k} = \frac{N_{m,k}}{N_{all}}, \quad (5)$$

where $C_{m,k}$ = identification coverage rate for dimension k under modality m
$N_{m,k}$ = number of posts in which experience dimension k is identified under modality m
N_{all} = total number of reviews

Subsequently, the proposed analytical framework was applied to the collected review data in the study cities, and user experience outcomes of the included urban greenspace sites were examined through statistical and spatial analysis. Descriptive statistics were computed for each identified experience dimension, and their spatial distributions across sites were explored through qualitative spatial analysis. Pearson correlation analysis was further applied to examine the relationships among experience dimensions and to identify co-occurring experience bundles.

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-661-2026

4. Results and Discussion

4.1 Characteristics of Collected Online Review Data

Our collected final dataset comprised 75,081 reviews from 228 urban greenspace sites across the two study cities, of which 58,828 reviews from 136 sites were located in Amsterdam and 16,253 reviews from 92 sites in Rotterdam. In terms of content composition, 76.6% of reviews contained text only, 18.9% contained both text and images, and 4.5% contained images only (Figure 2).

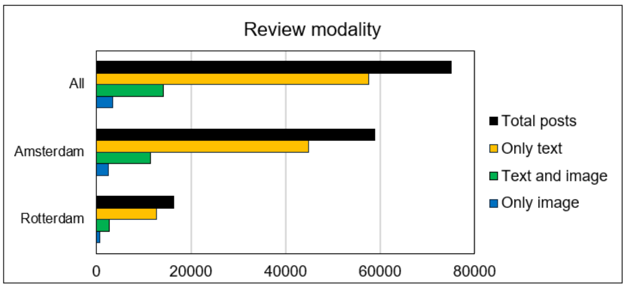


Figure 2. Distribution of review modalities

The review structural characteristics were broadly consistent across both cities (Table 4). The average review length was 78.7 characters or 14.0 words, indicating that the review is predominantly composed of short texts, and the analytical task may thus be characterized as short-text classification. Reviews containing images included an average of 2.4 images per post.

Table 4. Structural characteristics of the review dataset

	Mean Character count	Mean word count	Mean image count per post with image
All	78.69	14.00	2.44
Amsterdam	77.58	13.80	2.41
Rotterdam	82.93	14.73	2.57

With respect to temporal distribution, the annual volume of posted reviews increased substantially in the early years of the observation period, reaching its peak in 2019. Subsequently, the number of new reviews per year stabilized at a broadly consistent level, a trend likely attributable in part to the disruptions associated with the COVID-19 pandemic (Figure 3).

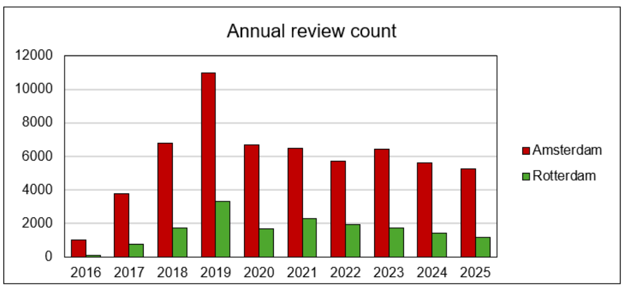


Figure 3. Temporal distribution of review data

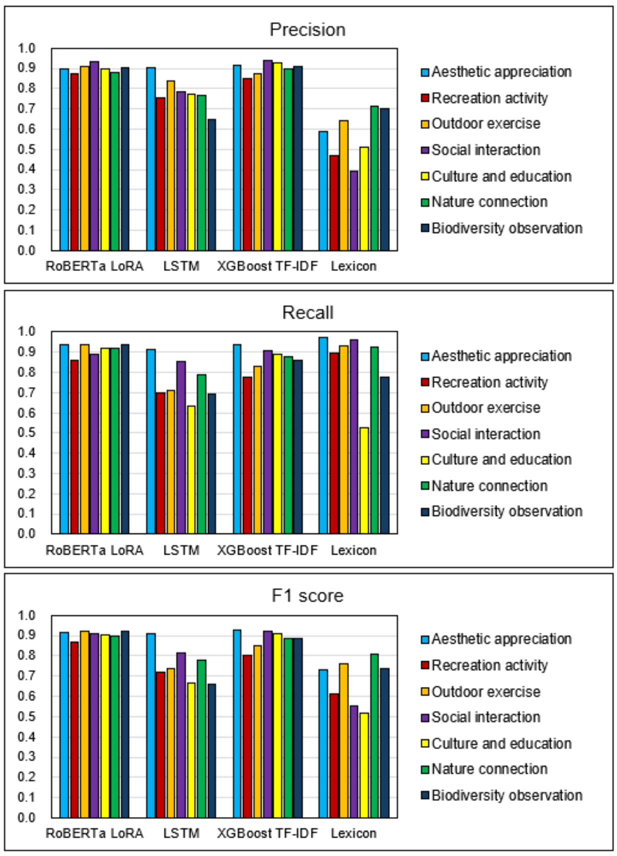
4.2 Performance Evaluation of the Analytical Framework

The overall model evaluation results are presented in Table 5. The proposed LoRA fine-tuned RoBERTa model demonstrates strong performance, achieving a precision of 0.90, recall of 0.91, F1 score of 0.91, and accuracy of 0.94. We consider Recall of particular importance for our analytical task, as many experiential expressions are implicit and would be overlooked by methods relying on surface-level lexical features. In this respect, the proposed model outperforms other widely used deep learning and lexicon-based methods by margins ranging from 5.2% to 20.7%, reflecting superior contextual awareness and better-balanced overall performance.

Table 5. Performance comparison between the proposed method and baseline approaches

	Precision	Recall	F1 Score	Accuracy
RoBERTa LoRA	0.90	0.91	0.91	0.94
LSTM	0.78	0.76	0.75	0.85
XGBoost TF-IDF	0.90	0.87	0.88	0.92
Lexicon	0.57	0.86	0.67	0.83

Detailed comparison results for the proposed analytical framework and baseline methods across all seven experiential dimensions are presented in Figure 4. The LSTM method performed weakly across all four metrics, likely attributable to its comparatively limited capacity for modelling long-range dependencies, which constrains its ability to capture complex semantic relationships relative to Transformer-based architectures. The XGBoost TF-IDF method achieved relatively high precision but underperformed compared to the proposed framework on other metrics, most notably recall, suggesting that its classification is more conservative and that it may miss more nuanced or implicit experiential expressions. The lexicon-based method exhibited the weakest overall performance, with the lowest precision among all methods, indicating a notable tendency towards false positive classifications potentially arising from the absence of contextual understanding. Comparison of model performance across the seven experience dimensions also reveals that while the proposed RoBERTa model achieved more robust results, alternative methods may achieve comparable performance on specific dimensions, particularly ones that are more lexically explicit, such as aesthetic appreciation that may be associated with more identifiable marker terms.



https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-661-2026

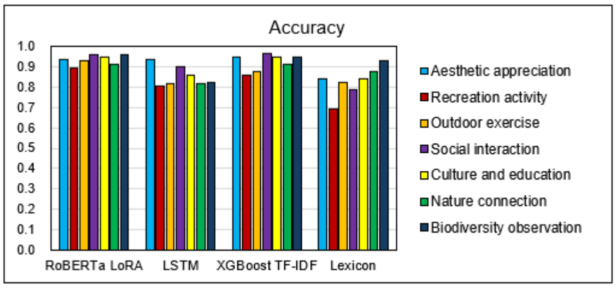


Figure 4. Performance comparison across experiential dimensions

4.3 Performance Evaluation of Multimodal Approach

The identification coverage rate was used to quantify and compare the performance of the adopted multimodal analytical framework and text-only and image-only approaches (Figure 5). The multimodal approach relative to the text-only approach improved coverage across all seven urban greenspace experience dimensions by at least 3.1%, with particularly notable gains exceeding 20% for biodiversity observation experience and culture and education experience. Compared to the image-only approach, the multimodal approach achieved substantially higher coverage across all dimensions, with improvements of at least 159.8%. This may be partially explained by the relatively modest proportion of reviews containing images in the dataset.

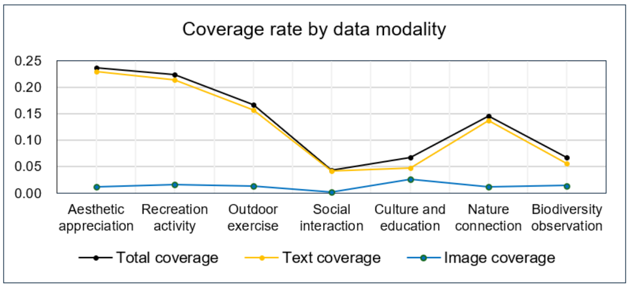


Figure 5. Identification coverage rate of text-only, image-only, and multi-modal approaches across experience dimensions

Inter-modal agreement between the image and text modalities was further examined using the subset of reviews containing both modality types (Table 6). Results indicate that, for most experience dimensions, agreement between modalities was relatively high but not identical, suggesting that the two modalities capture partly distinct information. Notably, agreement fell below 75% for aesthetic appreciation experience and recreation activity experience, implying that image and text data may capture related but meaningfully different insights for these dimensions, therefore holding considerable complementary value. An exception was social interaction experience, for which inter-modal agreement was markedly higher, potentially indicating that a single modality may already offer relatively comprehensive coverage of socialization-related experiences. This finding warrants further empirical validation in future research.

Table 6. Agreement between image and text modalities

	Fusion gain over text	Fusion gain over image	Agreement between image and text
Aesthetic appreciation	3.3%	1961.0%	74.3%
Recreation activity	4.5%	1254.5%	74.9%
Outdoor exercise	6.1%	1154.0%	80.5%
Social interaction	3.1%	2988.2%	95.1%
Culture and education	43.0%	159.8%	85.6%
Nature connection	6.7%	1062.6%	80.8%
Biodiversity observation	20.0%	377.4%	88.4%

4.4 Statistical and Spatial Analysis of User Greenspace Experience

Applying the developed multimodal NLP analytical framework to the full review dataset, we quantified user experience profiles across the urban greenspace sites in both study cities. Descriptive statistics are reported in Table 7. Aesthetic appreciation emerged as the most frequently reported experience dimension, followed by outdoor exercise and nature connection, each reported by approximately 15–20% of users. By contrast, culture and education experience and social interaction experience were the least commonly reported dimensions.

Table 7. User urban greenspace experience conditions across seven dimensions in the study cities

	min	mean	max	SD
Aesthetic appreciation	0	0.21	0.71	0.16
Recreation activity	0	0.14	0.5	0.11
Outdoor exercise	0	0.19	0.8	0.14
Social interaction	0	0.03	0.31	0.05
Culture and education	0	0.04	0.5	0.09
Nature connection	0	0.17	0.8	0.15
Biodiversity observation	0	0.06	0.6	0.09

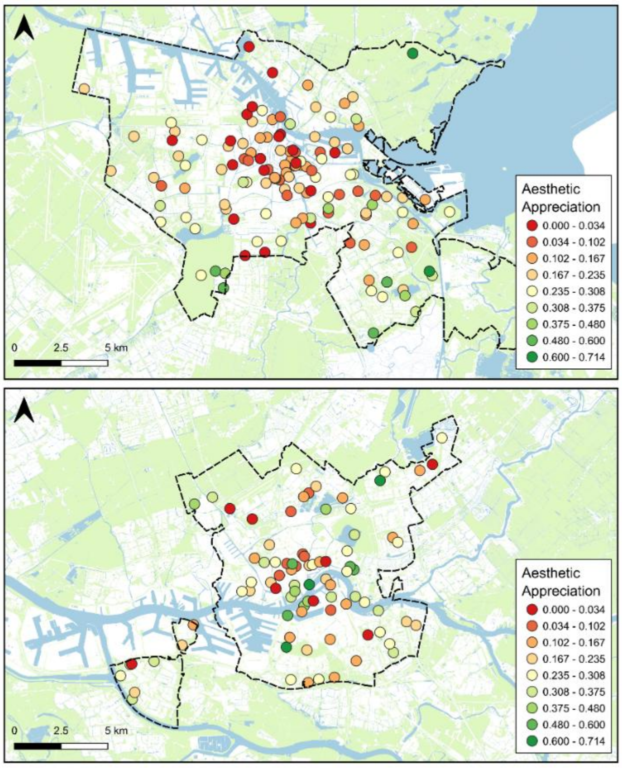


Figure 6. Spatial patterns analysis of user aesthetic appreciation experience

Qualitative spatial analysis of representative experience dimensions reveals distinct geographical patterns. For aesthetic appreciation (Figure 6), urban greenspace located in city centers generally attracted lower proportions of user-reported aesthetic experiences, whereas peripheral greenspaces received comparatively stronger aesthetic appreciation. Rotterdam's urban greenspace exhibited comparatively higher levels of aesthetic appreciation overall compared to Amsterdam, and the spatial distribution of high-aesthetic greenspace sites was more even across the city.

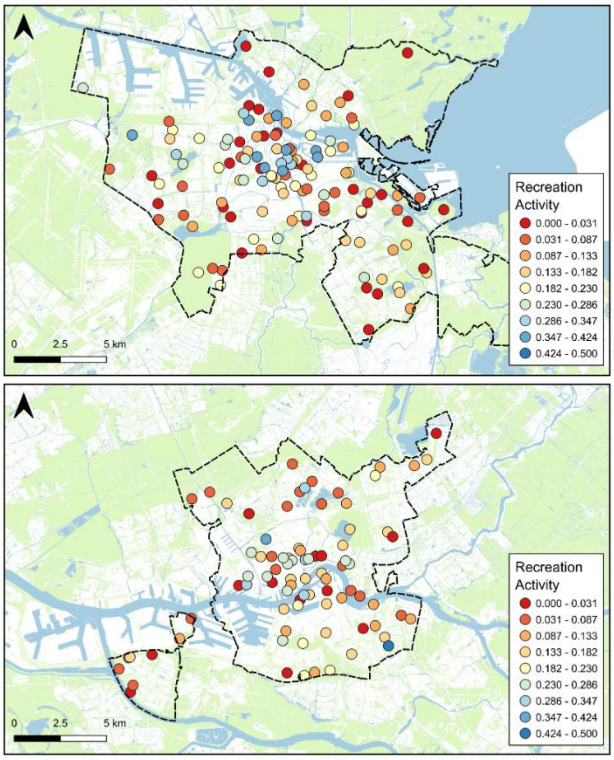


Figure 7. Spatial patterns analysis of user recreation activity experience

In comparison, recreation activity (Figure 7) displayed a different spatial pattern. Urban greenspaces in central urban areas were associated with higher proportions of reported recreational experience, particularly in central Amsterdam, while greenspaces in peripheral areas generally recorded lower recreation experience, though with exceptions.

Spatial distribution of outdoor exercise experience (Figure 8) was less continuous and more localized, with considerable variation both in central and peripheral areas. This pattern likely reflects the impact of site-specific facilities and infrastructure factors supporting exercise-related experiences in urban greenspace.

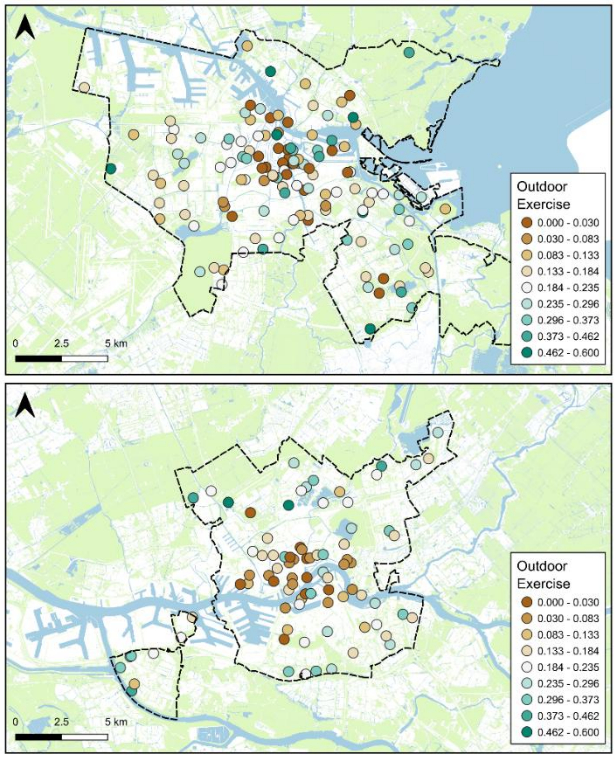


Figure 8. Spatial patterns analysis of user outdoor exercise experience

Pearson correlation analysis (Figure 9) was applied to examine the interrelationship among the seven greenspace experience dimensions, and to identify experience bundle patterns. Statistically significant moderate positive correlations were

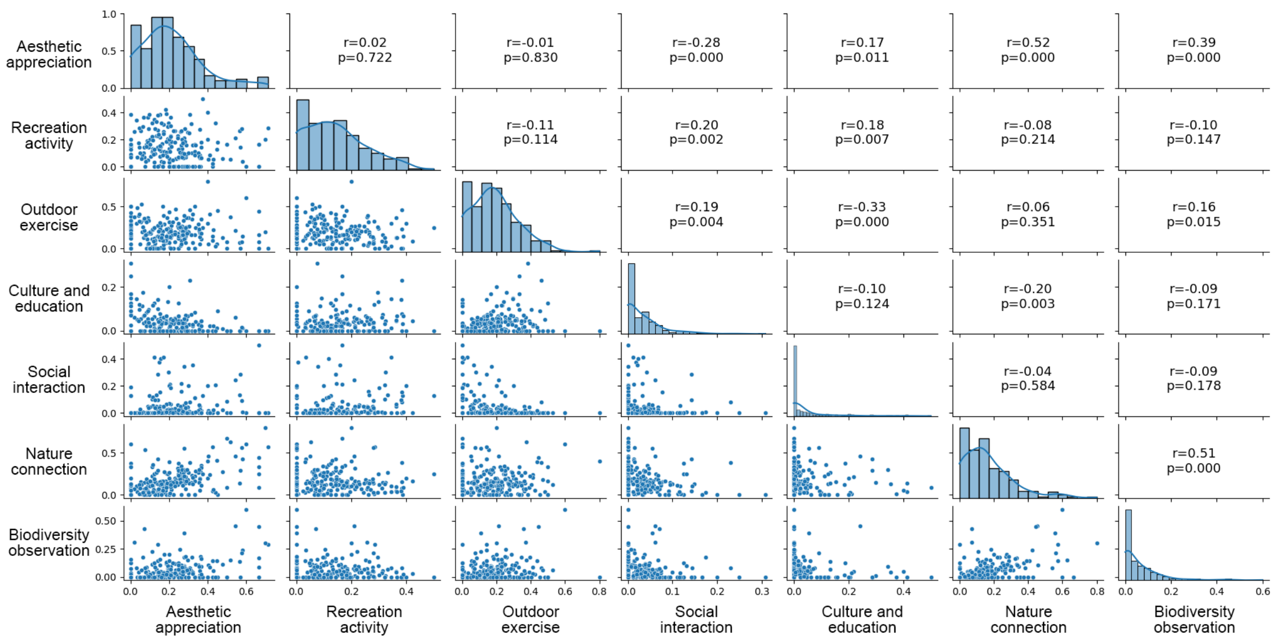


Figure 9. Pearson correlation analysis of user experience dimensions

found among aesthetic appreciation, nature connection, and biodiversity observation experiences, suggesting the existence of a coherent nature-oriented experience bundle in urban greenspace in which these dimensions tend to co-occur. Recreation activity, social interaction, culture and education, and outdoor exercise experiences also exhibited statistically significant weak positive correlations, indicating an activity- and use-oriented experience bundle. Analysis also revealed important experience trade-offs. Negative correlations were identified between social interaction and both aesthetic appreciation and nature connection experiences, suggesting that socialization-oriented and nature-oriented experiences may exert mutually inhibiting effects, which warrants further empirical investigation. Additionally, negative correlation between outdoor exercise and culture and education experiences exists, which points to a potential divergence between active-use and cultural-use patterns in greenspace.

5. Conclusion

Understanding user experiences in urban greenspace is important for supporting evidence-based urban greenery planning and management. Traditional survey methods are often resource-intensive, while UGC-based data-driven approaches, while offering promising alternative approaches, are also subject to challenges relating to data limitations, data analysis difficulties, and multimodal data integration. In response to these methodological constraints, the current study proposed a novel approach that draws on online review data and multimodal NLP methods to extract multidimensional urban greenspace experiential insights from user-generated text and image content.

Validation results demonstrate that the proposed LoRA fine-tuned RoBERTa model achieves more robust extraction and analysis of user insights compared to conventional deep learning and lexicon-based methods, exhibiting greater capacity to process the contextual information of user experiences in greenspace embedded in online review data. When combined with the CLIP vision-language model in a multimodal approach, the framework captures a more complete picture of user experiences than either modality alone, particularly supplementing text-only analysis with meaningful gains in the coverage of user's biodiversity observation and culture and education experience dimensions in greenspace.

Employing Amsterdam and Rotterdam as case studies, the analysis revealed statistical and spatial heterogeneity in user urban greenspace experiences and identified key experiential bundles, highlighting meaningful synergies and trade-offs across experience dimensions. These findings underscore the complexity of user experience in urban greenspace and carry positive implications for informing more responsive and human-centric urban greenery planning and management. In addition to practical insights, this study develops and validates a generalizable methodological framework for extracting detailed, multidimensional insights from multimodal online review data, offering a practical analytical tool for future planners and policymakers.

Several limitations of the study should be acknowledged. Our dataset captures only users who visited urban greenspace and posted reviews, limiting its representativeness of the general public. The POI-based urban greenspace selection may not cover all green space types. Additionally, despite the use of advanced multimodal methods, the inherent subjectivity of individual experiences introduces unavoidable analytical limitations. Combining and triangulating UGC data-based approaches with traditional survey methods represents a promising avenue for future research to address these constraints.

References

Dunkel, A., 2015: Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, *142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

Gelashvili, V., Martínez-Navalón, J. G., DeMatos, N., & de Brito Correia, M., 2024: Technological transformation: The importance of E-WOM and perceived privacy in the context of opinion platforms. *Technological Forecasting and Social Change*, *205*, 123472. https://doi.org/10.1016/j.techfore.2024.123472

Ghermandi, A., & Sinclair, M., 2019: Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*, *55*, 36–47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

Goodchild, M. F., & Li, L., 2012: Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

Greaves, F., Pape, U. J., King, D., Darzi, A., Majeed, A., Wachter, R. M., & Millett, C., 2012: Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: An observational study. *BMJ Quality & Safety*, *21*(7), 600–605. https://doi.org/10.1136/bmjqs-2012-000906

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W., 2021: *LoRA: Low-Rank Adaptation of Large Language Models* (arXiv:2106.09685). arXiv. https://doi.org/10.48550/arXiv.2106.09685

Huai, S., & Van de Voorde, T., 2022: Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. *Landscape and Urban Planning*, *218*, 104307. https://doi.org/10.1016/j.landurbplan.2021.104307

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V., 2019: *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M. J., Lupp, G., Richardson, E. A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M., Heinrich, J., & Fuertes, E., 2017: Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environmental Research*, *158*, 301–317. https://doi.org/10.1016/j.envres.2017.06.028

Paget, S., 2025: Local Consumer Review Survey. *BrightLocal*. https://www.brightlocal.com/research/local-consumer-review-survey-2025/

Wang, Z., Yang, Y., Nijhuis, S., & van der Spek, S., 2025: Understanding human-environment interaction in urban spaces with emerging data-driven approach: A systematic review of methods and evidence. *Cities*, *167*, 106346. https://doi.org/10.1016/j.cities.2025.106346

Xiang, Z., Du, Q., Ma, Y., & Fan, W., 2017: A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Zhao, X., Huang, H., Lin, G., & Lu, Y., 2025: Exploring temporal and spatial patterns and nonlinear driving mechanism of park perceptions: A multi-source big data study. *Sustainable Cities and Society*, *119*, 106083. https://doi.org/10.1016/j.scs.2024.106083

Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., & Qian, J., 2023: Multimodal Sentiment Analysis With Image-Text Interaction Network. *IEEE Transactions on Multimedia*, *25*, 3375–3385. https://doi.org/10.1109/TMM.2022.3160060