This is the Accepted Manuscript version of an article published by Elsevier in the journal *Computers, Environment and Urban Systems* in 2026.
Cite as: Liu, H., Yang, S., Abdelrahman, M., Zhu, Y., Wei, X., Biljecki, F. (2026): Inferring urban functions from Google Maps reviews: A multi-scale, multi-modal and cross-city approach. *Computers, Environment and Urban Systems*, 129: 102475. doi: 10.1016/j.compenvurbsys.2026.102475

Inferring urban functions from Google Maps reviews: A multi-scale, multi-modal and cross-city approach

Haixiao Liu[a], Sijie Yang[a], Mahmoud Abdelrahman[a], Yihan Zhu[a], Xiaobing Wei[a,c], Filip Biljecki[a,b,*]

[a] *Department of Architecture, National University of Singapore, 4 Architecture Dr, Singapore, 117566, Singapore, Singapore*
[b] *Department of Real Estate, National University of Singapore, 15 Kent Ridge Dr, Singapore, 119245, Singapore, Singapore*
[c] *Department of Geography, Ghent University, Krijgslaan 281, S8, Ghent, 9000, Belgium*

Abstract

Characterising and classifying urban functions is a long-standing research focus in urban studies and plays a critical role in urban management and community renewal. However, traditional point-of-interest (POI) categories rely on predefined labels that are often inconsistent across cities and may not fully capture how places are described, represented, or experienced in user-generated data. Further, point-based representations are highly sensitive to spatial aggregation scales, which limits their ability to capture areal functional characteristics and relative differences in POI activity intensity. To address these challenges, we propose a unified framework that, for the first time, leverages place reviews from Google Maps, a form of user-generated geographic information, as a previously untapped POI-linked extended data stream for urban functional inference and classification. Specifically, we employ pre-trained BERT and Vision Transformer models to embed textual and visual information from place reviews, enabling POIs in Singapore and Hong Kong to be represented and clustered within a shared functional embedding space. We then incorporate the weighted volume of place reviews as an indicator of relative POI activity intensity to construct category intensity vectors for spatial units, and demonstrate their effectiveness through cross-city similarity matching tasks. Finally, urban functional classification is conducted across three spatial scales: a 1 km hexagonal grid, administrative areas, and traffic analysis zones (TAZs), using graph neural networks combined with k-means clustering, producing results that preserve spatial continuity and robustness. The proposed framework provides a data-driven approach that highlights the value of place reviews as a complementary data source to conventional POIs and offers a reliable urban functional classification that works across different cities.

Keywords:
Urban functional classification, Multimodal representation learning, Graph neural networks, Volunteered geographic information (VGI)

1. Introduction

Identifying areal functional characteristics in cities has long been a central focus of urban analysis (Harrison and Dourish, 1996). As outlined in Burgess's concentric zone

*Corresponding author: Filip Biljecki, Email: filip@nus.edu.sg
Email address: filip@nus.edu.sg (Filip Biljecki)

model, cities have traditionally been segmented into commercial, residential, and industrial areas radiating from a central core (Burgess, 2015). However, as cities evolve and diversify, areal functions grow increasingly complex, blurry, and interdependent. Emerging functional areas exhibit unique characteristics to each city and cannot be captured by a single functional label (Hu et al., 2021). Moreover, urban functions are shaped not only by land use and the physical environment, but also by how places are actually used and perceived by people. Place reviews provide user-generated evidence of how places are used, described, and experienced in everyday contexts (Liu et al., 2020a). For example, reviews of a café frequently mentioning "remote working", "power sockets", and "long stays" often indicate that the venue functions more like a co-working space rather than a social media–oriented café located in a commercial core. Therefore, under such complex conditions, there is a pressing need for methods that can not only represent areal characteristics, but also infer user-described patterns of urban space usage from complementary data sources. This capability is crucial for robust representations of urban functions and their associated spatial dynamics.

Point-of-interest (POI) data can be used to characterise urban functional areas and differs fundamentally from remote sensing and plan-based approaches (Xia et al., 2020). POIs represent specific points with defined locations, names, and functional attributes (Psyllidis et al., 2022). They often emphasise locations with concentrated commercial activities, dining, educational institutions, entertainment facilities, or other functions linked to human behaviour and social needs (Yuan et al., 2012; Zhou et al., 2022b). However, these studies often implicitly assume that the predefined POI categories directly represent the actual functions of urban areas. In practice, POIs belonging to the same category may perform different functions across urban contexts, while different combinations of POI categories may collectively serve similar functional roles. As a result, accurately identifying the effective functions of POIs and aggregating them into stable, areal functional representations remains a persistent challenge.

Recent studies show that, driven by the rapid growth of location-based services (Wu et al., 2018), the spatial distribution of POIs is a key indicator for characterising areal functions and assessing local characteristics (Qian et al., 2021). Major map providers, government agencies, and crowdsourcing platforms are primary sources of POI data (Niu and Silva, 2020). Many studies use POI data from Amap (Zhu and Shi, 2022), Baidu Maps (Yao et al., 2017), Google Maps (Shi et al., 2023), and OpenStreetMap (Liu and Long, 2016) to obtain location, name, and category attributes and infer areal functions from a bottom-up perspective. Existing studies represent POIs as high-dimensional vectors, incorporating factors such as quantity, distribution, density, and spatial interactions (Gao et al., 2017; Yao et al., 2017; Yan et al., 2017). These methods not only reduce data-collection costs but also provide a refined understanding of areal functional patterns. They offer activity-related signals and support the classification of areal functions (Liu et al., 2017; Chen et al., 2026). On the other hand, supplementary data associated with POIs, such as place reviews obtained from Google Maps, often contain rich multimodal information, including text, images, and ratings (Mathayomchan and Taecharungroj, 2020), but remain underexplored. These reviews provide a user-centred perspective on urban functions, revealing how places are described and experienced rather than relying only on expected or nominal POI functions.

In recent years, the use of visual and textual data in urban research has advanced significantly, enabling the analysis of multimodal information from place reviews. Such advances include applying deep learning models to detect urban elements using object detection and semantic segmentation (Zhao et al., 2019). Meanwhile, advances in natural language processing, including topic modelling and pre-trained language models such as BERT, have enabled studies using social media text to explore themes including urban sentiment (Ghahramani et al., 2021), city image (Zhu et al., 2025), street vitality (Chen et al., 2024), and community disaster response (Zhou et al., 2022a). Similarly, Yan et al. (2017) integrated images and text descriptions to enhance comparative analyses of areal functions.

In urban functional classification tasks, spatial units are often connected through adjacency, distance, or network-based relationships rather than simple Euclidean structures, making it essential to account for neighbourhood feature interactions (Verburg et al., 2004; Wang et al., 2025b). Graph neural networks (GNNs) address this issue by explicitly modelling neighbourhood relationships and learning spatial representations from graph topology, thereby improving areal functional classification (Hu et al., 2021; Kong et al., 2024). However, this assumption may vary depending on the choice of spatial aggregation units (Chen et al., 2022). This issue is particularly pronounced when point-based information, such as POIs and place reviews, is aggregated into areal units, as different zoning schemes can induce variations in data density and the resulting embedding representations (Deng et al., 2024). Given the sensitivity introduced by spatial partitioning, it is therefore necessary to evaluate the robustness of urban functional classification results under different spatial aggregation strategies.

This study aims to develop an urban functional classification framework that leverages place reviews associated with POIs to extract functional representations of urban areas and applies GNNs to areal classification. This study makes three main contributions. First, we introduce place reviews as a novel data source for classifying and mapping urban functions and develop a multimodal feature fusion approach. Specifically, we extract user-generated textual and visual content from place reviews obtained through Google Maps and integrate them to enable the identification of POI categories across different urban contexts. Given that place reviews are still a nascent and emerging form of urban data and user-generated geographic information, more broadly we also contribute to the discourse on their usability in urban studies. Second, by combining refined and unified POI categories with review-based weighting and TF-IDF, we construct an areal representation that captures both relative POI activity intensity and category uniqueness. We validate its effectiveness by identifying functionally similar areas across cities. Third, we conduct a multi-scale urban functional classification analysis. By incorporating neighbourhood learning through GNNs, we demonstrate that the proposed representation exhibits strong robustness across different spatial scales, and provides insights for comparative analysis across different countries and regions. Figure 1 presents a conceptual framework that connects earlier POI-based studies with our proposed approach. It illustrates the methodological evolution from representing areal functions using POIs, to incorporating check-in data as proxies for human activity, and ultimately to our method, which infers POI functions from place reviews and constructs an areal representation that captures relative POI activity intensity across functional categories, thereby enabling more fine-grained, multi-scale, and cross-city urban functional classification.

The remainder of this paper is organised as follows: Section 2 reviews related work on urban functional classification, POI-based representation, and multimodal graph-based spatial analysis. Section 3 details the datasets and methodologies employed in this study. Section 4 presents POI category inference using place reviews, introduces an areal functional representation approach, and examines multi-scale urban functional classification. Section 5 discusses the findings, implications, and the limitations of the study. Finally, Section 6 concludes the paper and outlines directions for future research.

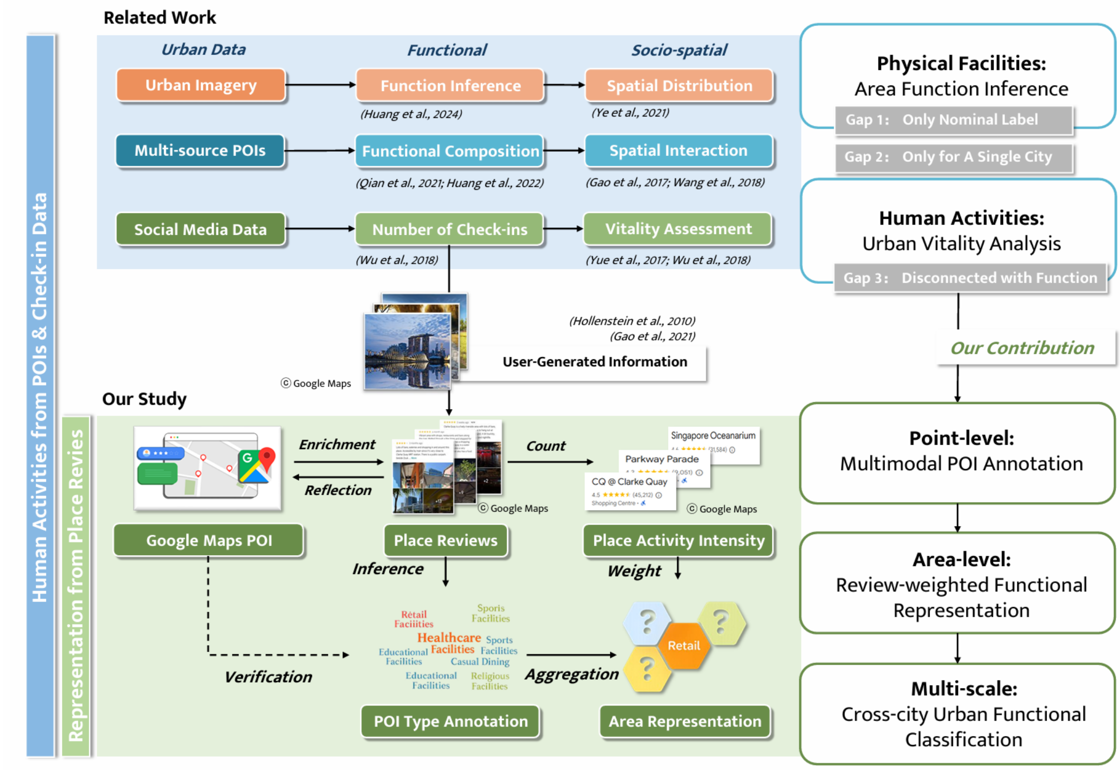


Figure 1: Conceptual framework linking urban function, POI semantics, and urban functional classification.

2. Related work

2.1. Urban functional classification

Cities around the world are rapidly expanding, giving rise to increasingly heterogeneous areas with diverse functions (Luca et al., 2023). As urban areas are closely intertwined with their spatial and social structures, defining and identifying them has long been a central concern in urban planning and serves as the foundation for exploring urban functional classifications (Du et al., 2021). Traditional functional classifications often rely on statistical data, such as questionnaires and expert analyses (Smith, 1965). At the same time, several studies have used census data to examine human activities from economic and social perspectives, capturing patterns of urban vitality and function (Schneider and Woodcock, 2008). However, these approaches are constrained by the limitations of statistical data, being time-consuming, labour-intensive, and potentially subjective (Yoshihara and Yoneoka, 2014).

With the advancement of computer science and geographic information science, data-driven classification methods have gradually replaced traditional empirical judgment in urban functional classification (Bibri, 2021). Research based on remote sensing imagery has long attracted significant attention, as classification results are fundamental to numerous environmental and urban applications and can provide detailed information about surface features (Mehmood et al., 2022). For example, Zhang et al. (2018) employed high-resolution remote sensing imagery to identify urban functional areas, achieving fine-grained spatio-temporal classification. Several studies have also utilised nighttime light data to analyse human activity, thereby revealing patterns of urban vitality and functional differentiation (Fang et al., 2022).

However, these studies primarily emphasise physical characteristics while often overlooking social and economic dimensions, thereby limiting their ability to fully capture urban functions (Wang et al., 2025a). Urban functions can be defined as geographic units composed of various interconnected elements within a city (Dijkstra et al., 2019). These elements are spatially organised and interlinked through flows of materials, capital, and information, providing spaces for living, production, recreation, and commercial activities (Castells, 2020). Therefore, a more comprehensive classification of urban functions should consider both static physical characteristics and dynamic human activities (Zhang et al., 2020). Such an integrated perspective is essential for understanding the spatial interactions among different urban functions (Yang et al., 2024).

The rise of big data and information and communication technologies (ICTs) has facilitated new approaches to understanding urban functions through social perception (Niu and Silva, 2020; Gao et al., 2021). By leveraging emerging data sources such as mobility trajectories and social media, researchers have increasingly explored the representation of urban characteristics and socio-spatial structures, achieving encouraging results (McKenzie and Romm, 2021). Moreover, social media provides a unique perspective for capturing spatio-temporal preferences and human behaviour (Martí et al., 2019; Yu et al., 2020), as these platforms generate vast amounts of geotagged content daily (Ilieva and McPhearson, 2018). In addition to spatial and temporal cues, geotagged textual and visual data offer rich semantic information about user interactions with the environment (Wang et al., 2018; Huang et al., 2021). Shen and Karimi (2016) used geotagged check-in data to represent functional urban streets and infer urban land use. Soliman et al. (2017) used Twitter data to characterise urban functional categories, and Olson et al. (2021) developed methods based on Yelp reviews to understand collective urban identity and track neighbourhood characteristic changes over time.

Meanwhile, POI data, derived from mapping services or crowdsourced platforms, have become a valuable resource for mapping socioeconomic activities and delineating detailed urban functional categories, owing to their precise spatial locations and rich semantic attributes (Liu et al., 2020b). More importantly, POI data can capture the dynamic, real-time nature of urban functions, revealing the diversity and co-existence of different facility types, and thereby providing deeper insights into a region's socioeconomic profile (Kang et al., 2021).

2.2. POI-Based functional representation

With the exponential growth of location-based services, POI data collected from cities worldwide can be leveraged to understand urban functional composition (Psyllidis et al., 2022). POIs offer valuable insights into the characteristics and functions of specific urban areas through comprehensive metadata such as categories and descriptions (Yue et al., 2017; Sun et al., 2023).

Recent research has primarily utilised the numerical distribution and attributes of POI data to represent regional characteristics, resulting in two main analytical approaches: feature frequency–based methods and semantic analysis–based methods (Liu et al., 2023). The former statistically extracts POI densities and categories to classify functional areas, whereas the latter incorporates natural language processing techniques to extract features and mine high-level semantic information (Xu et al., 2022). However, these approaches often overlook the internal spatial correlations among POIs, potentially leading to the underutilisation of valuable spatial information. Yao et al. (2017) drew inspiration from Google's Word2Vec model to learn semantic relationships between POI categories. By analysing the co-occurrence relationships between POIs, the model projects them into a high-dimensional vector space, positioning semantically similar POIs closer together. Building upon this foundation, Place2Vec was developed to reduce the loss of spatial contextual information and enhance the capacity for regional representation (Yan et al., 2017; Zhai et al., 2019).

While these methods perform well in identifying land use from POIs, they still struggle to capture the dominant functional roles of certain locations, such as large shopping malls or hospitals, because semantic similarity of POI categories alone may not accurately reflect actual socioeconomic activities, including their relative importance within the urban system (Zhai et al., 2019). The notion of relative importance is particularly crucial for urban functional classification and for POI-based applications such as identifying dense commercial and residential clusters in city centres (Luo et al., 2021). When POIs serve as the primary data source, class imbalance and spatial heterogeneity between hot and cold spots can also lead to substantial identification bias (Wu et al., 2021). A direct and practical manifestation of this issue is that large hospitals attract far more patients than small clinics, and newly developed shopping malls tend to generate greater local consumption potential than older ones (Listorti et al., 2023). However, previous studies have often treated these entities as functionally similar and assigned them to the same category.

To address these challenges, several studies have employed probabilistic topic modelling techniques such as TF–IDF to identify key POIs and delineate urban functional zones (Yuan et al., 2012). This approach identifies global frequency patterns and filters distinctive POIs; however, it has limited ability to distinguish whether a POI is more popular or more frequently visited than other POIs of the same category. Other studies have incorporated additional data sources, such as social media or check-in data, as proxies for visit intensity and urban vitality around POIs (Hollenstein and Purves, 2010; Li et al., 2022). Nevertheless, social media data also have inherent limitations: only a small portion of posts are geotagged, and their spatial coverage is highly uneven, leading to potential sampling bias in representing user populations (Middleton et al., 2018). Moreover, the limited positional accuracy of geotagged content makes it difficult to precisely match social media activities with specific POIs, thereby constraining the ability to attribute activity-related signals to individual places.

Compared with the limitations of previous studies, place reviews obtained through Google Maps, as an extension of POI data, have emerged as a promising data source. Unlike traditional POI location information, place reviews provide much richer user-generated evaluation content, including ratings, the number of reviews, as well as textual and visual materials contributed by users (Leiras and Eusébio, 2024). In contrast to check-in data from conventional social media platforms such as Twitter or Instagram, place reviews obtained through Google Maps are explicitly associated with specific POIs, thereby reducing errors caused by inaccurate geotagging and irrelevant content (Chuang et al., 2016).

However, Google Maps reviews have never been used to solve this problem. Further, existing studies using such data have primarily focused on the textual content, such as using review texts to evaluate dining satisfaction and user preferences (Li and Hecht, 2021), explore cultural expressions (Rabiei-Dastjerdi et al., 2022), and assess the distribution and performance of public libraries (Khan and Loan, 2022). In contrast, limited attention has been paid to other valuable elements contained in the crowdsourced data associated with POI reviews, such as user-uploaded images. Moreover, the category information assigned to POIs in Google Maps reviews lacks an official definition, which may lead to inconsistencies when comparing data across different countries or cities.

2.3. Multimodal and graph-based methods for spatial analysis

Recent advances in machine learning and computer vision have greatly enhanced the ability to process multimodal data that combine textual and visual information (Ye et al., 2021). Progress in natural language processing (NLP) has made it possible to capture subtle nuances in online discourse, including sentiment analysis and topic modelling (Devlin et al., 2019). A recently developed language representation model, Bidirectional Encoder Representations from Transformers (BERT), is based on a deeply trained neural architecture and has demonstrated outstanding performance across a wide range of complex tasks (Koroteev, 2021; Zhou et al., 2024). Niu and Xing (2024) fine-tuned the BERT model to classify sentiment in online restaurant reviews in Beijing and further analysed the spatial distribution of sentiments. On the other hand, deep learning architectures, such as convolutional neural networks (CNNs) and transformer models, have also contributed to more accurate and fine-grained interpretations of visual content (Wang and Lu, 2024). These approaches have been applied to various types of imagery in urban analysis, including remote sensing images (Mehmood et al., 2022), street-level imagery (Biljecki and Ito, 2021; Huang et al., 2024), and crowdsourced photos (Fan et al., 2025).

These approaches have also facilitated research on effectively integrating heterogeneous data sources through embedding-based methods within multimodal frameworks (Jenkins et al., 2019). In the context of place reviews, the primary modalities are user-generated text and images. Previous studies have explored multimodal deep learning techniques that jointly model textual and visual information to some extent (Srivastava et al., 2019). Chen et al. (2025) extracted textual representations of human activities from street-view imagery and integrated them with remote sensing and POI data, providing an effective approach for improving the identification of urban functional zones (UFZs). Similarly, Yan et al. (2024) used a large language model (LLM) to generate textual descriptions for each satellite image as complementary semantic information, and developed UrbanCLIP, a text-enhanced urban area analysis framework based on contrastive language–image pretraining.

In the context of urban functional classification, as Tobler's First Law of Geography posits that spatial proximity influences the degree of correlation between entities (Tobler, 1970), underscoring the importance of modelling spatial dependencies among regions. GNNs are particularly well-suited for this task. In complex urban environments, GNNs demonstrate superior efficiency in integrating multimodal data for urban functional analysis (Tao et al., 2025). Graph-based approaches are also advantageous for this type of analysis, as they can capture intricate dependencies within non-Euclidean spaces and learn representations from both node attributes and graph topology (Wu et al., 2020). Building on these advantages, GNNs have been successfully applied in various urban analytics tasks, including traffic prediction (Xue et al., 2022), streetscape classification and prediction (Kong et al., 2024; Lei et al., 2024), and the identification of emerging urban structures (Fan et al., 2024). Overall, GNNs are capable of capturing spatial dependencies and functional interactions across multiple scales, showing significant potential for fusing multimodal features and improving the accuracy of urban functional classification.

However, because GNN-based approaches explicitly model spatial relationships based on predefined areal units, their performance may vary under different spatial aggregation schemes. Existing studies have noted that, when point-based data such as POIs are aggregated into areal representations, variations in spatial scale can influence both data distribution and the resulting embeddings (Fan and Thakur, 2023). Accordingly, recent research has begun to emphasise the importance of examining cross-scale consistency in urban classification, in order to better understand the extent to which learned representations reflect underlying urban patterns rather than artefacts of specific spatial configurations (Qing et al., 2024; Ma et al., 2026).

This paper integrates multimodal analysis of place reviews with GNN modelling to enhance cross-city functional classification. By exploiting the spatial relational structures inherent in POIs, as well as the textual and visual information contained in POI-based place reviews, and by examining the effects of multiple spatial scales, the proposed method provides a more robust assessment of cross-city functional classification.

3. Data & Methodology

3.1. Study area and data

Singapore and Hong Kong are selected as the two case study cities. As major port-based trading centres and highly urbanised global hubs, they share comparable geographic scale, urban density, and functional complexity. Despite these similarities, they differ markedly in spatial organisation, urban morphology, culture, climate, urban planning strategies, and degrees of functional mixing. Therefore, a comparative analysis using POI-based place reviews in these two contexts enables the clustering of urban spaces with semantically similar functions, forms, or atmospheres under distinct development models, thereby enriching sample diversity for cross-city functional classification research.

POI-based place review data were collected from Google Maps in October 2024 and include extensive metadata. The raw data were cleaned through three steps. First, we verified the geographic coordinates in the metadata and removed locations outside the two study cities. Second, as the review text contains timestamps, the analysis was restricted to reviews from the most recent year, and records prior to October 2023 were excluded. This constraint ensures that the dataset reflects the current function while mitigating potential bias introduced by newly opened establishments with limited review data. Finally, we retained only locations containing both textual and visual information to support subsequent multimodal embedding fusion. After data cleaning, the Singapore dataset contains place reviews for 10,492 POIs (comprising 1,881,545 textual reviews and 601,275 images), while the Hong Kong dataset contains place reviews for 10,204 POIs (comprising 586,971 textual reviews and 565,616 images). Each POI record includes five core attributes: POI name, hierarchical category labels, geographic coordinates (latitude and longitude), and the corresponding POI-based user-generated textual reviews and images. In addition, the number of reviews received by each POI within this year was used as an input for constructing the relative POI activity intensity indicator in the subsequent areal representation. Figure 2 illustrates a representative excerpt of place review data for ION Orchard in Singapore, an iconic shopping mall.

Figure 2: Excerpts of place review data: ION Orchard, a shopping mall in Singapore, reviewed in Google Maps.

In this study, POIs broadly capture everyday consumption, leisure activities, economic production, public services, and socio-cultural functions in urban environments. However, the original metadata-based single-label ”main category” labels provided by Google Maps, which are available for both cities, exhibit substantial inconsistency, with 906 unique labels in Singapore compared with 806 in Hong Kong, as well as limited overlap between the two sets. Moreover, these labels are often noisy and overly fine-grained. For example, restaurants may be subdivided into Chinese, Indian, and other cuisine types, which is unnecessary for urban functional classification, rendering them unsuitable as a unified taxonomy for comparative analysis.

Nevertheless, a coarse external reference is still useful for assessing whether the learned clusters remain broadly consistent with conventional POI semantics. We therefore manually reclassified the 906 unique labels from Singapore and 806 unique labels from Hong Kong into a harmonised taxonomy consisting of 13 high-level categories (*L1_types*): Accommodation, Art & Culture, Business, Community, Education, Entertainment, Food & Beverage, Healthcare, Recreation, Religious, Retail, Tourism, and Transportation.

It is important to note that these *L1_types* are not treated as ground truth labels for actual human activity. They are derived from existing Google Maps metadata and therefore inherit some limitations of metadata-based taxonomies. In this study, they serve only as a coarse metadata-derived reference benchmark for evaluating broad semantic consistency between the learned multimodal clusters and conventional POI categories. The objective of the clustering is not to reproduce the *L1_types*, but to identify finer-grained review-derived functional patterns from textual and visual evidence.

In addition, this study employs three types of spatial aggregation units:

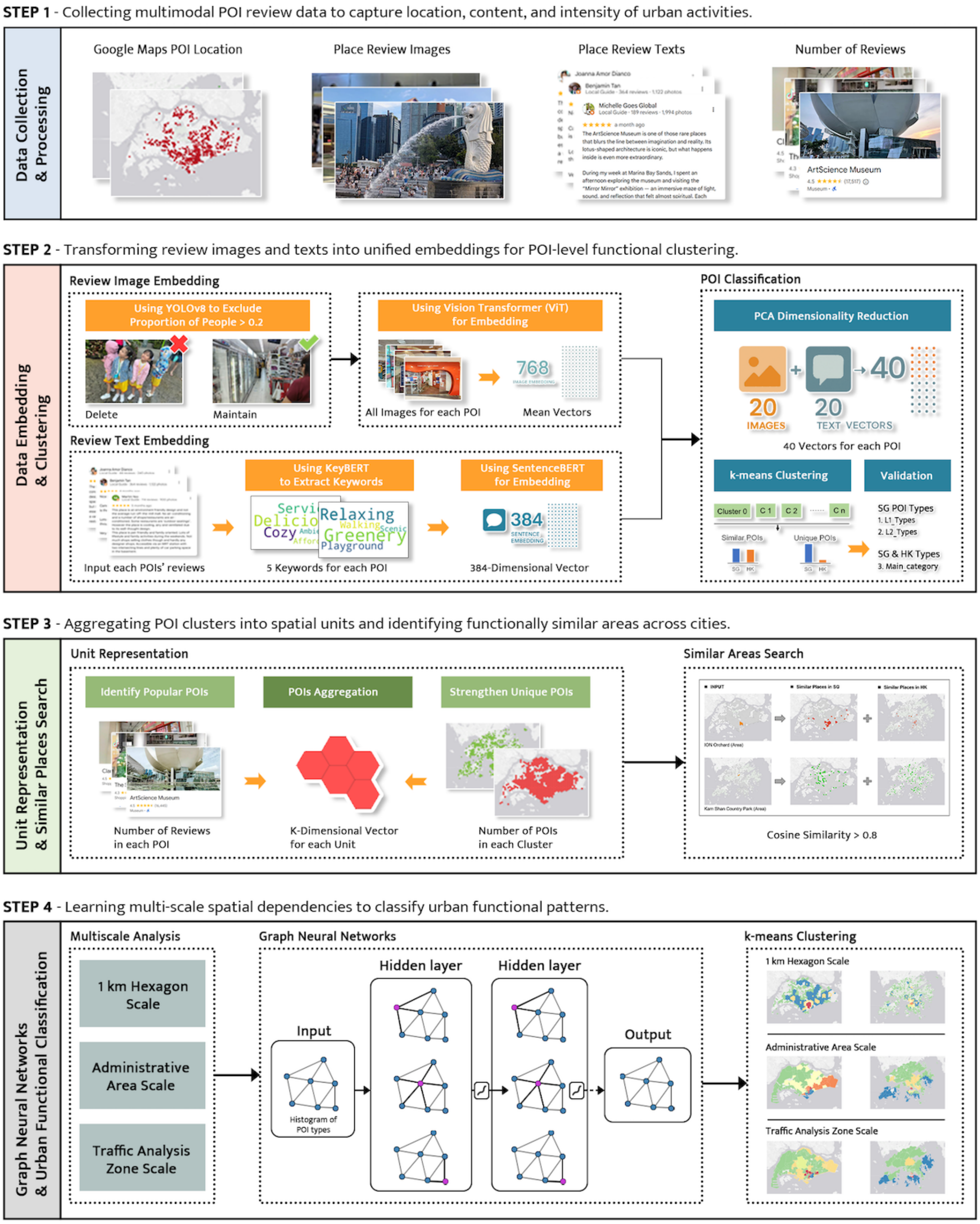
Hexagonal grids: grids generated in GIS, with an average cell diameter of approximately 1 km (roughly corresponding to H3 resolution 8).

Administrative areas: Master Plan 2019 Subzone Boundaries for Singapore (data.gov.sg) and 2021 Tertiary Planning Units (TPUs) for Hong Kong (data.gov.hk). Owing to differences in administrative definitions, both units are referred to as administrative areas hereafter.

Traffic analysis zones (TAZs): zones delineated using major and secondary road networks, with road data obtained from OpenStreetMap and processed in a GIS environment.

3.2. Methodology

The framework of the proposed model consists of four consecutive stages, as shown in Figure 3: (1) data collection and processing; (2) extraction of representative feature vectors for each POI via visual feature extraction, semantic feature extraction, feature fusion, and clustering; (3) development of an areal representation framework that characterises the functional profile of POI collections and identifies similar areas across cities; and (4) neighbourhood learning on area feature vectors to delineate and classify urban functional areas.



1. All initial data in Step 1 are obtained through Google Maps.
2. YOLOv8 is a real-time object detection model for identifying objects in images. (https://github.com/ultralytics/ultralytics)
3. Vision Transformer (ViT) is a deep learning model for extracting visual feature representations from images. (https://github.com/google-research/vision_transformer)
4. BERT is a language model for learning contextual representations from text. (https://github.com/google-research/bert)
5. Graph Neural Networks (GNNs) are models for learning representations from graph-structured data. (https://github.com/pyg-team/pytorch_geometric)

Figure 3: A framework for POI functional inference, areal representation, and urban functional classification based on Google Maps reviews.

3.2.1. Embedding representation model and POI category clustering

The embedding representation model extracts textual and visual features to construct a unified multimodal embedding for each POI-based place. Textual features are derived by applying KeyBERT to extract keywords from all user comments associated with each

POI. In similar tasks, the number of keywords is typically selected within the range of 3–10. In this study, we conducted preliminary experiments on a subset of POIs by varying the number of keywords from 3 to 10 and performing manual inspection. We found that using five keywords provides a stable representation while avoiding excessive redundancy, yielding a concise and representative summary of dominant semantic signals. These keywords are subsequently encoded into a 384-dimensional vector using Sentence-BERT. Visual features are obtained by encoding each image with a Vision Transformer to produce a 768-dimensional vector; these image embeddings are then averaged per POI.

The textual and visual embeddings are subsequently concatenated. However, high dimensional embeddings often contain redundant and noisy information, which can negatively affect distance-based clustering methods (Poggio et al., 2017). To mitigate the curse of dimensionality and improve clustering performance, we apply principal component analysis (PCA) to reduce both textual and visual embeddings. Dimensionality reduction preserves the most informative variance while enhancing computational clustering efficiency. To maintain balance between modalities, both textual and visual embeddings are reduced to the same dimensionality.

We further evaluated a range of dimensions (10–50) and found that the first 20 principal components capture the majority of variance in both textual and visual embeddings. Additional components provide only marginal improvements while potentially introducing noise. This dimensionality also ensures a compact and balanced representation when combining multimodal features. The resulting 40-dimensional multimodal vectors are then clustered using k-means, yielding k clusters that represent distinct POI categories.

3.2.2. Areal functional representation

The study areas were tessellated into different spatial units, such as equal-area hexagons, administrative areas, and traffic analysis zones, and each unit was represented by a k-dimensional feature vector, This vector captures the weighted composition of POI categories within each unit. Specifically, two weighting indices are applied: relative POI activity intensity, derived from review count, and uniqueness, quantified using term frequency–inverse document frequency (TF-IDF). The weighted values are then averaged by the number of POIs in the corresponding category, yielding the average intensity of each category within the unit.

The relative POI activity intensity weight for a given POI was calculated as the ratio of its number of place reviews over the past year to the mean review count of its inferred functional cluster. For example, a POI with 200 reviews in a year, whose cluster's average is 100 reviews, receives a weight of 2.0. This cluster-level scaling means that the indicator does not directly use raw review volume as activity intensity; instead, it measures whether a POI is relatively more or less active than other POIs within the same functional cluster. This procedure reduces inherent differences in review frequency across POI categories, such as restaurants versus playgrounds.

Uniqueness was computed using the IDF framework to attenuate the influence of ubiquitous POI categories and emphasise those that are rare yet characteristic of specific areas (e.g., large hospitals or shopping malls). The IDF for category c is defined as:

$$\mathrm{IDF}(c) = \log\left(\frac{H}{1+n_i}\right) \tag{1}$$

where:

H is the total number of hexagons.

n_i is the number of hexagons containing at least one POI of category i.

The addition of 1 in the denominator ensures numerical stability by preventing division by zero if a category does not appear in any hexagon.

Higher IDF implies that category i is rare or less common across the entire area. Consequently, hexagons containing rare categories receive higher weights.

Lower IDF indicates that category i is common and widely distributed across hexagons. Such categories are therefore considered less distinctive. Next, by calculating the cosine similarity between each hexagon, the similarity of each area can be obtained.

where the weight of category i within unit u is given by:

$$w_{u,i} = \frac{1}{N_i} \sum_{j \in C_{u,i}} \mathrm{RAI}_j \times \mathrm{IDF}_i \tag{2}$$

Here, N_i represents the number of POIs belonging to category i within unit u; $C_{u,i}$ denotes the set of POIs in unit u that belong to category i; RAI_j is the relative activity intensity of POI j, calculated based on its relative review volume; and IDF_i quantifies the uniqueness of category i by reducing the influence of common POI categories.

Ultimately, each unit will be represented as a k-dimensional vector. The final representation of each spatial unit u is defined as:

$$\mathbf{R}_u = [\, w_{u,1}, w_{u,2}, \ldots, w_{u,k} \,] \tag{3}$$

3.2.3. Graph neural networks and functional classification

Feature vectors derived from place review embeddings provide a novel representation of POIs, and units constructed from these representations can offer initial insights into urban functional classification. However, accounting for spatial continuity and inter-area relationships is essential for robust functional zoning and urban structure delineation (Hu et al., 2021). Accordingly, a two-layer graph convolutional network (GCN) was implemented to learn neighbourhood dependencies among units. Each GCN layer updates a unit's k-dimensional feature vector by aggregating information from adjacent cells, thereby integrating multimodal attributes with spatial topology to produce more discriminative area embeddings. This mechanism enables the model to capture spatial autocorrelation and neighbourhood effects that are not accounted for in non-spatial clustering approaches.

For graph construction in the GNN model, the KNN parameter was set to 6 based on the neighbourhood structures and the average number of contiguous neighbours across the different spatial scales, providing a consistent neighbourhood definition for cross-scale comparison. The network outputs refined k-dimensional representations that encapsulate both intrinsic characteristics and contextual relationships. The dimensionality is preserved to align with the original feature space and support downstream analysis. Finally, the learned embeddings are clustered using k-means, yielding spatially continuous urban functional zones.

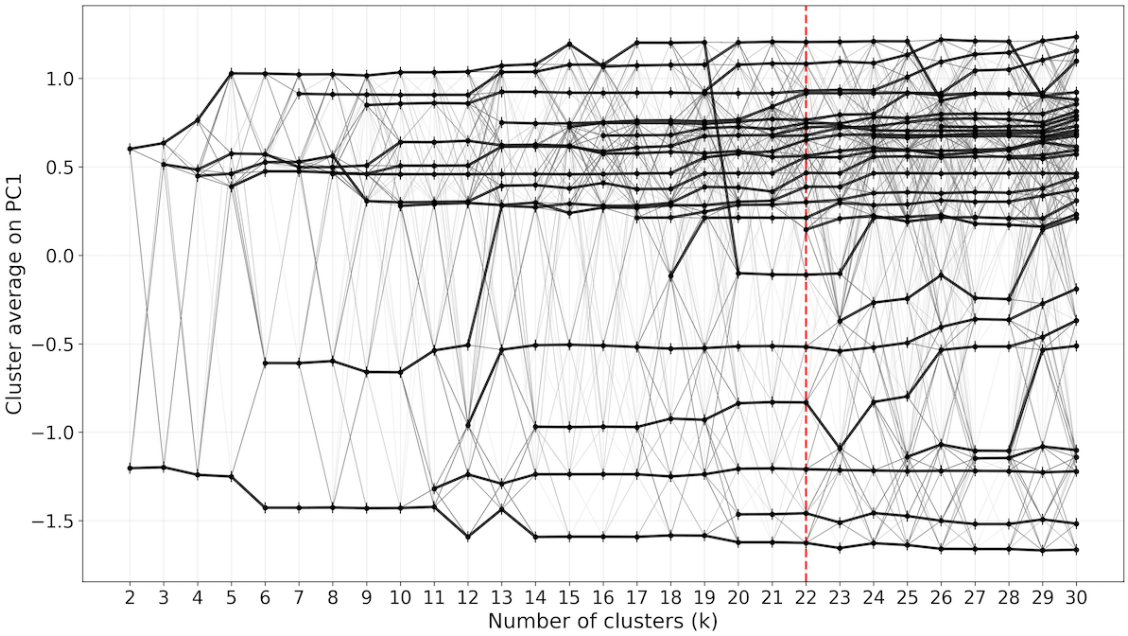
At the same time, numerous studies have demonstrated that the choice of analytical units can influence results. We therefore considered the scale and zonal effects induced by differences in spatial aggregation (Wong, 2004; Kajosaari, 2024). To examine the robustness of our results across different spatial scales, we conducted our analyses at three types of spatial scales: regular 1 km hexagonal grids, officially defined administrative areas, and traffic analysis zones (TAZs) delineated by road networks.

4. Results

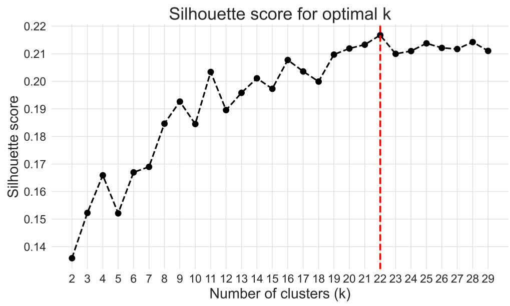
4.1. Embedding representation and POI category clustering

We applied k-means clustering to the vectorised multimodal embeddings derived from place reviews for all POIs in both cities. Although k-means requires the manual specification of the number of clusters, its computational efficiency makes it well suited for large-scale datasets. To determine an appropriate number of clusters, we employ a clustergram (Schonlau, 2002), the silhouette score, and the elbow method. The clustergram provides a hierarchical visualisation of clustering structures across different values of k. In each clustering iteration, cluster centres are projected onto the first principal component of the original data, yielding a one-dimensional representation for each cluster. The data points are then ordered according to the first principal component of their assigned clusters and plotted against the number of clusters ranging from 2 to 30, forming the clustergram.

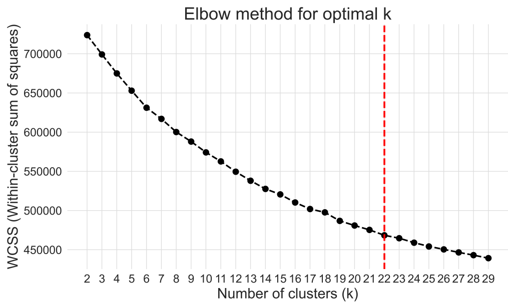
Figure 4 presents the clustering validation results, including the clustergram, the silhouette score, and the elbow method. Overall, the results indicate that the clustering structure becomes relatively stable and well separated at 22, where the silhouette score reaches its highest value. Therefore, we select k=22 as the optimal number of clusters. Figure 5 shows the distribution of POIs across the 22 clusters in Singapore and Hong Kong, respectively. This joint clustering framework enables a direct comparison between the two cities under a unified POI category system.



(a) Clustergram



(b) Silhouette score



(c) Elbow method

Figure 4: Cluster validation visualisation for multimodal POI embeddings from $k = 2$ to 30, including clustergram, silhouette score, and elbow method results.

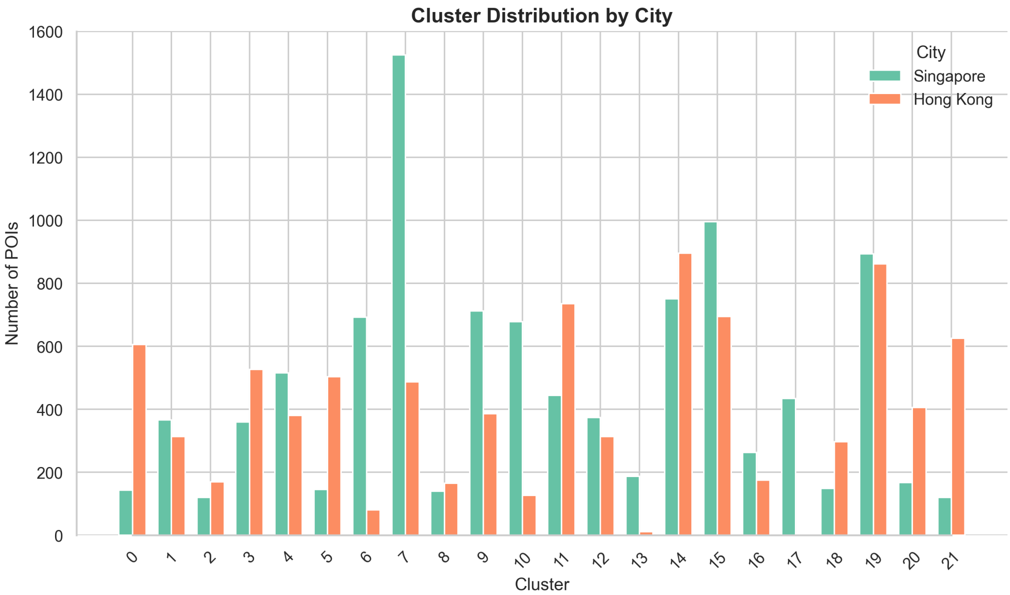


Figure 5: The number of POIs in Singapore and Hong Kong in each cluster.

To evaluate the effectiveness of the multimodal clustering approach, we conduct an ablation study. Specifically, clustering is performed on text embeddings and image embeddings both before and after dimensionality reduction, as well as on the proposed multimodal representations. The performance is assessed using five commonly adopted metrics for unsupervised learning, including Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI), Homogeneity, Completeness, and V-measure (Pedregosa et al., 2011; Lu and Uddin, 2024).

Among these metrics, ARI, AMI, Homogeneity, Completeness, and V-measure are evaluated by comparing the clustering results with the harmonised metadata-derived *L1_types* reference labels. They provide a coarse external benchmark for assessing whether the review-derived multimodal clusters remain broadly consistent with conventional POI semantics. As shown in Table 1, the proposed multimodal method achieves the highest agreement with the harmonised POI reference taxonomy across all five external metrics. However, these metrics should be interpreted as measures of semantic consistency with a harmonised POI reference taxonomy, rather than as validation against actual human activity. Because our multimodal framework is inherently unsupervised, perfect alignment with the *L1_types* is neither expected nor necessarily desirable. The observed agreement suggests that the learned clusters retain broad semantic consistency with conventional POI categories, while still allowing finer-grained review-derived distinctions beyond the reference taxonomy.

Table 1: Ablation study results of different embedding strategies.

Method	Variables	ARI	AMI	Homo.	Comp.	V-meas.
BERT	Text	0.1908	0.4325	0.4877	0.4308	0.4575
ViT	Image	0.2789	0.5300	0.5006	0.4354	0.4657
BERT + PCA	Text	0.1674	0.3845	0.4372	0.3867	0.4104
ViT + PCA	Image	0.3390	0.5337	0.5005	0.4405	0.4686
BERT + ViT + PCA	Multi	**0.3552**	**0.5887**	**0.5713**	**0.4852**	**0.5247**

To support cluster interpretation, Table 2 reports the top three metadata-derived *L1_types* reference categories for each cluster in both cities. Our analysis reveals that clusters 7, 11, 15, and 19 exhibit a *Food & Beverage* proportion exceeding 97%, demonstrating the complete dominance of this category within these clusters. To further investigate the factors influencing the clustering, we employed the XGBoost + SHAP method and randomly selected place images from both cities for illustration. Clusters 7, 11, 15, and 19 were used as representative examples. Figure 6 displays the results of the XGBoost + SHAP analysis and the distribution of clusters in Singapore and Hong Kong along with examples of place review images.

Table 2: The proportion of the top 3 L1_types in each cluster in both cities.

Cluster 0		**Cluster 1**		**Cluster 2**	
Food & Beverage	32.9%	Food & Beverage	72.4%	Healthcare	58.3%
Retail	15.4%	Entertainment	19.1%	Recreation	28.3%
Recreation	11.9%	Education	1.9%	Retail	5.0%
Cluster 3		**Cluster 4**		**Cluster 5**	
Accommodation	85.8%	Retail	85.3%	Recreation	51.0%
Retail	8.6%	Food & Beverage	5.2%	Tourism	21.4%
Entertainment	1.4%	Recreation	3.1%	Entertainment	6.9%
Cluster 6		**Cluster 7**		**Cluster 8**	
Education	92.8%	Food & Beverage	98.7%	Education	25.0%
Community	3.0%	Retail	1.0%	Healthcare	20.7%
Art & Culture	2.0%	Business	0.1%	Recreation	14.3%
Cluster 9		**Cluster 10**		**Cluster 11**	
Recreation	81.2%	Food & Beverage	52.7%	Food & Beverage	98.4%
Tourism	11.8%	Retail	27.9%	Retail	0.9%
Art & Culture	1.5%	Healthcare	18.9%	Art & Culture	0.5%
Cluster 12		**Cluster 13**		**Cluster 14**	
Art & Culture	32.1%	Recreation	92.0%	Retail	18.9%
Religious	16.3%	Business	2.7%	Education	17.6%
Tourism	15.5%	Education	2.7%	Food & Beverage	15.7%
Cluster 15		**Cluster 16**		**Cluster 17**	
Food & Beverage	99.5%	Religious	97.0%	Recreation	97.2%
Business	0.2%	Tourism	0.8%	Entertainment	1.8%
Retail	0.2%	Community	0.8%	Education	0.5%
Cluster 18		**Cluster 19**		**Cluster 20**	
Recreation	32.9%	Food & Beverage	97.3%	Retail	89.8%
Education	12.1%	Retail	1.1%	Food & Beverage	3.6%
Food & Beverage	12.1%	Community	0.7%	Community	1.8%
Cluster 21					
Healthcare	58.3%				
Education	15.8%				
Retail	7.5%				

Cluster 7

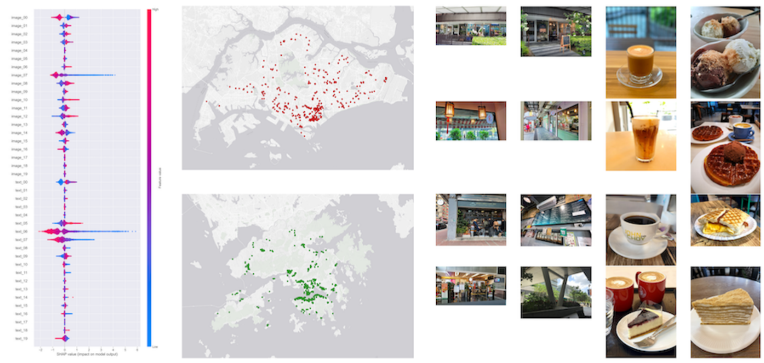
Hawker and Street Food Facilities
hawker centre, food court, cheap eats

XGBoost feature contributions are mainly derived from highly ranked image and text feature vectors, indicating that visual and textual information play equally important roles in classifying casual dining venues within this cluster.

Spatial distribution shows that these venues are present across most areas, with a notable share located in residential neighbourhoods, suggesting a function closely embedded in residents' daily lives rather than destination-oriented dining.

Review Images are characterised by street-level storefronts, menu boards, casual interiors, and visually diverse local dishes, reflecting informal, affordable, and culturally diverse dining environments.

Cluster 11



Cafe and Light Dining Facilities
Cafe, bakery, tea shop

XGBoost feature contributions are mainly driven by overall image feature vectors with additional contributions from a subset of prominent text vectors, indicating that visual information plays a dominant role while selected textual features provide more fine-grained insights.

Spatial Distribution shows a strong concentration in city centres, with a subset of venues located in neighbourhood hubs, suggesting a closer association with areas of high commercial activity, consistent with cafés functioning as social and daytime consumption spaces.

Review images prominently feature beverages, minimalist interior designs, and seating-oriented scenes, emphasising relaxed atmospheres and short-duration visits.

Cluster 15

Casual Dining Facilities
Local food, ethnic cuisine, casual dining

XGBoost feature contributions are primarily driven by text feature vectors, with several showing particularly strong influence, indicating that textual information plays a greater role in identifying this cluster, while visual features provide secondary support.

Spatial distribution resembles that of Cluster 7, with venues evenly distributed across both city centres and residential areas, suggesting that this cluster also represents everyday consumption-oriented dining facilities.

Review images highlight relatively well-presented storefronts, brighter interior designs, and seating-oriented layouts, reflecting a more formal dining style compared with highly informal casual eateries.

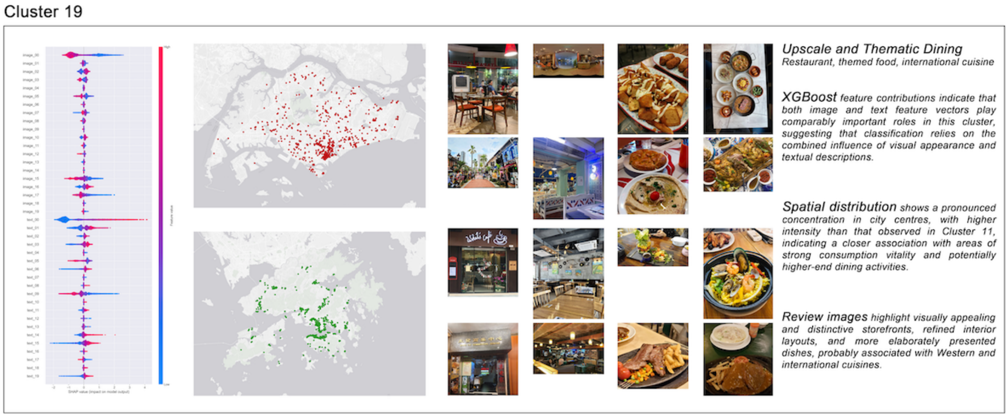


Figure 6: XGBoost + SHAP visualisation for four representative clusters.

By examining the category ratios within each cluster and manually inspecting images from both cities, we identified the distinctive characteristics of the 22 clusters and assigned each a descriptive label. Table 3 lists the names of these 22 clusters, which are grouped into eight major categories.

Table 3: Naming for 22 clusters.

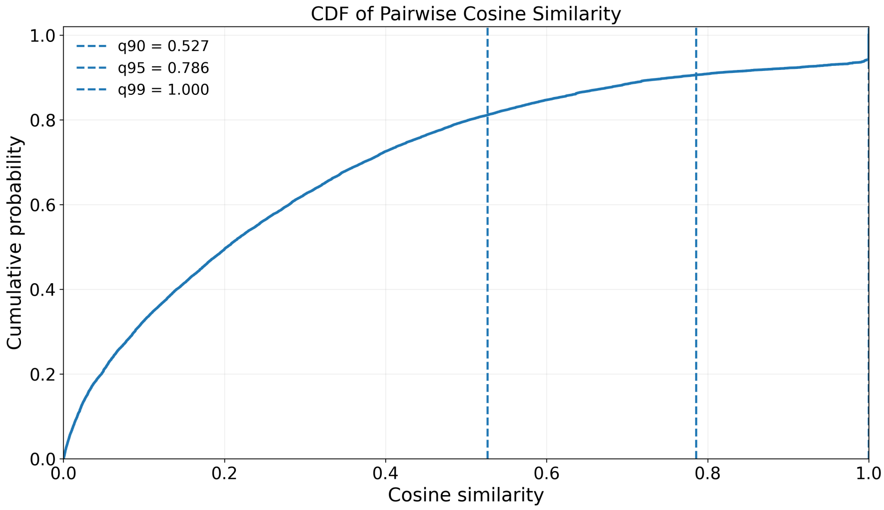
Cluster	Name	Supplementary Tags
Group A: Retail and Mixed-Use Commercial Hubs		
0	Mixed Commercial Facilities	retail, dining, shopping mall
4	Diverse Retail Facilities	retail, fashion, electronics
10	Daily Convenience Retail Facilities	supermarket, convenience store, daily goods
20	Integrated Retail and Leisure Complexes	retail, entertainment, mall
Group B: Entertainment and Recreational Attractions		
1	Nightlife and Entertainment Facilities	bar, nightlife, karaoke
5	Outdoor Leisure and Tourist Attractions	recreation, park, waterfront
9	Green and Recreational Spaces	recreation, community park, green space
Group C: Sports, Wellness, and Community life		
2	Fitness and Wellness Facilities	recreation, sports, gym
8	Specialised Sports and Interest Facilities	clinic, sports school, dance
13	Community Sports Facilities	recreation, sports court, basketball
17	Neighbourhood Playground Facilities	recreation, community, playground
Group D: Hospitality and Short-Term Stay		
3	Hospitality Facilities	hotel, accommodation, short-term stay
Group E: Food and Dining Diversity		
7	Hawker and Street Food Facilities	hawker centre, food court, cheap eats
11	Cafe and Light Dining Facilities	cafe, bakery, tea shop
15	Casual Dining Facilities	local food, ethnic cuisine, casual dining
19	Upscale and Thematic Dining Facilities	restaurant, theme food, international cuisine
Group F: Cultural and religious institutions		
12	Cultural and Heritage Facilities	gallery, museum, heritage site
16	Religious Facilities	worship, church, mosque
Group G: Educational services		
6	Educational Facilities	school, kindergarten, library
14	Integrated Community Facilities	retail, education, school
18	Mixed-use Local Service Facilities	school, park, clinic
Group H: Healthcare and Professional Services		
21	Healthcare and Professional Service	healthcare, clinic, therapy

4.2. Hexagonal functional representation and similarity measurement

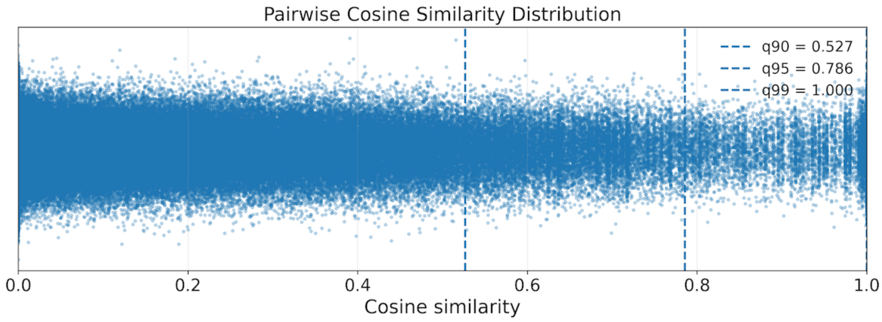
To verify whether our unit representations can effectively capture local functional patterns, we first conducted preliminary tests using hexagons as examples, such as identifying areas with similar functions.

The hexagonal functional representation is constructed from three components: the 22 review-derived POI categories obtained from the previous clustering step, relative POI activity intensity derived from review counts, and the spatial frequency of POIs across hexagons, which reflects the global uniqueness of each category. For the relative POI activity-intensity weight, we calculate the ratio of a POI's review count to the average review count of its cluster and then apply a logarithmic transformation. The resulting values are truncated at the 95th percentile, with any value exceeding this threshold set to the 95th percentile value (P95). This procedure mitigates the long-tail phenomenon and avoids treating raw review volume as a direct measure of activity intensity. Meanwhile, we employ TF-IDF to compute the frequency with which POIs in each cluster appear within the hexagons, thereby representing the uniqueness of each cluster.

Next, a histogram representing the 22 categories is computed for each hexagon as the final feature vector. By calculating the cosine similarity between pairs of hexagons, we identify similar places within the city. We set the cosine similarity threshold to 0.8, whereby pairs of areas with similarity values greater than 0.8 are considered functionally similar. The choice of this threshold is informed by the empirical distribution of pairwise cosine similarities across all spatial units. As shown in Figure 7, the cumulative distribution function (CDF) indicates that the 95th percentile (q95) is approximately 0.786. This threshold suggests that only a small proportion of area pairs exhibit similarity values above this level, corresponding to the upper tail of the distribution. Therefore, a threshold of 0.8 effectively captures those rare pairs with highly aligned functional compositions while excluding the majority of weak or moderate similarities. Figure 8 illustrates these similar urban functional areas within and across the two cities.



(a) Cumulative distribution function (CDF) of pairwise cosine similarity across all spatial units.



(b) Empirical distribution of pairwise cosine similarity across all spatial units.

Figure 7: Cumulative distribution function and pairwise cosine similarity distribution across spatial units.

We verified the histogram results for the 22 categories computed for each area. Using Orchard Road, Singapore's most prosperous downtown district, as the input, all areas with a cosine similarity exceeding 0.8 are retrieved, yielding comparable areas in both Singapore and Hong Kong. The results demonstrate that this method effectively identifies similar functional areas; for example, in Singapore, the similar areas include other central areas such as Chinatown, Clarke Quay, and Bugis, as well as commercial centres like Jurong East, Jurong Point, and HarbourFront, while in Hong Kong, the corresponding areas include well-known commercial areas such as Tsim Sha Tsui, Causeway Bay, and Sha Tin. Similarly, when Tai Mo Shan Country Park in Hong Kong is used as input, the method retrieves comparable areas in Singapore, including the Singapore Zoo and major parks such as East Coast Park.

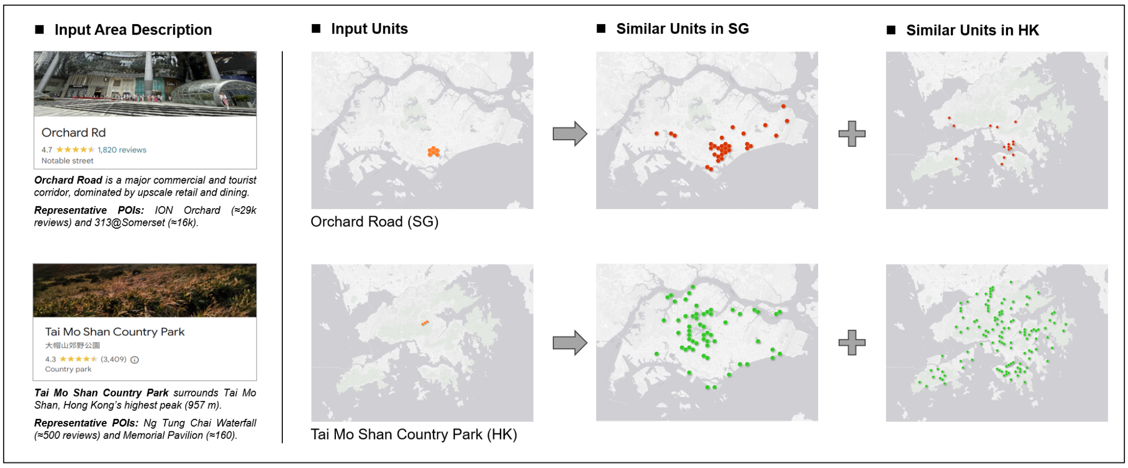


Figure 8: Identification of functionally similar areas within and across Singapore and Hong Kong. Basemap: (c) OpenStreetMap contributors.

4.3. Graph neural networks and multi-scale urban functional classification

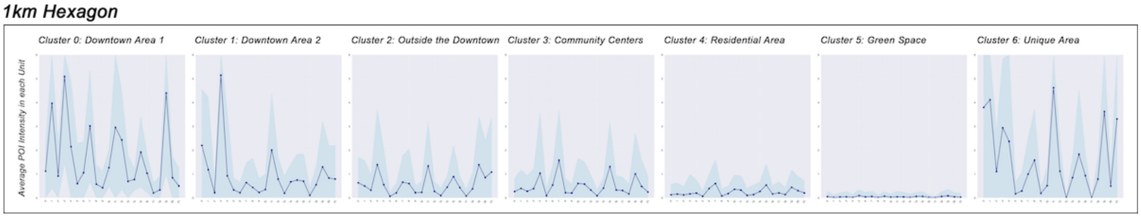
The integration of GNN and k-means clustering classified all regions into 7 distinct clusters, with the validation results reported in the Appendix. We present the results across three spatial scales: 1 km hexagons, administrative areas, and traffic analysis zones (TAZs).

Figure 9 illustrates the average relative activity intensity of the 22 POI categories across seven clusters at three spatial scales. In this figure, the x-axis represents POI categories ranging from 0 to 21, and the y-axis represents the corresponding relative intensity. This visualisation highlights differences in relative POI activity intensity among the clusters. Although the same clusters exhibit broadly similar characteristics across different spatial scales, subtle distinctions emerge due to variations in the number and categories of POIs captured at each spatial resolution. Descriptive names were assigned to each cluster to reflect their geographic locations and functional roles.

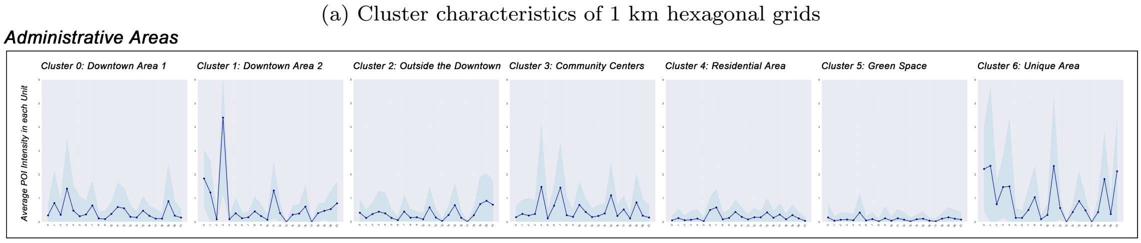
Cluster 0, designated "Downtown Area 1," exhibits the highest relative intensity at the 1 km hexagonal scale, driven by prominent nightlife venues, hotel accommodations, and upscale theme restaurants. At the administrative area and traffic analysis zone (TAZ) scales, Cluster 0 demonstrates moderate intensity without any standout POI categories.

Clusters 1 through 5 maintain consistent POI profiles across all spatial scales. Notably, Cluster 1 exhibits high relative activity intensity in hotel and short-term accommodation features at all three scales, while Cluster 5 displays lower intensity and sparse POI counts, predominantly located on the urban periphery or adjacent to natural green spaces.

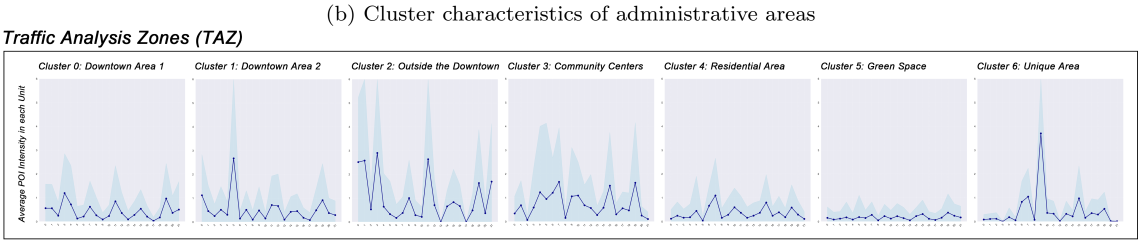
Cluster 6 exhibits distinct functional characteristics and spatial distributions across all three scales, setting it apart from other clusters. At the 1 km hexagonal scale, Cluster 6 comprises cafes and bakeries located in the centre of Hong Kong. At the administrative area scale, it appears exclusively in Hong Kong, with spatial distribution and functional characteristics similar to the 1 km hexagon. Conversely, at the traffic analysis zone (TAZ) scale, it occurs only in Singapore, corresponding to the city's major green and recreational areas within the central park region.



(a) Cluster characteristics of 1 km hexagonal grids



(b) Cluster characteristics of administrative areas



(c) Cluster characteristics of traffic analysis zones

Figure 9: Cluster characteristics across three spatial scales (relative POI activity intensity of the 22 POI categories).

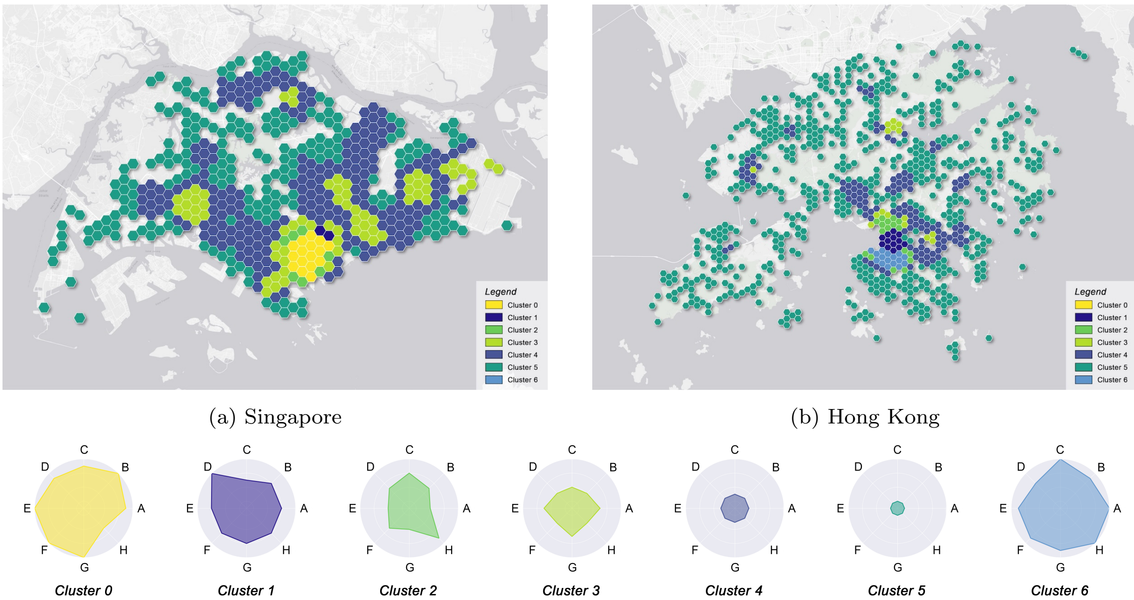
Figure 10, Figure 11, and Figure 12 illustrate the spatial distribution of the clustering results across the two cities. Specifically, Figure 10 presents the clustering results using 1 km hexagons as the analytical unit, while Figure 11 and Figure 12 show the results based on administrative areas and traffic analysis zones (TAZs), respectively.

To enhance the interpretability of POI characteristics in the results, we aggregated the feature values of 22 category vectors into a higher-level grouping structure. As established in Table 3, these categories were organised into eight functional groups. For each cluster, we generated radar charts to visualise the rankings of the aggregated group values. To facilitate cross-cluster comparison of the same group, we converted the group value into discrete scores from 1 to 7, where the highest feature value across clusters was assigned a score of 7 and the lowest a score of 1. Intermediate values were mapped using integer scores between 2 and 6.

Figure 10 illustrates the hexagon-scale clustering results, wherein the cores of Singapore and Hong Kong (Clusters 0 and 6, respectively) occupy distinct clusters. This distinction arises because Singapore's Cluster 0 demonstrates high category feature value in hotels, short-term accommodations, and upscale restaurants, whereas Hong Kong's Cluster 6 is driven by light dining venues, including cafés and bakeries. Both cities also share Cluster 1 as a secondary downtown cluster, characterised by significantly higher hotel intensity compared to other categories. Additionally, Cluster 2 emerges as a sub-centre cluster in both cities, reflecting a balanced mix of POI functions without dominance by any single category. Clusters 3 and 4 correspond to local neighbourhood centres; both are dominated by casual dining and hawker food facilities, although Cluster 3 exhibits higher feature value than Cluster 4, explaining its identification as community centres. Cluster 5 represents predominantly park and natural landscape areas, with some vacant land exhibiting minimal POI presence.

Administrative areas provide clear delineation of localised community units. As shown in Figure 11, Singapore exhibits a more cohesive community structure compared to the hexagonal scale results. The entire city centre, including Sentosa, Chinatown, and Bugis, is classified as Cluster 0, reflecting a unified urban core with distinct functional identity, though with lower feature value than observed at the 1 km hexagonal scale. The Newton district, known for its hawker centres, corresponds to Cluster 2, which represents areas characterised by mixed residential and local food-related activities. In contrast, Hong Kong displays a more fragmented pattern. Cluster 0 occupies only a small portion of its urban core, while Cluster 2 is predominantly distributed across large residential zones. Within central Hong Kong, Clusters 0, 1, and 6 delineate high-density functional areas: Clusters 0 and 1 are dominated by hotels and accommodation-related POIs, whereas Cluster 6, unique to Hong Kong, is characterised by light dining venues, consistent with the patterns observed at the hexagonal scale.

At the traffic analysis zone (TAZ) scale (Figure 12), Singapore exhibits a clear yet fragmented pattern of functional clustering. The commercial core (Cluster 0) is divided into several non-contiguous units, effectively delineating key commercial corridors such as Orchard Road and Clarke Quay while also revealing spatial fragmentation within the urban core. Consistent with the administrative area results, the Merlion precinct is classified as Cluster 1, encompassing recreational, park, and waterfront functions. Cluster 3 captures the city's community centres, where dining facilities are concentrated, though the overall feature value is lower than that of the commercial core. In contrast, Hong Kong presents a more compact and topographically constrained pattern. Cluster 1 corresponds to major country parks and natural landscapes, highlighting the city's distinctive terrain and abundant green spaces. Cluster 3 also represents community centres, though these areas are smaller and more scattered compared to their Singaporean counterparts.

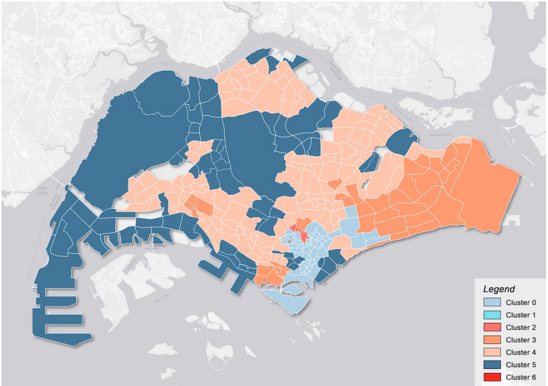


(a) Singapore

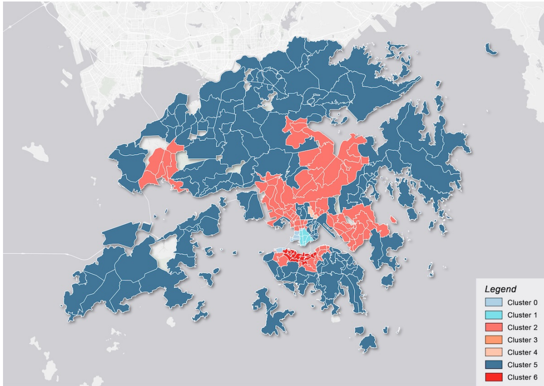
(b) Hong Kong

(c) Intensity radar chart of 8 groups(A-H) across different clusters.

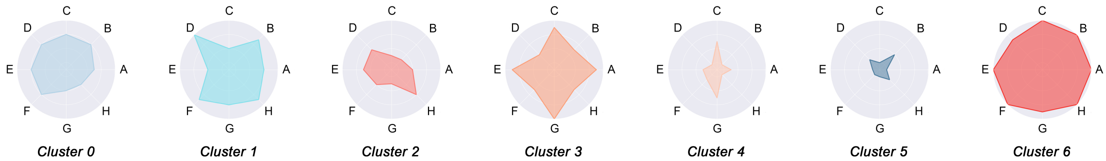
Figure 10: GNN + Clustering results across cities (1 km hexagon).



(a) Singapore

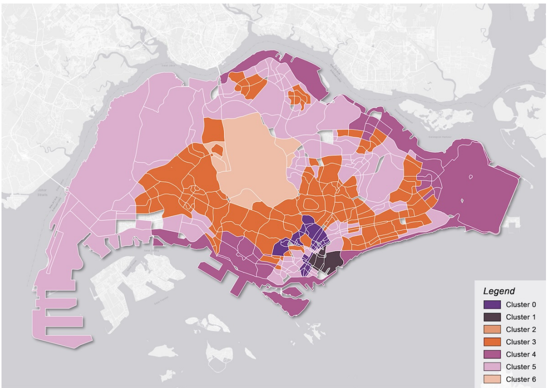


(b) Hong Kong

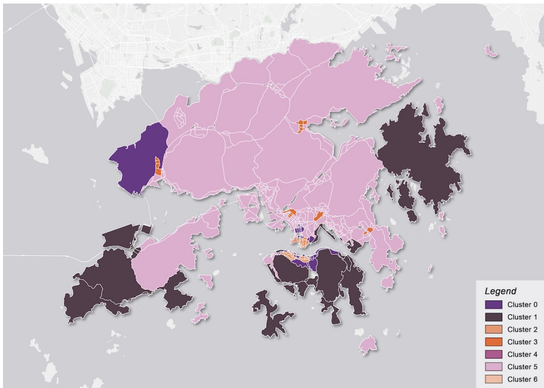


(c) Intensity radar chart of 8 groups(A-H) across different clusters.

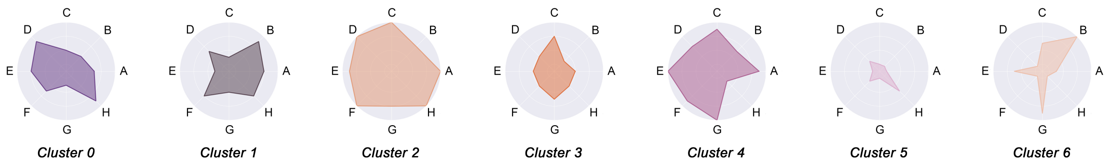
Figure 11: GNN + Clustering results across cities (Administrative areas).



(a) Singapore



(b) Hong Kong



(c) Intensity radar chart of 8 groups(A-H) across different clusters.

Figure 12: GNN + Clustering results across cities (Traffic analysis zones).

5. Discussion

This section synthesises our methodological contributions and their implications for urban planning practice. Figure 13 provides an overview of our multi-layer analytical framework, illustrating how different components of our approach support specific planning applications across scales.

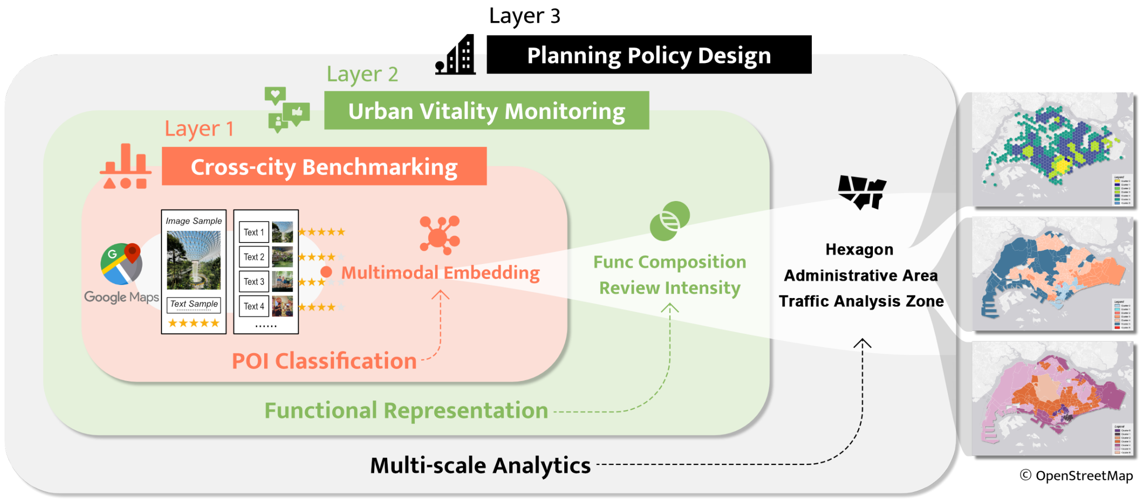


Figure 13: From place review data to planning applications: a multi-layer synthesis.

5.1. Multimodal POI classification for cross-city benchmarking

We demonstrate that review-derived POI embeddings provide a useful representation for analysing urban functions. By integrating unsupervised learning over textual and visual signals, the multimodal embeddings achieve the strongest semantic agreement with the harmonised metadata-derived *L1_types* reference labels for Singapore and Hong Kong when compared with single-modality approaches. The approach integrates heterogeneous data, including text, images, and check-ins, into vector embeddings within a unified neural space, addressing the limitation of relying solely on traditional POI names or categories. This capability is particularly valuable for Google Maps data, where POI functional categories are often irregular and unreliable.

We leverage place review data to aggregate POIs from both cities into a unified classification scheme, enabling cross-city functional comparison. As shown in Table 3, our 22 clusters depart from conventional taxonomies by drawing finer distinctions from text semantics and image scenes, for example, food- and dining-related venues are separated into at least four categories reflecting different price levels and use contexts. Figure 5 illustrates one divergence: Cluster 7 contains substantially more POIs in Singapore than in Hong Kong, corresponding to Singapore's hawker centres and highlighting distinctive culinary infrastructure. For reporting clarity, we consolidated the 22 clusters into eight top-level categories. Overall, our approach integrates visual environmental cues and user-generated textual evidence to refine POI functional categories, supporting planning strategies that consider user-described place functions rather than relying only on administrative or metadata-based designations.

5.2. Relative POI activity intensity for areal functional representation

When aggregating POIs to spatial units, we go beyond traditional frequency-based models by using place review counts to construct POI-specific relative activity-intensity weights. This allows the representation to distinguish POIs that are relatively more active within comparable functional contexts. Review counts provide POI-linked activity-related signals, but they are not used as direct measurements of actual human activity. Instead, they are scaled within each inferred functional cluster to construct relative POI activity intensity. Our innovation uses review counts to monitor both the relative level and dynamics of urban function over flexible time windows, enabling responsive planning interventions. Given rapid POI turnover, we restrict the analysis to reviews from the most recent year.

We validate our approach by computing cosine similarity on hexagon-level representations to identify similar functional areas across cities, demonstrating that POI aggregation effectively reveals urban regions with homogeneous functional structures. Review counts serve as indicators of relative POI activity intensity, helping distinguish areas where specific functional categories are more strongly represented in POI-linked review activity from areas where they are less represented. To mitigate the dominance of extreme values in heavy-tailed distributions, a logarithmic transformation is applied to this relative intensity, and the transformed distribution is truncated at the 95th percentile (Gao et al., 2017).

5.3. Multi-scale spatial analysis for planning policy design

We introduce GNNs combined with k-means clustering to account for spatial context and interactions between neighbouring units. We evaluate the approach at three spatial scales, including 1 km hexagons, administrative areas, and traffic analysis zones (TAZs). Discrepancies across scales underscore the importance of scale-sensitive policy design.

The hexagonal scheme, with uniform cell geometry, reveals spatial gradients in relative POI activity intensity, providing a consistent baseline for analysing urban spatial structure, which is valuable for location-based service planning and retail site selection. By contrast, administrative areas emphasise inter-community differences and expose distinct governance patterns, for example, the linear subdivision in eastern Singapore versus the core-diffusion pattern in Hong Kong, which are valuable for policy coordination and resource allocation.

The TAZ scale yields distinct clustering patterns reflecting road network organisation. Many suburban residential areas are classified as green or open-space clusters, particularly in Hong Kong, reflecting sparse road networks and larger TAZs with low POI density variation. Urban cores, with dense road networks and diverse POI distributions, are subdivided into multiple clusters. This scale-dependent pattern offers insights for transportation infrastructure investment and traffic management.

Across all three scales, city centres are partitioned into multiple distinct clusters, with some clusters appearing only in one city, providing evidence of unique functional areas and their potential value as urban landmarks. To the best of our knowledge, this is the first study using Google Maps review data to reveal systematic differences in functional composition and relative POI activity intensity between urban cores and suburban areas. This granular differentiation provides actionable intelligence for planners: identifying landmark functional zones, guiding place-making initiatives, benchmarking commercial district performance, and informing targeted regeneration strategies that preserve local character while promoting economic vitality.

5.4. Limitations and practical implications

However, several limitations should be considered when applying this framework to urban planning practice. First, place reviews are user-generated and inherently subjective and noisy. Factors such as merchants posting overly positive reviews or users submitting malicious or misleading feedback may introduce information bias. These issues are not unique to place reviews but are common to other forms of volunteered geographic information (VGI) and user-generated geographic information, such as Twitter data, which typically exhibit an irreducible subjective nature that constrains more precise assessment (Khosravi Kazazi et al., 2024). Second, many POIs lacking review data were excluded from the dataset. As a result, discrepancies may exist between the functional characteristics inferred from place reviews and the actual land-use functions of POIs. In addition, areas with similar functional compositions in review data may be mistakenly identified as similar urban areas. For example, POIs located within airports are often dominated by restaurants and public amenities, which may lead them to be misclassified as community-oriented spaces, while office-dominated environments such as central business districts (CBDs) are less adequately represented. Third, as the proposed framework relies entirely on unsupervised learning, its applicability to different research questions and planning contexts requires careful justification and, where possible, validation against local ground-truth data.

Despite these limitations, the proposed method provides urban planners with a scalable, cost-effective, and data-driven complement to traditional surveys and land-use records. A major drawback of earlier approaches that relied on Twitter data as a proxy for check-ins is the frequent absence of reliable metadata, which constrains functional interpretation and spatial accuracy (Fonte et al., 2017). In contrast, place reviews obtained through Google Maps are directly linked to specific POIs, enabling rapid functional assessment across diverse urban contexts. This linkage facilitates cross-city benchmarking and supports iterative policy refinement as review data accumulates over time. The approach is particularly valuable for planning applications that require frequent updates (e.g., monitoring commercial district vitality) or large-scale spatial coverage (e.g., regional functional mapping), where traditional field surveys may be prohibitively expensive or time-consuming.

6. Conclusion

This study represents the first effort to apply place review data for the inference of POI categories and urban functional classification. Specifically, this paper introduces a novel multimodal approach for clustering and interpreting POIs using text and image data from place reviews, and evaluates its semantic consistency against a harmonised metadata-derived POI reference taxonomy. Beyond this novelty, it demonstrates the feasibility and broader potential of place reviews as a valuable data source for urban studies. By integrating heterogeneous multi-source data, we construct an areal functional representation that better captures user-described place functions and activity patterns in real-world urban environments. This approach overcomes the static, subjective, and oversimplified drawbacks of traditional land-use classifications and POI datasets, while also identifying areas with mixed functions and indistinct boundaries. Importantly, the clustering framework does not require manual annotations for model training, which supports transferability across cities with different POI metadata structures. It also enables cross-city comparative analyses and offers broad applicability and scalability for urban renewal and precision management.

Additionally, we use place review counts and POI uniqueness weighting to construct areal POI category vectors. This representation captures both the functional structure of areas and the relative activity intensity of inferred functions, thereby facilitating the identification of similar areas within and across cities. The feasibility of this approach is validated by computing cosine similarity between a given input unit and all other areas. By integrating GNN-based neighbourhood learning with k-means clustering, we more effectively model spatial interactions and preserve spatial continuity. Overall, this approach provides a practical solution for urban functional classification.

We evaluated the framework in both Singapore and Hong Kong. The experimental results demonstrate its effectiveness and flexibility. Even under imbalanced data distributions, the framework produces interpretable and semantically consistent functional clusters both within and between cities. Furthermore, through controlled experiments at three spatial scales, we gained insights into functional gradients, community distributions, and traffic organisation. These comparisons further validate the robustness of the classification results.

Looking ahead, future research could leverage place review data from multiple years to examine the dynamic evolution of urban functions over time, such as urban expansion and changes in community vitality. In addition, rating information contained in place reviews could be incorporated as a representative indicator of place satisfaction. These data and indicators are critical for urban management and planning initiatives, including transportation improvements, community renewal, and social equity.

CRediT authorship contribution statement

Haixiao Liu: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sijie Yang:** Conceptualization, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Mahmoud Abdelrahman:** Conceptualization, Data collection & curation, Methodology, Writing – review & editing. **Yihan Zhu:** Conceptualization, Data collection & curation, Writing – review & editing. **Xiaobing Wei:** Conceptualization, Methodology, Writing – review & editing. **Filip Biljecki:** Conceptualization, Methodology, Project administration, Funding acquisition, Writing – review & editing, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

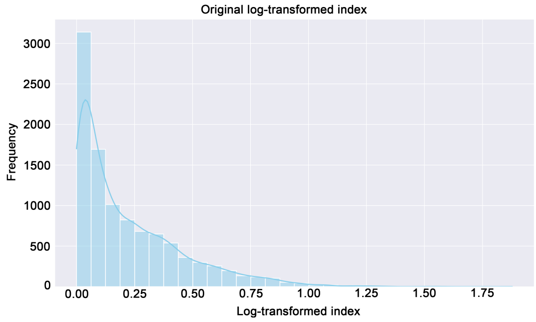
During the preparation of this work, the authors used ChatGPT 5.1 to improve the language of the manuscript and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgements

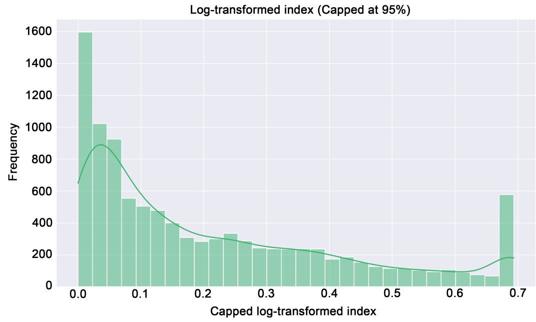
We thank our colleagues at the NUS Urban Analytics Lab for the discussions, and especially Yang Chen and Xiucheng Liang for their valuable suggestions. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133. SY is supported by the NUS Research Scholarship (NUSGS-CDE DO IS AY22&L GR-SUR0600042). XW acknowledges doctoral support from the China Scholarship Council (Grant No. 202106270030), Ghent University (Grant No. BOFCHN2021001101), and EIT Urban Mobility.

Appendix A. Additional Data Description

Figure A.14 displays histograms and kernel density estimations (KDEs) for Singapore and Hong Kong, before and after truncation.

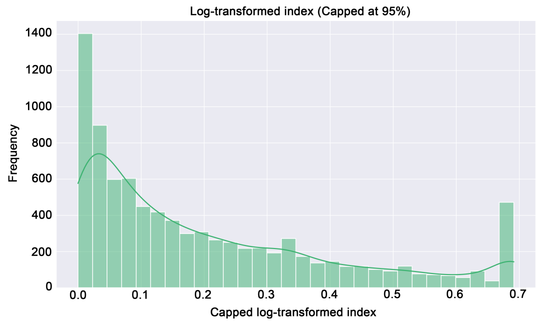


(a) Original Review Index in Singapore



(b) Capped Review Index in Singapore

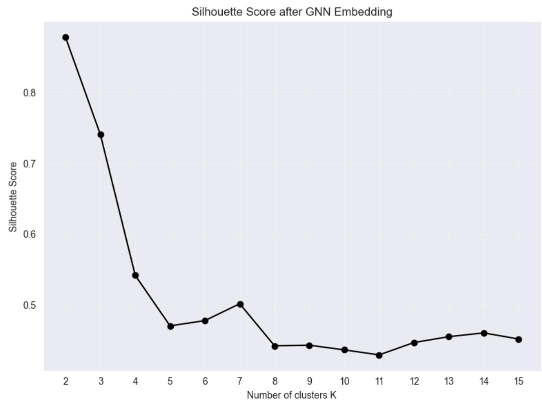
(c) Original Review Index in Hong Kong



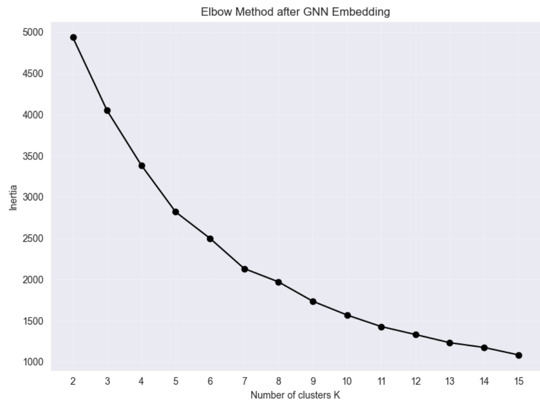
(d) Capped Review Index in Hong Kong

Figure A.14: Distribution comparison of index_log before and after 95% value capping in Singapore and Hong Kong.

Figure A.15 presents the evaluation results for the clustering analysis in Section 4.3. We finally selected 7 clusters based on a combined assessment of the silhouette score and elbow method. First, the silhouette score shows a local optimum at 7 clusters, while solutions with fewer than four clusters may be insufficient to capture the differences in urban functional classifications. More importantly, the elbow curve becomes noticeably flatter after 7 clusters.



(a) Silhouette score



(b) Elbow method

Figure A.15: Validation of the k-means clustering results using the silhouette score and elbow method.

Figure A.16 shows a three-level classification cascade that transforms Google Maps image–text reviews into interpretable urban functional areas. POIs are first embedded and clustered into 22 review-derived groups, then aggregated into spatial units as 22-dimensional vectors. A GNN model then learns neighbourhood relationships and classifies spatial units into seven functional types, making the framework both interpretable and adaptable to different cities or zoning systems.

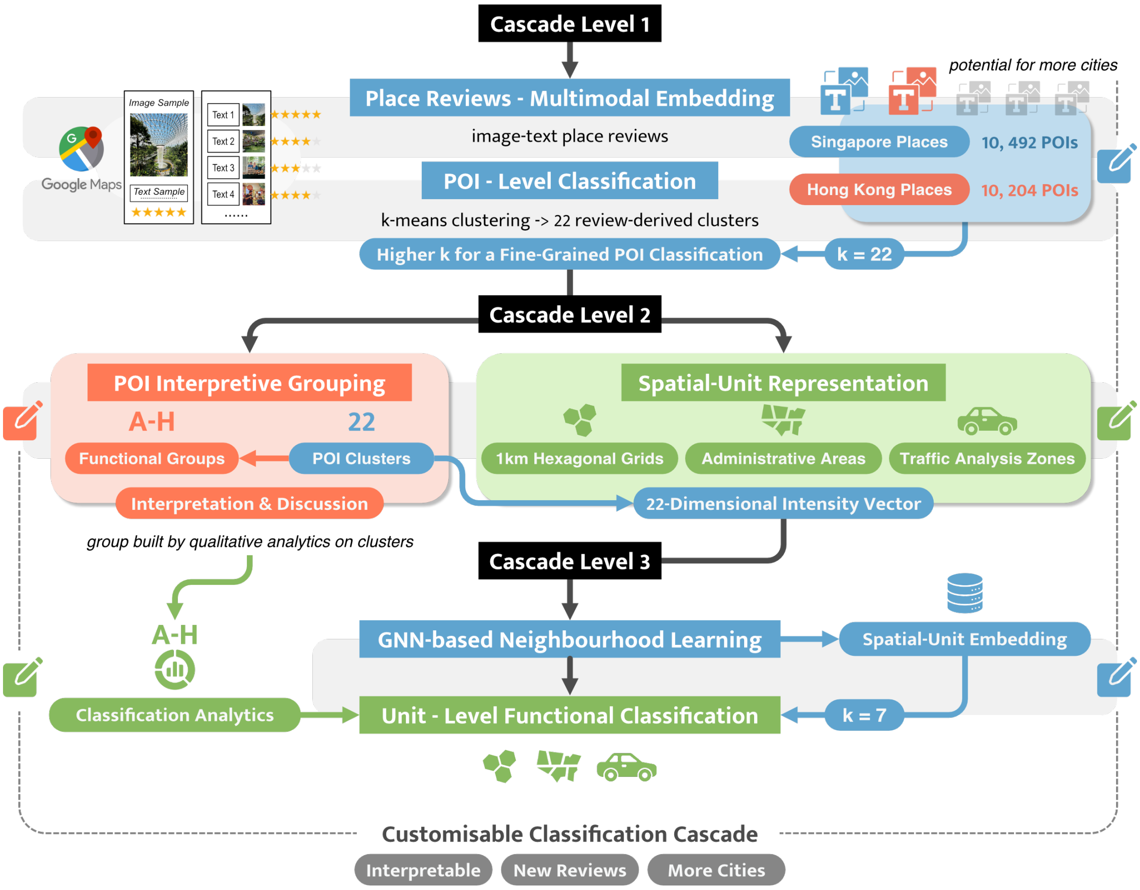


Figure A.16: Interpretable and customisable classification cascade from place reviews to urban functional areas.

References

Bibri, S.E., 2021. Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning. Computational Urban Science 1, 8.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and gis: A review. Landscape and Urban Planning 215, 104217.

Burgess, E.W., 2015. The growth of the city: an introduction to a research project, in: The city reader. Routledge, pp. 212–220.

Castells, M., 2020. Space of flows, space of places: Materials for a theory of urbanism in the information age, in: The city reader. Routledge, pp. 240–251.

Chen, J., Fan, R., Niu, H., Xu, Z., Yan, J., Song, W., Feng, R., 2025. A unified multimodal learning method for urban functional zone identification by fusing inner-street visual–textual information from street-view and satellite images. International Journal of Applied Earth Observation and Geoinformation 142, 104685.

Chen, M., Cai, Y., Guo, S., Sun, R., Song, Y., Shen, X., 2024. Evaluating implied urban nature vitality in san francisco: An interdisciplinary approach combining census data, street view images, and social media analysis. Urban forestry & urban greening 95, 128289.

Chen, X., Ye, X., Widener, M.J., Delmelle, E., Kwan, M.P., Shannon, J., Racine, E.F., Adams, A., Liang, L., Jia, P., 2022. A systematic review of the modifiable areal unit problem (maup) in community food environmental research. Urban Informatics 1, 22.

Chen, Y., Yang, X., Li, X., Yang, L., Cheng, S., Tang, G., Biljecki, F., 2026. Vertical 15-minute city: Modeling urban density and functional mix with multi-source geospatial data. Cities 169, 106516.

Chuang, H.M., Chang, C.H., Kao, T.Y., Cheng, C.T., Huang, Y.Y., Cheong, K.P., 2016. Enabling maps/location searches on mobile devices: Constructing a poi database via focused crawling and information extraction. International journal of geographical information science 30, 1405–1425.

Deng, H., Liu, K., Feng, J., 2024. Understanding the impact of modifiable areal unit problem on urban vitality and its built environment factors. Geo-Spatial Information Science , 1–17.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pp. 4171–4186.

Dijkstra, L., Poelman, H., Veneri, P., et al., 2019. The EU-OECD definition of a functional urban area. volume 10. Oecd Publishing Paris, France.

Du, S., Du, S., Liu, B., Zhang, X., 2021. Mapping large-scale and fine-grained urban functional zones from vhr images using a multi-scale semantic segmentation network and object based approach. Remote Sensing of Environment 261, 112480.

Fan, C., Yang, Y., Mostafavi, A., 2024. Neural embeddings of urban big data reveal spatial structures in cities. Humanities and Social Sciences Communications 11, 1–15.

Fan, J., Thakur, G., 2023. Towards poi-based large-scale land use modeling: spatial scale, semantic granularity, and geographic context. International Journal of Digital Earth 16, 430–445.

Fan, R., Niu, H., Xu, Z., Chen, J., Feng, R., Wang, L., 2025. Refined urban informal settlements mapping at agglomeration-scale with the guidance of background-knowledge from easy-accessed crowdsourced geospatial data. IEEE Transactions on Geoscience and Remote Sensing .

Fang, G., Gao, Z., Tian, L., Fu, M., 2022. What drives urban carbon emission efficiency?– spatial analysis based on nighttime light data. Applied Energy 312, 118772.

Fonte, C.C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J.J., Bayas, J.C.L., See, L., Vatseva, R., 2017. Assessing vgi data quality. Mapping and the citizen sensor , 137–163.

Gao, S., Janowicz, K., Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. Transactions in GIS 21, 446–467.

Gao, S., Liu, Y., Kang, Y., Zhang, F., 2021. User-generated content: A promising data source for urban informatics, in: Urban informatics. Springer, pp. 503–522.

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

Harrison, S., Dourish, P., 1996. Re-place-ing space: the roles of place and space in collaborative systems, in: Proceedings of the 1996 ACM conference on Computer supported cooperative work, pp. 67–76.

Hollenstein, L., Purves, R., 2010. Exploring place through user-generated content: Using flickr tags to describe city cores. Journal of Spatial Information Science , 21–48.

Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., Gong, X., 2021. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. Computers, Environment and Urban Systems 87, 101619.

Huang, H., Yao, X.A., Krisp, J.M., Jiang, B., 2021. Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. Computers, Environment and Urban Systems 90, 101712.

Huang, W., Wang, J., Cong, G., 2024. Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. International Journal of Geographical Information Science 38, 1414–1442.

Ilieva, R.T., McPhearson, T., 2018. Social-media data for urban sustainability. Nature Sustainability 1, 553–565.

Jenkins, P., Farag, A., Wang, S., Li, Z., 2019. Unsupervised representation learning of spatial data via multimodal embedding, in: Proceedings of the 28th ACM international conference on information and knowledge management, pp. 1993–2002.

Kajosaari, A., 2024. Scale dimensions in public participation gis: an overview for planning and research. GeoJournal 89, 197.

Kang, C., Fan, D., Jiao, H., 2021. Validating activity, time, and space diversity as essential components of urban vitality. Environment and Planning B: Urban Analytics and City Science 48, 1180–1197.

Khan, A.M., Loan, F.A., 2022. Exploring the reviews of google maps to assess the user opinions about public libraries. Library Management 43, 601–615.

Khosravi Kazazi, A., Rahmani, Y., Amiri, F., Hosseinali, F., 2024. A school of thought on vgi challenges: A literature review. Papers in Applied Geography 10, 53–68.

Kong, B., Ai, T., Zou, X., Yan, X., Yang, M., 2024. A graph-based neural network approach to integrate multi-source data for urban building function classification. Computers, Environment and Urban Systems 110, 102094.

Koroteev, M.V., 2021. Bert: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943 .

Lei, B., Liu, P., Milojevic-Dupont, N., Biljecki, F., 2024. Predicting building characteristics at urban scale using graph neural networks and street-level context. Computers, Environment and Urban Systems 111, 102129.

Leiras, A., Eusébio, C., 2024. Perceived image of accessible tourism destinations: a data mining analysis of google maps reviews. Current Issues in Tourism 27, 2584–2602.

Li, H., Hecht, B., 2021. 3 stars on yelp, 4 stars on google maps: a cross-platform examination of restaurant ratings. Proceedings of the ACM on Human-Computer Interaction 4, 1–25.

Li, X., Li, Y., Jia, T., Zhou, L., Hijazi, I.H., 2022. The six dimensions of built environment on urban vitality: Fusion evidence from multi-source data. Cities 121, 103482.

Listorti, E., Pastore, E., Alfieri, A., 2023. How to direct patients to high-volume hospitals: exploring the influencing drivers. BMC Health Services Research 23, 1269.

Liu, H., Xu, Y., Tang, J., Deng, M., Huang, J., Yang, W., Wu, F., 2020a. Recognizing urban functional zones by a hierarchical fusion method considering landscape features and human activities. Transactions in GIS 24, 1359–1381.

Liu, W., Wu, W., Thakuriah, P., Wang, J., 2020b. The geography of human activity and land use: A big data approach. Cities 97, 102523.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., Hong, Y., 2017. Classifying urban land use by integrating remote sensing and social media data. International Journal of Geographical Information Science 31, 1675–1696.

Liu, X., Long, Y., 2016. Automated identification and characterization of parcels with openstreetmap and points of interest. Environment and Planning B: Planning and Design 43, 341–360.

Liu, Y., Ding, J., Fu, Y., Li, Y., 2023. Urbankg: An urban knowledge graph system. ACM Transactions on Intelligent Systems and Technology 14, 1–25.

Lu, H., Uddin, S., 2024. Unsupervised machine learning for disease prediction: a comparative performance analysis using multiple datasets. Health and Technology 14, 141–154.

Luca, D., Terrero-Davila, J., Stein, J., Lee, N., 2023. Progressive cities: Urban–rural polarisation of social values and economic development around the world. Urban Studies 60, 2329–2350.

Luo, Y., Liu, Q., Liu, Z., 2021. Stan: Spatio-temporal attention network for next location recommendation, in: Proceedings of the web conference 2021, pp. 2177–2185.

Ma, M., Zhao, H., Wang, J., Lai, Y., 2026. Urban multi-scale information graph and fusion model. Urban Informatics 5, 7.

Martí, P., Serrano-Estrada, L., Nolasco-Cirugeda, A., 2019. Social media data: Challenges, opportunities and limitations in urban studies. Computers, Environment and Urban Systems 74, 161–174.

Mathayomchan, B., Taecharungroj, V., 2020. "how was your meal?" examining customer experience using google maps reviews. International Journal of Hospitality Management 90, 102641.

McKenzie, G., Romm, D., 2021. Measuring urban regional similarity through mobility signatures. Computers, Environment and Urban Systems 89, 101684.

Mehmood, M., Shahzad, A., Zafar, B., Shabbir, A., Ali, N., 2022. Remote sensing image classification: A comprehensive review and applications. Mathematical problems in engineering 2022, 5880959.

Middleton, S.E., Kordopatis-Zilos, G., Papadopoulos, S., Kompatsiaris, Y., 2018. Location extraction from social media: Geoparsing, location disambiguation, and geotagging. ACM Transactions on Information Systems (TOIS) 36, 1–27.

Niu, H., Silva, E.A., 2020. Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. Journal of Urban Planning and Development 146, 04020007.

Niu, K., Xing, Y., 2024. Opinion distribution: spatial sentiment analysis of online restaurant reviews through bert model and gis, in: Fourth International Conference on Signal Processing and Machine Learning (CONF-SPML 2024), SPIE. pp. 208–218.

Olson, A.W., Calderón-Figueroa, F., Bidian, O., Silver, D., Sanner, S., 2021. Reading the city through its neighbourhoods: Deep text embeddings of yelp reviews as a basis for determining similarity and change. Cities 110, 103045.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. the Journal of machine Learning research 12, 2825–2830.

Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., Liao, Q., 2017. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. International Journal of Automation and Computing 14, 503–519.

Psyllidis, A., Gao, S., Hu, Y., Kim, E.K., McKenzie, G., Purves, R., Yuan, M., Andris, C., 2022. Points of interest (poi): A commentary on the state of the art, challenges, and prospects for the future. Computational urban science 2, 20.

Qian, J., Liu, Z., Du, Y., Liang, F., Yi, J., Ma, T., Pei, T., 2021. Quantify city-level dynamic functions across china using social media and pois data. Computers, Environment and Urban Systems 85, 101552.

Qing, Y., Wu, H., Qi, K., Gui, Z., Liu, Y., Li, Z., Li, R., 2024. Integrating street-view imagery and points of interest for refining population spatialization: a case study in wuhan city. Sustainable Cities and Society 115, 105883.

Rabiei-Dastjerdi, H., McArdle, G., Aghajani, M.A., 2022. User-generated data in cultural mapping: Analyzing google point of interest reviews in dublin. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4, 107–112.

Schneider, A., Woodcock, C.E., 2008. Compact, dispersed, fragmented, extensive? a comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. Urban studies 45, 659–692.

Schonlau, M., 2002. The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses. The Stata Journal 2, 391–402.

Shen, Y., Karimi, K., 2016. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. Cities 55, 9–21.

Shi, Z., Silvennoinen, H., Chadzynski, A., von Richthofen, A., Kraft, M., Cairns, S., Herthogs, P., 2023. Defining archetypes of mixed-use developments using google maps api data. Environment and Planning B: Urban Analytics and City Science 50, 1607–1623.

Smith, R.H., 1965. Method and purpose in functional town classification. Annals of the Association of American Geographers , 539–548.

Soliman, A., Soltani, K., Yin, J., Padmanabhan, A., Wang, S., 2017. Social sensing of urban land use based on analysis of twitter users' mobility patterns. PloS one 12, e0181657.

Srivastava, S., Vargas-Munoz, J.E., Tuia, D., 2019. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. Remote sensing of environment 228, 129–143.

Sun, K., Hu, Y., Ma, Y., Zhou, R.Z., Zhu, Y., 2023. Conflating point of interest (poi) data: A systematic review of matching methods. Computers, Environment and Urban Systems 103, 101977.

Tao, Y., Liu, W., Chen, J., Gao, J., Li, R., Wang, X., Zhang, Y., Ren, J., Yin, S., Zhu, X., et al., 2025. A graph-based multimodal data fusion framework for identifying urban functional zone. International Journal of Applied Earth Observation and Geoinformation 136, 104353.

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. Economic geography 46, 234–240.

Verburg, P.H., de Nijs, T.C., van Eck, J.R., Visser, H., de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems 28, 667–690.

Wang, L., Long, T., Jiang, W., Adam, E., Wen, C., Jiao, W., He, G., 2025a. Economic well-being assessment: a review of traditional and remote sensing approaches. International Journal of Digital Earth 18, 2504137.

Wang, S., Zhao, C., Jiang, Q., Zhu, D., Ma, J., Sun, Y., 2025b. Application of graph convolutional neural networks and multi-sources data on urban functional zones identification, a case study of changchun, china. Sustainable Cities and Society 119, 106116.

Wang, W., Lu, Z., 2024. Few-shot bronze vessel classification via siamese fourier networks. Scientific Reports 14, 18011.

Wang, Z., Ye, X., Lee, J., Chang, X., Liu, H., Li, Q., 2018. A spatial econometric modeling of online social interactions using microblogs. Computers, Environment and Urban Systems 70, 53–58.

Wong, D.W., 2004. The modifiable areal unit problem (maup), in: WorldMinds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the association of American geographers 1904–2004. Springer, pp. 571–575.

Wu, C., Ye, X., Ren, F., Du, Q., 2018. Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in shenzhen, china. Cities 77, 104–116.

Wu, H., Lin, A., Clarke, K.C., Shi, W., Cardenas-Tristan, A., Tu, Z., 2021. A comprehensive quality assessment framework for linear features from volunteered geographic information. International Journal of Geographical Information Science 35, 1826–1847.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S., 2020. A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems 32, 4–24.

Xia, C., Yeh, A.G.O., Zhang, A., 2020. Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five chinese megacities. Landscape and Urban Planning 193, 103669.

Xu, Y., Zhou, B., Jin, S., Xie, X., Chen, Z., Hu, S., He, N., 2022. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. Computers, Environment and Urban Systems 95, 101807.

Xue, J., Jiang, N., Liang, S., Pang, Q., Yabe, T., Ukkusuri, S.V., Ma, J., 2022. Quantifying the spatial homogeneity of urban road networks via graph neural networks. Nature Machine Intelligence 4, 246–257.

Yan, B., Janowicz, K., Mai, G., Gao, S., 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts, in: Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 1–10.

Yan, Y., Wen, H., Zhong, S., Chen, W., Chen, H., Wen, Q., Zimmermann, R., Liang, Y., 2024. Urbanclip: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web, in: Proceedings of the ACM Web Conference 2024, pp. 4006–4017.

Yang, X., Bo, S., Wang, J., 2024. Classifying urban functional zones by integrating the homogeneity and structural similarity of pois. Journal of Urban Planning and Development 150, 04024052.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model. International Journal of Geographical Information Science 31, 825–848.

Ye, C., Zhang, F., Mu, L., Gao, Y., Liu, Y., 2021. Urban function recognition by integrating social media and street-level imagery. Environment and Planning B: Urban Analytics and City Science 48, 1430–1444.

Yoshihara, H., Yoneoka, D., 2014. Understanding the statistics and limitations of large database analyses. Spine 39, 1311–1312.

Yu, F., Cui, L., Guo, W., Lu, X., Li, Q., Lu, H., 2020. A category-aware deep model for successive poi recommendation on sparse check-in data, in: Proceedings of the web conference 2020, pp. 1264–1274.

Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and pois, in: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 186–194.

Yue, Y., Zhuang, Y., Yeh, A.G., Xie, J.Y., Ma, C.L., Li, Q.Q., 2017. Measurements of poi-based mixed use and their relationships with neighbourhood vibrancy. International Journal of Geographical Information Science 31, 658–675.

Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z.R., Gu, C., 2019. Beyond word2vec: An approach for urban functional region extraction and identification by combining place2vec and pois. Computers, Environment and Urban Systems 74, 1–12.

Zhang, X., Du, S., Wang, Q., Zhou, W., 2018. Multiscale geoscene segmentation for extracting urban functional zones from vhr satellite images. Remote Sensing 10, 281.

Zhang, X., Du, S., Zheng, Z., 2020. Heuristic sample learning for complex urban scenes: Application to urban functional-zone mapping with vhr images and poi data. ISPRS Journal of Photogrammetry and Remote Sensing 161, 1–12.

Zhao, W., Bo, Y., Chen, J., Tiede, D., Blaschke, T., Emery, W.J., 2019. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and openstreetmap (osm). ISPRS Journal of Photogrammetry and Remote Sensing 151, 237–250.

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., Cai, H., Abedin, J., Mandal, D., 2022a. Victimfinder: Harvesting rescue requests in disaster response from social media with bert. Computers, Environment and Urban Systems 95, 101824.

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., et al., 2024. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. International Journal of Machine Learning and Cybernetics , 1–65.

Zhou, Y., Yang, G., Yan, B., Cai, Y., Zhu, Z., 2022b. Point-of-interest recommendation model considering strength of user relationship for location-based social networks. Expert Systems with Applications 199, 117147.

Zhu, L., Shi, F., 2022. Spatial and social inequalities of job accessibility in kunshan city, china: Application of the amap api and mobile phone signaling data. Journal of Transport Geography 104, 103451.

Zhu, Y., Brenken, M., Biljecki, F., Zhang, Y., 2025. Unveiling the meaning in the image of the city: A novel approach using place reviews and large language models. Environment and Planning B: Urban Analytics and City Science , 23998083251369143.