Digital Intelligence PRESS

Journal of Digital Frontier

https://doi.org/10.67541/jdf2605

Research Article **Volume 1, Issue 1 (July 2026)** *Open Access*

Algorithm for Identifying Micro-Update Potential in Urban Corner Spatial Environment Design by Integrating Multi-Source Data

Jing Zhang *

College of Arts, Shandong Jianzhu University, Jinan, Shandong, China

***Corresponding author:** Jing Zhang. Email: outlook_8A9201AAC2554F63@outlook.com

Abstract

Aiming at the problems of multi-source data heterogeneity, insufficient environmental representation, and insufficient spatial correlation description in the potential recognition of urban corner space micro-renewal, a potential recognition method based on multi-source data and spatial relationship enhancement learning is proposed. By integrating remote sensing images, street view images, POI data, road network data, building form data, land use data, and population activity data, multi-scale environmental features are constructed, and a cross-modal attention mechanism and a multi-relation graph attention network are introduced to realize heterogeneous feature interaction and neighborhood spatial dependence modeling. On 1,216 urban corner space samples, the accuracy, F1-score, ROC-AUC, and PR-AUC of the proposed model reach 0.887, 0.881, 0.946 and 0.921 respectively, which is 3.3 percentage points higher than that of the standard GAT. Ablation, cross-region migration, and perturbation experiments further verify the effectiveness, generalization ability and robustness of the model. This study can provide data-driven technical support for automatic screening of urban corner space, prioritization of micro-update and fine governance.

Keywords

Urban corner space; Multi-source data fusion; Cross-modal attention; Graph attention network; Micro update potential identification

Received: 31 May 2026; Revised: 17 Jul 2026; Accepted: 29 Jul 2026; Published: 21 Aug 2026

1. INTRODUCTION

As urban development shifts from large-scale expansion to stock space optimization, renewal efforts increasingly target smaller, scattered, and underutilized micro-spaces [1]. Urban corner spaces—such as residual land at intersections, building setbacks, under-bridge areas, neighborhood edge open spaces, and fragmented plots—are widely embedded in high-density built environments. Though limited in individual area, these spaces closely affect residents' daily travel, public activities, and environmental quality. Through small-scale, low-disturbance micro-renewal, they can improve public space supply and activate inefficient stock without large-scale demolition. Thus, efficiently identifying corner spaces with high renewal value and determining their priority has become a critical task in fine-grained urban governance.

Current identification methods primarily rely on field surveys, remote sensing interpretation, and indicator-based evaluation [2],[3]. Field surveys are labor-intensive and subjective; remote sensing data capture physical structure but miss pedestrian-scale environmental experience; indicator-based systems

require predefined weights, while renewal value actually emerges from interactions among physical conditions, functional configuration, traffic connectivity, and activity demand. Traditional single-source approaches struggle to capture this complexity, limiting identification granularity and cross-regional applicability [4],[5],[6],[7].

Recent advances in urban sensing and spatial big data offer richer information for micro-scale identification. Remote sensing, street view imagery, POI data, road networks, and population activity data collectively describe urban environments across physical, visual, functional, and behavioral dimensions. However, simply increasing data volume does not guarantee improved performance. These sources differ significantly in structure, resolution, and sampling mode. Direct feature concatenation leads to scale mismatches and redundancy, while high-dimensional visual features may overshadow functionally significant variables [8]. Effective multi-source fusion must suppress redundancy while preserving complementary information, requiring sophisticated cross-modal interaction rather than simple data superposition [9],[10],[11].

Moreover, most machine learning studies treat each spatial unit independently, neglecting urban relational structure. In reality, a corner space's renewal potential is substantially influenced by adjacent population activity, road connectivity, and public services. Two physically similar spaces may differ drastically in renewal value due to their surrounding contexts. Recent graph neural networks offer tools for modeling spatial relations, yet conventional graphs rely solely on geometric distance or fixed adjacency, failing to simultaneously represent geographical, transport, and functional connections in urban space [12].

Another key challenge is the tension between rankability and explainability. Indicator systems are interpretable but subjective; deep learning models are accurate but black-box. For practical renewal decisions, a method must not only achieve high classification accuracy but also provide traceable explanations of how environmental factors and spatial relations affect the outcome [13],[14].

To address these challenges, this paper proposes a micro-renewal potential identification framework integrating multi-source urban data. Remote sensing, street view, POI, road network, building form, land use, and population activity data are unified into micro-scale analysis units. Multi-scale coding and independent modal coding reconcile data differences, while a cross-modal attention mechanism enables dynamic interaction among environmental information sources. Candidate corner spaces and surrounding units are constructed as a spatial graph incorporating distance, road connectivity, and functional associations. Graph attention propagation captures dependencies between target spaces and their urban context. The output includes potential probabilities, grades, and continuous ranking scores, forming a complete pipeline from data input to spatial learning and potential judgment.

2. URBAN CORNER SPACE MULTI-SOURCE DATA FUSION AND ENVIRONMENTAL FEATURE CONSTRUCTION METHOD

Multi-source urban datasets such as remote sensing images, street view images, POI data, road network data, building form data, land use data, and population activity data are constructed to complete spatial coordinate unification, scale matching, outlier processing, missing data repair and spatial unit division. Feature extraction methods were designed for different data modes, including extracting semantic features of street view and remote sensing environment based on deep vision network, extracting features of building density, road accessibility, function mixing and public service coverage based on spatial statistics, and extracting features of population activity intensity and spatial vitality

based on kernel density estimation and spatial-temporal statistics. A multi-modal feature matrix with a uniform spatial index is further constructed. Aiming at the problem that traditional direct feature stitching is difficult to deal with scale differences and information redundancy of multi-source data, multi-scale spatial coding, feature attention weighting and cross-modal feature alignment mechanism are introduced to form a comprehensive environmental feature representation of urban corner space that can simultaneously express spatial form, environmental quality, functional supply, traffic connection and usage vitality. Provide high-quality input for subsequent potential identification models.

Urban corner space is usually characterized by small area, irregular boundary, fuzzy spatial function, strong heterogeneity of surrounding environment, and so on. Therefore, this paper organizes multi-source urban data from the five dimensions of spatial form, visual environment, functional supply, traffic connection and crowd activity, and maps remote sensing images, street view images, POI, road network, building contour, land use and population activity data to a unified spatial unit. At the level of data processing, coordinate reference system transformation, spatial clipping, outlier elimination and missing value repair are completed first, and then unified spatial index is established by combining regular grid with actual plot boundary, so that data from different sources, different granularity and different time scales can be matched in the same analysis unit. Considering that corner spaces have obvious small-scale characteristics, this paper adopts a regular grid with a side length of 50 m as the basic computing unit and counts the surrounding environment attributes at three neighborhood scales of 100 m, 300 m and 500 m, so as to retain the local spatial morphological characteristics and neighborhood functional environment information at the same time. For the residual units with an area less than 30% of the basic grid area, they are merged with the neighboring units with the highest degree of spatial contiguity and consistent land use attributes, thus reducing the influence of too small spatial units on the stability of statistical features.

Different data sources have obvious differences in spatial resolution, update period and original expression form, so feature splicing cannot be carried out directly. In this paper, quantile truncation is first implemented for continuous variables, and abnormal observations lower than 1% quantile and higher than 99% quantile are replaced by corresponding quantile values respectively. For continuous variables with less than 10% missing ratio, neighborhood weighted interpolation is used, and for variables with higher missing ratio but clear spatial correlation, spatial nearest neighbor estimation is used to repair. Then standardized transformation is used to eliminate the dimensional differences of different indicators, and the calculation form is as follows [15],[16]:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{1}$$

Where x_{ij} represents the value of the i-th spatial unit on the original feature of the j-th item, μ_j and σ_j represent the mean and standard deviation of the feature in all valid spatial units respectively, and z_{ij} is the standardized feature value. After the processing, all continuous variables are transformed into a unified numerical space with mean value close to 0 and standard deviation close to 1, so as to avoid unreasonable weight shift in the subsequent fusion process of indicators with large magnitude differences such as road density, building area and population activity intensity.

Table 1. Multi-source urban data and spatial unified parameter Settings

Type of data	Original spatial granularity	Uniform statistical scale	Main information dimensions	Initial number of features
Remote sensing image	10 m	50 m grid	Vegetation, hardening, water bodies, land cover	24
Street view image	Sampling points ranged from 20 to 50 m	50 m grid	Green view rate, sky rate, interface continuity, environment perception	32
POI data	Element of point	300 m neighborhood	Function density, function mixing degree, public service supply	18
Road network	Element of line	300 m neighborhood	Road network density, node density, accessibility	12
Buildings and land use	Element of surface	50 ~ 300 m	Building density, volume intensity, land mix	20
Population activity data	100 ~ 500 m	300 m neighborhood	Flow intensity, period fluctuation, activity persistence	16

The data organization shown in Table 1 avoids the mechanical transformation of all data to exactly the same original resolution but determines the statistical scale according to the spatial scope of different urban elements. Remote sensing images and architectural forms mainly reflect the corner space ontology and its immediate environment, so the statistics are carried out at the scale of 50 m. POI, roads and population activities reflect more surrounding services and spatial connections, and their impacts usually span a single plot, so 300 m neighborhood is used for aggregation. According to the above design, 122 dimensions of basic environmental variables can be initially obtained for each spatial unit, among which visual environment related features account for about 45.9%, spatial form and functional supply features account for about 31.1%, and traffic and population activity features account for about 23.0%, so that different environmental dimensions retain sufficient information expression ability. At the same time, it is avoided that a single data mode forms an absolute advantage in the number of dimensions.

In terms of remote sensing and streetscape visual environment feature construction, this paper extracts information from two levels of top-down surface structure and pedestrian perspective environment perception respectively. A pre-trained visual backbone network is used to extract deep semantic features from remote sensing images, and explicit indicators such as vegetation coverage, proportion of impervious surface, proportion of water and proportion of bare land are calculated at the same time. For the case that a spatial unit contains N_i valid pixels, the k-th type land cover ratio is defined as follows:

$$R_{ik} = \frac{1}{N_i} \sum_{p=1}^{N_i} I\left(c_p = k\right) \tag{2}$$

Where R_{ik} represents the proportion of the k-th type of surface elements in the i-th spatial unit, c_p represents the semantic category corresponding to the p-th pixel, and $I(\cdot)$ is the indicator function, which takes 1 when the pixel belongs to the target category, and 0 otherwise. Through this index, complex image information can be converted into a structured variable with clear environmental

meaning. For example, the proportion of vegetation reflects the ecological basis that can be used for greening, the proportion of impervious surface reflects the degree of hard environment, and the proportion of bare land can help judge whether there is inefficient use of space [17],[18]. Compared with using only deep visual features, explicit semantic ratio can enhance feature interpretation and reduce the problem of visual network over-sensitivity to local texture differences.

Street view data uses fixed sampling intervals to obtain road perspective images in the analysis area, and identifies main visual elements such as vegetation, sky, buildings, roads, walls and facilities through semantic segmentation network. For spatial units with multiple street view images, a single image is not directly selected, but area weighting and direction averaging are used to form comprehensive visual features. The calculation of visual proportion indicators such as green vision rate is as follows:

$$G_i = \frac{\sum_{m=1}^{M_i} P_{im}^{\text{veg}}}{\sum_{m=1}^{M_i} P_{im}^{\text{all}}} \tag{3}$$

Where G_i represents the comprehensive green viewing rate of the i-th spatial unit, M_i represents the number of effective street view images matching the spatial unit, P_{im}^{veg} represents the number of vegetation pixels in the m-th image, and P_{im}^{all} represents the number of all effective pixels of the image. The same method can further obtain indicators such as sky visibility rate, building interface rate and road proportion. For the case that there are multiple viewing directions in a spatial unit, the results of different directions are averaged to reduce the bias caused by single view occlusion. Through the joint expression of remote sensing images and street view images, the physical coverage structure "from top to bottom" and the actual visual experience "from human to environment" can be depicted at the same time, reducing the information loss caused by single view environment perception.

Building form and land use information are mainly used to represent the spatial compactness and functional background of the built environment in which the corner space is located. In this paper, building coverage, average building height, building area density, building boundary complexity, and land use mix degree are calculated within the base grid and its neighborhood. Among them, the building coverage is calculated as follows:

$$B_i = \frac{\sum_{q=1}^{Q_i} A_{iq}^{\text{build}}}{A_i} \tag{4}$$

Where B_i represents the building coverage of the i-th spatial unit, A_{iq}^{build} represents the projected area of the q-th building that intersects the spatial unit inside the unit, Q_i represents the number of intersecting buildings, and A_i represents the total area of the spatial unit. This index can quantitatively reflect the intensity of building around corner space. When the building coverage rate is high, the space often has strong physical enclosure and land use constraints; When the coverage rate is low and the proportion of bare land is high, it indicates that there may be a larger margin of environmental adjustment in the space. The degree of land use mixing is calculated by information entropy:

$$H_i = -\frac{1}{\ln K} \sum_{k=1}^{K} p_{ik} \ln p_{ik} \tag{5}$$

Where H_i represents the land use mixing degree in the neighborhood of the i-th spatial unit, K represents the number of land types, and p_{ik} represents the proportion of the land area of the k-th type in the total land area of the neighborhood. After normalization, the value of H_i ranges from 0 to 1, and the closer the value is to 1, the more balanced the spatial distribution of different functions such as commercial, residential and public services is. Compared with the direct use of land use categories, this index can express the spatial function composite degree in the form of continuous variables, which is more suitable for the subsequent unified feature learning.

POI and road network supplement the external environmental information of urban corner space from the perspectives of functional supply and traffic connection respectively. Since POI belongs to typical point event data and its direct quantity is easily affected by the boundary of statistical units, this paper uses the kernel density method to obtain the continuous spatial distribution [19],[20]. For the k-th type of urban facilities, the kernel density is expressed as:

$$D_{ik} = \frac{1}{h^2}\sum_{j=1}^{n_k} K\left(\frac{d_{ij}}{h}\right) \tag{6}$$

Where D_{ik} represents the kernel density of the i-th spatial unit corresponding to the k-th POI, n_k represents the total number of POI of this type in the study area, d_{ij} represents the distance between the center of the spatial unit and the j-th POI, h represents the bandwidth of the kernel function, and $K(\cdot)$ represents the kernel function. In this paper, we estimate the density of POIs for commercial services, public services, transportation facilities, leisure and entertainment, and life services respectively, and then calculate the comprehensive function supply index. Compared with simply counting the number of POIs in the 300-m buffer, kernel density can reduce the mutation caused by administrative boundary and grid boundaries and make the functional supply characteristics have a more continuous spatial expression.

Road network characteristics include road length density, intersection density, road class composition, and network accessibility. Among them, the road density is calculated according to the ratio of the total length of the road in the neighborhood to the neighborhood area, while the local network accessibility is constructed according to the shortest network distance between the spatial unit and the surrounding road nodes. For the i-th spatial unit, its comprehensive accessibility can be expressed as:

$$A_i^{\mathrm{road}} = \frac{1}{J_i}\sum_{j=1}^{J_i} \exp\left(-\frac{d_{ij}^{\mathrm{net}}}{\tau}\right) \tag{7}$$

Where A_i^{road} represents the accessibility of the road network, J_i represents the number of road nodes that can be connected in the specified neighborhood, d_{ij}^{net} represents the shortest network distance between the spatial unit and the j-th road node, and τ is the distance attenuation parameter. This expression enables the closer road nodes to contribute more to accessibility, while the influence of the distant nodes is gradually weakened, which can reflect the degree of connection between space and urban transportation system more realistically than the simple Euclidean distance.

Population activity data are used to supplement the actual state of use that cannot be directly reflected by material spatial data. In this paper, aggregated mobile location data, thermal data or crowd activity intensity with time labels are uniformly summarized by hour, and the daily average activity

intensity, peak activity intensity, day-night difference and activity fluctuation degree are calculated respectively. Considering the large differences in population base in different regions, the activity data are first converted into the relative intensity of spatial units, and then the coefficient of variation is used to describe the temporal stability:

$$CV_i = \frac{s_i}{\bar{a}_i} \tag{8}$$

Where CV_i represents the fluctuation coefficient of crowd activity in the i-th spatial unit, s_i represents the standard deviation of activity intensity in the study period, and $\bar{a}_i$ represents the average activity intensity in the same period. When CV_i is small, it indicates that the space maintains a relatively stable use demand in different periods. When its value is high, it indicates that there is a clear peak-valley difference in space use. By combining the average activity intensity with fluctuation degree, we can distinguish the continuous active space from the short-term aggregation space and avoid misjudging the actual use characteristics based on the thermal data of a single time point.

Table 2. Characteristic system and quantitative scope of urban corner space environment

Dimensions of environment	Representative characteristics	Original quantization form	Unified processing scope	Dimension of output
Form of space	Building coverage, bare land rate, and boundary complexity	Ratio/index	About −3 ~ 3 after normalization	20
Visual environment	Green view rate, sky rate, interface continuity	0 ~ 1 ratio value	0 ~ 1 and standardization	32
Supply of functions	POI density, function mixing degree, and service coverage	Units /km², entropy value	0 ~ 1 or normalized	18
Transportation link	Road density, node density, and network accessibility	km/km², index	0 ~ 1 or normalized	12
Use of vitality	Mean activity intensity, peak, and day-night differences	Relative strength	0 ~ 1 and standardization	16
Remote sensing semantics	Characteristics of vegetation, hardening, water body, bare soil deep layer	Proportion/eigenvector	0 ~ 1 and standardization	24

Table 2 further shows that this study does not reduce "environmental quality" to a certain comprehensive score but retains independent quantitative characteristics of different environmental dimensions. Among the 122 dimensions of initial variables, visual and remote sensing environment have 56 dimensions, accounting for 45.9%; There are 50 dimensions of spatial form, functional supply and transportation connection, accounting for 41.0%; The characteristics of population activities are 16 dimensions, accounting for 13.1%. This dimensional configuration can maintain the full expression of spatial physical environment information while introducing functional and behavioral data to supplement it. All proportional variables retain a clear physical meaning of 0 to 1, and distance, density, and quantity variables are standardized after truncation of outliers, so that different variables can enter a unified feature space without losing their original environmental interpretation.

Because 122-dimensional basic features come from different data modes, direct splicing is easy to produce three problems: first, the same kind of information is repeated, for example, the proportion of

remote sensing vegetation and the green view rate of street view have a certain correlation; Secondly, the difference in the number of different modal dimensions will lead to the numerical advantage of high-dimensional modes. Third, the same spatial environmental factors have different scopes of action at different scales. To this end, this paper adds multi-scale spatial coding after basic feature extraction. For any basic environmental feature, the statistical values of 50 m ontology scale and 100 m, 300 m and 500 m neighborhood are calculated respectively, and the aggregation is completed by learnable scale weights:

$$\tilde{x}_{ij} = \sum_{s=1}^{S} \alpha_{js} \, x_{ij}^{(s)} \tag{9}$$

Where $\tilde{x}_{ij}$ represents the multi-scale fusion result of the j-th feature of the i-th spatial unit, $x_{ij}^{(s)}$ represents the statistical value of the feature at the s spatial scale, S represents the number of scales, this paper takes 4, α_{js} represents the corresponding scale weight, and satisfies that the sum of all scale weights is 1. In this processing method, local environmental variables such as building coverage and green view rate can retain more small-scale information, while variables with strong neighborhood effect such as POI density and road connection can adaptively absorb large-scale features, so as to avoid information loss caused by artificially fixing the same spatial scale.

On the basis of multi-scale coding, feature attention mechanism is further used to compress redundant environmental information. Firstly, the features of different modes are projected to the unified hidden space. For the m-th data mode, the projection process is expressed as:

$$h_i^{(m)} = \phi\left(W_m x_i^{(m)} + b_m\right) \tag{10}$$

Where $x_i^{(m)}$ represents the original feature vector of the i-th spatial unit from the m-th data mode, W_m and b_m represent the linear mapping matrix and bias vector of the mode respectively, $\phi(\cdot)$ represents the nonlinear activation function, and $h_i^{(m)}$ represents the unified dimension representation after mapping. This process only undertakes the function of feature space alignment and does not directly predict micro-update potential, so it can ensure that this section and subsequent potential identification models are functionally independent from each other.

In order to reduce the influence of low-quality or highly redundant modes on comprehensive environmental characteristics, this paper calculates the normalized attention weight according to the internal information of each mode:

$$\beta_i^{(m)} = \frac{\exp\left(e_i^{(m)}\right)}{\sum_{r=1}^{M} \exp\left(e_i^{(r)}\right)} \tag{11}$$

Where $\beta_i^{(m)}$ represents the fusion weight of the m-th data mode in the i-th spatial unit, $e_i^{(m)}$ represents the feature importance obtained by nonlinear transformation of the mode, and M represents the total number of data modes. After normalization, the sum of all modal weights is 1. When a certain space lacks clear street view images, or population activity data coverage is insufficient, the corresponding mode can obtain a relatively low fusion weight, so as to reduce the interference caused by data quality differences on the environmental representation results. Compared with fixed weight fusion, this method can dynamically adjust the information contribution according to the actual data conditions of different corner space.

Table 3. Structural differences of different feature integration methods

Feature integration mode	Dimension of input	Scale adaptive capacity	Redundancy suppression capability	Missing modal adaptation	Dimension
Original features are directly spliced	122	No	low	low	122
Splicing after standardization	122	No	low	low	122
Multi-scale feature coding	122×4	Yes	medium	low	122
Modal projection fusion	122	Part of	medium	medium	128
Multi-scale attention fusion	122×4	Yes	high	high	128

Table 3 illustrates the main differences between the feature construction method in this paper and simple data stitching from the structural level. If the four-scale expansion of 122-dimensional features is directly implemented, the theoretical input scale will reach 488 dimensions, which not only increases the calculation burden of subsequent models, but also introduces a large number of highly correlated variables in the neighborhood. In this paper, multi-scale compression is first completed within the modes, and then each mode is mapped to a unified hidden space. Finally, the environment representation is controlled to 128 dimensions, which reduces the feature scale by about 73.8% compared with direct four-scale splicing. More importantly, this compression is not a simple principal component dimension reduction, but retains the environmental semantics of different data modes, and performs weight learning at two levels of scale and mode respectively, so it can give consideration to computational efficiency and environmental information integrity.

After completing the mapping of each modal space, this paper constructs a cross-modal environmental feature alignment mechanism to reduce the expression bias of the same environmental attribute in different data sources. Specifically, for the six modes of remote sensing, street view, POI, road, building land use and population activities, the projection vectors are unified to the same feature dimension, and the embedding space of different data sources is constrained by mode mean centralization. The cross-modal difference can be expressed as:

$$L_{\text{align}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{m=1}^{M} \left\| h_i^{(m)} - \bar{h}_i \right\|_2^2 \tag{12}$$

Where L_{align} represents the cross-modal feature alignment loss, N represents the number of effective urban edge and corner spatial units, M represents the number of data modes, $h_i^{(m)}$ represents the projection features of the i-th spatial unit under the m-th mode, and $\bar{h}_i$ represents the average representation of all modal features of the spatial unit. This constraint does not require different modes to be completely consistent, but inhibits non-environmental differences caused by data scale, dimension and sampling mode, so that environmental states such as "rich vegetation", "dense function" and "convenient transportation" can form relatively stable representations in the common feature space.

Finally, the six types of modal features are weighted and integrated according to the adaptive

weights to form the comprehensive environment representation of urban corner space:

$$F_i = \sum_{m=1}^{M} \beta_i^{(m)} h_i^{(m)} \tag{13}$$

Where F_i represents the comprehensive environmental feature vector finally formed in the corner space of the i-th city, $\beta_i^{(m)}$ represents the adaptive weight of the m-th data mode, and $h_i^{(m)}$ represents the feature representation after the unified coding of the corresponding mode. In this paper, the final feature dimension is controlled to 128 dimensions, and the vector does not directly contain the micro-update level or manual evaluation results but only describes the ontology of the corner space and the surrounding environment state, so as to ensure that the subsequent potential identification process can be completed independently on the basis of unified, objective and sufficient information input.

3. MULTI-SOURCE FEATURE FUSION ALGORITHM AND MODEL ARCHITECTURE FOR MICRO-UPDATING POTENTIAL IDENTIFICATION

After completing the construction of environmental characteristics of multi-source urban data, it is further necessary to solve the problems such as nonlinear interaction between different modal information, spatial dependence between adjacent spatial units, and unbalanced distribution of samples of different potential levels. Therefore, this paper constructs a multi-source feature fusion model for the identification of urban corner space micro-renewal potential and adopts the calculation process of "modal feature coding, cross-modal interaction, spatial relationship propagation, deep residual fusion, potential prediction, and grade output" as a whole. Different from the method that directly inputs multi-source features into the traditional classifier, this paper firstly retains the independence of different information channels such as visual environment, spatial form, functional supply, traffic connection and usage vitality, completes nonlinear coding within the mode, and then learns the correlation between environmental elements through the cross-modal attention mechanism. Then, the urban corner space and its surrounding spatial units are represented as spatial graph structure, and the Euclidean proximity, road connection and functional similarity relations are jointly encoded as graph edge weights. Finally, multi-layer feature aggregation is completed through residual fusion network, and the adaptive loss function considering class imbalance is used to output the probability of micro-update potential. This architecture enables the model not only to judge the physical and functional conditions of a corner space itself, but also to identify its structural relationship with the surrounding block environment, thus forming the transformation from unit attribute recognition to spatial association recognition.

For the multi-mode environment features formed in section 2, the unified fully connected network is not directly used for processing, but independent coding branches are established for different types of features. Let the m-th modal feature of the edge and corner space of the i-th city be x_i^m. After the m-th independent encoder, the corresponding hidden space expression can be obtained:

$$e_i^m = \mathrm{LN}[\phi(W_2^m \phi(W_1^m x_i^m + b_1^m) + b_2^m)] \tag{14}$$

Where, x_i^m represents the input feature of the i-th spatial unit belonging to the m-th data mode, W_1^m and W_2^m represent the weight matrix of the two-layer coding network respectively, b_1^m and b_2^m are bias parameters, $\phi(\cdot)$ represents the nonlinear activation function, GELU function is used in this paper, and $\mathrm{LN}(\cdot)$ represents the layer normalization. e_i^m is the result of modal coding. The

independent coding method can avoid premature mixing of remote sensing, street view, POI, and other different data in the initial stage of the model, so that all kinds of data form stable internal modality representations. Considering the balance between the model size and the number of samples in the corner space, each branch is finally uniformly mapped to the 64-dimensional hidden space to establish a consistent calculation dimension for the subsequent cross-modal interaction.

Table 4. Main structure and characteristic dimensions of the multi-source micro-update potential identification network

Level of model	Form of input	Core processing mode	Dimension of output	Main functions
Modal coding layer	Multiple types of environmental characteristics	Double layer MLP+GELU+LayerNorm	64/ mode	The internal nonlinear features of the modes are extracted
Cross-modal interaction layer	Multimodal coding vector	Long cross attention	128	Establish dynamic correlation of environmental elements
Space map propagation layer	Node characteristics + adjacency	Multi-head graph attention propagation	128	Learn about neighborhood spatial dependence
Residual fusion layer	Graph features + original fusion features	Residual mapping + gated fusion	128	Shallow information is retained and feature attenuation is suppressed
Potential prediction layer	Integrated spatial representation	Fully connected +Softmax	3	Output low, medium, and high potential probabilities

Table 4 presents the main structure of the model from inputs to potential grade outputs. Compared with directly constructing large-scale deep networks, this paper controls each modal coding dimension at 64 dimensions, expands it to 128 dimensions after completing cross-modal interaction, and keeps 128 dimensions unchanged in the process of spatial propagation and residual fusion, so that the features of the middle layer have a stable scale. Based on the calculation of five main environmental modes, the modal coding stage can form a local hidden space representation of 5×64, while the cross-modal attention layer does not simply form a 320-dimensional splicing vector, but dynamically aggregates different environmental information through attention weights, and finally compresses it to 128 dimensions. In this way, multi-source information can be retained, and the rapid growth of parameter scale caused by continuous splicing can be avoided, and the risk of overfitting can be structurally reduced.

After modal independent coding, this paper introduces a multi-head cross-modal attention mechanism to enable different types of urban environmental information to form adaptive associations according to specific spatial states. For the encoded modal feature matrix E_i, the query matrix, key matrix and value matrix are respectively mapped:

$$Q_i = E_i W_Q,\ K_i = E_i W_K,\ V_i = E_i W_V \tag{15}$$

Where E_i represents the feature matrix composed of all modal coding results of the i-th spatial unit, and W_Q, W_K and W_V represent the learnable mapping matrices of query, key and value respectively. The cross-modal attention result is calculated as follows:

$$Z_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \tag{16}$$

Where, Z_i represents the environmental features after cross-modal interaction, d_k represents the dimension of the key vector, and the square root scaling term is used to prevent the excessive concentration of Softmax gradient caused by the excessive inner product of high-dimensional vectors. This mechanism can learn the dynamic correlation between environmental attributes such as "spatial form, environmental quality," "functional supply-population vitality," and "traffic accessibility, intensity of use." For example, for the corner space formed by road edges, road accessibility and streetscape environment may have higher information contribution; For the scattered space inside the residential area, the surrounding population activities and the allocation of public services may form a stronger correlation, so it is difficult to fully represent the differences of different types of corner space by using fixed weights.

In order to further prevent a high-dimensional mode from taking advantage in attention calculation for a long time, this paper adds adaptive mode gating after cross-mode interaction. The gating weight of the m-th mode is defined as follows:

$$g_i^m = \frac{\exp\left(w_g^T e_i^m + b_g\right)}{\sum_{r=1}^{M} \exp\left(w_g^T e_i^r + b_g\right)} \tag{17}$$

Where g_i^m represents the adaptive weight of the i-th spatial unit corresponding to the m-th mode, w_g and b_g are the trainable parameters of the gated network, and M represents the total number of modes. The sum of the modal weights is 1 after normalization. Different from the data-level feature quality control in section 2, the weight here is backpropagated by the potential recognition task. Its purpose is not to complete data preprocessing, but to make the recognition network automatically change the information source on which the decision depends under different spatial conditions, so as to realize task-driven mode selection.

Only considering the attributes of a single corner space is still not enough to judge its actual micro-update value. For example, the actual regeneration potential of two Spaces with similar area, planting rate and hardening may be quite different if one of them is located between a high-vitality block and a public transport node and the other is located in a low-density urban periphery. Therefore, this paper further constructs the urban spatial adjacency graph $G = (V, E)$, in which each urban corner space or neighborhood analysis unit is regarded as a node, node features are composed of cross-modal fusion results, and the edges between nodes simultaneously consider spatial distance, road network relationship and functional correlation degree. In order to avoid rigid adjacency relations caused by a single distance threshold, this paper transforms the three types of relations into continuous edge weights [21]. The spatial distance relationship is defined as follows:

$$w_{ij}^d = \exp\left(-\frac{d_{ij}^2}{2\sigma_d^2}\right) \tag{18}$$

Where w_{ij}^d represents the distance weight of spatial unit i and j, d_{ij} represents the Euclidean distance between the centroids of two units, and σ_d is the distance attenuation parameter. With the increase of spatial distance, the action weight between nodes decreases rapidly in Gaussian form. The weight of road connection is expressed as:

$$w_{ij}^{r} = \exp\left(-\frac{l_{ij}}{\sigma_r}\right) \tag{19}$$

Where l_{ij} represents the shortest road network distance between two spatial units, and σ_r is the road distance attenuation coefficient. This weight can distinguish the different situations of "a short straight-line distance but blocked by urban expressways, walls and so on" and "a slightly longer straight-line distance but directly connected by the road network".

The degree of functional correlation is obtained according to the cosine similarity of POI function vectors around the two nodes [22],[23]:

$$w_{ij}^{f} = \frac{p_i^T p_j}{\| p_i \|_2 \| p_j \|_2} \tag{20}$$

Where p_i and p_j represent the urban function composition vectors of nodes i and j, the closer w_{ij}^{f} is to 1, the more similar the functional environment of the two spaces. On this basis, the three types of spatial relations are unified to form comprehensive edge weights:

$$w_{ij} = \lambda_d w_{ij}^{d} + \lambda_r w_{ij}^{r} + \lambda_f w_{ij}^{f} \tag{21}$$

Where w_{ij} represents the comprehensive correlation degree of node i and node j, λ_d, λ_r and λ_f correspond to the fusion coefficients of spatial distance, road network and functional relationship respectively, and the sum of the three is 1. In the training process, the three coefficients are updated by learnable parameters, so as to avoid artificially specifying the importance of fixed spatial relations.

Table 5. Relationship types and edge construction methods of urban spatial adjacency graph

Graph relationship type	The main basis	Initial adjacency condition	Range of weights	Effect of expression
Adjacency by distance	Euclidean distance of unit centroid	Within 500 m	0 ~ 1	Describe the geographical proximity effect
Road connection	Shortest road network distance	It can be reached within 800 m	0 ~ 1	Describe real traffic connections
Function association	The POI constitutes the similarity	Similarity ≥0.60	0.60 ~ 1	Describe functional environment linkages
Integrated adjacency	Three types of relationship weighted fusion	The comprehensive weight is ≥0.20	0.20 ~ 1	Establish the final graph propagation edge

The multi-relation mapping method shown in Table 5 makes the connection between spatial units no longer depend entirely on administrative boundaries or geometric adjacency. Distance adjacency mainly describes local spatial continuity, road connection is used to express the actual connection formed by urban transport network, and functional correlation allows some spaces that are not directly adjacent but have similar urban use environment to establish weak connection. 500 m and 800 m are used as the initial geographical and road neighborhood search ranges respectively, and only the effective edges with a comprehensive weight of no less than 0.20 are retained in the end, so the noise caused by distant weakly associated nodes on graph propagation can be reduced. Compared with the traditional K-nearest neighbor graph, this method can more truly express the spatial action mechanism of the

coexistence of "geometric proximity, network accessibility and functional correlation" in urban space.

After the spatial graph is constructed, the multi-head graph attention network is used to propagate the node features in the neighborhood. For node i and its adjacent node j, the attention coefficient between them is first calculated:

$$e_{ij} = \mathrm{LeakyReLU}\left[a^T\left(Wz_i \parallel Wz_j \parallel w_{ij}\right)\right] \tag{22}$$

Where z_i and z_j represent the features of the two nodes after cross-modal fusion respectively, W represents the linear mapping matrix, a represents the attention parameter vector, $\parallel$ represents vector splicing, and w_{ij} represents the aforementioned integrated spatial edge weight. By directly introducing the weight of spatial relationship into attention calculation, the graph network can determine the intensity of information transmission according to the similarity of node features and the real spatial connection at the same time. The normalized graph attention coefficient is:

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(e_{ik}\right)} \tag{23}$$

Where α_{ij} represents the final information contribution ratio of node j to node i, and $\mathcal{N}_i$ represents the set of effective adjacent nodes of node i. After neighborhood aggregation, the node status of layer $l+1$ can be expressed as:

$$h_i^{(l+1)} = \phi\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\,W^{(l)}h_j^{(l)}\right) \tag{24}$$

Where $h_i^{(l+1)}$ is the spatial representation of node i at layer $l+1$, and $W^{(l)}$ is the propagation weight matrix of layer l. Through two-layer graph attention propagation, the environment information of the direct neighborhood and the second-order neighborhood can be obtained simultaneously in the target corner space. In order to avoid node feature convergence caused by too deep propagation layers, this paper does not continue to overlay a large number of graph convolutional layers but enters the residual fusion module after two layers of propagation.

Deep space propagation is prone to the problem that the original local features are covered by neighborhood information. Therefore, residual connection and gated fusion are introduced after graph attention module in this paper. Let the graph propagation result be h_i^G, and the cross-modal original fusion feature be z_i. Firstly, the two dimensions are unified through linear mapping, and then the residual representation is obtained:

$$r_i = \mathrm{LN}\left(h_i^G + W_r z_i\right) \tag{25}$$

Where r_i represents the node representation after residual fusion, and W_r represents the dimension mapping matrix. In order to further control the information ratio between local attributes and neighborhood attributes, a gating factor is introduced [24],[25]:

$$\gamma_i = \sigma\left[W_\gamma\left(h_i^G \parallel z_i\right) + b_\gamma\right] \tag{26}$$

Where γ_i represents the fusion gating coefficient of nodes, $\sigma(\cdot)$ is the sigmoid function, whose value ranges from 0 to 1, and W_γ and b_γ represent trainable weight and bias respectively. The final synthesis is expressed as:

$$f_i = \gamma_i \odot h_i^G + (1 - \gamma_i) \odot z_i \tag{27}$$

Where f_i represents the spatial characteristics of urban edges and corners used for final potential identification, and $\odot$ represents element-by-element multiplication. When the surrounding environment of a certain corner space has a strong influence on its potential judgment, the model can learn a large weight of neighborhood information. When the state of the target space itself is more discriminative, the retention ratio of local cross-modal features is correspondingly increased. Therefore, the same local-neighborhood fusion mode can be avoided in all Spaces.

Urban corner space samples often have obvious category imbalance. In the actual built-up area, a large number of corner Spaces may be in the state of low or medium renewal potential, while the number of high potential samples with obvious renewal value is relatively limited. If ordinary cross-entropy training is directly adopted, the model tends to obtain high overall accuracy by biasing toward the majority categories but reduces the detection ability of high potential space. Therefore, this paper synthesizes category frequency weight and difficult sample adjustment mechanism to construct weighted Focal Loss. For the i-th sample, the loss function is defined as follows:

$$L_i = -\omega_{y_i}\left(1 - p_{i,y_i}\right)^{\eta} \log p_{i,y_i} \tag{28}$$

Where y_i represents the true potential category of the sample, p_{i,y_i} represents the predicted probability of the model for the true category, ω_{y_i} represents the weight of the corresponding category, and η is the adjustment coefficient of difficult samples. The category weight is calculated as the inverse ratio of the effective number of samples:

$$\omega_c = \frac{N}{C \cdot N_c} \tag{29}$$

Where N represents the total number of training samples, C represents the number of potential grade categories, which is set as 3 in this paper, and N_c represents the number of classes c samples. When the number of high-potential category samples is small, weight loss is increased accordingly, so that the model is sufficiently sensitive to a few high-value samples in the training process. The Focal adjustment coefficient reduces the dominant effect of a large number of easily classified samples on the gradient, so that the parameter update focuses more on the samples with fuzzy boundaries and difficult to distinguish.

Table 6. Key model configurations for the potential identification phase

Configuring the project	Set the value	Position of action	Purpose of design
Number of multi-head cross-modal attention heads	4	Cross-modal fusion layer	Learning the interaction of different environmental elements
Figure attention network layer number	2	Spatial relationship propagation layer	Obtain first-order and second-order neighborhood information
Figure the number of attention heads	4	Space propagation layer	Improve the stability of neighborhood relationship expression
Hidden layer unified dimension	128	Fusion and graph propagation stage	Control the parameter scale and unify the feature scale
Dropout ratio	0.30	Attention and full connectivity layer	Reduce overfitting
Focal adjustment coefficient	2.0	Function of loss	Strengthen hard to classify sample learning

The parameters in Table 6 are mainly used to control model complexity and information propagation depth. Four-head attention is used for cross-modal and graph space propagation, so that the model can learn environment association in parallel in different feature subspaces. The spatial map propagation is limited to two layers, focusing on obtaining the first order and second-order environmental relations around the target space, without over-deep neighborhood diffusion, so as to reduce the risk of over-smoothing under the heterogeneous conditions of urban space. In the fusion stage, 128-dimensional hidden vector is uniformly used, and Dropout of 0.30 is adopted in the attention layer and the full connection layer, so that the model can control parameter redundancy while maintaining the ability of multi-source information expression. The Focal adjustment coefficient is set to 2.0, so that the gradient of a large number of easily identified samples is moderately attenuated, and the training focus is turned into a few samples with high potential and complex samples near the boundary of potential level.

In the output stage, the 128-dimensional comprehensive space after the fusion of residuals represents the input potential prediction head. If the potential level includes low potential, medium potential and high potential, the corresponding predicted probability is:

$$p_{ic} = \frac{\exp(w_c^T f_i + b_c)}{\sum_{k=1}^{C} \exp(w_k^T f_i + b_k)} \tag{30}$$

Where p_{ic} represents the predicted probability that the corner space of the i-th city belongs to the c-th potential level, and w_c and b_c represent the weight and bias corresponding to the c-th prediction node respectively, $C = 3$. The final level of the model is determined by the corresponding category of the maximum posterior probability:

$$\hat{y}_i = \arg \max_{c \in \{1,2,3\}} p_{ic} \tag{31}$$

Where, $\hat{y}_i$ is the level of micro-updating potential predicted by the model. In order to avoid further ranking between the same grade space caused by only output discrete categories, this paper also constructs a continuous potential index:

$$S_i = \sum_{c=1}^{3} v_c \, p_{ic} \tag{32}$$

Where S_i represents the continuous micro-updating potential score of the i-th spatial unit, and v_c represents the ordered assignment corresponding to different potential levels, which can be set to 0, 0.5 and 1 respectively. Therefore, the theoretical value of S_i ranges from 0 to 1, and the closer it is to 1, the higher the comprehensive probability that the space is recognized by the model as a high-potential update object. Discrete potential grades are suitable for forming city-scale classification maps, while continuous potential scores can be used for ranking priorities within the same hierarchical space.

In addition to the identification results, the model also establishes explanatory outputs oriented to urban environmental design decisions. For a single spatial sample, the SHAP idea is used to estimate the marginal contribution of input environmental characteristics to the change of potential score, which is basically expressed as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \tag{33}$$

Where ϕ_j represents the contribution value of the j-th environmental feature to the model output, F represents the complete feature set, S represents any feature subset excluding feature j, and $f(\cdot)$ represents the potential prediction function. When $\phi_j > 0$, it indicates that this environmental factor makes space more inclined to be identified as high micro-renewal potential; When $\phi_j < 0$, it indicates that it has an inhibitory effect on potential judgment. At the same time, cross-modal attention weight can be used to judge the information contribution of different data modes in specific corner space, while graph attention weight can track which adjacent spatial units have a great impact on the potential judgment of target space. Thus, a multi-layer explanation chain of "original environmental factors-modal contribution-neighborhood relations-potential results" is formed, rather than relying only on a final classification label.

In general, the model in this paper structurally divides the fusion of multi-source urban environment information and urban spatial relationship learning into two continuous but functionally different stages: The cross-modal attention mechanism is responsible for answering how to combine different environmental elements, the graph attention propagation is responsible for answering how the surrounding urban environment affects the target corner space, and the residual gating further determines the relative contribution of local attributes and neighborhood attributes in the final recognition. On this basis, category-weighted Focal loss improves the training target for the relative scarcity of high-potential samples. Probability output and continuous potential index support both grade recognition and spatial sorting, and SHAP, modal attention and graph attention jointly provide interpretable information. Therefore, a complete algorithm chain from multi-source environmental feature input to potential probability output is formed in this section, while the specific identification accuracy, module contribution, parameter sensitivity, cross-region generalization ability and robustness are left to be independently tested in the subsequent experimental verification section to avoid content repetition between model method design and experimental result analysis.

4. EXPERIMENT DESIGN, ALGORITHM VERIFICATION AND RESULT ANALYSIS

The typical urban built-up area is selected as the experimental area, and the urban corner space sample data set is constructed, and the benchmark labels are established based on on-site investigation, expert judgment, existing updated cases or cross-validation of multi-source evidence. According to the training set, validation set and test set, the space is divided independently to prevent data leakage caused by adjacent spatial samples. Benchmark algorithms such as Logistic Regression, Random Forest, XGBoost, MLP, CNN, multi-modal fusion model and typical graph neural network are set up to compare with the improved model in this paper. Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC and Kappa were used to comprehensively evaluate the recognition performance. Further ablation experiments were carried out to remove modules such as street view, remote sensing, POI, population activity, spatial map structure and cross-modal attention, respectively, to test the actual contributions of various data sources and algorithm modules. Comparison experiments of different fusion strategies were carried out to verify the performance differences between early fusion, late fusion and the cross-modal fusion mechanism in this paper. Hyperparameter sensitivity experiments, different spatial scale experiments and random repeated experiments were carried out to test the stability of the model. The generalization ability of the algorithm in different urban areas or different urban environments is verified by cross-regional migration experiments. The robustness of the model is evaluated by noise disturbance, missing data and partial mode missing experiments. Finally, the identification results are mapped to GIS space to form a classification map of urban corner space micro-renewal potential, and the model judgment results and key influencing factors are explained in combination with typical high, medium and low potential space cases to verify the consistency between the algorithm results and the real space renewal requirements.

In order to verify the effectiveness of the proposed multi-source feature fusion model in the identification of micro-renewal potential of urban corner space, this study selects the urban central built-up area with typical characteristics of high-density built environment as the experimental area. The research scope is about 38. 6km².The space also includes old residential areas, commercial blocks, traffic facilities, public service facilities and the transition zone between old and new urban areas. It can cover a variety of corner space forms such as road corner space, building boundary space, bridge space, neighborhood residual space and scattered and inefficient open space. According to the spatial unit extraction and data cleaning method established in section 2, 1468 candidate corner Spaces were initially obtained, and 1216 effective experimental samples were finally formed after area threshold screening, repeated space merging, data integrity checking and field verification. In order to reduce the label bias caused by a single subjective evaluation, the benchmark label was formed by integrating on-site environmental investigation, existing urban renewal cases, multi-source environmental indicators, and independent evaluation by three experts with urban design and planning background. Experts first made judgments according to low, medium and high levels respectively. When there were grade differences, they combined space environment photos, functional gaps and actual use status for a second verification, and finally obtained 492 low potential, 451 medium potential and 273 high potential samples, accounting for 22.45% of the high potential samples, reflecting a relatively obvious category imbalance characteristic. This is consistent with the fact that the space with high micro-renewal priority value in real cities is relatively limited.

In order to evaluate the consistency of artificial labels, Fleiss' Kappa was used for consistency test based on the independent judgment results of experts. The calculation form is as follows:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{34}$$

Where κ represents the consistency coefficient of multiple evaluators, P_o represents the actual observed agreement rate of experts, and P_e represents the random agreement rate calculated according to the marginal probability of each level. The closer κ is to 1, the higher the degree of agreement among different evaluators is. In the experimental data, the agreement rate of the first independent judgment of the three experts is 87.6%, and Fleiss 'Kappa reaches 0.821, indicating that the labels have high consistency. For the 151 spatial units with obvious differences, the field photos, POI supply, surrounding population activities and existing updated records are verified again, so as to reduce the influence of training label noise on algorithm performance evaluation as much as possible.

Different from the conventional random division, the urban spatial samples have significant spatial autocorrelation. Therefore, this paper adopts spatial block method for independent data division, divides the study area into training area, validation area and test area according to continuous spatial block, and sets 150 m buffer separation zone between the boundaries of different data subsets. The final training set contains 730 samples, the validation set contains 243 samples, and the test set contains 243 samples, with a ratio of about 60:20:20. In the division process, the overall proportion of different potential classes is basically consistent, while the same continuous plot and highly overlapping neighborhood are prohibited from entering different data subsets.

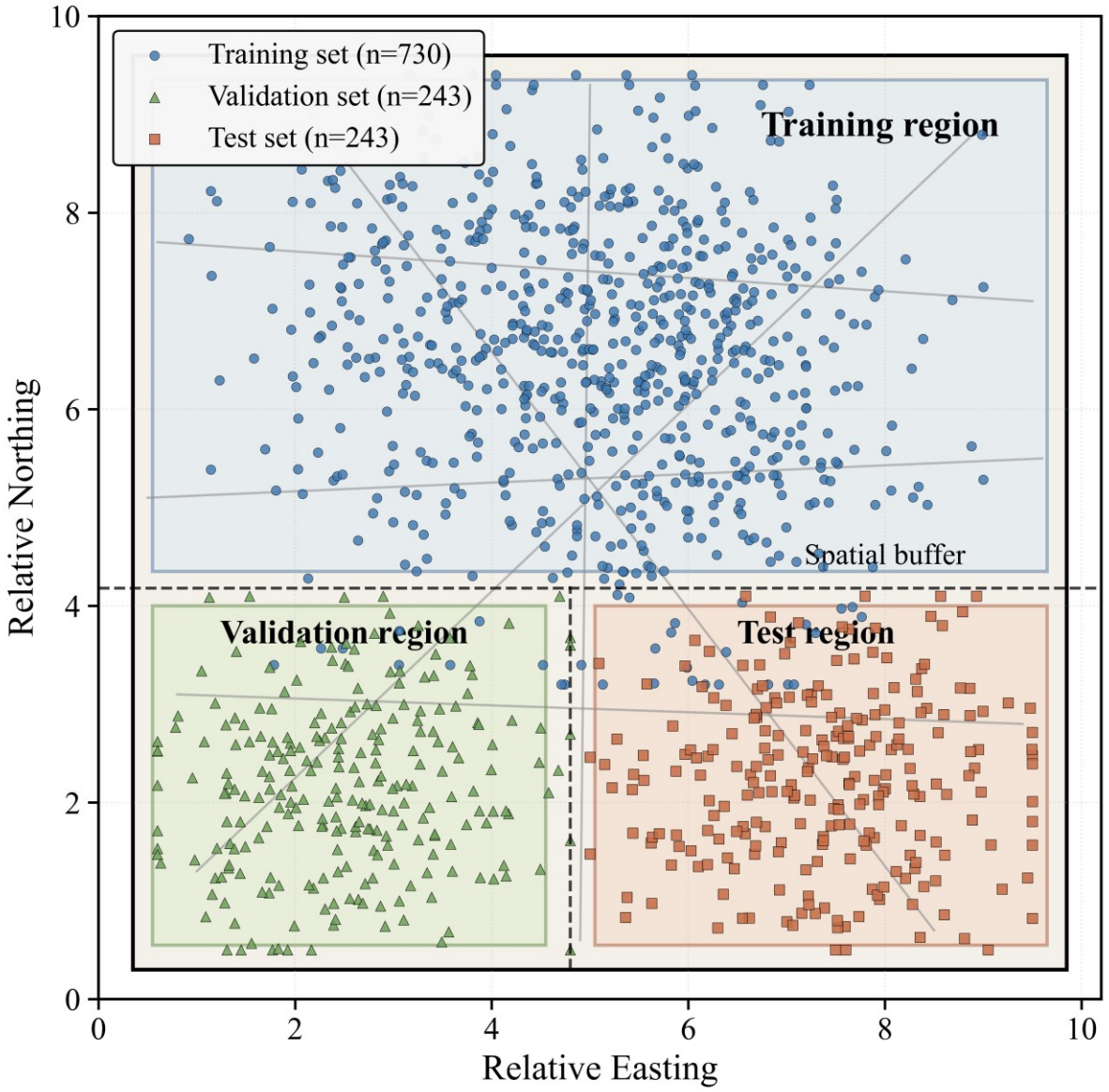

Figure 1. Distribution of urban corner space samples and independent spatial division of training set, validation set and test set

As shown in Figure 1, the training samples are mainly distributed in the central and northern continuous spatial regions of the study area, the validation area is located in the southwest, and the test area is mainly located in the southeast and several independent regions. The training samples are mainly distributed in the central and northern continuous space blocks of the study area, the validation area is located in the southwest, and the test area is mainly located in the southeast and several independent blocks. Compared with completely random division, this independent spatial division increases the difficulty of the test task, but it can more truly evaluate the recognition ability of the model in the face of "new spatial regions not involved in training". Therefore, the following models use exactly the same data division conditions for comparison.

In order to systematically evaluate the recognition ability of different algorithms for micro-update potential level, this paper sets Logistic Regression (LR), Random Forest (RF), XGBoost, MLP, CNN, multimodal fusion network, standard GAT and the model of this paper as comparison algorithms. LR is

used to represent the traditional linear classification method, RF and XGBoost represent the typical integrated machine learning method, MLP is used to evaluate the simple nonlinear feature learning ability, CNN is used to evaluate the local high-dimensional environment feature learning effect, and multi-mode fusion network is used to more truly evaluate the model in the face of "source data fusion ability, GAT is used to evaluate the recognition performance of common spatial graph attention networks. All deep models were trained under the same training set, validation set and test set. AdamW was used as the optimizer, the initial learning rate was set to 0.001, the batch size was 32, the maximum training round was set to 300, and the strategy of stopping early when the validation set had no improvement for 25 consecutive rounds was adopted. All deep learning models were trained independently and repeatedly 10 times, initialized with different random seeds, and the average result of the test set was used as the final performance index.

The classification results were comprehensively evaluated by Accuracy, macro average Precision, macro average Recall, macro average F1-score, ROC-AUC, PR-AUC and Cohen's Kappa. Accuracy is defined as the proportion of correctly classified samples in all test samples:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} I\left(\hat{y}_i = y_i\right) \tag{35}$$

Where N represents the total number of test samples, y_i represents the true category of the i-th sample, $\hat{y}_i$ represents the model prediction category, and $I(\cdot)$ is the indicator function, which takes 1 if the prediction is correct, and 0 otherwise.

For class c potential samples, the precision rate and recall rate are calculated as follows:

$$Precision_c = \frac{TP_c}{TP_c + FP_c} \tag{36}$$

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \tag{37}$$

Where TP_c represents the number of samples that really belong to class c and are correctly predicted, FP_c represents the number of samples that actually do not belong to class c but are incorrectly predicted to be in this category, and FN_c represents the number of samples that really belong to class C but are not correctly identified by the model. Since the number of samples of the three categories is not completely balanced, this paper mainly adopts the macro average index evaluation model, so that the high-potential minority class has the same evaluation weight as the other categories. The macro average F1-score is defined as:

$$\frac{1}{C}\sum_{c=1}^{C} \frac{2Precision_c Recall_c}{Precision_c + Recall_c} \tag{38}$$

Where C represents the number of potential grade categories, and this paper takes 3. This index can consider both precision rate and recall rate and avoid a large number of low and medium potential samples to mask the identification error of high potential categories.

Kappa coefficient is further used to evaluate the consistency of prediction results relative to random classification:

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{39}$$

Where, p_o represents the actual prediction accuracy rate of the model, and p_e represents the random consensus probability obtained according to the marginal distribution of the true label and the predicted label. ROC-AUC mainly reflects the overall discrimination ability of the model under different classification thresholds, while PR-AUC pays more attention to the comprehensive relationship between Precision and Recall under the condition of minority class identification, so it can more sensitively evaluate the model performance under the condition of relatively scarce high-potential spatial samples. The test results of different models are shown in Table 7.

Table 7. Comprehensive performance comparison of different potential identification algorithms

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC	Kappa
LR	0.724	0.716	0.702	0.706	0.801	0.752	0.576
RF	0.781	0.775	0.762	0.767	0.859	0.817	0.667
XGBoost	0.804	0.799	0.785	0.791	0.883	0.842	0.702
MLP	0.816	0.809	0.798	0.803	0.892	0.851	0.719
CNN	0.824	0.817	0.806	0.811	0.898	0.858	0.732
Multi-mode fusion network	0.847	0.842	0.832	0.836	0.915	0.880	0.765
GAT	0.858	0.853	0.844	0.848	0.924	0.892	0.781
Model of this paper	0.887	0.884	0.879	0.881	0.946	0.921	0.824

As can be seen from Table 7, with the enhancement of environmental feature fusion ability and spatial relationship modeling ability, the recognition performance of different algorithms shows a continuous improvement trend. The F1-score of LR is only 0.706, indicating that there is an obvious nonlinear relationship between the potential level and the environmental characteristics, and it is difficult to effectively distinguish the simple linear decision boundary. The F1-score of RF and XGBoost increased to 0.767 and 0.791 respectively, indicating that ensemble learning can capture a certain nonlinear environment combination relationship. MLP and CNN further improve to 0.803 and 0.811, but their performance improvement is limited because different environmental modes and urban spatial adjacency are not explicitly considered.

The F1-score of the multi-modal fusion network reached 0.836, 3.3 percentage points higher than that of the common MLP, indicating that distinguishing different data modes and integrating them can help extract more effective environmental representations. GAT is further improved to 0.848, indicating that neighborhood spatial relations have additional contributions to the judgment of urban corner spatial potential. The F1-score of the model in this paper reaches 0.881, which is 3.3 percentage points higher than that of GAT and 4.5 percentage points higher than that of multi-modal fusion network. ROC-AUC reached 0.946, 2.2 percentage points higher than GAT; PR-AUC reached 0.921, 16.9 percentage points higher than LR. Especially in the case of unbalanced categories, PR-AUC still exceeds 0.92, indicating that the model does not obtain high accuracy by preferring most categories, but can effectively identify

a relatively limited number of high-potential corner Spaces.

To observe the misjudgment relationships among the three types of potential Spaces, a confusion matrix is constructed for the 243 independent test samples. The number of low-potential, medium-potential and high-potential samples in the test set is 98, 90 and 55, of which 91, 77 and 48 were correctly identified by the model, respectively.

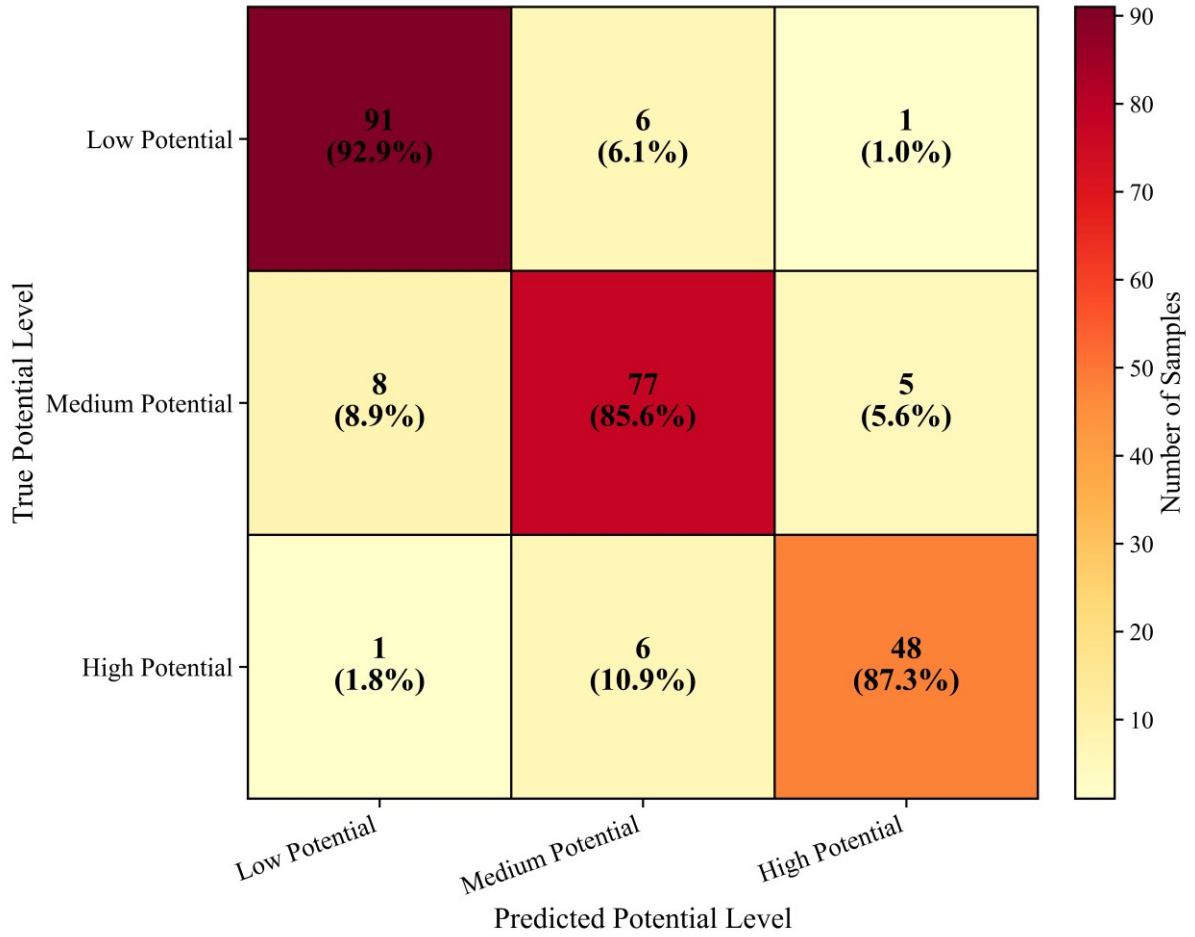

Figure 2. Potential rank confusion matrix of the proposed model on the test set

The confusion matrix in Figure 2 can be set as follows: among the real low-potential samples, 91 are correctly identified as low potential, 6 are misjudged as medium potential, and 1 is misjudged as high potential; Among the true medium potential samples, 8 were misjudged as low potential, 77 were correctly identified, and 5 were misjudged as high potential. Among the true high potential samples, 1 was misjudged as low potential, 6 were misjudged as medium potential, and 48 were correctly identified. Thus, the recall rate of high-potential categories reaches 87.3%, while the proportion that is directly misjudged as low-potential is only 1.8%. Most classification errors occur between adjacent levels, that is, between low-medium and high-high, while the number of serious misjudgments across two levels is only 2, accounting for 0.82% of all test samples. This result shows that the model output has good order, and even if misjudgment occurs, it is more manifested as a small deviation between the boundary levels, rather than the judgment with completely opposite potential nature.

In order to test whether the model improvement is statistically significant and not only caused by random initialization, the F1-score and ROC-AUC generated by 10 independent repeated training are

tested for significance. For the index difference obtained by the two models under the condition of the same random seed, the paired test statistic is used:

$$t = \frac{\bar{d}}{s_d/\sqrt{R}} \tag{40}$$

Where $\bar{d}$ represents the average performance difference between the proposed model and the comparison model in R repeated experiments, s_d represents the standard deviation of the difference, and R represents the number of independent repeated experiments, which is taken as 10 in this paper. For ROC-AUC, the DeLong method is also used for supplementary tests, so as to determine whether the difference in ROC curve area between the two models is significant.

Table 8. The significance test of the performance difference between the proposed model and the main benchmark model

Model of comparison	F1 difference	p value of F1 test	ROC-AUC difference	The AUC test p value
XGBoost	+0.090	<0.001	+0.063	<0.001
MLP	+0.078	<0.001	+0.054	<0.001
Multi-mode fusion network	+0.045	0.002	+0.031	0.001
GAT	+0.033	0.006	+0.022	0.007

Table 8 shows that the F1-score improvement of the model in this paper compared with the four main comparison models has reached a statistically significant level. Relative to the nearest standard GAT, the average F1-score increased by 0.033, corresponding (p=0.006), and the ROC-AUC increased by 0.022, corresponding (p=0.007). This shows that the performance improvement caused by cross-modal interaction, multi-relation space graph and residual gating mechanism is not caused by random training fluctuations. Compared with the multi-modal fusion network without spatial relationship propagation, the F1-score increases by 0.045, which further indicates that the spatial potential of urban edge and corner is not purely determined by the characteristics of a single space, and its adjacent urban environment also contains effective identification information.

In order to further clarify the specific contribution of different data sources and core algorithm modules to the final performance, ablation experiments were carried out under the condition of keeping other network structures and training parameters unchanged, and street view data, remote sensing data, POI data, population activity data, spatial map propagation module and cross-modal attention module were removed respectively. The results of the ablation experiments are shown in Table 9.

Table 9. Multi-source data and ablation experimental results of the core algorithm module

Model setup	Accuracy	F1-score	ROC-AUC	F1 decreases in the relative full model
The complete model	0.887	0.881	0.946	—
Remove the Street View data	0.869	0.863	0.931	0.018
Remove remote sensing data	0.873	0.867	0.934	0.014
Remove the POI data	0.875	0.869	0.937	0.012
Population activity data were removed	0.868	0.862	0.930	0.019
Remove the spatial map structure	0.855	0.850	0.921	0.031
Remove cross-modal attention	0.860	0.855	0.925	0.026

From Table 9, it can be seen that both data modes and algorithm modules make substantial contributions to potential identification, but there are obvious differences in the degree of contribution. After removing the population activity data, the F1-score decreases from 0.881 to 0.862, down 1.9 percentage points, which has the largest impact in the single data mode, indicating that whether the urban corner space has continuous actual use demand is an important basis to distinguish the renewal potential. After removing street view data, F1 decreases by 1.8 percentage points, indicating that pedestrian scale visual environment can supplement interface continuity, green vision status and environmental perception information that remote sensing images cannot directly provide. In contrast, the 1.2 percentage point decrease in F1 after the removal of POI data, although relatively small, still indicates that peripheral functional supply has irreplaceable information value.

The influence of the algorithm module is more obvious. When the spatial map structure is completely removed and only single-spatial multimodal features are retained for classification, the F1-score decreases to 0.850, 3.1 percentage points lower than the full model, which is the largest performance degradation among all ablation Settings. This indicates that for the corner space, the surrounding spatial relationship is not a simple auxiliary variable, but an important structural information that affects the micro-renewal potential. After removing cross-modal attention, F1-score decreases by 2.6 percentage points, indicating that visual, morphological, functional and activity data are not independent of each other, and dynamic interaction can obtain more sufficient environmental combination information than direct feature splicing.

In order to further judge the advantages of the cross-modal mechanism in this paper, four methods are set for comparison: early fusion, late fusion, common attention fusion and cross-modal fusion in this paper. Early fusion directly connects all features in the input layer, and later fusion weights the prediction results after training different modal branches respectively. Ordinary attention fusion only calculates modal weights without explicitly establishing cross-modal feature interaction.

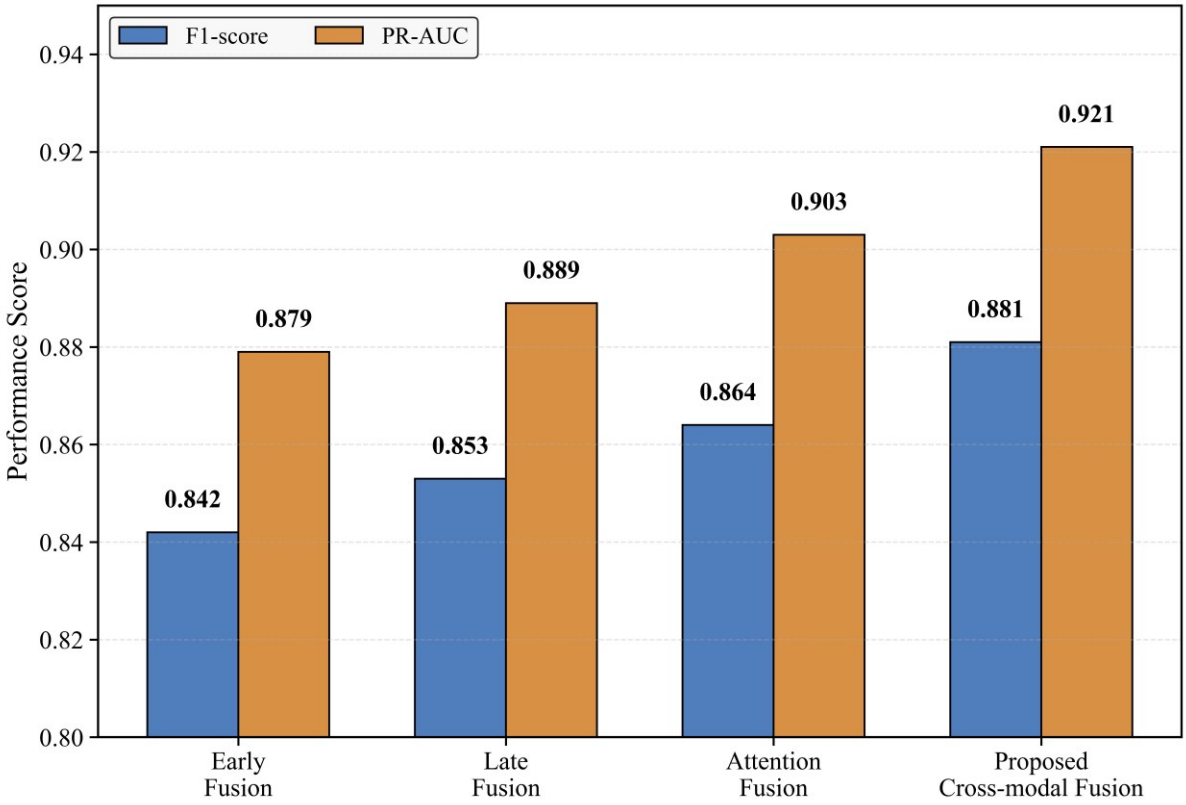

Figure 3. Comparison of F1-score and PR-AUC under different multi-source feature fusion strategies

In Figure 3, the grouped bar graph can be used to represent different fusion methods, in which the F1-score of early fusion is 0.842, and the PR-AUC is 0.879. The late fusion was 0.853 and 0.889, respectively. The general attention fusion was 0.864 and 0.903, respectively; In this paper, the cross-modal fusion reaches 0.881 and 0.921 respectively. From the early fusion to the proposed method, F1-score increases by 3.9 percentage points cumulatively, and PR-AUC increases by 4.2 percentage points. Although the independent expression of each mode is retained in the later fusion, the combination of different information only occurs in the prediction layer, so it is difficult to capture the composite environmental modes such as “low green vision rate but high population activity” and “low service supply but high traffic accessibility”. Ordinary attention fusion can adjust the importance degree of different modes, but it still mainly solves the problem of “which mode is more important”. The method in this paper further learns "how different modes interact", so it obtains more obvious performance advantages.

In spatial relationship modeling, the neighborhood scale directly determines how much surrounding environment information can be obtained by each node. If the neighborhood scope is too small, the model may not be able to obtain the complete block environment. If the range is too large, a large number of distant nodes with weak association with the target space will be introduced. To this end, the maximum neighborhood search radius is set to 100, 300, 500, 800 and 1000 m respectively, and the sensitivity experiment is carried out with other consistent conditions.

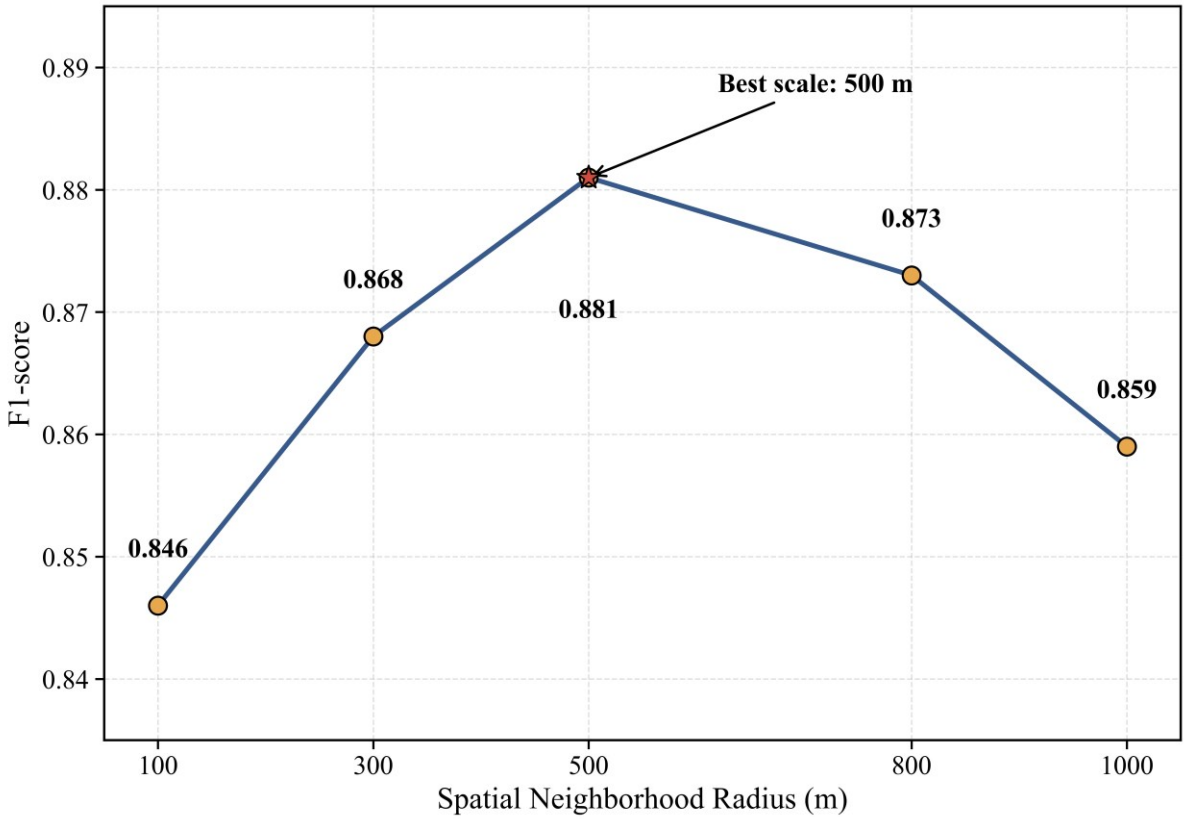

Figure 4. Impact of spatial neighborhood scale change on model F1-score

As shown in Figure 4, the model performance continued to improve as the neighborhood scale increased from 100 meters to 500 m, increasing by 2.2 percentage points from 100 meters to 300 meters and by another 1.3 percentage points from 300 m to 500 m. This shows that too small a neighborhood is not enough to completely describe the traffic, service and activity background of the neighborhood where the corner space is located. When the neighborhood continues to expand to 800 m and 1000 m, the F1-score decreases to 0.873 and 0.859 respectively, indicating that excessive expansion of spatial propagation range will introduce distant units with large functional differences into the neighborhood, thus weakening the discrimination of local spatial structure. Therefore, 500 m is able to strike a good balance between the scope of local association and the integrity of spatial information.

In addition to the spatial scale, the joint sensitivity analysis is conducted on the number of graph network layers, hidden layer dimensions and the number of attention heads. When the number of graph propagation layers increases from 1 to 2, the F1-score increases from 0.864 to 0.881, but decreases to 0.872 and 0.858 after increasing to 3 and 4 layers respectively, indicating that the two-layer propagation can cover the main neighborhood structure, and the node characteristics are prone to excessive smoothing when the graph network is deepened. When the hidden layer dimension is set to 64, 128 and 256, the corresponding F1-score is 0.866, 0.881 and 0.879 respectively, among which the performance of 128 dimension is the highest, while 256 dimension does not continue to produce substantial improvement, indicating that too high dimension increases the number of model parameters. When the number of attention heads increases from 2 to 4, the F1-score increases from 0.871 to 0.881, and when it increases to 8, it is only 0.878. Therefore, four heads of attention are finally adopted to maintain a balance between computational complexity and recognition ability.

Considering that the deep learning model may be affected by random initialization and small batch

sample order, 10 independent random repeated experiments were further conducted. The average value of F1-score of the model in this paper is 0.879, the standard deviation is 0.006, the lowest value is 0.869, and the highest value is 0.888. The mean value of ROC-AUC was 0.945, and the standard deviation was 0.004. The 95% confidence interval of the repeated experiment results is calculated according to the following equation:

$$\bar{x} \pm t_{0.975,R-1} \frac{s}{\sqrt{R}} \tag{41}$$

Where $\bar{x}$ represents the mean value of repeated experiment indexes, s represents the standard deviation of samples, R represents the number of repeated experiments, and $t_{0.975,R-1}$ represents the critical value corresponding to 95% confidence level when the degree of freedom is $R-1$. Therefore, the 95% confidence interval of F1-score is about 0.875-0.883, and the interval width is only 0.008, indicating that the model has good training stability under different random initialization conditions.

In order to test whether the model relies too much on the data distribution of a single urban area, the cross-regional migration method is further used to evaluate the generalization ability. According to the urban form and development age, the study area is divided into central high-density area A, old mixed function area B and new city area C. In the cross-region experiment, the target region labels are not used at all in the training stage, and the model trained by the source region is directly used for the target region prediction. At the same time, multi-regional joint training is set as a supplementary condition.

Table 10. Model generalization performance under cross-region migration

Area of training	Area of test	F1-score	ROC-AUC	F1 variation is tested within relative regions
A	A	0.881	0.946	—
A	B	0.846	0.918	−0.035
A	C	0.831	0.906	−0.050
B	C	0.838	0.911	−0.043
A+B	C	0.857	0.925	−0.024

Table 10 shows that when the model is directly moved from the central urban area A to the old area B, F1-score decreases by 3.5 percentage points; When moving to the new urban area with greater differences in urban form, C decreases by 5.0 percentage points, indicating that there is still some data distribution deviation among different built environments. However, even if the target region label is not used at all, the cross-region F1-score remains above 0.831, indicating that the model does not learn a spatial pattern unique to a single region. When the joint training of A and B regions is adopted, the F1-score in region C increases to 0.857, 2.6 percentage points higher than that of single region training, indicating that expanding the training coverage of urban spatial types can effectively improve the generalization ability of the model.

For actual urban data applications, problems such as cloud pollution from remote sensing images, street view occlusion, POI registration omission and insufficient coverage of population activity data are difficult to be completely avoided. Therefore, the robustness of the model is further evaluated

through noise disturbance and data missing experiments. The continuous feature noise is added with zero-mean Gaussian disturbance according to a certain proportion of the original standard deviation, and the input features are randomly masked according to the specified proportion in the random missing experiment, and the data repair method in section 2 is used for processing. In addition, the extreme cases of complete unavailability of the street view mode and the population activity mode are simulated separately.

Table 11. Model robustness under the condition of data perturbation and missing modes

Condition of disturbance	F1-score	ROC-AUC	Relative decline of F1
No disturbance	0.881	0.946	—
5% characteristic noise	0.874	0.940	0.007
10% characteristic noise	0.864	0.932	0.017
20% of random features are missing	0.849	0.918	0.032
The street view mode is completely missing	0.842	0.912	0.039
The population activity mode is completely missing	0.836	0.905	0.045

It can be seen from Table 11 that the F1-score only decreases by 0.7 percentage points under the condition of 5% characteristic noise, and by 1.7 percentage points under the condition of 10% noise, indicating that the internal coding of the mode, the attention weight and the residual structure can suppress the influence of local abnormal data to a certain extent. When 20% of the input features are missing at random, the F1-score remains 0.849, which is 3.2 percentage points lower than the complete data. Compared with the random missing, the full mode missing has a more obvious impact on the model, and the F1-score decreases by 4.5 percentage points when the population activity data are completely unavailable, which is consistent with the importance of population activity characteristics in the ablation experiment mentioned above. However, the model still maintains an F1-score of 0.836, indicating that the multi-source structure can provide certain compensation through other modes when a certain kind of data is not available, and has good engineering adaptability.

After the model performance test is completed, all the effective spatial units are input into the trained model, and GIS spatial mapping is carried out according to the continuous potential index defined in section 3. Among the 1216 urban corner Spaces, 288 Spaces were finally identified as high potential, accounting for 23.68%; 456 with medium potential, accounting for 37.50%; Low potential 472, accounting for 38.82%. The high-potential space is not randomly distributed in the study area, but mainly along the edge of high-density old residential areas, the periphery of rail transit and bus transfer nodes, the transition area of commercial-residential functions, and the neighborhoods with insufficient public service facilities but high population activity intensity. The low-potential space is more distributed in the peripheral low-density built areas, the surrounding closed industrial facilities and the single functional area with low population activity level.

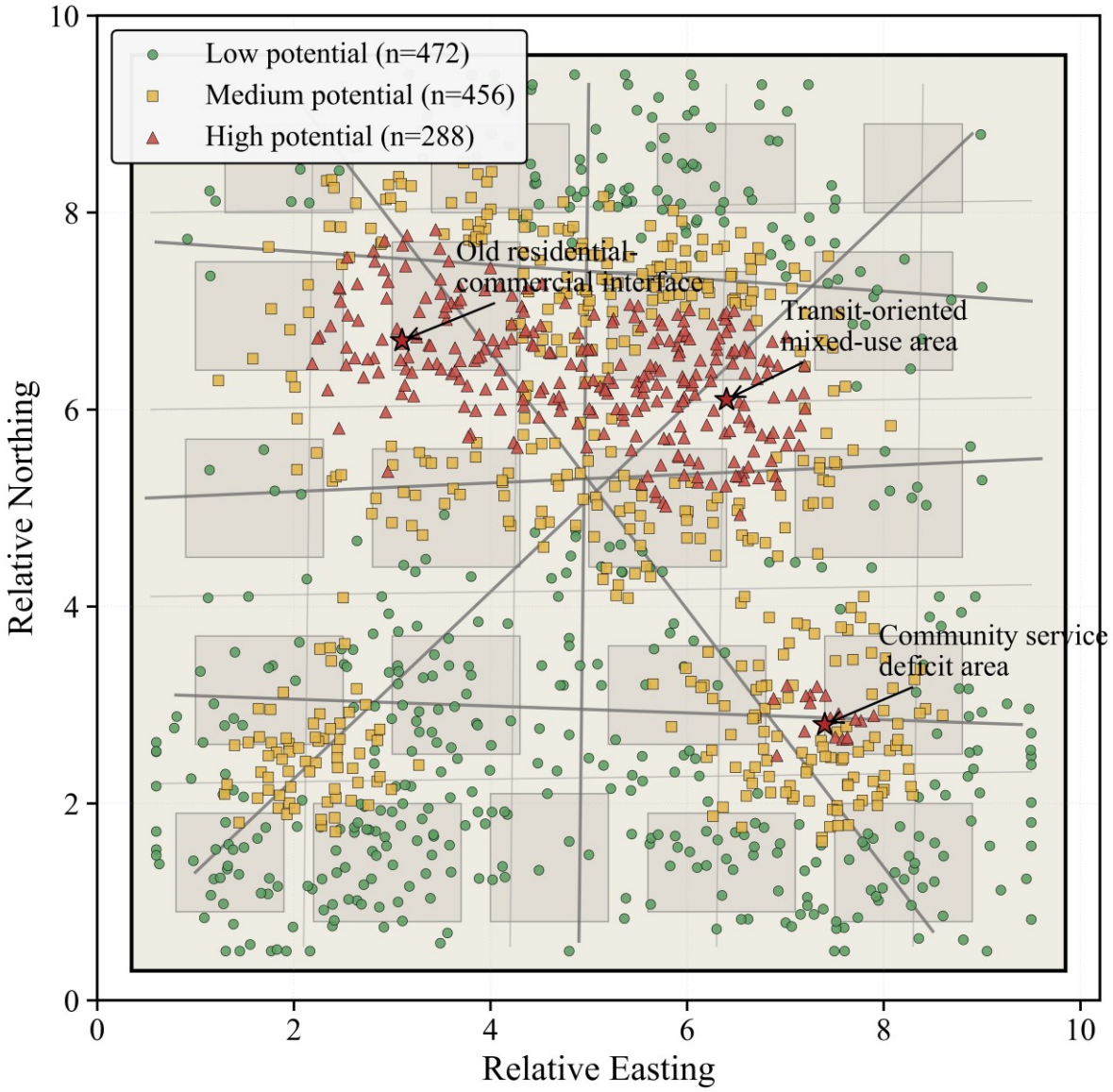

Figure 5. GIS spatial distribution of urban edge and corner space micro-renewal potential level

Figure 5 suggests hierarchic expression from low to high according to the continuous potential index, and simultaneously marks the low, medium and high potential Spaces. Among them, the high potential space shows obvious local agglomeration characteristics, rather than a simple monotonic relationship with the distance from the city center. For example, although some central urban Spaces have high population density and POI density, their potential index does not reach a high level due to the good environmental quality and sufficient supply of public space. On the contrary, some Spaces located at the junction of secondary roads and residential blocks, although small in area, have high population activity, weak green environment and obvious public service gap at the same time, so they get high potential score. This indicates that what the model identifies is the renewal opportunity formed by the joint action of multiple environmental conditions, rather than simply repeating the spatial distribution of population density or land value.

In order to test the degree of consistency between the model output and the realistic update demand, typical Spaces were extracted from the three potential levels of high, medium and low for field

backtracking. The high-potential case A is located between the old residential area and the commercial interface of the community, with a space area of about 860m². SHAP analysis shows that population activity intensity, public service gap, road accessibility and low green vision rate are the main positive contributing factors. From the perspective of the actual environment, this space has a clear demand for walking and community activities, but the current space utilization efficiency is low, so the model judgment is more consistent with the site renewal demand.

Medium potential case B is located in the back of the commercial block, with a high intensity of population activities, but the space area is only about 360m², and there are relatively complete leisure facilities around, and the building enclosure degree is high. The probability of low, medium and high potential output of the model is 0.163, 0.674 and 0.163, respectively. This case illustrates that the model is able to identify offsetting relationships between environmental factors, rather than directly equating "high footfall" with "high renewal potential".

Low-potential case C is located at the edge of a single functional area in the urban periphery. The probability of low potential output for the model is 0.884. According to the on-site investigation, although space has the conditions for physical transformation, the actual demand for short-term use is insufficient. If the ranking is only based on "large idle area", the efficiency of public investment may be low. This shows that the model in this paper can consider both the spatial transformation conditions and the potential use demand, so as to reduce the misjudgment caused by relying solely on physical environment indicators.

Further, the high potential identification results of the model are cross verified with the final judgment of experts and the existing micro-update project space. Among 273 high potential Spaces judged by experts, the model correctly identified 238, and the high potential recall rate was 87.2%. Of the 46 corner space cases that were implemented or included in the renewal plan, 41 were identified by the model as having high potential, with a coverage rate of 89.1%. At the same time, among the 288 high-potential Spaces identified by the model, 43 of them have not been included in the list of existing renewal projects, but 31 of them are considered by experts in the subsequent on-site review to have a clear necessity for environmental improvement and practical use basis, accounting for 72.1% of this part of the sample. This result shows that the model can not only reproduce the existing manual judgment but also has the ability to find the potential renewal space that has not been noticed by the traditional project screening process.

It can be seen from the comprehensive experimental results that the Accuracy, F1-score, ROC-AUC and PR-AUC of the proposed model on the independent test set reach 0.887, 0.881, 0.946 and 0.921 respectively, which is better than the traditional machine learning, multi-layer perceptron, multi-mode fusion network and standard graph attention model. The ablation experiment further proves that spatial relationship propagation and cross-modal attention contribute 3.1 and 2.6 percentage points to the F1-score respectively, while the data of street view, remote sensing, function supply and population activity all show varying degrees of irreplaceability. The spatial scale experiment shows that the 500m neighborhood has better comprehensive recognition performance, and the standard deviation of F1-score obtained by random repeated experiment is only 0.006, indicating that the model training results are relatively stable. Under the condition of cross-regional migration, the F1-score remains between 0.831 and 0.857, indicating that the method has certain spatial generalization ability. In the case of 10% noise and 20% missing random features, the F1-score still reaches 0.864 and 0.849 respectively, which further verifies the adaptability of the model to the incompleteness of the actual multi-source urban data. GIS spatial mapping and typical case backtracking show that the model identification results can better

correspond to the real environment state of urban corner space and the actual micro-update demand, thus forming a complete evidence chain composed of "algorithm accuracy verification, module validity test, stability test, generalization and robustness test, spatial application verification". This paper provides a reliable experimental basis for the priority identification of urban corner space micro-update driven by multi-source data.

5. CONCLUSION

Facing the problems of large number of corner Spaces, complex types, heterogeneous environmental information and limited manual screening efficiency under the background of urban stock renewal, this paper constructs a micro-renewal potential identification method integrating multi-source data such as remote sensing images, street view images, POI, road network, building form, land use and population activities. This study converts data of different scales and structures into jointly computable environmental features through unified spatial index. On this basis, multi-scale coding and cross-modal interaction are used to establish a comprehensive expression of spatial form, visual environment, functional supply, traffic connection and actual use state. Furthermore, the urban spatial proximity, road connection and functional association are introduced into the graph structure learning process, so that the potential judgment is extended from the static attribute analysis of a single spatial unit to the joint identification of "the spatial state itself, the surrounding environmental conditions and the urban functional connection". This method can extract information with potential discrimination significance from complex urban environmental data and provides a data-driven technical path for large-scale automatic identification of urban corner space.

The experimental results show that the Accuracy, F1-score, ROC-AUC and PR-AUC of the proposed method on the independent test set reach 0.887, 0.881, 0.946 and 0.921, respectively. It is better than Logistic Regression, Random Forest, XGBoost, MLP, CNN, multimodal fusion network and standard GAT. Compared with the standard GAT, which has the closest performance, the F1-score of the model in this paper is further improved by 3.3 percentage points, and the ROC-AUC is further improved by 2.2 percentage points, indicating that the introduction of cross-modal interaction and multi-relation spatial connection on the basis of general graph structure learning can effectively enhance the potential recognition ability. The ablation results further show that the F1-score decreases by 3.1 and 2.6 percentage points respectively after removing the spatial graph structure and the cross-modal attention mechanism, indicating that both the spatial neighborhood relationship and the co-expression between different environmental modes are important sources of performance improvement. Street view, remote sensing, and population activity data also show different degrees of complementary effects, among which population activity and street view information have a more obvious impact on the identification results, indicating that the renewal value of urban corner space is not only related to physical space conditions, but also affected by the actual use demand and pedestrian scale environmental quality.

In terms of model reliability, the proposed method shows good parameter stability, data adaptability, and spatial migration capability. The standard deviation of F1-score in 10 random repeated experiments is only 0.006, indicating that the model is not sensitive to the random initialization of parameters and the order of training samples. Under the condition of 5% and 10% feature noise, the F1-score still remains at 0.874 and 0.864 respectively and still reaches 0.849 under the condition of 20% random feature missing, reflecting a certain compensation ability between multi-source information. In the cross-regional experiment, the model can still obtain an F1-score above 0.831 in the new urban area

that does not participate in the training, which is further improved to 0.857 after the multi-regional joint training, indicating that what the model learns is not the local data law of a single urban area, but has a certain ability of cross-spatial environment transfer. At the same time, the neighborhood scale experiment shows that the spatial relationship range of about 500 m can better take into account the local environmental correlation and block scale functional connection, too small range is easy to cause insufficient neighborhood information, and too large range will introduce environmental noise with weak spatial connection with the target.

From the perspective of method innovation, the main contribution of this paper is not to simply increase the types of data, but to form a set of multi-level information organization and calculation mechanism suitable for the identification of urban corner space micro-renewal. Multi-source heterogeneous urban data can be expressed jointly through unified spatial units, so that visual environment, spatial form, facility supply and actual activities can participate in the calculation in the same recognition framework. The cross-modal attention mechanism breaks through the traditional fusion method of fixed weight or direct splicing and enables the model to dynamically learn the combination relationship between different environmental factors according to the specific spatial state. Based on the spatial map structure jointly established by distance proximity, road accessibility and functional association, it further extends the traditional method of constructing neighborhood only based on Euclidean distance and enables the urban spatial network relationship to enter the process of potential identification. On this basis, the residual gating structure dynamically adjusts the spatial ontology information and neighborhood propagation information and improves the identification ability of high-potential minority samples by combining the class imbalance constraint, thus forming an integrated algorithm framework for multi-source environmental data, spatial relationship and potential grade prediction.

At the same time, this paper combines the discrete potential grade with the continuous potential index, so that the identification results can not only answer which level of renewal potential a certain space belongs to, low, medium and high, but also support further ranking among the Spaces of the same grade. GIS mapping results show that high-potential space is mainly concentrated in local areas of cities with strong population activities, insufficient environmental quality, gaps in public services and good transportation links, rather than simply concentrated in urban centers or high-density areas. The high potential Spaces identified by the model are highly consistent with expert judgment and existing micro-renewal projects and can further find some Spaces that have not yet entered the list of existing projects but have obvious improvement needs after on-site review. This shows that the application value of the algorithm is not limited to the classification accuracy itself, but can further serve the urban corner space census, automatic screening of candidate projects and sequencing of construction time, and transform the preliminary screening process that originally relied on a large number of on-site survey and expert experience into a repeatable and quantifiable calculation process, thus providing technical support for the fine allocation of limited renewal resources.

For urban environmental design and refined governance, the method in this paper can also combine the model identification results with the contribution of environmental factors, so that the causes of the formation of high-potential space can be traced to a certain extent. The model can not only output the potential level but also identify the main contributing factors from the aspects of visual environment, population activities, service supply, traffic connections and surrounding spatial influence. Therefore, in practical application, the potential index can be further used as the basis for project screening, and the key environmental factors can be used as the direction of subsequent design intervention, so as to form a continuous workflow of "space discovery, potential ranking, problem diagnosis, and design

response". Compared with the ranking method only using the comprehensive evaluation index, this model can improve the degree of connection between the identification results and the specific environmental design decisions and also help to gradually shift the urban micro-renewal from the empirical driven case selection to the refined decision supported by multi-source data.

There is still room for further improvement in this paper. First of all, the update frequency of data sources in different cities is not consistent, and there may be time misalignment among remote sensing, POI, street view and population activity data. In the future, continuous time-series remote sensing, periodic street view and time-sharing population activity data can be further introduced to construct spatio-temporal joint representation, so that potential identification can be extended from static evaluation to dynamic change monitoring. Secondly, there are differences in spatial scale, block texture, land use system and public space demand in different cities. In the future, domain adaptive, transfer learning and self-supervised pre-training methods can be combined to reduce the dependence of new city applications on large-scale manual annotation samples and enhance the universality of the model under different urban forms and data conditions.

In addition, the existing potential identification is mainly based on the judgment of the environmental state that has been formed, while the urban micro-renewal itself has obvious dynamic feedback characteristics. Therefore, subsequent research can combine annual or quarterly multi-temporal data with the dynamic update mechanism of graph structure to build a dynamic graph model that can continuously update node status and spatial correlation and continuously correct the prediction results through the actual project data before and after the update. After more cities and more samples of actual projects are obtained, the joint optimization between potential identification and renewal costs, implementation difficulty, public benefits and construction constraints can be further studied, so that the model can be further developed from “identifying which Spaces are worthy of renewal” to “which Spaces are prioritized for renewal under resource constraints”. In general, the multi-source data fusion and spatial relationship enhancement potential identification method established in this paper provides an achievable algorithm basis for intelligent discovery and micro-updating assisted decision-making of urban corner space and also provides a new technical idea for the transformation of urban stock space governance from manual investigation to data-driven, from static screening to dynamic intelligent decision-making.

Abbreviations

AdamW, Adaptive Moment Estimation with Weight Decay;

AUC, Area Under the Curve;

CNN, Convolutional Neural Network;

F1, F1-score;

GAT, Graph Attention Network;

GELU, Gaussian Error Linear Unit;

GIS, Geographic Information System;

Kappa, Cohen’s Kappa;

LN, Layer Normalization;

LR, Logistic Regression;

MLP, Multilayer Perceptron;

POI, Point of Interest;

PR, Precision-Recall;

RF, Random Forest;

ROC, Receiver Operating Characteristic;

SHAP, SHapley Additive exPlanations;

XGBoost, eXtreme Gradient Boosting.

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **J.Z.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original draft, Writing – Review & Editing,

Visualization, Supervision, Project administration.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, **J.Z.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used DeepSeek for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Chen, L., Cheng, Y., Zhou, Z., & Wen, Y. (2025). The Impact of Differences in Renovation Models of Abandoned Boiler Rooms on Community Vitality—A Case Study of Shenyang, China. *Buildings*, *15*(11), 1807. https://doi.org/10.3390/buildings15111807

[2] Álvarez-Martínez, J. M., Lugonja, T. N., Valdés, A., Le Barbier, J. G., Suárez, M. P., Romero, G. H., ... & Jiménez-Alfaro, B. (2025). Four decades of remote sensing for monitoring terrestrial ecosystems: A global review and future challenges. *Science of Remote Sensing*, 100341. https://doi.org/10.1016/j.srs.2025.100341

[3] Zhou, Z., Deng, F., Nie, J., Che, F., Guo, Y., & Wang, S. (2026). Research Progress in Algal Bloom Early Warning Technologies for Lakes: Methodological Evolution, Framework Development, and Adaptation to Cold and Arid Region Lakes. *Applied Sciences*, *16*(15), 7469. https://doi.org/10.3390/app16157469

[4] Zhao, T., Chen, G., Pang, C., Seenoi, P., Papukdee, N., & Busababodhin, P. (2025). Hybrid convolutional neural network-graph attention network-gradient boosting decision tree model for seismic impedance inversion prediction. *Journal of Seismic Exploration, 34*(5), 81-98. https://doi.org/10.36922/JSE025310051

[5] Zhao, T., Chen, G., Suraphee, S., Phoophiwfa, T., & Busababodhin, P. (2025). A hybrid TCN-

XGBoost model for agricultural product market price forecasting. *PLoS One, 20*(5), e0322496. https://doi.org/10.1371/journal.pone.0322496
[6] Qutaishat, D., & Li, S. (2025). Multimodal Spatiotemporal Deep Fusion for Highway Traffic Accident Prediction in Toronto: A Case Study and Roadmap. *ISPRS International Journal of Geo-Information*, *14*(11), 434. https://doi.org/10.3390/ijgi14110434
[7] Zhao, T., Chen, G., & Busababodhin, P. (2026). An intelligent expert system for logistics disruption prediction and mitigation in global supply chains: Integrating graph-based risk inference with ensemble forecasting. *Scientific Reports*. https://doi.org/10.1038/s41598-026-61617-0
[8] Sowmya, K., & Basthikodi, M. (2026). CGP-Net: A Cross-Modal Attention Network with Graph Priors for Integrated Multi-Omics Disease Subtyping. *International Journal of Computer Information Systems and Industrial Management Applications*, *18*(14s), 401-415. https://doi.org/10.70917/ijcisim-2026-4239
[9] Zhao, T., Chen, G., Pang, C., Li, L., & Busababodhin, P. (2026). Forecasting global agricultural trade imbalances using a hybrid deep learning and gradient boosting framework. *Discover Computing, 29*(1), 443. https://doi.org/10.1007/s10791-026-10367-8
[10] Wang, M., Zhang, Q., Liu, X., Zhang, J., Yu, F., Zhang, X., & Zhao, R. (2025). Research progress on multimodal data fusion in forest resource monitoring. *Frontiers in Plant Science*, *16*, 1710618. https://doi.org/10.3389/fpls.2025.1710618
[11] Zhao, T., Chen, G., Li, J., Pang, C., Seenoi, P., Papukdee, N., & Busababodhin, P. (2026). XGBoost-Deep Residual (FedXGB-ResNet) collaborative porosity prediction in a federated learning framework. *Journal of Seismic Exploration, 35*(2), 201. https://doi.org/10.36922/JSE025440094
[12] Zhao, T., Chen, G., Pang, C., Seenoi, P., Papukdee, N., Busababodhin, P., & Du, Y. (2025). Hybrid convolutional neural network-graph attention network-gradient boosting decision tree model for seismic impedance inversion prediction. *Journal of Seismic Exploration, 34*(5), 81-98. https://doi.org/10.36922/JSE025310051
[13] Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications–Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815
[14] Zhao, T., Chen, G., Pang, C., Seenoi, P., Papukdee, N., & Busababodhin, P. (2025). Time-lapse earthquake difference prediction based on physics-informed long short-term memory coupled with interpretability boosting. *Journal of Seismic Exploration, 34*(3), 25. https://doi.org/10.36922/JSE025310049
[15] Zhang, L., Liang, T., Wei, X., & Wang, H. (2024). An improved indicator standardization method for multi-indicator composite evaluation: A case study in the evaluation of ecological civilization construction in China. *Environmental Impact Assessment Review*, *108*, 107600. https://doi.org/10.1016/j.eiar.2024.107600
[16] Mukhametzyanov, I. (2023). On the conformity of scales of multidimensional normalization: An application for the problems of decision making. Decision Making: Applications in Management and Engineering, 6(1), 399-341. https://doi.org/10.31181/dmame05012023i
[17] Kuang, W. (2019). Mapping global impervious surface area and green space within urban environments. Science China Earth Sciences, 62(10), 1591-1606. https://doi.org/10.1007/s11430-018-9342-3
[18] Abass, K., Buor, D., Afriyie, K., Dumedah, G., Segbefi, A. Y., Guodaar, L., ... & Gyasi, R. M.

(2020). Urban sprawl and green space depletion: Implications for flood incidence in Kumasi, Ghana. *International Journal of Disaster Risk Reduction*, *51*, 101915. https://doi.org/10.1016/j.ijdrr.2020.101915
[19] Liang, W., Ahmad, Y., & Mohidin, H. H. B. (2024). Spatial pattern and influencing factors of tourism based on POI data in Chengdu, China: W. Liang et al. Environment, development and sustainability, 26(4), 10127-10143. https://doi.org/10.1007/s10668-023-03138-8
[20] Lu, C., Pang, M., Zhang, Y., Li, H., Lu, C., Tang, X., & Cheng, W. (2020). Mapping urban spatial structure based on poi (point of interest) data: A case study of the central city of Lanzhou, China. *ISPRS International Journal of Geo-Information*, *9*(2), 92. https://doi.org/10.3390/ijgi9020092
[21] Izadi, M., & Saeedi, P. (2012). Robust weighted graph transformation matching for rigid and nonrigid image registration. *IEEE Transactions on Image Processing*, *21*(10), 4369-4382. https://doi.org/10.1109/TIP.2012.2208980
[22] 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*(4), 825-848. https://doi.org/10.1080/13658816.2016.1244608
[23] Qin, Q., Xu, S., Du, M., & Li, S. (2022). Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest. *International Journal of Digital Earth*, *15*(1), 2468-2494. https://doi.org/10.1080/17538947.2022.2160841
[24] Xing, R., Cheng, R., Huang, J., Li, Q., & Zhao, J. (2023). Learning to understand the vague graph for stock prediction with momentum spillovers. *IEEE Transactions on Knowledge and Data Engineering*, *36*(4), 1698-1712. https://doi.org/10.1109/TKDE.2023.3310592
[25] Zhang, L., Lin, G., Wei, L., & Kou, Y. (2024). Feature subset selection for multi-scale neighborhood decision information system via mutual information: L. Zhang et al. *Artificial Intelligence Review*, *57*(1), 15. https://doi.org/10.1007/s10462-023-10626-w

APA Citation

Zhang, J. (2026). Algorithm for Identifying Micro-Update Potential in Urban Corner Spatial Environment Design by Integrating Multi-Source Data. *Journal of Digital Frontier, 1*(1), 106-143. **https://doi.org/10.67541/jdf2605**