Proceedings of the CONF-MLA 2026 Symposium: Explainable Computing, Modeling & Data Science in Complex Systems
DOI: 10.54254/2755-2721/2026.GF37457

Construction of Urban Functional Zones Identification Model Based on TF-IDF and Public Perception-Weighted Algorithms

Shuo Yang

School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China
yangshuo9006@outlook.com

Abstract. Urban neighborhoods used to be fairly straightforward—residential here, commercial there. That is less and less true, and planners need ways to keep up. This paper tries to build one such tool, tested on Pudong New Area in Shanghai. We combine a TF-IDF weighting scheme with survey-based perception scores and feed both into a POI-based functional-zone classifier. TF-IDF handles the spatial-rarity side of the problem. If a POI category clusters heavily in a particular grid cell but barely appears elsewhere in the city, the algorithm gives it a high weight. A cargo terminal near an airport, for instance, is rare city-wide and gets flagged; a convenience store, which is everywhere, does not. The perception scores, borrowed from Zhao (2008), get at something else entirely. They reflect how strongly residents mentally connect different facility types with particular urban functions. Metro stations top the scale at 1.0. Residential compounds sit at 0.15—a gap that frequency-density counting, which gives every POI the same weight, has no mechanism to detect. We gridded Pudong into 1,672 cells of 1 km × 1 km. The baseline frequency-density method scored 66.74% overall accuracy. Our model reached 86.33%. Transportation zones and mixed-use areas saw the biggest gains, which makes sense because those are the zone types where brute-force counting is most prone to error. We also mapped single- and mixed-function zones across the district. The practical takeaway is that folding perception data into a TF-IDF framework yields a noticeably sharper read of urban functional structure.

Keywords: Urban functional zones, POI, TF-IDF, public awareness, frequency-density model

1. Introduction

The diversification of urban functions is one of the more conspicuous results of rapid urbanization, and it has turned functional-zone identification into a question that planners and researchers cannot easily set aside. The conventional response, for the better part of two decades, was to lean on statistical yearbooks, cadastral records, and satellite imagery [1]. These sources do reveal macro-level patterns, but they come with constraints that are hard to ignore. Update cycles tend to be long. Spatial resolution is coarse enough to smooth over important intra-district variation. Perhaps most critically, none of these datasets captures how the people who live and work in an area actually

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GF37457

perceive its function. The rapid spread of Point-of-Interest (POI) data has opened a different avenue. A POI record contains a category label and a coordinate pair, and commercial mapping platforms refresh these records on a rolling basis, so the resulting datasets track the contemporary urban landscape at a granularity that decadal censuses simply cannot match [2, 3].

Most existing POI-based studies rely on what is commonly called the frequency-density method, which amounts to counting the POIs in each category within a spatial unit and assigning the dominant function to whichever category has the highest share [4, 5]. This works well enough when an area serves a single, clear-cut purpose, but it runs into trouble in more complicated settings. Mixed-function zones and transit hubs are cases in point. A busy metro interchange may contain dozens of small restaurants and shops that numerically overshadow a handful of bus terminals and subway entrances; the frequency-density approach would then label the area "commercial" even though its defining feature, from a planning standpoint, is transportation. There is also a less obvious but equally important shortcoming. Urban functions are not determined by physical facilities alone; they are also shaped by how people perceive and make use of those facilities [3, 6]. A method blind to this social dimension risks producing classifications that look neat on a map but do not match lived experience.

Those two problems are what led us to the model in this paper. We are calling it a Public Perception-Weighted TF-IDF scheme, which is a mouthful, but the logic behind it is not complicated. TF-IDF has been around since Salton and Buckley (1988); in information retrieval it measures how important a particular word is to a particular document by weighing within-document frequency against rarity across the whole corpus [7]. Transplanting this to geography turns out to be remarkably natural. Grid cells become "documents." POI categories become "words." A category that piles up in one cell but is scarce across the rest of the city earns a high score, which is the kind of diagnostic signal that frequency-density counting throws away because it treats every POI as interchangeable. We then add perception scores adapted from Zhao (2008) [6], which quantify how strongly survey respondents link different facility types to specific urban functions. On that scale, metro stations come in at 1.0 and residential compounds at 0.15—the spread is large enough that ignoring it noticeably degrades classification accuracy. Combining TF-IDF with these perception weights lets us score POI categories on both their spatial concentration and the degree to which residents actually pay attention to them [3].

For the study site we chose Pudong New Area in Shanghai. We cut the region into 1,672 grid cells of 1000 × 1000 m and classified functional zones from Amap POI data using the proposed model. We then compared results against the conventional frequency-density method. Our model hit 86.33% overall accuracy; the baseline managed 66.74%. Transportation facility zones saw the widest jump (from 53.49% up to 90.65%) and mixed-function zones went from 49.37% to 91.63%.

Three contributions come out of this work. First, the TF-IDF adaptation gives a principled, information-retrieval-grounded way to weigh POI categories, replacing the ad hoc thresholds that most frequency-density studies rely on. Second, adding perception scores means the pipeline picks up on which facilities people mentally flag as important, not just which ones are physically present —so the classification ends up with a layer of social information that pure counting misses. Third, the Pudong application generates concrete spatial evidence about where mixed-function areas form and under what conditions, which is relevant for land-use coordination and facility planning.

DOI: 10.54254/2755-2721/2026.GF37457

2. Literature review

2.1. Overview of urban functional zone identification methods

What function does a particular piece of urban territory serve? Is it residential, commercial, a mix of both, or something else entirely? The question matters for planning, land management, and governance, and the toolbox available for answering it has changed considerably over the past two decades.

Earlier work drew primarily on statistical yearbooks, official land-use maps, and remote-sensing imagery, all operating at a relatively macro scale [1]. These sources do capture broad functional patterns, but infrequent updating and limited spatial resolution constrain what they can reveal about a city's internal organization. As location-based services and open-data platforms spread in the 2010s, a new possibility emerged: using POI data to infer urban functions at much finer granularity. Each POI record carries a category label and a pair of coordinates. Because commercial mapping platforms update these records regularly, the resulting datasets mirror the contemporary urban landscape far more closely than decadal censuses do [2, 3].

Within the family of POI-based methods, the frequency-density approach has been the workhorse. Its logic is simple: compute the relative share of each POI category in a spatial unit and assign that unit the function whose share is highest [4]. Simplicity, of course, is a virtue when dealing with large datasets, and in areas where one function predominates the method performs adequately. Its limitations surface in more heterogeneous settings. Because every POI is treated as carrying equal functional weight, the method cannot distinguish between a POI that is genuinely diagnostic of an area's purpose and one that merely happens to be common. Consider a neighborhood near a major rail terminus: dozens of convenience stores and noodle shops will outnumber the station entrances and bus bays, yet it is the transport infrastructure that defines the area's role in the metropolitan network. The frequency-density count, blind to this distinction, may well assign a "commercial" label. Situations of this kind have been documented in several studies and are widely acknowledged as a bottleneck for classification accuracy in mixed-function and transit-oriented areas [5, 8].

2.2. Adaptation of the TF-IDF algorithm for spatial analysis

TF-IDF was formalized by Salton and Buckley (1988) as a way to decide which words matter most in a document. The intuition is simple enough: if a word appears often in one document but shows up rarely across the broader collection, it probably tells you something important about that document's topic. Local frequency (the TF part) captures whether a term is prominent in a given context; corpus-wide scarcity (the IDF part) filters out words that pop up everywhere and therefore carry little discriminating power. The combination has proved durable across a range of applications, from search-engine ranking to keyword extraction.

Geographic information scientists recognized, starting in roughly the mid-2010s, that the same logic translates naturally into urban space. Spatial units (grid cells, census tracts, blocks) stand in for documents; POI categories stand in for words [9, 10]. A category with a dense local presence in one grid cell but a thin spread across the rest of the city will earn a high score, marking it as a reliable fingerprint of that cell's function. The practical payoff is clearest for specialized zones. A transport hub, for example, may be outnumbered in raw POI count by commercial categories, but its airport terminals and rail entrances are city-wide rarities, and TF-IDF picks up on that rarity. A growing body of studies has applied this or closely related scoring schemes to urban-function tasks [8].

DOI: 10.54254/2755-2721/2026.GF37457

2.3. The role of public perception in urban functional zone identification

A city's functional structure cannot be read off its physical inventory alone. How residents perceive and interact with the built environment matters, too [3]. Two neighborhoods may contain an identical set of facilities yet feel entirely different to the people who use them, especially if one is anchored by a highly visible transit station and the other is not [1, 3]. Zhao (2008) gave early quantitative shape to this intuition by surveying residents about how they associated various facility types with urban functions [6]. Transportation facilities came out on top in terms of perceived functional importance; residential areas sat near the bottom, perhaps because housing forms the backdrop of daily routines rather than a destination people actively seek out. The gaps, Zhao suggested, had to do with how frequently and how conspicuously people encounter a facility in the course of ordinary life.

Liu et al. (2015) generalized this idea into a "social sensing" framework, arguing that urban analysis should weigh the objective distribution of facilities alongside residents' subjective readings of them [3]. In practice, that means attaching social weights to POI categories so that a metro station, say, carries more classificatory influence than a residential compound, even when the latter is more numerous. The logic is straightforward: if residents register certain facility types as strong functional markers, then a classification that ignores those perceptual signals is leaving useful information on the table [6, 11].

2.4. Mixed-function zones and urban spatial structure

Mixed-function zones pose a distinctive challenge. When several POI categories appear in comparable proportions within a single spatial unit, there is no obvious "winner," and conventional frequency-density methods tend to either misclassify or oversimplify the result [5, 8]. These mixed-use areas do not arise at random. Their formation tracks the logic of Transit-Oriented Development (TOD): a major transport node draws commercial, leisure, and public-service activities into its orbit, and the result is a layered functional landscape that no single label can fairly describe [12, 13].

Empirical work bears this out. The most commonly documented mixed-function types are commercial–transportation, commercial–leisure, and commercial–scenic-tourist combinations, all clustering around high-accessibility nodes. Du et al. (2021), using multi-source data, found much the same and observed that weighting mechanisms outperform raw-frequency approaches where functional complexity is high [8].

2.5. Research gaps and study contributions

The literature surveyed above points to a recurring tension. Frequency-density methods and their variants treat every POI as carrying the same functional weight, which works passably in homogeneous districts but degrades where multiple functions overlap and where a handful of rare facilities (say, an airport terminal amid a sea of convenience stores) actually define an area's role [8, 10]. That is a methodological gap. There is also a conceptual one: most existing frameworks stop at counting what is physically present, without asking whether residents even notice those facilities or associate them with a particular urban function. The social dimension of functional identity, in other words, goes largely unmeasured. And while TF-IDF and public-perception weighting have each been explored independently, bringing them together under a single model remains an underexplored option [8, 10, 11].

DOI: 10.54254/2755-2721/2026.GF37457

The present study steps into that space. Building on the retrieval-theoretic logic of Salton and Buckley (1988) [7] and the social-sensing perspective of Goodchild (2007) and Liu et al. (2015) [1, 3], we propose a Public Perception-Weighted TF-IDF model that scores each POI category along two axes, spatial distinctiveness and perceived social importance, before fusing the scores into a composite classification. Pudong New Area provides the empirical ground for testing the model, with the conventional frequency-density approach serving as the benchmark.

3. Methodology

3.1. Study area overview and data sources

Pudong New Area was our study site (Figure 1). Geographically it occupies the eastern bank of the Huangpu River, extending to the East China Sea coast. As both a National New Area and the nucleus of the China (Shanghai) Pilot Free Trade Zone, it contains subregions that look almost nothing like each other. Lujiazui is high-rise finance. Zhangjiang is a science-and-technology cluster built around labs and startups. Lingang, down at the southern end, is still filling in—large tracts remain under development. That kind of internal variety is what makes Pudong a good proving ground for a functional-zone model. If commercial towers, residential blocks, port infrastructure, and open farmland all coexist within one administrative unit, a classification method that works here should also work in less complicated cities. We divided the study area into 1 km × 1 km grid cells, which gave us 1,672 units.

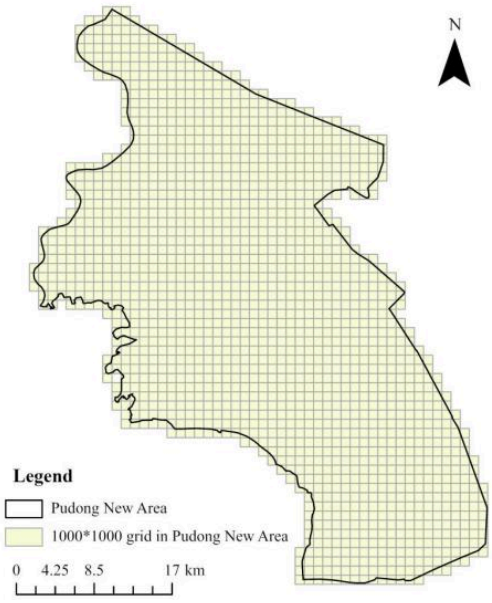

Figure 1. Study area

POI data came from the Amap platform. Each record includes a place name, a category tag, and geographic coordinates, spanning facility types from restaurants and banks to subway stations and hospitals. We removed duplicates and dropped records with missing fields.

Following the Code for Classification of Urban Land Use and Planning Standards of Development Land (GB50137–2011) and conventions from prior studies, we consolidated Amap's original fine-grained categories into seven functional classes: commercial; education and training;

DOI: 10.54254/2755-2721/2026.GF37457

residential; scenic and tourist; leisure and entertainment; public service; and transportation facilities. The full mapping is shown in Table 1.

Table 1. Reclassification of POI categories

Reclassified Category	Amap Primary Category	Amap Secondary Category
Commercial Zone	Food and Beverage Services	Chinese restaurants, international restaurants, cafés, dessert shops, tea houses, and others
	Shopping and Consumption	Supermarkets, convenience stores, department stores, shopping centers, home furnishing stores, building materials markets, and others
	Financial Institutions	Banks, ATMs, insurance services, and others
	Companies and Enterprises	Companies, factories, agricultural and forestry ranches, and others
	Automobile Services	Gas stations, vehicle maintenance, vehicle inspection stations, automobile sales, and others
	Business and Residential Services	Office buildings, industrial parks, commercial-residential complexes, and others
	Daily Services	Beauty salons, intermediary agencies, laundries, logistics services, and others
Education and Training Zone	Science, Education and Culture	Kindergartens, primary schools, secondary schools, higher education institutions, vocational and technical schools, training institutions, and schools
Residential Zone	Business and Residential Services	Residential communities, dormitories, villas, and others
Scenic Zone	Tourist Attractions	Scenic spots, parks, historical sites, museums, zoos, aquariums, and others
Leisure and Entertainment Zone	Leisure and Entertainment	Cinemas, KTVs, bars, chess and card rooms, internet cafés, theaters, amusement parks, and others
	Sports and Fitness	Gyms, swimming pools, football fields, basketball courts, tennis courts, and others
	Hotels and Accommodation	Star-rated hotels, budget hotels, serviced apartments, guesthouses, and others
Public Service Zone	Healthcare Services	General hospitals, specialized hospitals, clinics, pharmacies, and others
	Science, Education and Culture	Libraries, museums, technology museums, art galleries, exhibition halls, and others
	Daily Services	Parks, post offices, logistics facilities, public utilities, and others
	Tourist Attractions	Temples, squares, and others
Transportation Facility Zone	Transportation Facilities	Bus stations, subway stations, parking lots, ports and terminals, railway stations, airports, and others

3.2. Public perception-weighted TF-IDF model

Two weighting layers make up the identification framework (Figure 2). The first uses TF-IDF to measure how spatially distinctive each POI category is within a given grid cell. The second introduces public-perception scores that reflect how much different facility types register in

DOI: 10.54254/2755-2721/2026.GF37457

residents' awareness—a different question from how many of those facilities happen to exist. We average the two layers into a composite score for each POI category in each spatial unit, then use the composite to infer the dominant function and assign a zone label (classification rules are in Section 3.3).

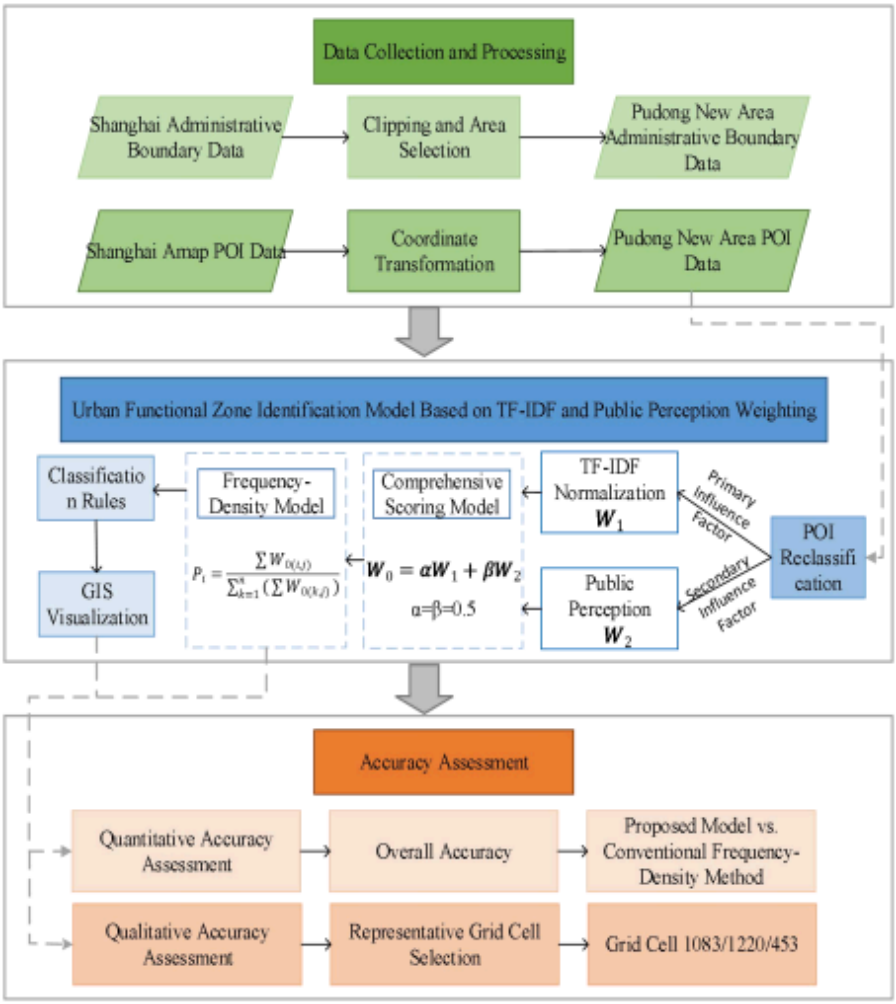

Figure 2. Research framework

The point of combining them is to get the model to consider both what is physically on the ground and what residents actually notice, instead of relying on either one in isolation. The subsections below lay out the mechanics.

3.2.1. TF-IDF weighting

TF-IDF serves as the first-level weighting factor. It originated in text mining, where researchers have been using it for decades to sort out which words carry the most information in a given document. We borrow the same logic for geographic analysis. Each grid cell is treated as a "document" and each POI category as a "term." If a POI category appears often in a particular cell but is scarce across the rest of the city, it gets a high weight. If it appears everywhere in roughly equal densities, it gets a low one.

The term frequency (TF) of POI category i within spatial unit j is defined as:

DOI: 10.54254/2755-2721/2026.GF37457

$$T_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \tag{1}$$

where TF_ij is the term frequency of POI category i in spatial unit j, n_ij is the count of POIs of category i in unit j, and the denominator is the total POI count across all categories in unit j.

The inverse document frequency (IDF) of category i is:

$$I_i = \log \frac{N}{D_i} \tag{2}$$

with N being the total number of spatial units and D_i the number of units containing at least one POI of category i. The raw TF-IDF weight is then:

$$F_{ij} = T_{ij} \times I_i \tag{3}$$

and its normalized form:

$$W_i = \frac{F_{ij} - F_{min}}{F_{max} - F_{min}} \tag{4}$$

A high normalized value means the category is locally prominent and globally uncommon. Bus terminals clustered near an airport are a good example, you will find very few of them elsewhere in the city, so they score high. Convenience stores, on the other hand, are distributed so evenly that their IDF is low no matter where you look.

3.2.2. Public perception weighting

The perception scores adopted here originate from Zhao (2008), who administered questionnaire surveys to quantify how strongly residents associate different types of urban facilities with distinct functional identities [6]. To align these scores with the seven-class POI framework used in the present study, the functional categories reported by Zhao were mapped onto corresponding POI classes. The matched scores were then incorporated as weighting parameters:We use perception scores from Zhao (2008), who surveyed residents about how strongly they associated various facility types with distinct urban functions [6]. Because Zhao's original functional categories do not perfectly match our seven-class POI framework, we had to map them over, matching each of Zhao's categories to the closest POI class we use. The perception score for POI category i is then:

$$W_i = C_i \tag{5}$$

where C_i denotes the perception score obtained through category mapping. Table 2 lists the resulting weights.

DOI: 10.54254/2755-2721/2026.GF37457

Table 2. Public perception weights of different functional zones

Functional Zone Type	Public Perception Weight
Transportation Facility Zone	1.0000
Commercial Zone	0.8146
Scenic Zone	0.7397
Education and Training Zone	0.6706
Leisure and Entertainment Zone	0.5286
Public Service Zone	0.4310
Residential Zone	0.1529

Transportation at 1.0 versus residential at 0.15, a factor of more than six. Folding these weights into the classification model counteracts the equal-treatment bias built into frequency-density methods. In practice, the labels end up closer to what a resident would tell you about an area if you asked.

3.2.3. Composite scoring model

The two weighting layers are fused through a weighted-average aggregation:

$$W_0 = \alpha W_1 + \beta W_2 \tag{6}$$

Here W_1 is the normalized TF-IDF weight and W_2 is the perception score. Because there is no strong a priori reason to favor one layer over the other, both are assigned equal influence (α = β = 0.5). The composite score W_0 thus reflects, for each POI category in each grid cell, an integrated measure of spatial distinctiveness and social salience—providing the quantitative foundation on which the subsequent classification step rests.

3.2.4. Frequency-density method

Converting composite scores into zone labels requires knowing how much each category dominates within a cell. We compute this as a proportion:

$$P_i = \frac{\sum W_{0(i,j)}}{\sum_{k=1}^{n}(\sum W_{0(k,j)})} \tag{7}$$

3.3. Classification criteria for urban functional zones

Zone labels follow from the frequency-density proportions using threshold rules that are standard in the urban-analytics literature. A cell where one POI category exceeds 50% of the total frequency-density value is labeled a single-function zone. When nothing exceeds 50%, the cell is classified as mixed-function. We subdivide further within that category. One POI category in the 20%–50% range, with everything else under 20%, makes it a function-dominated mixed-function zone. Two

DOI: 10.54254/2755-2721/2026.GF37457

categories both between 20% and 50% makes it dual-function-dominated. All categories below 20% gives a comprehensive functional zone. Cells without any POIs are recorded as no-data zones.

4. Results and analysis

4.1. Results of urban functional zone identification

After running the model on all 1,672 cells we produced the spatial distribution maps in Figure 2: panel (a) covers single-function zones, panel (b) covers mixed-function zones. Mapping and quantitative tabulation were done in ArcGIS Pro 3.2.

4.1.1. Analysis of single-function zones

Single-function zones account for the majority of cells. Commercial zones (A) dominate the count at 996 units, or 59.57% of the total, and they are not randomly scattered. They concentrate in Lujiazui, Zhangjiang, and Qiantan, the three subregions with the densest clusters of office towers, shopping complexes, and banking facilities. This is basically what you would expect from Pudong's economic geography, and it would have been worrying if the model missed it.

Transportation facility zones (G) number only 92 units (5.50%). They are geographically concentrated around a handful of major infrastructure sites. Shanghai Pudong International Airport anchors the eastern cluster. The Waigaoqiao port area accounts for the northern one. And the Longyang Road corridor, where the maglev terminal sits, forms a third. A frequency-density classifier tends to get these cells wrong because the restaurants and convenience stores that spring up around airports and ports easily outnumber the actual transport infrastructure.

No-data zones cover 19.02% of the grid. They turn up in the outer reaches of Lingang, where development has not yet filled in the land. They also appear inside large enclosed industrial sites, certain factory parcels in Jinqiao Industrial Park, for instance, and across wide stretches of green space or water. In all these cases the POI density is simply too low for the model to classify.

4.1.2. Analysis of mixed-function zones

Mixed-function zones show how different land uses pile on top of each other in Pudong's denser corridors. Commercial–transportation zones (AG) are the most common mixed type at 100 units (5.98%). They cluster around major metro interchange stations. Century Avenue and Lancun Road are the two clearest anchors. From these nodes, the AG cells extend outward in radial and, in places, corridor-like patterns that follow rail alignments. On the ground, these cells contain a tangle of commuter infrastructure, retail outlets, and restaurants, all drawn in by the transit access.

Commercial–leisure-entertainment zones (AE) and commercial–scenic-tourist zones (AD) are much rarer at 0.90% and 0.36% of all cells. They show up along the eastern Huangpu River waterfront and near Century Park, areas where restaurants and retail have grown alongside parks and recreational amenities. That co-location is probably not accidental; it looks like a product of urban-renewal strategies in core Pudong. The functional overlap in these cells, commerce, leisure, green space, seems to sustain a broader range of neighborhood activity, including outside regular business hours.

DOI: 10.54254/2755-2721/2026.GF37457

4.2. Accuracy assessment of urban functional zone identification results

4.2.1. Quantitative accuracy assessment

Overall accuracy (OA) was adopted as the primary evaluation metric. From the 1,672 classified cells, 213 were drawn as a validation sample, with the sampling design guided by proportional representation across zone types. Categories whose proportional share yielded fewer than 20 samples were bumped up to a floor of 20 units; categories with fewer than 20 total cells were included in full. This hybrid strategy safeguards randomness while preventing underrepresented classes from being assessed on too thin an evidence base.

Each sampled cell was cross-checked against high-resolution remote-sensing imagery and supplementary reference materials to determine whether the model-assigned label matched ground-truth conditions. Table 3 summarizes the results.

Table 3. Accuracy comparison between the proposed model and the frequency-density method

Functional Zone Type	Proposed Model (%)	Frequency-Density Method (%)
Commercial Zone	73.87	47.32
Education and Training Zone	84.70	73.77
Residential Zone	82.19	75.43
Scenic Zone	94.12	78.66
Leisure and Entertainment Zone	88.13	80.74
Public Service Zone	85.36	75.14
Transportation Facility Zone	90.65	53.49
Mixed-Function Zone	91.63	49.37
Overall Accuracy (OA)	**86.33**	**66.74**

At the aggregate level the proposed model reached 86.33% OA, a lead of 19.59 percentage points over the baseline's 66.74%. But the aggregate hides important variation. The widest gap opened in transportation facility zones, where accuracy jumped from 53.49% to 90.65%. Mixed-function zones followed a similar trajectory: 49.37% to 91.63%. These are categories defined by intense functional overlap, the kind of setting where a handful of rare but defining POI types (cargo terminals, metro entrances) get drowned out by the sheer numerical mass of restaurants and shops under a frequency-density regime. TF-IDF scoring, by design, counteracts that drowning-out effect, and the perception layer compounds the correction.

Even where the baseline already performed reasonably well (education zones, residential zones, public-service and leisure areas), the proposed model added 7 to 11 percentage points. The dual-weighting mechanism, in other words, does not help only in hard cases; it lifts the floor across the board. Whether the task is distinguishing a transport hub from its surrounding eateries or confirming that a school cluster is indeed educational, the composite score outperforms the raw count.

4.2.2. Qualitative accuracy assessment

Numbers, of course, only go so far. To ground the comparison in actual geography, three grid cells were examined more closely.

Grid Cell 1083 (Figure 3). The frequency-density method labeled this cell a public-service-dominated mixed-function zone; our model labeled it a transportation facility zone. Looking at the

DOI: 10.54254/2755-2721/2026.GF37457

POI roster, the area is overwhelmingly composed of airport cargo terminals, logistics enterprises, parking structures, and other aviation-related infrastructure tied to Shanghai Pudong International Airport. A scattering of hotels and service outlets does exist, but they are there to support airport operations and passenger convenience. Transportation is the dominant function here by any reasonable reading, and the proposed model's classification is consistent with that.

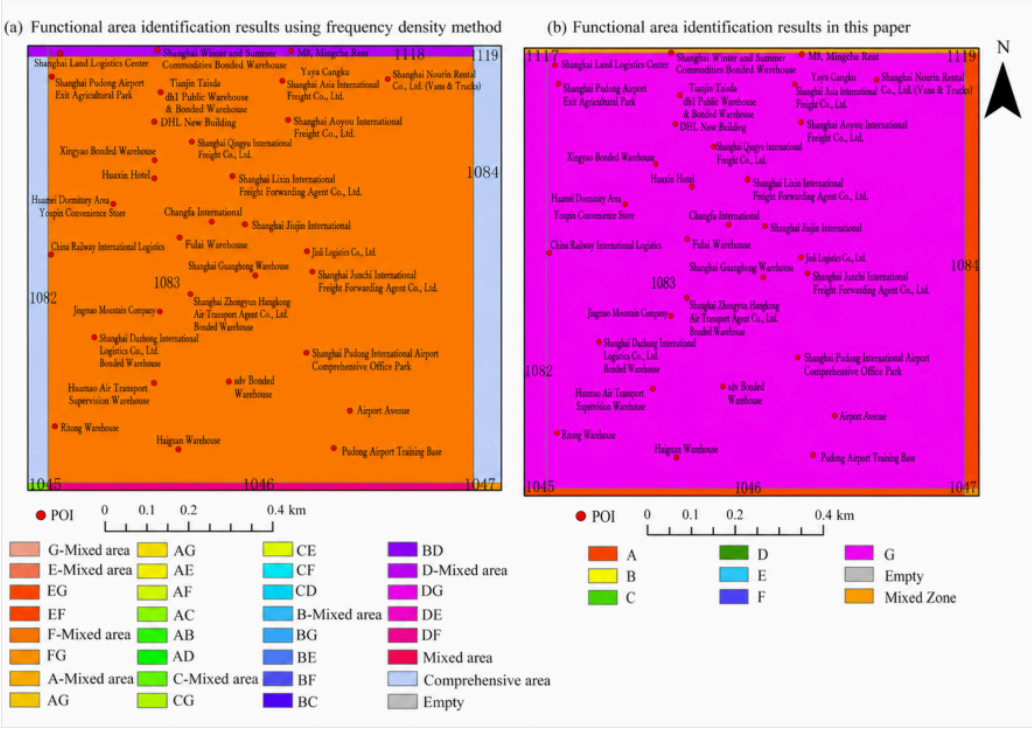

Figure 3. Comparison of urban functional zone identification results for Grid Cell 1083

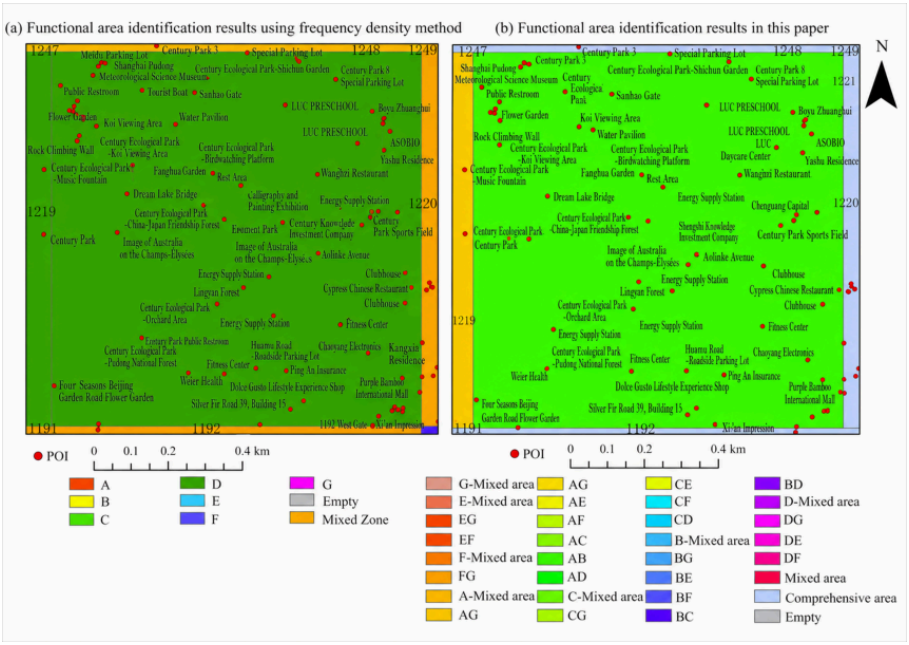

Figure 4. Comparison of urban functional zone identification results for Grid Cell 1220

DOI: 10.54254/2755-2721/2026.GF37457

Grid Cell 1220 (Figure 4). The baseline tagged this area as a scenic zone; our model called it a commercial–scenic mixed-function zone. On the ground, the cell encompasses Century Ecological Park alongside a substantial cluster of restaurants, experiential retail outlets, and business-service establishments. The commercial presence is not incidental; it has grown organically around the park's recreational draw. A mixed-function label captures that duality more faithfully than a single "scenic" tag.

Grid Cell 453 (Figure 5). The baseline assigned a leisure-and-entertainment label; our model recognized a mixed-function zone. The POI inventory includes commercial enterprises, agricultural suppliers, decoration firms, equipment vendors, residential settlements, and a modest number of eateries and leisure venues. No single activity dominates. The landscape is a patchwork, and a mixed-function classification mirrors that reality better.

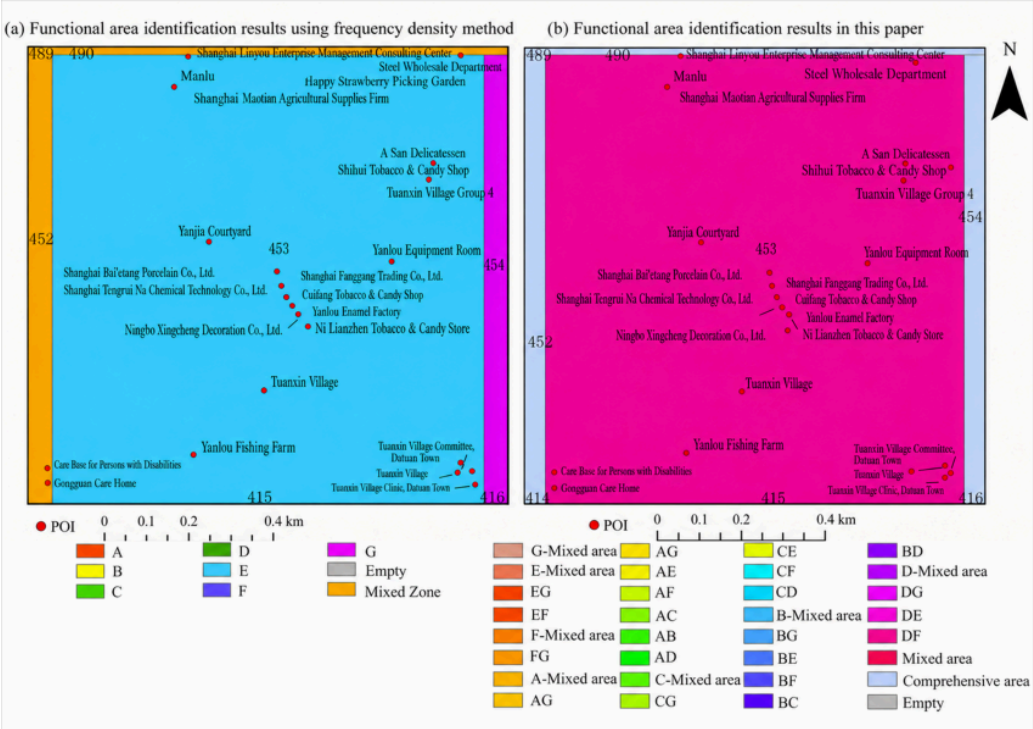

Figure 5. Comparison of urban functional zone identification results for Grid Cell 453

5. Discussion

Why does the proposed model outperform frequency-density counting? The answer, we believe, lies in how each method handles functional ambiguity. A frequency-density classifier is indifferent to the identity of POIs; it simply tallies them. When an airport cargo zone contains thirty restaurants but only five cargo terminals, the tally points to "commercial." The TF-IDF layer upends that logic by asking a different question: which POI categories are unusual here relative to the city as a whole? Cargo terminals, being rare city-wide, receive a high distinctiveness score; restaurants, being ubiquitous, do not. The perception layer pushes in the same direction. Transportation facilities carry the highest social importance in the Zhao (2008) survey, so the two weighting dimensions reinforce each other in cases where the frequency-density approach is most likely to err.

The Pudong results fit comfortably within urban-economic expectations. Dense commercial activity tracks the agglomeration cores of Lujiazui, Zhangjiang, and Qiantan; transport zones map onto gateway infrastructure; and the mixed-function clusters around metro interchanges echo the

DOI: 10.54254/2755-2721/2026.GF37457

predictions of Transit-Oriented Development (TOD) theory. None of this is surprising in itself. What is encouraging is that a relatively simple two-layer weighting scheme manages to reproduce these well-known spatial regularities, which suggests the model captures genuine functional structure rather than statistical noise.

Limitations deserve candid acknowledgment. The perception scores date from 2008. Pudong has changed substantially since then: Lingang barely existed as a built-up area at the time, and several metro lines have since opened, so the scores may not faithfully represent how today's residents read urban space. A fresh perception survey, ideally one embedded in a mobile-app format to reach a broader demographic, would be a valuable update. The reliance on Amap as the sole POI source is another constraint. Platform-specific coverage gaps and classification inconsistencies could bias results, and cross-referencing with Baidu Maps or OpenStreetMap would help. The 1 km grid resolution, too, trades detail for tractability; along campus edges and waterfront corridors, where land use shifts over a few hundred meters, a finer or adaptive mesh might reveal patterns the current grid smooths over.

6. Conclusions

This paper has proposed and empirically tested a model that weights POI data along two dimensions: spatial distinctiveness (via TF-IDF) and social importance as perceived by residents (via survey-derived perception scores). The Pudong New Area case demonstrates that the approach works. Overall accuracy reached 86.33% across 1,672 grid cells, nearly twenty percentage points above the 66.74% achieved by conventional frequency-density counting.

The gap was widest where frequency-density methods are weakest. Transportation facility zones jumped from 53.49% to 90.65%; mixed-function zones from 49.37% to 91.63%; commercial zones from 47.32% to 73.87%. All three are zone types defined not by sheer POI volume but by the presence of a few pivotal facility types, exactly the conditions under which TF-IDF's rarity logic and the perception layer's social weighting contribute the most.

The classified map of Pudong tells a coherent spatial story as well. Commercial zones anchor themselves to the Lujiazui–Zhangjiang–Qiantan corridor; transport zones track gateway infrastructure; mixed-function clusters ring metro interchanges in patterns consistent with TOD-driven agglomeration. That the model reproduces these regularities without being explicitly told about them lends ecological validity to the outputs and, at a practical level, gives planners a quantitative tool for land-use coordination and facility-siting decisions.

References

[1] Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

[2] Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 186–194.

[3] Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*(3), 512–530.

[4] Jiang, B., Yin, J., & Liu, Q. (2015). City as a complex adaptive system: The case of exploring the factors that determine urban growth. *Landscape and Urban Planning, 145*, 65–76.

[5] 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*(3), 446–467.

[6] Zhao, W., Li, Q., & Li, B. (2011). Extracting hierarchical landmarks from urban POI data. *Journal of Remote Sensing, 15*(5), 973–988.

DOI: 10.54254/2755-2721/2026.GF37457

[7] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513–523.
[8] 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.
[9] Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing urban land-use patterns by integrating Google TensorFlow with scene-level POI data. *Computers, Environment and Urban Systems, 73*, 66–78.
[10] 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*(8), 1675–1696.
[11] Zhu, D., Zhang, F., Wang, S., Li, Y., Zhang, A., Gao, Z., & Liu, Y. (2020). Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the Association of American Geographers, 110*(2), 408–420.
[12] Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment, 2*(3), 199–219.
[13] Liu, S., Zhou, C., Rong, J., Bian, Y., & Wang, Y. (2022). Concordance between regional functions and mobility features using bike-sharing and land-use data near metro stations. *Sustainable Cities and Society, 84*, 104010.