Public emotions and visual perception of historic district : An AI approach using social media data

Chunye Ma, Lan Wang
Affiliation:Tongji university
Email:mcy0503@126.com

Abstract
Historic districts possess unique visual landscape features that can influence public emotional experiences. Advances in artificial intelligence, particularly multimodal large models, enable integrated analysis of image and text data, offering new possibilities for understanding the complex relationship between visual environments and emotions. This study focuses on Nanluoguxiang, a historic district in Beijing, using 2,851 text reviews and 13,553 associated images collected from the social media platform Dianping between January 2024 and January 2025. An AI-based emotion analysis framework was developed, and XGBoost was applied to examine the associations between seven types of visual scenes and emotional responses. Results indicate that joy dominates public sentiment; commercial scenes are highly perceived; and architectural and streetscape elements are linked to positive emotions, while storefronts correlate with negative emotions.

Keywords: historic districts: public emotions: visual perception

1. Introduction
Historic districts can offer positive emotional experiences to the public, enhancing well-being, place attachment, and psychological restoration (Bornioli et al., 2018; Scopelliti et al., 2019; Reece et al., 2022). Among the various factors that influence such experiences, visual perception plays a particularly important role, as it represents one of the most immediate and intuitive ways through which individuals engage with urban heritage environments. Therefore, the visual landscape characteristics of historic districts—such as architectural forms, streetscapes, commercial elements, and natural features—may significantly shape how people feel in these settings.

However, existing research often treats emotional responses and visual landscape features as independent topics, lacking a unified framework that links the two. In particular, while numerous studies have examined either the aesthetic value of urban heritage or the restorative potential of urban spaces, the specific influence of visual landscapes in historic districts on public emotional responses remains underexplored.

Recent advances in Artificial Intelligence (AI), especially in the domains of computer vision, natural language processing, and multimodal learning, provide new opportunities to examine this relationship. AI techniques are capable of extracting detailed visual features from images and detecting emotional content in text, making it possible to systematically analyze large-scale user-generated data. Accordingly, this study proposes to construct an AI-based analytical framework to explore the relationship between visual landscape characteristics and public emotional experiences in historic districts. By leveraging multimodal social media data and interpretable machine learning models, this research aims to bridge the gap between environmental perception and emotional response in the context of heritage planning.

2. Case study
2.1 Study area
Nanluoguxiang is a renowned historic district in Beijing, China, known for its well-preserved siheyuan (courtyard residences) dating back to the Ming and Qing dynasties. It holds a prominent position in the field of world heritage studies. Following multiple rounds of conservation and renewal efforts, Nanluoguxiang now features a mix of commercial, residential, and cultural functions. Whether the current spatial form and its visual perception are beneficial to the public's emotional experience is an important issue for policymakers to consider.

2.2 Social media data on NLG
The review data for Nanluoguxiang on the Dianping platform is relatively high in both quality and quantity. As a well-known tourist destination in China, Nanluoguxiang has accumulated a large volume of social media reviews across various platforms, including Xiaohongshu, Weibo, Dianping, and Ctrip. However, platforms such as Xiaohongshu and Weibo contain a significant number of promotional accounts and non-authentic user comments, which may not accurately reflect the public's genuine emotional responses, and were therefore excluded from this study. Although Ctrip is one of China's major travel platforms, the volume of available data on

Nanluoguxiang is comparatively limited. Considering factors such as data volume, the proportion of authentic users, and content quality, Dianping was selected as the primary data source for this research.

This study collected all publicly available reviews of Nanluoguxiang from Dianping between January 1, 2024 and December 31, 2024, including text reviews, associated images, and timestamps. A total of 2,818 text reviews and 13,986 images were initially gathered. After data cleaning, the final dataset comprised 2,460 valid text reviews and 12,770 images. The average length of a review was 106.8 characters (standard deviation: 86.7), with a median length of 105 characters. Notably, approximately 25% of the reviews exceeded 126 characters, indicating that users were generally willing to provide detailed narratives, thus offering a strong semantic foundation for emotion analysis.

3. Methodology

3.1 Overview

Text reviews from social media help understand people's emotional experiences with the built environment (Ghahramani et al., 2021; Park et al., 2020), and images voluntarily shared on social media serve as a rich resource for understanding people's visual preferences for environments (Chen et al., 2020). Although some studies have attempted to qualitatively explore the relationship between visual perception and emotional experience (Yang & Zhang, 2024), no quantitative analysis framework for this relationship has been established. In this study, we use Nanluoguxiang, a typical historic district in Beijing, as a case study. This study develops a multimodal deep learning framework to examine the impact of 6 environmental scenes in historic districts on public emotional experiences.

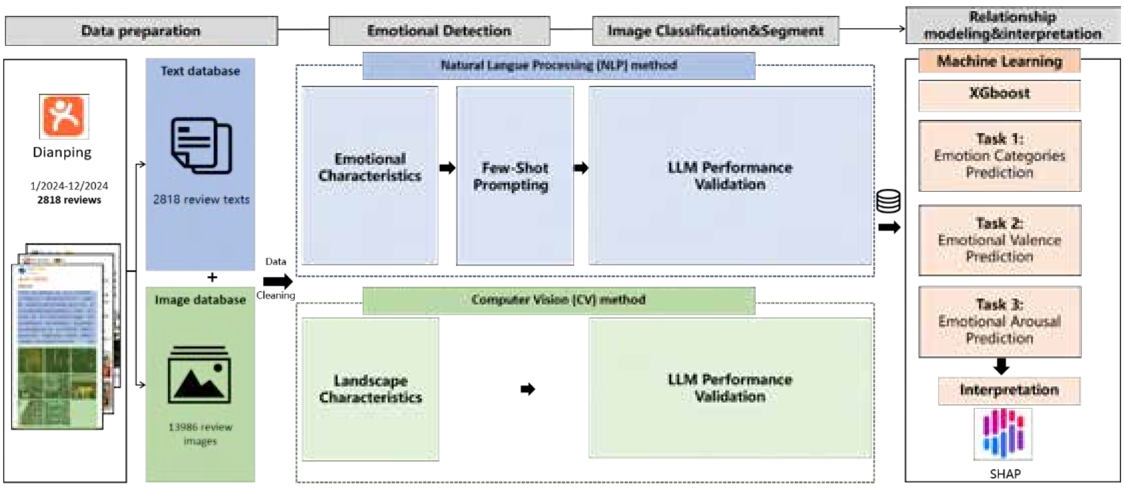

Fig. 1 Research Overview

3.2 Emotion Analysis Method

Aspect-Based Sentiment Analysis (ABSA) aims to identify specific spatial or managerial elements mentioned in user comments and determine their associated emotional tendencies. ABSA is a highly valuable technique in text analysis, as it allows the evaluation of sentiment across multiple dimensions within a single comment (Mehra, 2023; Viñán-Ludeña & Campos, 2024). Compared to overall sentiment classification, integrating ABSA into machine learning or regression models has been shown to significantly improve predictive performance (Li et al., 2023).

Based on the key dimensions typically emphasized in the conservation and renewal of historic districts, this study includes six analytical aspects:

- Material Form: Refers to the physical environment, including street surface design, street interfaces, and spatial morphology (e.g., sidewalks, building facades, architectural styles, and maintenance conditions).
- Commercial Functions: Encompasses commercial uses such as restaurants, entertainment, and retail businesses.
- Service Facilities: Includes public service amenities such as visitor centers, restrooms, parking areas, signage systems, and accessible infrastructure.
- Natural Environment: Covers elements such as vegetation, green spaces, water features, garden landscapes, and animals. Note: This does not include weather or air quality.

• Cultural Identity: Reflects the user's subjective appraisal and emotional identification with the district's historical value, cultural heritage, and ethnic character, including feelings of resonance, pride, and familiarity linked to heritage preservation and regional identity.
• Governance Mechanism: Relates to management practices such as crowd control, sanitation, shared bicycle regulation, ticketing systems, and public safety. ABSA enables fine-grained analysis and offers substantial value by providing insights for urban governance and decision-making.

In terms of emotional classification, this study draws on the basic universal emotions framework proposed by psychologist Paul Ekman, who identified seven core emotional categories: joy, sadness, anger, fear, surprise, disgust, and neutrality (Ekman, 1992). In this study, emotions are categorized into seven types: joy, pride, calmness, disappointment, fear, dissatisfaction, and mixed.

• Joy is defined as positive emotions such as happiness, relaxation, excitement, and satisfaction.
• Pride refers to expressions of admiration and identification with the district's historical or cultural background.
•Calmness indicates an absence of strong emotional fluctuation, often reflected in objective or neutral descriptions.
• Disappointment is characterized by a clear gap between expectation and actual experience.
• Fear involves concerns related to personal safety or social order.
• Dissatisfaction includes feelings of indifference, rejection, resistance, or negative judgment.
• Mixed emotion is used only when both positive (e.g., joy, pride) and negative (e.g., disappointment, fear, dissatisfaction) emotions are clearly expressed in the same comment with comparable intensity, making it difficult to determine the dominant emotional tendency.

3.3 Image Feature Extraction
Landscape type refers to the perceptual distinctions between one area and another, emphasizing the contextual differences that define distinct environments. Human perception responds more to holistic situations than to isolated elements, making landscape types—understood as cognitive situational units—more aligned with how people observe and experience their surroundings. Based on the principles of historic district conservation and renewal, and considering the typical scenes found in such areas, this study classifies visual landscape types into seven categories: historic architecture, streetscape, storefront façade, natural environment, signage and facilities, goods, and human activities.

In this study, a Mask2Former-based image segmentation model was employed to identify the visual landscape types embedded in user-generated images from social media. Mask2Former is a universal image segmentation framework capable of handling semantic segmentation, instance segmentation, and panoptic segmentation tasks (Cheng et al., 2022). The model was pretrained on the ADE20K dataset, which contains 150 categories of urban and natural scenes and broadly covers neighborhood elements and environmental features. ADE20K has been widely applied in studies of urban space recognition, landscape ecology, and visual perception (Zhou et al., 2019; Zheng et al., 2021; Zhao et al., 2023).

Compared with conventional convolutional neural network models such as U-Net or DeepLab, the key advantage of Mask2Former lies in its Transformer-based feature aggregation mechanism, which enables simultaneous modeling of both local details and global contextual relationships. This is particularly beneficial for interpreting images of historic districts, where spatial structures are complex and visually layered. In multi-label visual scenes, traditional models often struggle to differentiate highly co-occurring elements such as buildings, roads, and vegetation. Mask2Former, however, allows for fine-grained segmentation and accurate category recognition, thereby enhancing the detection of compositional structures in visual landscape types.

Accordingly, Mask2Former was selected for the interpretation of visual landscapes in historic districts. In this study, we extracted the top five segmented labels for each image, representing the dominant visual landscape types, and further analyzed their potential associations with emotional expressions identified in the corresponding textual comments.

3.4 Analytical Approach: Machine Learning Modeling and Explainability Techniques

In this study, XGBoost and SHAP were employed to analyze the relationship between emotional features and visual landscape characteristics. To investigate how users' emotional responses relate to the structural features of historic districts captured in images, we adopted Extreme Gradient Boosting (XGBoost), an ensemble learning method based on gradient-boosted decision trees. XGBoost is a supervised learning algorithm known for its strong capability in handling nonlinear relationships, robustness with high-dimensional sparse data, and high computational efficiency. It has been widely applied in urban research, affective computing, and geospatial data analysis, making it particularly suitable for modeling heterogeneous data that include image-derived landscape classifications and emotional variables.

To further interpret the contribution and directionality of each landscape feature in emotion prediction, we integrated the SHAP (SHapley Additive exPlanations) method. Rooted in game theory, SHAP assigns each feature a Shapley value, representing its marginal contribution to a particular prediction. It is widely recognized for its consistency and local accuracy in model interpretation. By applying SHAP analysis to the output of the XGBoost model, we identified key image-based structural features—such as the proportion of architectural elements, vegetation coverage, and frequency of human presence—that significantly influence emotional responses. Furthermore, SHAP enabled us to discern the positive or negative direction of each feature's impact on emotional outcomes.

4. Results

4.1 Overall Emotional Characteristics

The public's overall emotional experience in Nanluoguxiang is predominantly positive. Notably, the emotion of joy accounts for 68.7%, indicating that this historic district is generally successful in eliciting pleasant and satisfying emotional responses from visitors. However, negative emotions are still present to a certain extent. The proportion of dissatisfaction reaches 18.64%, followed by disappointment (5.89%) and calmness (5.44%). Emotions such as pride (0.97%) and fear (0.36%) appear infrequently and can be considered peripheral.

Monthly breakdowns of emotional distribution reveal notable seasonal variations. Joy remains consistently high during spring (March–May) and autumn (September–November), ranging from 67% to 73%. In contrast, its proportion drops significantly in the summer, reaching a yearly low of 61.24% in July. This decline is inversely correlated with a peak in dissatisfaction, which reaches 28.09% in the same month. This pattern may be attributed to increased tourist volumes during the summer holiday season, leading to greater spatial crowding and disruption of visitor rhythms, thereby influencing emotional perceptions.

By comparison, the milder climate and moderate foot traffic of spring and autumn provide more favorable spatial experiences, contributing to more positive emotional responses. These findings suggest that urban planners and tourism managers should pay particular attention to the emotional consequences of seasonal peaks in historic districts. During summer, strategies such as visitor flow control, enhanced service facilities, and spatial buffering may help alleviate the sense of overcrowding and emotional fatigue, ultimately improving emotional satisfaction and overall visitor experience.

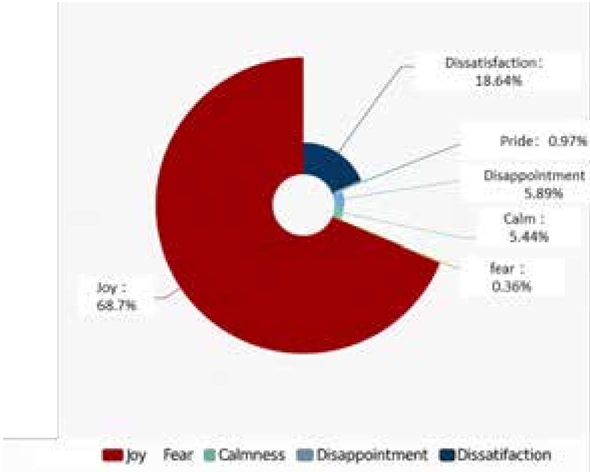

Fig.2 Distribution of Public Emotion Types in Nanluoguxiang

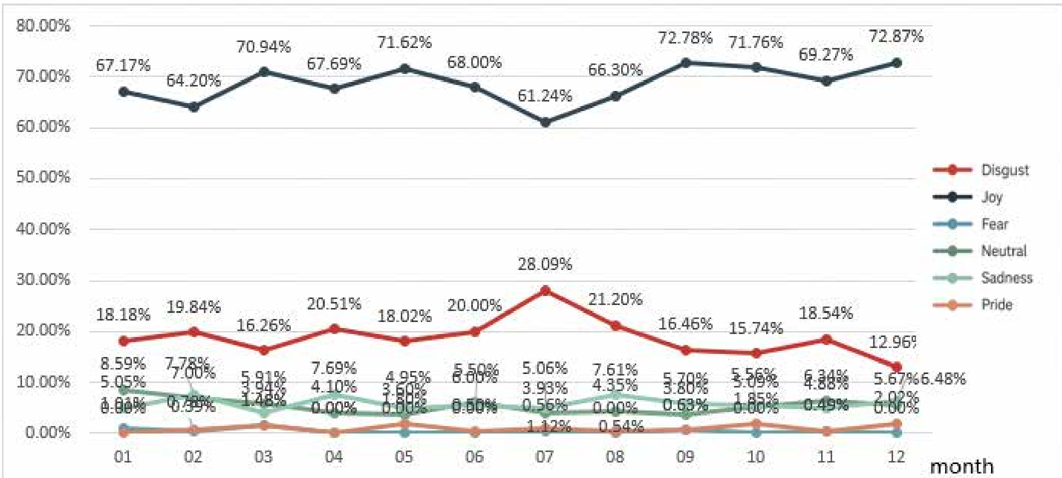

Fig.3 Annual Distribution of Public Emotion Types in Nanluoguxiang

4.2 Aspect-Based Emotional Characteristics

Overall, joy dominates across most aspects, reaching its highest proportion in the natural elements dimension (91.3%), followed by service facilities (82.5%) and functional formats (62.5%). This suggests that these aspects tend to elicit more positive emotional responses from the public. However, in the dimensions of management mechanisms and cultural identity, positive emotions are significantly diminished. The proportion of joy drops to 26.0% and 66.9%, respectively, in these two categories.

Notably, dissatisfaction accounts for as much as 52.0% under management mechanisms, the highest across all aspects, and considerably higher than its proportion under cultural identity (11.8%) and functional formats (24.8%). This indicates that institutional arrangements, regulatory measures, or managerial actions are more likely to provoke negative emotions among visitors. In addition, calmness also reaches a relatively high level under management mechanisms (18.8%), which may reflect a sense of passive acceptance or emotional neutrality among some visitors.

Among the less frequent emotions, disappointment is primarily associated with functional formats (5.5%) and cultural identity (4.8%). Pride appears slightly more often under cultural identity (11.8%) than in other aspects, suggesting that cultural elements can trigger a sense of emotional affiliation or resonance. Overall, this chart reveals substantial variation in emotional composition across different aspects, highlighting the need for differentiated optimization of landscape, functionality, and governance in historic districts based on their specific emotional effects.

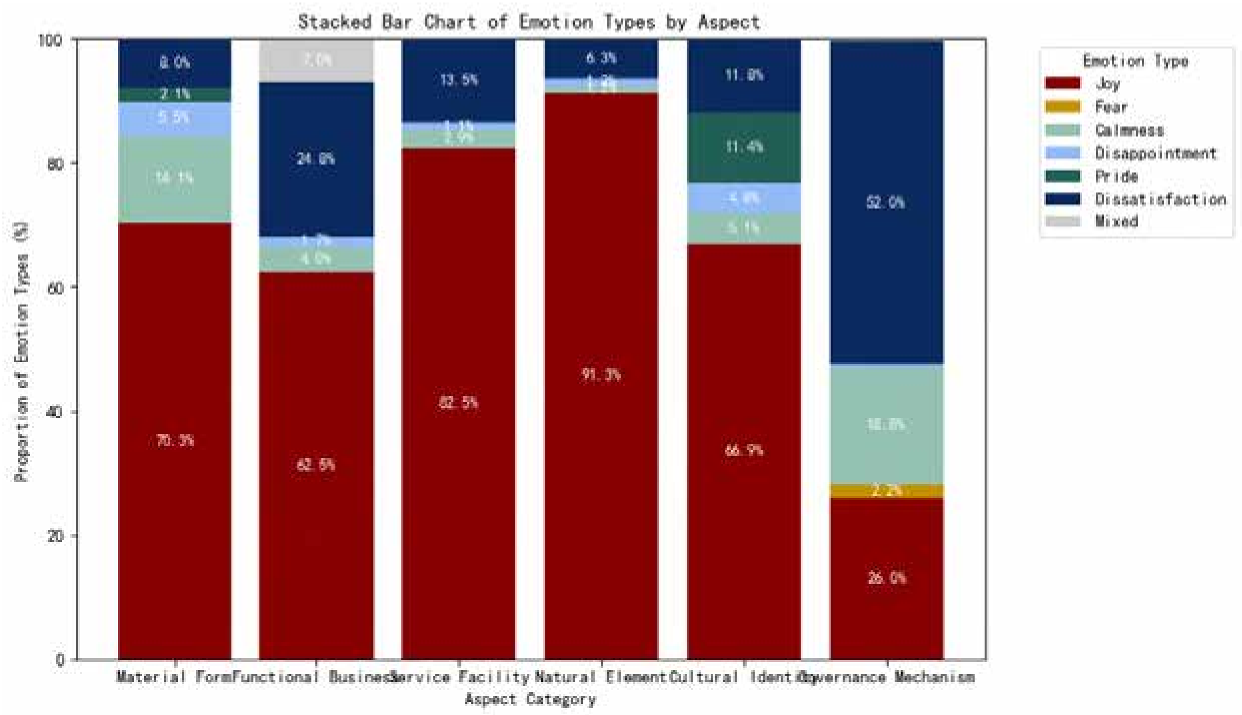

Fig.4 Aspect-Level Emotion Type Analysis

4.3 Visual Landscape Type Characteristics

Among the six categories, **commercial landscape** occupies the largest share at **48.5%**, suggesting that commercial scenes such as storefronts, signs, and retail displays are the most frequently photographed elements in this historic district. **Architectural landscapes** account for **23.7%**, indicating that built heritage and spatial enclosures remain important visual elements.

Streetscapes represent **15.2%** of the image content, while **facilities and signage** (8.7%) and **natural landscapes** (3.7%) appear less frequently. The **human activities** category accounts for only **0.3%**, implying either limited visible social interaction in the captured images or potential underrepresentation due to model classification limitations. Overall, the radar chart reflects a highly commerce-oriented visual landscape in Nanluoguxiang, with relatively low visibility of natural and sociocultural elements.

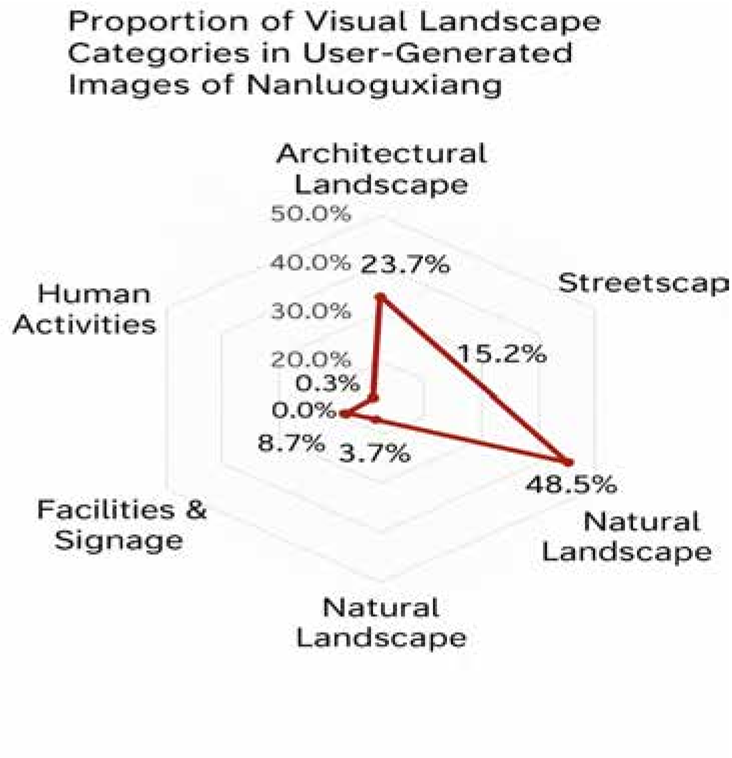

Fig.5 Proportion of Visual Landscape Categories in Nanluoguxiang

4.4 Relationship Between Emotional Characteristics and Visual Perception Types
The results are derived from an XGBoost (Extreme Gradient Boosting) model trained to predict public emotional responses based on visual landscape features extracted from images. To interpret the contribution of each feature to the model's output, SHAP (SHapley Additive exPlanations) values were employed. SHAP provides a theoretically grounded and visually interpretable method to quantify the marginal impact of each input variable on the prediction outcome.

Among the seven features, **Human Activities**, **Streetscape**, and **Historic Buildings** exhibit the widest spread of SHAP values, indicating they have the greatest influence on model output. High values of **Human Activities** and **Streetscape** (shown in red) are generally associated with positive contributions to the prediction, while higher levels of **Storefronts** and **Signage & Facilities** often reduce the predicted emotional positivity.

By contrast, **Natural Environment**, **Commercial Goods**, and **Signage & Facilities** show relatively narrow SHAP ranges, implying limited influence in the current model context—either due to low prevalence in the data or weaker emotional associations. This visualization provides a transparent global view of model interpretability, highlighting which physical elements in historical streetscapes are most strongly tied to emotional outcomes.

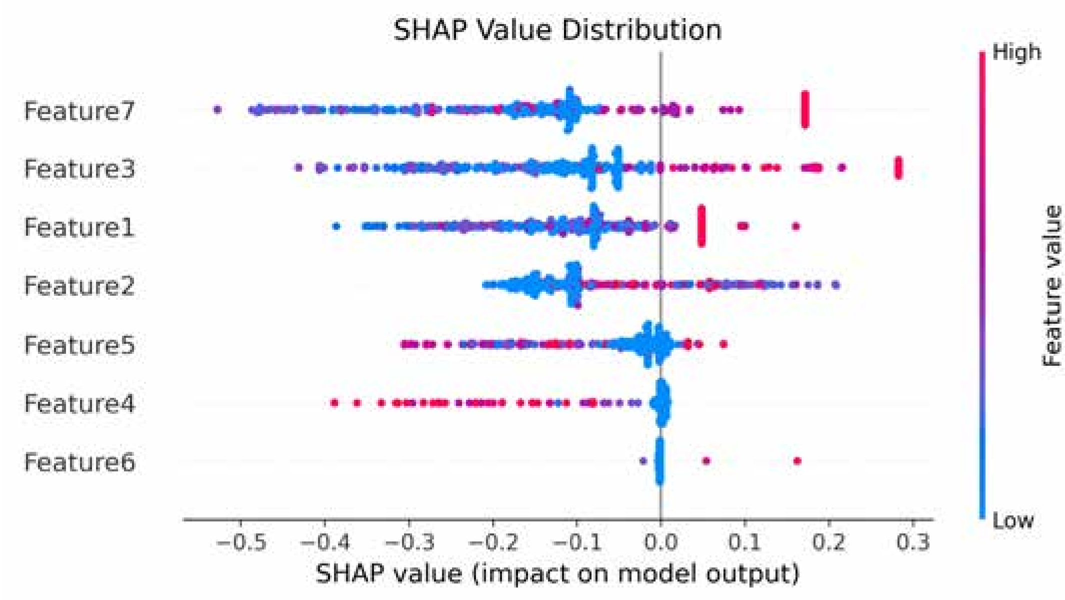

Fig.6 SHAP Value Distribution for Feature-Level Impact on Emotion Prediction Model

5. Discussion

This study demonstrates that text-based user comments and their accompanying images from social media platforms can serve as a valuable form of multimodal data for investigating the emotional effects of urban spaces. By integrating textual and visual content, the research reveals how public emotional responses are embedded within both narrative expressions and visual representations of historic districts.

Moreover, the findings highlight the potential of artificial intelligence (AI) in detecting and analyzing emotional experiences associated with urban environments. In particular, the AI models employed in this study were able to classify six distinct types of emotional responses with a relatively high degree of accuracy and consistency, illustrating their applicability to emotion analysis tasks in the urban and heritage planning context.

This research further confirms that it is feasible to integrate large language models (LLMs), computer vision techniques, and machine learning algorithms to jointly observe emotional signals and spatial landscape characteristics from social media data. Such integration enables a more comprehensive and scalable approach to understanding how people perceive and emotionally respond to historic urban spaces in real-world, user-generated contexts.

Nevertheless, certain limitations remain. A key constraint lies in the inherent biases of social media data. These datasets typically lack information on users' socioeconomic attributes, such as age, income, education level, or residential background, which may influence how different groups emotionally experience the same space. To address this limitation, future studies could incorporate experimental or survey-based methods to supplement and validate findings from digital trace data. Such triangulation would enhance the robustness and generalizability of emotion-oriented urban research.

6. Conclusion

This study is based on user-generated content (UGC) from the social media platform Dianping, using Nanluoguxiang (NLG), a representative historic district in Beijing, as a case study. It collected text reviews and associated images to construct an AI-driven framework for emotional analysis. Using XGBoost, the study investigates the relationship between six types of visual environmental scenes in the historic district and visitors' emotional experiences. The results reveal that specific visual scene types within historic districts have significant effects on public emotions. Meanwhile, the proposed multimodal framework proves effective in predicting the emotional impacts of visual landscapes in historic district settings.

References

Bornioli, A., Parkhurst, G. and Morgan, P.L. (2018) Psychological wellbeing benefits of simulated exposure to five urban settings: an experimental study from the pedestrian's perspective. Journal of Transport & Health, 9, pp.105–116.

Chen, M., Arribas-Bel, D. and Singleton, A. (2020) Quantifying the characteristics of the local urban environment through geotagged Flickr photographs and image recognition. ISPRS International Journal of Geo-Information, 9(4), p.264.

Ghahramani, M., Galle, N.J., Ratti, C. and Pilla, F. (2021) Tales of a city: Sentiment analysis of urban green space in Dublin. Cities, 119, p.103395.

Park, S.B., Kim, J., Lee, Y.K. and Ok, C.M. (2020) Visualizing theme park visitors' emotions using social media analytics and geospatial analytics. Tourism Management, 80, p.104127.

Reece, R., Bornioli, A., Bray, I., et al. (2022) Exposure to green, blue and historic environments and mental well-being: a comparison between virtual reality head-mounted display and flat screen exposure. International Journal of Environmental Research and Public Health, 15, p.9457. https://doi.org/10.3390/ijerph19159457.

Scopelliti, M., Carrus, G. and Bonaiuto, M. (2019) Is it really nature that restores people? A comparison with historical sites with high restorative potential. Frontiers in Psychology, 9, p.2742. https://doi.org/10.3389/fpsyg.2018.02742.

Yang, C. and Zhang, Y. (2024) Public emotions and visual perception of the East Coast Park in Singapore: A deep learning method using social media data. Urban Forestry & Urban Greening, 94, p.128285.