

Toward precision urban resilience: Integrating multimodal large language models with spatial analytics for fire risk governance

Dong Xie^a, Senhong Cai^a, Ke Wang^a, Ce Hou^b, Shaoqing Dai^c, Jingnan Huang^{a,d,*}

^a School of Urban Design, Wuhan University, Wuhan, Hubei, China

^b Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

^c School of Resource and Environmental Sciences, Wuhan University, Wuhan, Hubei, China

^d Hubei Habitat Environment Research Centre of Engineering and Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:

Urban fire risk
Multimodal large language model
Street view imagery
Urban resilience
Scenario simulation

ABSTRACT

Traditional urban fire-risk assessments rely heavily on static proxies, such as point-of-interest (POI) density, creating a semantic gap between macro-level urban indicators and micro-scale physical hazards. To bridge this gap, we develop a visual-semantic risk indicator system that uses multimodal large language models (MLLMs) to extract fine-grained fire-hazard features from street-view and remote-sensing imagery and integrates these features into a geographically weighted XGBoost (GW-XGBoost) model. Applied to Wuhan, China, the framework identifies localized physical hazards within broader patterns of urban intensity and distinguishes urban areas that appear similar in terms of static indicators but differ substantially in their micro-scale hazard conditions. Spatial interpretation reveals a clear outward shift in dominant risk predictors: fire risk in the urban core is associated primarily with functional overload, particularly the mixing of dining and residential uses, whereas risk in transitional areas is more strongly associated with infrastructure deficits and informal practices such as illicit e-bike charging. Scenario simulations indicate that interventions targeting such practices in transitional areas are projected to produce the greatest reductions in fire risk. By coupling direct visual-semantic sensing with spatially explicit modeling, this study provides an evidence-based framework for precision urban fire-risk governance.

1. Introduction

With the acceleration of urbanization, the global urban population density continues to increase, leading to a significant intensification of fire risks (Li et al., 2023). This not only poses a huge threat to public safety but also causes serious impacts on the sustainable development of cities (Hossain & Smirnov, 2023). According to the United Nations report, more than 150,000 people die globally each year from fire or burn-related injuries, with over 95% occurring in low- and middle-income countries, and these events are often accompanied by enormous economic losses and infrastructure damage (United Nations Office for Disaster Risk Reduction, 2019). For example, the major residential fire incident in Hong Kong's Wang Fuk Court residential area at the end of 2025 resulted in nearly 200 deaths and substantial economic losses, highlighting the potential hazards in high-density urban built environments (The New York Times, 2025). In this context, predicting fire occurrences through fire risk prediction models has become a research hotspot, aiming to identify potential risk factors and provide scientific

decision-making support for precision urban resilience, emergency management, and targeted resource allocation (Cui et al., 2024; Wang, Zhao, et al., 2021).

Existing fire risk prediction models mainly include statistical models, geographic information system (GIS) models, and machine learning (ML) models. Early models mostly adopted statistical methods, predicting based on historical fire data and environmental variables (Rohde et al., 2010). GIS-supported analysis and modeling play a potentially important role, and GIS-based analysis helps with risk identification, resource positioning and path selection (for firefighters and equipment), allocation of prevention facilities, and policy evaluation (Masoumi et al., 2019; Thakare & Tajne, 2025; Yu et al., 2018). In recent years, ML methods have been widely applied, as they can handle nonlinear relationships and improve prediction accuracy (Ahn et al., 2024). However, these models often rely on structured data, such as macro-statistical data or POI (point of interest) data. Due to neglecting the importance of visual data, the existing evaluation indicator systems are not comprehensive enough, while also ignoring the dynamism and

* Corresponding author at: School of Urban Design, Wuhan University, No. 8, Donghu South Road, Wuhan, 430072, Hubei, China.

E-mail address: huangjn73@whu.edu.cn (J. Huang).

complexity of visual environmental information itself (Liu et al., 2017). Regarding resilience governance, results based on these immutable indicators offer limited actionable guidance for micro-scale intervention, failing to address the dynamic vulnerabilities in high-density environments (Sun, 2026; Sun et al., 2025).

Factors in urban environments are closely linked to fire occurrences, and urban remote sensing (RS) and street view image (SVI) data provide a new data source for a comprehensive assessment of urban fire risks (Sijabat et al., 2025). RS images provide broad aerial environmental features, which can identify important factors affecting fire spread at the macro level, such as urban density and emergency exit conditions (Ibrahim et al., 2020; Kelly et al., 2012). SVIs provide a human-centered perspective, allowing attention to microscopic risk factors that trigger fires, such as cluttered wires or piles of flammable materials (Biljecki & Ito, 2021; Chacra & Zelek, 2017).

However, the environmental factors of fire occurrences are very complex and difficult to evaluate with a simple model or standard. For example, an area with many wires is prone to fires, while another area's roof materials are flammable and also prone to fires. It is difficult to systematically integrate these factors within a traditional research framework. Therefore, although we can observe elements that easily trigger fires from SVIs and RS images, there is still a lack of a powerful tool to effectively identify these diverse fire elements. The emergence of multimodal large language models (MLLMs) provides us with a powerful tool. The models not only have state-of-the-art object detection and recognition capabilities but can also use their emergent world knowledge to automatically reason and generalize factors of fire occurrences, achieving a more comprehensive summary of fires (Hou et al., 2025). Building upon this premise, this study explicitly posits two core scientific hypotheses:

- Hypothesis 1 (Semantic Explanatory Power): Multi-source urban imagery (street-view and remote sensing) embeds rich semantic risk features that possess strong explanatory power for fire hazards, and these nuanced environmental triggers can be effectively extracted and decoded by MLLMs.
- Hypothesis 2 (Spatial Heterogeneity): The statistical correlations and predictive efficacy of these diverse semantic risk features concerning fire events exhibit significant spatial heterogeneity across different urban environments.

To empirically test these hypotheses, this study proposes an interpretable geographically weighted ML analysis framework for mining fire risk elements based on multimodal data to quantify urban fire risk patterns. The main contributions of this study are: proposing an MLLM-based framework that uses multimodal data such as RS and SVI information for fine-grained scenario prediction of urban fire risks, and conducting natural language clustering based on fire risk texts to conduct comparative research on different fire risk types; technically, adopting an interpretable geographically weighted ML method to spatially quantify the risk driving mechanisms and differences of key features, strengthening the result interpretation path; identifying key visual environmental factors affecting fire occurrences and conducting governance scenario simulations, providing an important basis for similar research and fire prevention decisions.

2. Literature review

2.1. Factors causing urban fires

The inducing factors of urban fires are diverse, including climatic conditions, built environment, human factors, and socioeconomic factors (Zhang, Song, et al., 2025). Climatic factors, such as high temperature, drought, and wind speed, are the main drivers of fire spread, especially in extremely dry and hot areas (Flannigan & Harrington, 1988; Sufianto & Green, 2012). Wind speed and direction play a

dominant role in fire spread, owing to the heterogeneous spatial distribution of buildings, terrain, and vegetation within cities (Cao et al., 2025). Under intensified climate change, extreme weather events amplify the risk of urban fires (Yao et al., 2024).

In terms of the built environment, densely populated residential areas, old buildings, and narrow or obstructed public spaces are common high-incidence areas for fires (Chen et al., 2023; Elo et al., 2009; Zhang et al., 2023). Spatial variables such as building density and vegetation configuration have key impacts on regional fire spread (Shaham & Benenson, 2018). In addition, some spaces are high-risk places for fire occurrences, such as kitchens, ignition sources involving cooking equipment, absence of alarm systems, or response delays, all proven to be closely related to increased fire spread risks (Alshawawreh & Rush, 2024; Ardianto & Chhetri, 2019; Liu et al., 2025).

Human factors such as improper charging of electric bicycles, improper operation of electrical appliances, blockage of fire passages, and accumulation of flammable materials in safety passages further amplify fire risks (Hossain & Smirnov, 2023; Sufianto & Green, 2012). Among them, the number of fires caused by electric bicycles is increasing. For example, in 2021, electric bicycles in China caused nearly 18,000 fires, resulting in 57 deaths (Li et al., 2022). This is due to thermal runaway in the batteries of electric bicycles, which produces jet flames and further triggers fires (Yu et al., 2024). An easily overlooked urban fire factor is human-set outdoor billboards, which serve to attract consumption in commercial markets, but these billboards are mostly flammable and pose huge fire hazards. Fire hazards caused by building exterior wall billboards include easy ignition, fire spread, impact on personnel evacuation, and firefighting (Zhou, 2013).

Socioeconomic factors, such as poverty, low education levels, and population mobility, also affect the probability of fire occurrences, leading to insufficient fire prevention and extinguishing measures (Hu et al., 2019; Parente et al., 2024; Reilley et al., 2023). Socioeconomic status is negatively correlated with fire incidence, with higher fire rates in areas of lower socioeconomic status (Chhetri et al., 2010). Using social indices based on existing censuses, a comparative study of Brisbane (Australia) and Cardiff (UK) shows associations between social disadvantage and most types of fire incidents (Corcoran et al., 2011).

When predicting the occurrence of urban fires and drawing up contingency plans, meteorological conditions and socioeconomic factors are difficult to change. Therefore, the focus should be on the built environment and human factors, capturing microscopic details through visual risk factors to improve assessment accuracy. Although these variable factors have been proven to be causes of fire occurrences and beneficial for fire prevention, current literature has not incorporated them into predictive assessment models.

2.2. Methods for fire risk prediction

Traditional fire risk assessments have evolved from simple hazard mapping to complex predictive modeling, thanks to geospatial data acquisition, computational tool development, and the popularization of predictive algorithms (Iban & Sekertekin, 2022; Sun et al., 2025). Early studies focused on statistical methods combined with GIS overlays to associate fire events with environmental and predictive variables (Dong et al., 2018; Wang, Yuan, et al., 2021). For example, using Bayesian methods to generate spatial prediction maps considering factors like population density, building age, and meteorological conditions (Rohde et al., 2010), and urban fire risk modeling based on spatial regression methods (Ma et al., 2025). Although these methods are suitable for large-scale analysis, they often overlook fine-grained spatial heterogeneity and nonlinear interactions, leading to limited predictive capabilities in complex urban environments.

The emergence of RS technology marks significant progress, making the monitoring of large-scale environmental factors possible. RS images from sources like Landsat or high-resolution satellites have been widely used to simulate wildfire risks in urban fringe areas, revealing

correlations between urban expansion and fire occurrences (Xu, Li, et al., 2025). However, RS images alone are difficult to capture microscopic ground details, prompting integration with SVIs. SVIs have been used to integrate spatial vulnerability to quantify urban fire-prone areas (Xia, Zhang, et al., 2025), but have not yet been combined with RS images for fire risk assessment. Studies using these methods show higher accuracy in capturing multi-scale risks, but manual feature extraction remains labor-intensive and susceptible to subjective influence (Jin et al., 2020).

ML technologies address some limitations through automated pattern recognition and have been applied to scenarios such as post-disaster losses (Wang et al., 2022), forest fires (Thi Hang et al., 2024), and urban fire spread (Liu et al., 2025) due to their efficiency in handling high-dimensional data and capturing nonlinear relationships. In terms of algorithms, Song and Liu, based on multi-source data from prefecture-level and above cities in China, adopted five ML algorithms for modeling and comparison to quantify fire risks and explore their spatiotemporal distribution, with results showing that Extreme Gradient Boosting (XGBoost) performed best in predicting fire frequency and direct losses, with extremely high generalizability and reliability (Song & Liu, 2026). In addition, XGBoost has also been proven more accurate than multi-scale geographically weighted regression in predicting fire occurrences and losses (Yang et al., 2025). Despite these advances, integrating diverse data sources remains challenging, often requiring domain-specific preprocessing.

Furthermore, large language models (LLMs), as a class of deep

learning technologies, are increasingly applied in disaster prevention, particularly for information extraction, decision support, and predictive simulation (Xia, Tong, & Long, 2025; Xu, Ma, et al., 2025), enabling efficient, highly adaptive, and automated information integration (Chen et al., 2024). For example, LLMs can process social media data to monitor disaster events (such as fires, floods, and earthquakes) and generate emergency response strategies (Qian et al., 2024; Sun et al., 2024; Wang et al., 2024). However, conventional LLM pipelines remain separated from visual data acquisition and cannot directly process heterogeneous non-text data. By comparison, MLLMs solve the long-standing challenge of modality separation in disaster assessment by achieving complex, context-aware integration of visual and textual information (El-Niss et al., 2024; Fan et al., 2020).

In summary, traditional fire risk prediction models rely on statistical features, mostly on static or simple urban physical parameters, neglecting visual environmental information, lacking fusion consideration of macro factors and micro details in built environments, leading to limited applicability in high-density cities. These static parameters may reasonably predict fire occurrences, but cannot perform dynamic governance. In terms of research objects, only a few involve urban fire risk assessment and prediction studies (Cui et al., 2024). Urban fires are usually smaller in scale but cause greater damage, and due to their inherent dynamism and uncertainty, they are more difficult to predict in complex environments (Jin et al., 2020). In addition, to our knowledge, no studies have yet fused street view and RS with MLLM for urban fire assessment, leading to gaps in multimodal synergistic effects and

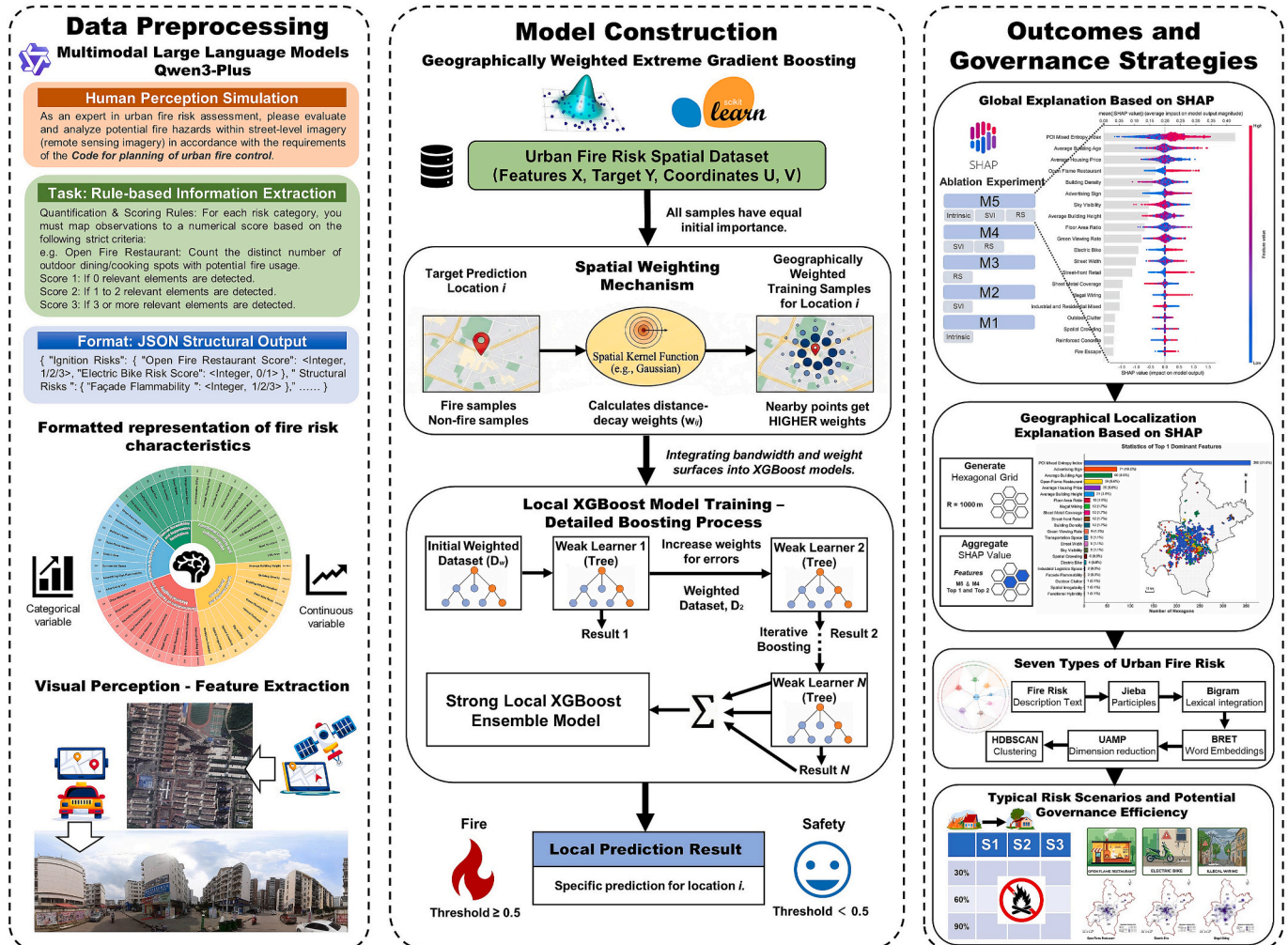


Fig. 1. Research framework.

interpretable predictions in this field.

3. Methodology

This study integrates multimodal data, including remote sensing and street view imagery, to conduct granular scenario-based predictions of urban fire risk within the MLLM framework. The research process comprises three primary stages (see Fig. 1): (1) Conducting prompt engineering and employing feature extraction techniques from MLLM to convert street view and remote sensing imagery into quantifiable risk features; (2) Incorporating urban fire risk features into a geographically weighted machine learning prediction model, trained under binary classification supervision to identify and explain key predictors; (3) Validating three representative governance actions through scenario simulation and formulating corresponding policy recommendations.

3.1. Research area and data overview

This study selects Wuhan, a national central city and the capital of Hubei Province, as the research area. By the end of 2024, its permanent resident population exceeded 13.8 million, characterizing it as a densely

populated and economically concentrated megacity in Central China with a high urbanization rate.

Wuhan's rapid urbanization has resulted in a complex built environment, comprising high-density residential zones, historic neighborhoods, emerging commercial centers, and expansive industrial areas. This functional diversity generates varied fire risk profiles, enhancing the city's representativeness for this research. These attributes establish Wuhan as an ideal case for investigating spatiotemporal patterns of urban fire risk. Fig. 2 illustrates the geographical location of Wuhan City, the spatial distribution of fire incidents (positive examples), and rule-based random negative examples.

The dataset relies on 39,129 fire incidents (Jan 2021–Sep 2025) provided by the local fire department as positive samples. As non-fire incidents cannot be recorded, we employed a rule-based pseudo-absence generation strategy to create negative samples. To strictly prevent the model's spatial distribution bias, we designed a three-step constrained sampling pipeline:

- Step 1: Sampling was strictly constrained to the functional urban built-up area, masking out water bodies, mountains, and farmland where urban fires logically do not occur.

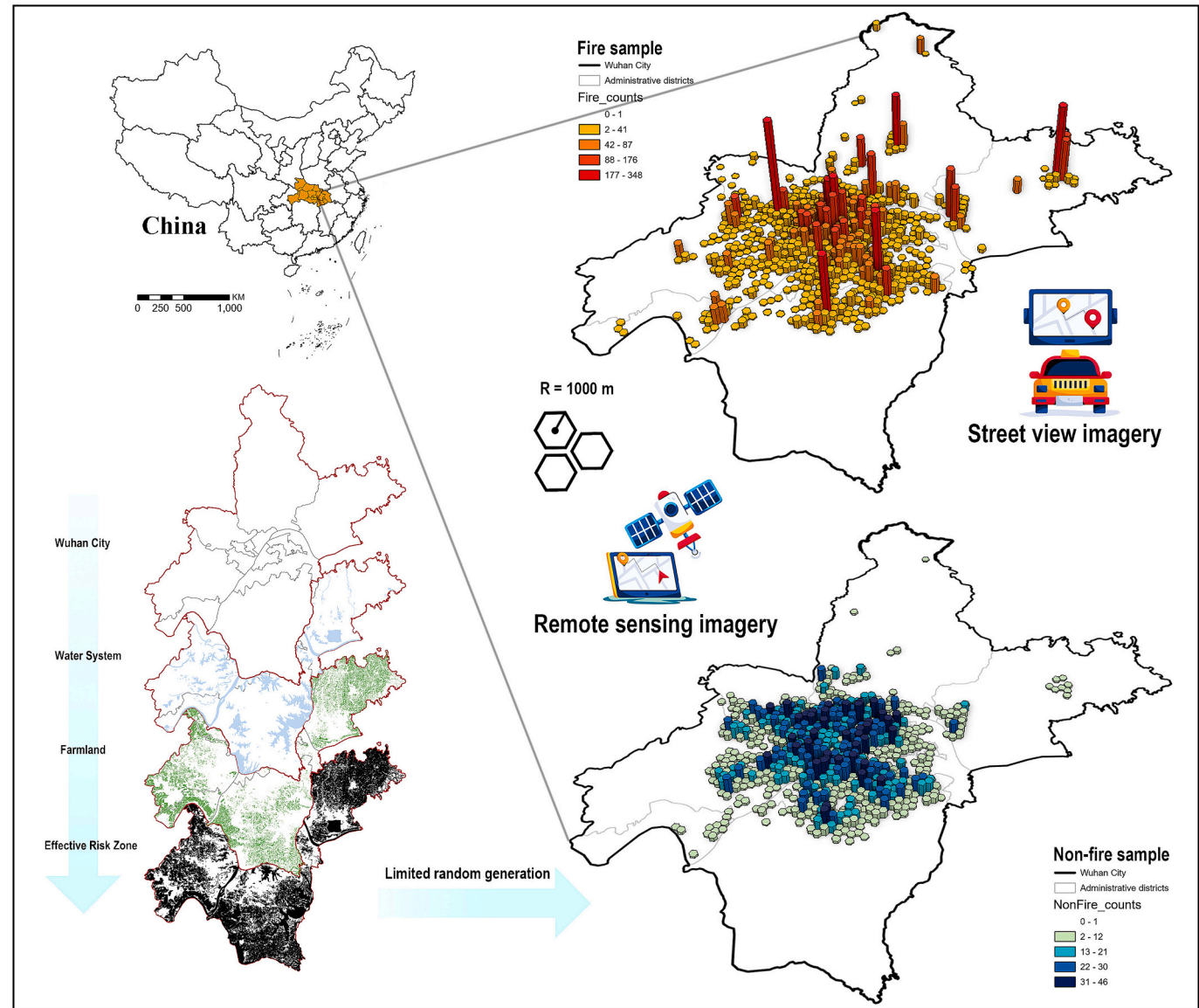


Fig. 2. The geographical location of Wuhan City and the distribution of samples.

- Step 2: Negative samples were generated following the Kernel Density Estimation (KDE) of the positive samples. By ensuring that negative samples share the same macro-spatial distribution as the fire incidents, we forced the model to learn the fundamental differences in micro-visual features rather than merely memorizing geographical coordinates or population densities.
- Step 3: A 75 m (standard spacing for valid street view images) exclusion buffer around any positive fire point was enforced. This ensures spatial independence and prevents information leakage, avoiding scenarios where a negative sample captures the identical street view or remote sensing imagery as a nearby positive sample.

Of the 39,129 recorded fire incidents, 17,189 positive samples with valid matched imagery were retained. While an initial 1:1 ratio was targeted, the final rigorous filtration—subject to the physical availability of Baidu Street View (BSV) panoramas (2019–2024)—yielded 8141 valid negative samples. To address the resulting class imbalance, a weighting strategy based on sample size was adopted during the subsequent model training phase, thereby imposing a higher penalty for incorrect predictions of negative samples (non-fire points). For each valid location, a 512×512 -pixel WorldView Level-18 remote-sensing image was acquired at a spatial resolution of 1.2 m per pixel, corresponding to an approximate ground coverage of 614×614 m. This spatial extent was selected to balance site-specific detail with the broader environmental context.

3.2. Qwen3-plus model and prompt engineering

We utilized Qwen3-plus, a state-of-the-art multimodal large language model (MLLM), as the core semantic reasoning engine. Unlike traditional computer vision models that classify discrete objects, MLLMs align visual features with linguistic concepts in a high-dimensional semantic space. This capability enables the interpretation of complex urban scenes—such as distinguishing dense wooden structures from generic buildings—which is critical for fire risk assessment. Qwen3-plus was selected for its superior performance in visual agent benchmarks (e.g., OSWorld) and its ability to integrate reasoning with perception. Furthermore, it performs exceptionally well in the context of research into urban environments in China, demonstrating a strong ability to interpret elements such as Chinese shop signs and advertisements.

To ensure precise feature extraction, we developed a dual-perspective prompt engineering framework (see Appendix A). This framework clearly defines the research context, environment, and objectives, and requires the MLLM, acting as an expert in urban planning and fire safety, to extract fire hazard features from street-view or remote-sensing imagery, and to produce formatted assessment results and justifications for each image. For street view images, prompts were designed to identify micro-scale fire hazards from a human-centric view; for remote sensing imagery, prompts focused on meso-scale environmental contexts. All results were output strictly in JSON format, thereby ensuring that the scores for each metric are comparable. The relevant pseudocode is provided in Supplementary S1.

To verify the reliability of the MLLM-based evaluation, a rigorous pre-validation procedure was conducted. First, we randomly selected 100 samples from both the positive and negative datasets (totaling 200 samples) and retrieved their corresponding street-view and remote sensing imagery. Second, two professionally trained researchers independently conducted manual scoring on these 200 samples. Any substantial discrepancies in their evaluations were arbitrated by a third expert. The inter-rater reliability between the initial two researchers demonstrated a high level of agreement (Cohen's Kappa = 0.815), satisfying the required quality threshold (Landis & Koch, 1977). Consequently, one set of these corroborated results was established as the human-annotated ground truth for the subsequent evaluation. Finally, the performance of the Qwen3-plus model was evaluated against these human annotations using rigorous statistical metrics

tailored to specific variable types: Quadratic Weighted Kappa (QWK) was employed for binary and ordinal variables—including intrinsically continuous variables (e.g., street width, sky visibility, and green viewing rate) that were discretized into standardized intervals to facilitate the calculation of the human-AI consistency matrix—while normalized confusion matrices were utilized for nominal (one-hot encoded) variables.

The results indicated that approximately 68% of the ordinal variables achieved a consistency score above 0.6, with 40% exceeding 0.8. Furthermore, the overall accuracy for the multi-class one-hot variables reached 63.9%. The results show that MLLM is capable of effectively understanding the content of street-view and remote-sensing imagery and extracting the fire risk indicators relevant to the study. For detailed results, see Appendix B.

3.3. Feature extraction and indicator system construction

This study constructed a comprehensive indicator system comprising 46 explicit features and 7 latent semantic concepts, derived from distinct multimodal data streams. These features are categorized into five analytical dimensions: Functional Conflicts and Industrial Risks, Spatial Density and Form Effects, Deterioration of Facilities and Material Fragility, Human Activity Load, and Rescue Accessibility and Suppression Restrictions (see Fig. 3 for definitions and Fig. 4 for distributions). The 46 explicit features consist of 40 visual features extracted by the MLLM and 6 static macroscopic urban indicators.

3.3.1. MLLM-derived visual features and latent concepts

For the visual environment, the Qwen3-plus model extracted 40 structural features (28 from street view and 12 from remote sensing imagery). To comprehensively capture the complexity of the urban fabric, these features were strictly encoded into four statistical types: binary variables (presence/absence), ordinal variables (graded intensity ratings), nominal variables (unordered multi-class categories, one-hot encoded), and continuous variables (e.g., street width, sky visibility). To ensure methodological transparency, the explicit operational definitions, specific grading criteria, and encoding schemes for all 40 MLLM-derived visual indicators are comprehensively detailed in Supplementary S2.

Beyond the structured visual data, we further leveraged the MLLM's reasoning capabilities by analyzing its descriptive text outputs. By applying text embedding and clustering techniques to the rationale texts generated by the model, we extracted 7 latent risk concept labels. This approach allows the indicator system to capture subtle, qualitative risk contexts that are often missed by traditional rigid quantification.

3.3.2. Integration of static urban indicators

To robustly benchmark the micro-visual features against the macro-urban context, we incorporated 6 standard macroscopic urban environment indicators, aggregated within a 100 m buffer around each sample:

- Built Environment: Floor area ratio (FAR), building density (BD), height, and age were derived from the Comprehensive Multi-Attribute Building (CMAB) dataset (Zhang, Zhao, & Long, 2025).
- Socioeconomic Factors: Housing prices were generated via spatial interpolation of market data (source: soufun.com.cn).
- Functional Mix: We quantified land-use diversity using Shannon Entropy based on ten fire-relevant Point of Interest (POI) categories acquired from the Amap (source: lbs.amap.com) API. A higher entropy value reflects a more complex and functionally mixed urban environment.

3.4. GW-XGBoost and spatial SHAP explanation

Standard global models rely on the assumption of spatial



Fig. 3. Urban fire risk indicator system.

stationarity, ignoring that risk predictors vary significantly across diverse urban fabrics (spatial heterogeneity). For instance, building density may be a critical hazard in informal settlements (urban villages), but acts differently in planned high-rise districts.

To capture this spatial non-stationarity, we employed Geographically Weighted XGBoost (GW-XGBoost), a method proven effective in similar spatial contexts (Grekousis, 2025). By integrating geographical weighting theory into an ensemble learning framework, GW-XGBoost constructs locally weighted sub-models for each spatial unit. This approach enables the precise fitting of nonlinear relationships between multimodal semantic features and fire risk, adapting dynamically to the local urban context.

XGBoost is an efficient algorithm based on gradient-boosted decision trees, excelling in non-linear fitting and capturing feature interactions

(Chen, 2016). To transition the model from a global to a local perspective, we introduce geographical coordinates (u_i, v_i) . For each target prediction point j within the study area, a locally weighted objective function is constructed. The core of GW-XGBoost lies in assigning spatial weights w_{ij} to training samples, determined by the Euclidean distance d_{ij} between the sample i and the target point j . The modified local objective function $\mathcal{L}_j^{(t)}$ is expressed as Formula (1):

$$\mathcal{L}_j^{(t)} = \sum_{i=1}^n w_{ij} \cdot \left[l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) \right] + \Omega(f_t) \quad (1)$$

Within this framework, the contribution (weight) of sample point i to model training increases with its proximity to target point j . Furthermore, the loss function l incorporates an inverse class frequency

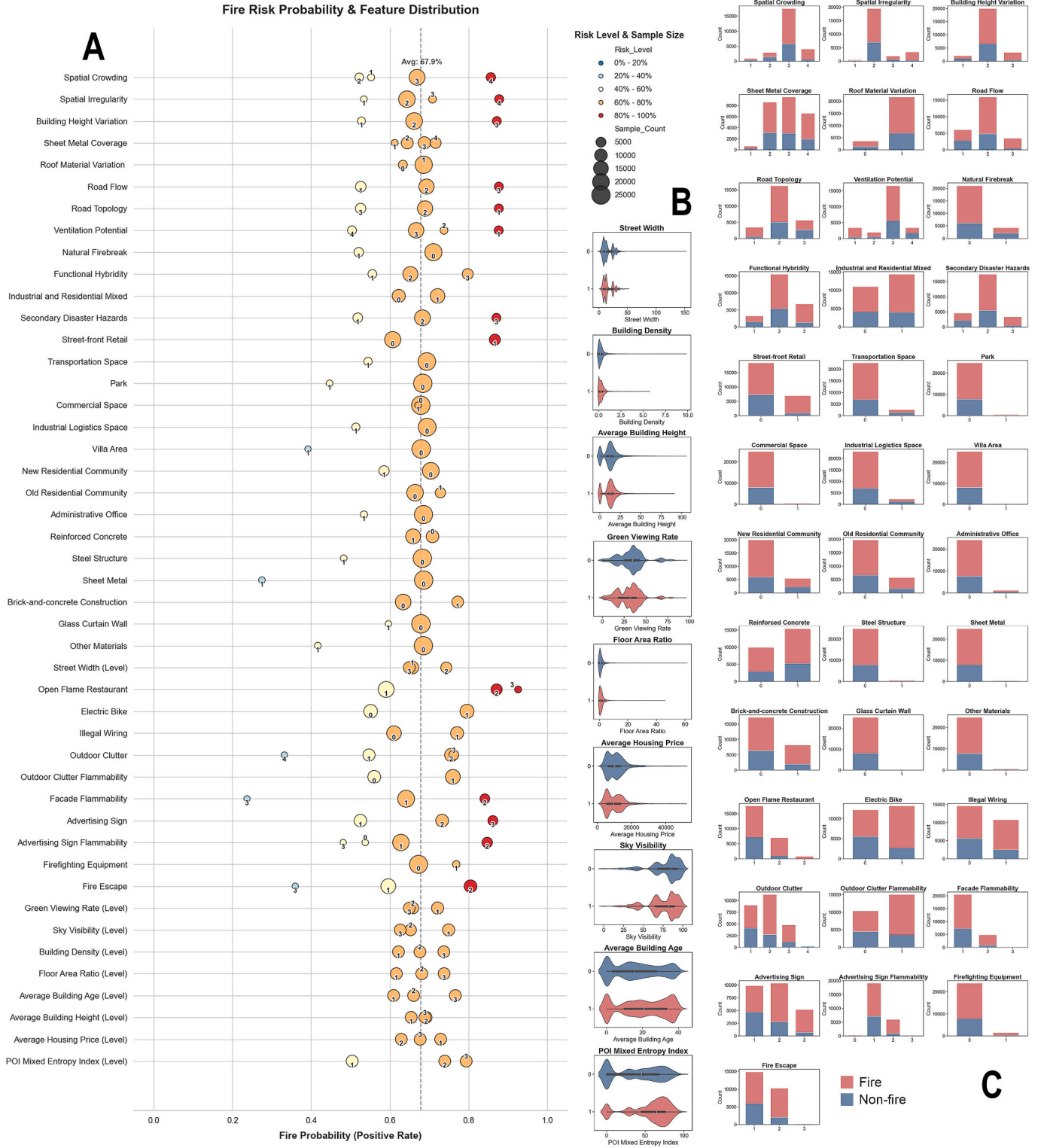


Fig. 4. Distributions of urban fire-risk features: (A) feature-level fire probability and sample size; (B) distributions of selected continuous features for fire and non-fire samples; and (C) class composition of categorical and ordinal features.

weighting method to balance the number of positive and negative samples. The spatial weight w_{ij} is computed via the Adaptive Gaussian Kernel Formula (2):

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2b^2}\right) \quad (2)$$

Among these, b denotes the bandwidth, a key hyperparameter governing the degree of spatial smoothing. The optimal bandwidth b_{opt} in this study was determined using Optuna, an automated hyperparameter optimisation software framework specifically designed for machine learning.

We introduce Shapley Additive Explanations (SHAP) values to deconstruct the black-box mechanism of each local model. SHAP is an

interpretability method for understanding the contribution of each feature to the prediction of a machine learning model. The method is inspired by cooperative game theory and treats each feature as a contributor to the model for additive interpretation (Lundberg & Lee, 2017). This enables us to generate spatial feature contribution maps, visualizing the variation in marginal effects of specific risk factors across different areas of Wuhan. For the local model at location j , the contribution ϕ_j^k of feature k is calculated using Formula (3):

$$\hat{y}_j = \phi_j^0 + \sum_{k=1}^M \phi_j^k \quad (3)$$

3.5. Semantic vector space-based risk scenario clustering

To extract stable fire risk patterns from the high-dimensional, unstructured risk description texts obtained via MLLM, this study constructs a semantic clustering framework based on BERTopic. The process primarily comprises the following three components.

Step 1: The text corpus processed using the Bigram model (Simons et al., 1997) is mapped to a high-dimensional semantic space. Let $D = \{d_1, d_2, \dots, d_N\}$ denote a set of N fire risk descriptions. Each document d_i is mapped to a fixed-dimensional vector $v_i \in \mathbb{R}^m$ via a pre-trained model (BAAI/bge-large-zh-v1.5). To mitigate sparsity in the high-dimensional space while preserving local topological structure, the UMAP algorithm (Draganov et al., 2023) projects vectors onto a low-dimensional space ($d = 8$). UMAP optimizes layout by minimizing cross-entropy between the fuzzy simplicial set in the high-dimensional space and the low-dimensional representation. Its objective function C is defined as Formula (4):

$$C = \sum_{i \neq j} \left(p_{ij} \log \frac{p_{ij}}{q_{ij}} + (1 - p_{ij}) \log \frac{1 - p_{ij}}{1 - q_{ij}} \right) \quad (4)$$

where p_{ij} denotes the local connection probability between samples i and j in the high-dimensional space, while q_{ij} represents the corresponding probability in the low-dimensional embedding space. This process ensures that the semantic proximity relationships in the risk description are preserved following dimensionality reduction.

Step 2: Within a low-dimensional semantic space, this study employs the HDBSCAN algorithm to identify risk clusters. To address noise in urban environmental data, HDBSCAN incorporates the mutual reachability distance d_{mreach} (Neto et al., 2017) to enhance clustering robustness, as shown in Formula (5):

$$d_{mreach}(a, b) = \max\{\text{core}_k(a), \text{core}_k(b), d(a, b)\} \quad (5)$$

where $\text{core}_k(x)$ denotes the k -nearest neighbour distance at point x , while $d(a, b)$ represents the Euclidean distance. This metric effectively separates dense semantic cores from sparse noise points, ultimately generating initial micro-topics.

Step 3: To extract representative fire risk keywords (such as colour-coated steel tiles and flying wires) from the clustering results, we employed the class-based TF-IDF algorithm (Imbe et al., 2021). This approach treats each cluster as a single document, calculating the importance score $W_{t,c}$ for term t within cluster c as Formula (6):

$$W_{t,c} = tf_{t,c} \times \log \left(1 + \frac{A}{f_t} \right) \quad (6)$$

where $tf_{t,c}$ denotes the frequency of term t within cluster c , f_t represents the total frequency of that term across all clusters, and A is the average number of terms across all clusters.

Based on the $W_{t,c}$ values, we conducted a hierarchical clustering of the initial themes, ultimately converging upon seven distinct categories of urban fire risk scenarios exhibiting significant semantic differences.

4. Results

4.1. GW-XGBoost model performance

4.1.1. Comparative performance analysis

To validate the predictive capability of the proposed GW-XGBoost, we benchmarked it against five standard algorithms: global XGBoost, random forest (RF), support vector machine (SVM), multilayer perceptron (MLP), and logistic regression (LR). Performance was evaluated using AUC, Accuracy, F1-score, Precision, and Recall.

Fig. 5A presents the ROC curves for all models. GW-XGBoost demonstrated superior generalization, achieving the highest AUC of 0.868. This performance significantly outperforms the global XGBoost baseline (AUC = 0.842) and traditional classifiers like SVM (0.774) and MLP (0.751), highlighting the advantage of modeling spatial non-stationarity over assuming a homogeneous risk landscape. Fig. 5B details the comparative metrics. The proposed model consistently led in Accuracy (>0.77) and Precision (>0.80), indicating high reliability in minimizing false positives. Crucially, it achieved the highest F1-Score (>0.83). As the harmonic mean of precision and recall, a high F1-score confirms that GW-XGBoost strikes an optimal balance between sensitivity and reliability—a critical trait for risk assessment. While SVM and RF exhibited marginally higher Recall, their compromised Precision and lower AUC suggest a tendency toward over-prediction (high false-positive rates). In contrast, GW-XGBoost offers a robust and operationally viable framework for identifying high-risk urban scenarios.

4.1.2. Ablation study

To rigorously quantify the contribution of multimodal features, we designed a systematic ablation study comprising five configurations (M1–M5, see Fig. 6).

The Proxy Effect: A pivotal finding is that the pure visual model (M4, AUC = 0.808) achieved performance statistically equivalent to the traditional static model (M1, AUC = 0.802). This parity is highly significant from a practical standpoint. Traditional risk models (M1) heavily rely on granular socioeconomic data (e.g., precise housing age, internal building metrics) that are often costly, confidential, or unavailable in many developing cities. The fact that M4 rivals M1 demonstrates that semantic features extracted from ubiquitous open-source imagery can effectively “encode” these latent variables. Consequently, our proposed visual framework serves as a robust, low-cost substitute for data-heavy traditional methods, enabling high-fidelity risk assessment even in information-poor environments.

Complementarity in Full Fusion: While M4 proves the substitutability of visual data, the full fusion model (M5) achieved the state-of-the-art (SOTA) performance (AUC = 0.834, F1 = 0.841), surpassing all sub-models. This confirms that while visual and static data share some information, they also capture orthogonal dimensions of risk: static data defines the rigid structural baseline, while visual data captures dynamic, transient hazards. The integration of both is necessary for maximum precision, but M4 stands alone as a viable solution for scalability.

4.2. Predictors of urban fire risk: global and local interpretations

4.2.1. Global feature importance ranking

Fig. 7A illustrates the global importance and directional impact of features in the fusion model (M5), while Fig. 7B visualizes the cumulative contribution of these factors across individual samples.

The POI Mixed Entropy Index emerged as the paramount predictive indicator, confirming that functionally heterogeneous zones—characterized by the chaotic interweaving of residential and commercial activities—present severe fire management challenges. Alongside Average Building Age (#2) and Average Housing Price (#3), these metrics form the socioeconomic bedrock of vulnerability, demonstrating that older, economically disadvantaged areas

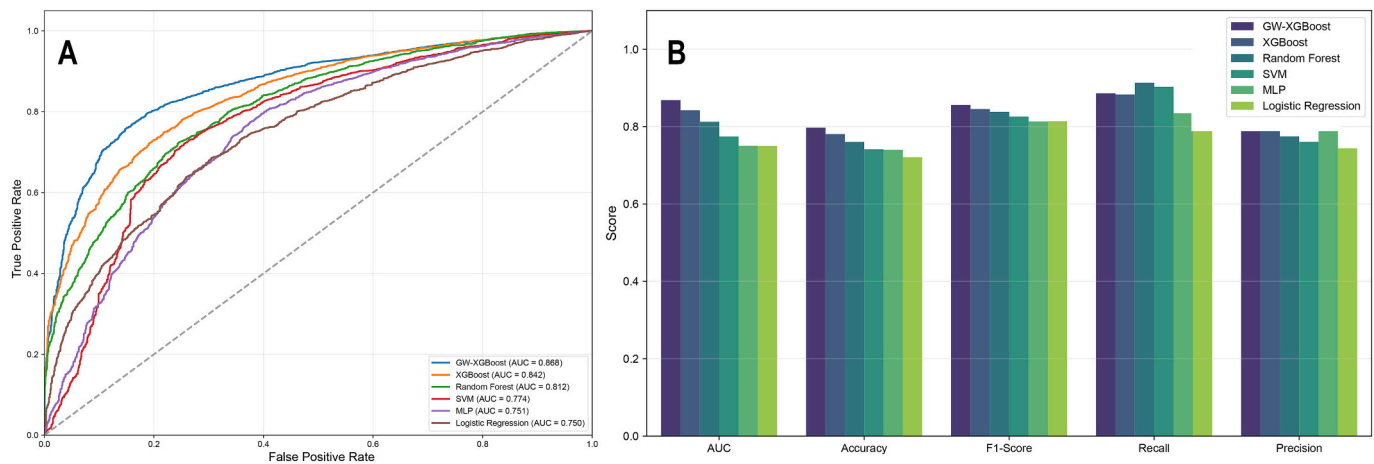


Fig. 5. Predictive performance of GW-XGBoost and the baseline models: (A) receiver operating characteristic (ROC) curves; and (B) comparison of AUC, accuracy, F1-score, recall, and precision.

structurally amplify hazard susceptibility.

MLLM-derived visual semantic features successfully penetrated the top tier of importance, rivaling traditional static metrics. For instance, Open Flame Restaurant (#4) surpassed classic indicators like FAR, acting as a potent visual proxy for elevated ignition probabilities. Furthermore, features such as Advertising Sign (#6), Electric Bike (#11), and Illegal Wiring (#15) effectively pinpointed specific physical hazards and spatial informality. The model perceptively links this visual disorder directly to high fuel loads, evacuation obstructions, and infrastructure decay in high-density neighborhoods.

Interestingly, Sheet Metal Coverage (#14) exhibited a counter-intuitive negative correlation with fire risk. Rather than merely reflecting material fragility, its prevalent high-density occurrence typically signifies suburban logistics or industrial zones where human activity intensity—and thus ignition frequency—is significantly lower than in city centers. This highlights the model's capacity to disentangle actual ignition probability from mere material vulnerability based on the underlying spatial logic of urban functions.

Fig. 7B illustrates the cumulative risk composition, revealing a “synergistic lock-in” effect in high-risk areas. Extreme vulnerability is rarely driven by a single factor, but rather by the simultaneous overlay of high functional mixing, aging infrastructure, and dense visual hazards. Crucially, the heatmap identifies a tipping point: even when risks are theoretically suppressed by favorable static conditions (e.g., moderate street widths or higher housing prices), the overwhelming accumulation of localized visual hazards (e.g., dense open-flame restaurants and clutter) can still propel the predicted risk to critical levels.

4.2.2. Non-linear thresholds and interactive effects

To deconstruct the complex predictive patterns, Fig. 8 identifies critical tipping points and coupling effects between static socioeconomic baselines and dynamic visual cues. Static attributes define the structural boundaries of risk through distinct non-linear thresholds. For instance, Average Housing Price exhibits a critical safety threshold around 15,000 RMB/m², below which communities are hypersensitive to hazards due to chronic management deficits. Similarly, Building Age peaks in risk at 20–25 years, reflecting structures pre-dating modern fire codes. Crucially, their interaction reveals a compounded vulnerability: physical obsolescence acts as a severe risk multiplier primarily when paired with economic disadvantage, confirming that low-income, aging neighborhoods form the core of structural fire risk.

Against this static backdrop, MLLM-derived visual features function as active risk amplifiers that trigger predicted risk surges upon surpassing specific density thresholds. The agglomeration of Open Flame Restaurant introduces non-linear risk spikes, suggesting that dining

density easily overwhelms local safety capacities. Furthermore, visual proxies for spatial disorder—such as Advertising Sign—interact with Building Age to maximize risk on older structures, highlighting how external fuel loads exacerbate inherent structural fragility. Similarly, the mere visual detection of Electric Bike clusters acts as an immediate risk catalyst. This hazard is significantly more acute in lower-income zones, pointing to a critical governance gap regarding informal charging in affordable housing sectors.

Finally, the synergistic analysis uncovers a “density paradox” governed by physical environmental conditions. The coupling of high building density and poor ventilation drastically elevates risk, physically aligning with the urban canyon effect of heat accumulation and smoke entrapment. Conversely, high floor area ratio sites often display lower risk scores, as they typically represent newer, highly regulated vertical developments. This counter-intuitive finding demonstrates that vertical density in planned, modern communities is inherently safer than the horizontal sprawl of unregulated settlements, establishing that the quality and regulation of urban form are far stronger predictors of fire resilience than mere density.

4.2.3. Spatial distribution of dominant predictors

Although the Holistic Model (M5) achieves the state-of-the-art predictive performance as shown, the dominant influence of static socioeconomic features (e.g., housing price, building age) tends to mask the granular contributions of visual perception attributes.

To decouple granular visual risk indicators from the overshadowing predictive influence of dominant socioeconomic variables, we mapped the spatial distribution of primary risk predictors for both the holistic model (M5) and the visual-only model (M4). A comparative analysis between Fig. 9A and Fig. 9C reveals a distinct divergence in spatial logic. The M5 results exhibit a pattern of structural homogeneity, where the POI Mixed Entropy Index exerts monolithic dominance, covering 51.6% of the grid cells. This confirms that functional mixing serves as the overarching macro-predictor, defining the potential risk zones at a strategic level. In stark contrast, the M4 model exposes high spatial heterogeneity, successfully revealing the distinct distribution patterns of visual proxies that are otherwise masked by the powerful signal of static attributes in the fusion model.

The M4 analysis further uncovers three distinct spatial morphologies of micro-scale risk hidden within the urban fabric. First, Advertising Sign (the top predictive feature, 29.2%) exhibits a linear extension pattern, stretching along major arterial corridors and secondary streets rather than being confined to commercial nodes. This highlights visual clutter as a systemic facade risk along the urban movement network. Second, Open Flame Restaurant (17.9%) forms dense, discrete, clustered

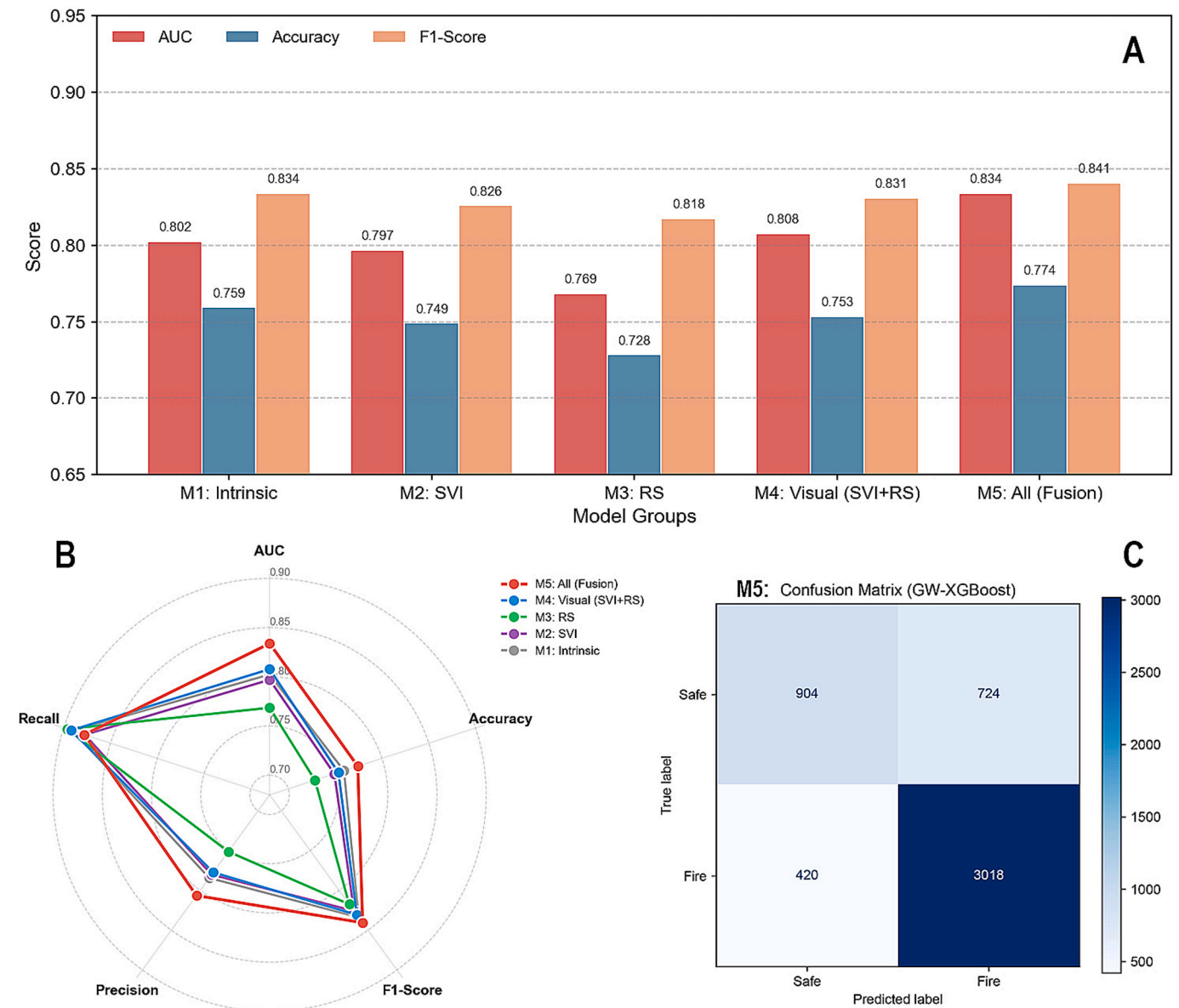


Fig. 6. Ablation analysis of five feature configurations: (A) comparison of AUC, accuracy, and F1-score; (B) radar-chart comparison of AUC, accuracy, F1-score, precision, and recall; and (C) confusion matrix for the full-fusion Model M5.

hotspots within high-density residential interiors and backstreets, spatially pinpointing proxies for active ignition hazards embedded within the living environment. Finally, as shown in the results, Electric Bike (~13%) displays a diffuse, stochastic spatial pattern. Unlike fixed physical facilities, their diffused presence (indicated by green patches) reflects the dynamic encroachment of informal behaviors—often associated with illicit charging and evacuation blockage—that transcends specific functional zoning and permeates the transitional spaces of older neighborhoods.

4.2.4. Visualization of local effects of key features

This section focuses on the spatial interpretability of the Visual-Only Model (M4) to uncover independent physical risk associations. As shown in Fig. 10, a critical divergence emerges between risk intensity and spatial extent among the dominant predictors. Advertising Sign, despite being the second most intense feature on average, dominates the spatial landscape in terms of coverage. High-risk values form a continuous, network-dependent structure along arterial roads and commercial corridors. This linear distribution confirms that visual clutter is a systemic

hazard: in high-density commercial zones, excessive signage indicates a potential widespread “vertical fire load” that spans the entire network.

In contrast, Open Flame Restaurant exhibits a distinct morphology of clustered intensity. While its spatial coverage is less extensive than billboards, their contribution to risk is highly concentrated in specific mixed-use areas. The dense red clusters reveal that in these localized zones—typically older residential-commercial interfaces—dining venues serve as strong visual proxies for active ignition hazards. This comparison clarifies the statistical nuance: while billboards represent a ubiquitous, chronic condition of the urban facade, open-flame dining represents an acute, site-specific threat driven by functional agglomeration.

Beyond these primary predictors, secondary features reveal distinct micro-spatial patterns. Electric Bike displays a stochastic permeability, with risk points scattered deeply into residential interiors rather than being confined to main streets. This reflects the diffuse encroachment of informal charging behaviors into living spaces. Furthermore, environmental proxies effectively map structural vulnerability: Sky Visibility shows low values that precisely outline the high-density urban core,

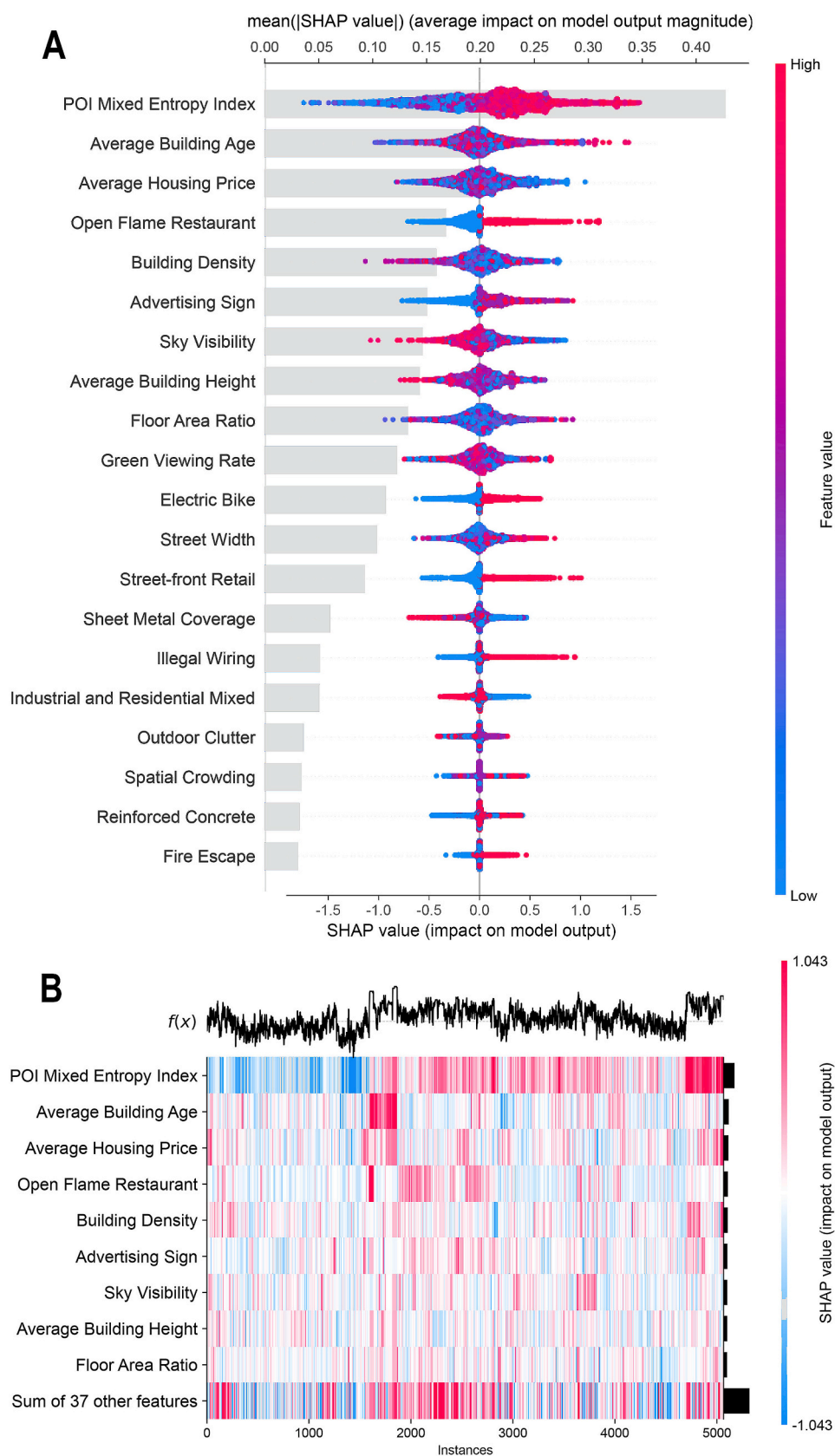


Fig. 7. Global SHAP interpretation for Model M5: (A) mean absolute SHAP importance ranking and beeswarm distribution of the 20 highest-ranking features; and (B) sample-level SHAP heatmap showing cumulative feature contributions.

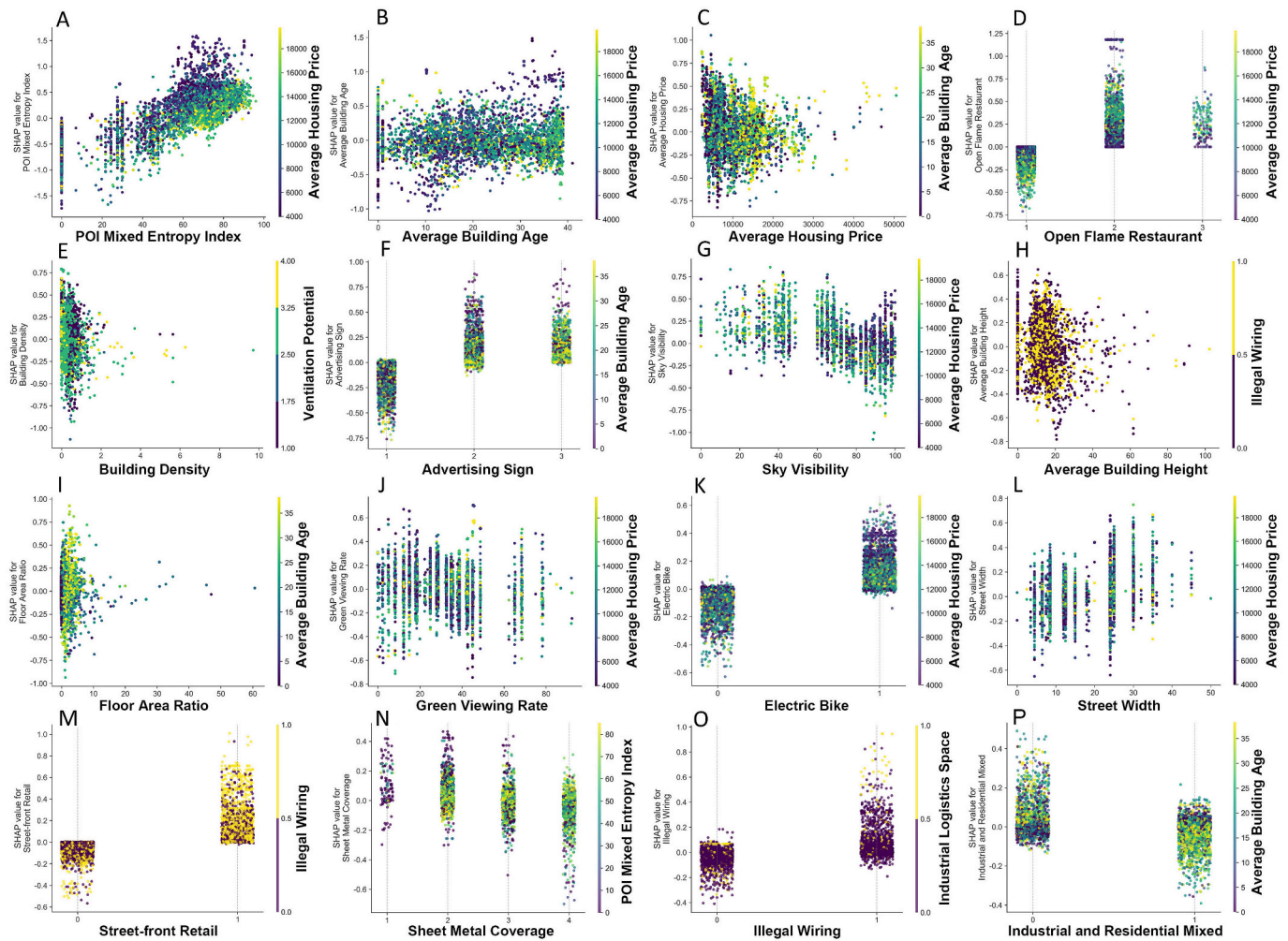


Fig. 8. SHAP dependence and interaction plots for the 16 highest-ranking features in Model M5 (A–P); point colors indicate the interacting feature represented by the corresponding color bar.

serving as a proxy for urban canyon effects and insufficient fire control separation. Conversely, Sheet Metal Coverage shows high predictive importance in the urban fringe, identifying temporary structures with low fire resistance. Collectively, these visual proxies reconstruct a multi-dimensional risk landscape, distinguishing between the extensive spread of facade hazards, the intensive concentration of ignition sources, and the structural fragility of the built environment.

4.3. Semantic clustering and pattern decoupling via MLLM

4.3.1. Identification of semantic risk archetypes

The urban fabric comprises distinct semantic clusters, each characterized by unique risk profiles. While the GW-XGBoost model effectively addresses spatial non-stationarity by adapting to local geographical variances, it fails to capture semantic non-stationarity—a phenomenon where identical visual cues imply completely different risks depending on their functional contexts.

Beyond structured feature extraction, the MLLM generated descriptive reasoning for each sample, offering a rich “semantic layer” of risk interpretation. To decode the underlying patterns of fire risk in micro-scenarios, we synthesized the textual outputs from both street view and remote sensing imagery into a unified corpus for analysis. Utilizing the BERTopic framework, we initially identified 192 micro-topics, which were subsequently aggregated via hierarchical clustering into seven distinct fire risk archetypes (Fig. 11). As detailed in Appendix C, each archetype represents a stable, recurring urban scenario defined by

unique spatial functions and hazard profiles, providing a qualitative taxonomy for the subsequent quantitative analysis.

4.3.2. Contextual dependency of risk predictors across sub-models

To decode the spatially stratified risk mechanisms, we trained independent GW-XGBoost sub-models for each semantic archetype. The comparative analysis of performance metrics (Fig. 12A and Fig. 12C) and feature importance heatmaps (Fig. 12B and Fig. 12D) reveals a profound contextual dependency of fire risk predictors. Regarding predictive performance, a distinct “U-shaped” reliability pattern emerges. Cluster 1 (Low-Density Open Spaces) achieves high AUC scores due to its simple spatial structure and lack of visual noise. Conversely, Cluster 7 (Typical Urban Villages) achieves the highest F1-scores (>0.90 in M5). This suggests that while urban villages are chaotic, their visual signatures—characterized by extreme density and occlusion—are highly distinctive and consistent, allowing the model to capture their risk patterns with high fidelity. Intermediate clusters (e.g., Cluster 2, Cluster 3), representing transitional urban forms, show slightly higher variance, reflecting the ambiguity of their risk profiles.

The feature importance heatmaps further validate that different urban scenes are characterized by distinct dominant threat indicators. In high-density living environments (Cluster 5 & Cluster 7), the M4 heatmap (Fig. 12D) identifies Open Flame Restaurant (SHAP >0.20) and Sky Visibility as the paramount visual predictors. This indicates that in these spatially constrained zones, the primary threat is strongly associated with the embedding of active ignition sources within urban canyons,

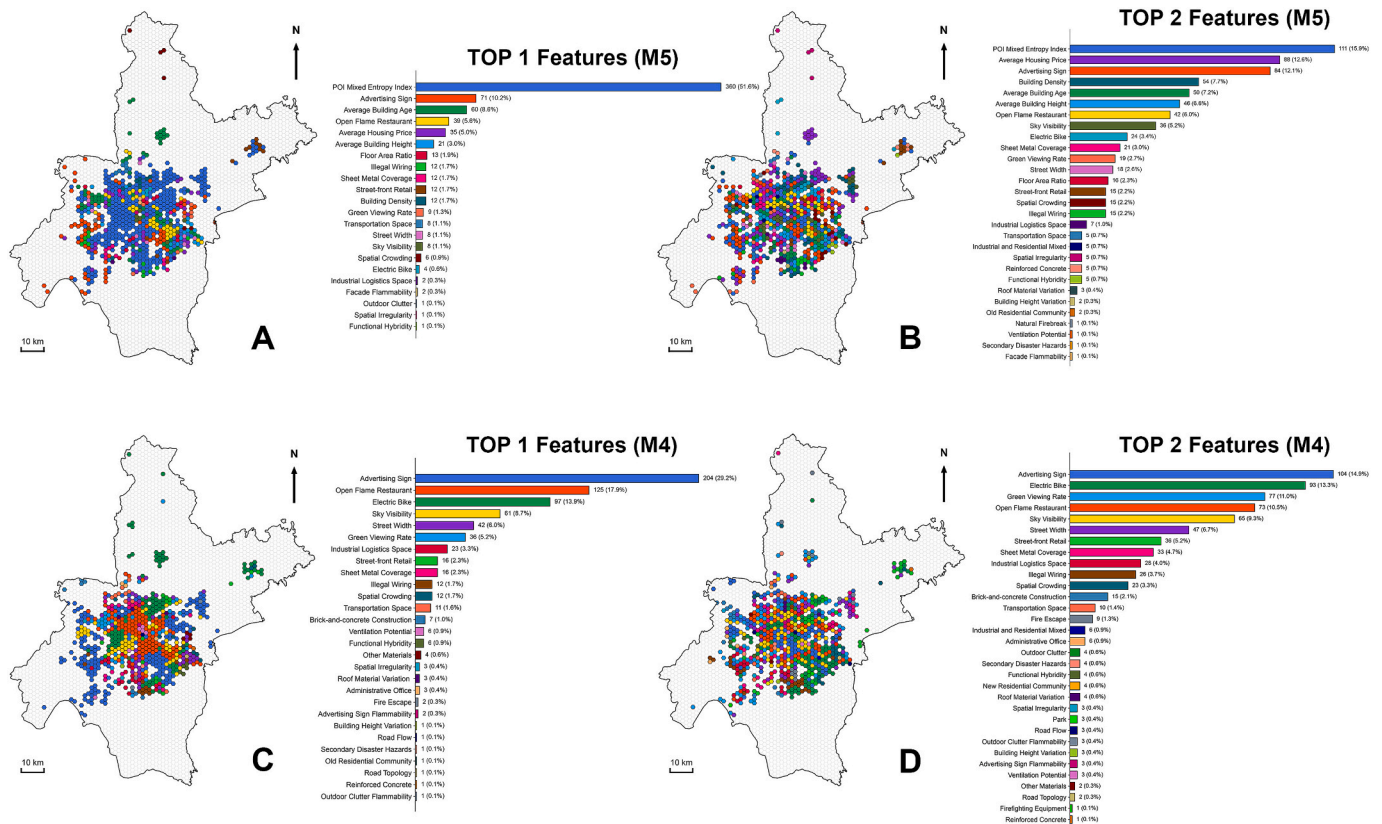


Fig. 9. Spatial distributions and proportional composition of the first- and second-ranked dominant predictors: (A) first-ranked predictors in Model M5; (B) second-ranked predictors in Model M5; (C) first-ranked predictors in Model M4; and (D) second-ranked predictors in Model M4.

where ventilation is poor and evacuation is difficult. While Electric Bike and Illegal Wiring act as secondary aggravating factors, the overwhelming risk is structurally linked to density and functional misuse. In contrast, for Cluster 6 (Industrial-Residential Interface), the M5 heatmap (Fig. 12B) shows that POI Mixed Entropy exerts a monolithic influence (SHAP = 0.20), confirming that the disorganized juxtaposition of industrial operations and residential life is a fundamental indicator of fire risk. Meanwhile, Cluster 2 is uniquely predicted by Advertising Sign and construction-related cues, pointing to temporary hazards during urban renewal.

To visualize the spatial allocation of risk archetypes and their underlying predictive indicators across different administrative scales, we constructed a Sankey diagram (Fig. 13). This diagram maps the logical flow from Wuhan's 13 administrative districts (Left) to the seven identified semantic risk archetypes (Middle), and finally to the five aggregated predictor categories derived from the 46 features (Right).

The flow analysis reveals a distinct core-periphery divergence in risk composition. For central urban districts, risk flows predominantly converge into Cluster 4 and Cluster 5. This indicates that the core built-up environment is primarily threatened by compound risks in high-density spaces, where structural fire hazards are critically exacerbated by secondary risks, such as obstructed evacuation routes and spatial overcrowding. Conversely, distal districts show a strong tendency toward Cluster 2 and Cluster 7. This pattern suggests that vulnerability in these transitional zones is strongly associated with the structural friction between new urban expansion and legacy renovations. Specifically, the chaotic interface between modern construction sites and informal settlements during the rapid urban renewal process correlates with an elevated probability of uncontrolled fire propagation.

4.4. Scenario simulation: from prediction to risk regulation

To translate predictive capability into actionable governance strategies, we established a scenario-based sensitivity analysis framework augmented by Monte Carlo simulations. We simulated risk reduction effects under three governance scenarios (30%, 60%, and 90% reduction) for three modifiable features that are operationally feasible: Open Flame Restaurant, Electric Bike, and Illegal Wiring. Independent parallel experiments ($N = 10$) were conducted to ensure statistical robustness (Fig. 14). The global simulation results reveal that Electric Bike exhibits the highest risk sensitivity; at 90% remediation coverage, the aggregate city-wide predicted risk decreases by 4.87%, identifying informal charging as the primary active visual proxy for intervention. In comparison, regulating Open Flame Restaurant and Illegal Wiring yielded moderate (−4.03%) and lower (−1.72%) marginal gains, respectively. Notably, the simulated risk reduction across all features follows a strictly linear trajectory, indicating that current governance has not yet reached the threshold of diminishing returns—higher investment corresponds to proportional simulated safety dividends. The minimal error bars further confirm the statistical stability of these simulated macro-level benefits.

However, global averages often mask the spatial non-stationarity of the model's sensitivity to these proxies. To resolve the “where to govern” question, we mapped the risk reduction potential surface based on the 90% intervention scenario (Fig. 15). A striking spatial divergence emerges between risk types. The high-yield zones for Open Flame Restaurant manifest a “polycentric insular” structure, strictly anchored to core commercial nodes (e.g., Jiangnan Road, Chu River and Han Street). This suggests that dining-related risks are not diffuse but are “vertically locked” within high-density commercial-residential hybrids, necessitating a precision-strike enforcement strategy focused on these discrete islands. Conversely, the risk-reduction potential of electric-bike interventions forms a “contiguous annular belt” encircling the city

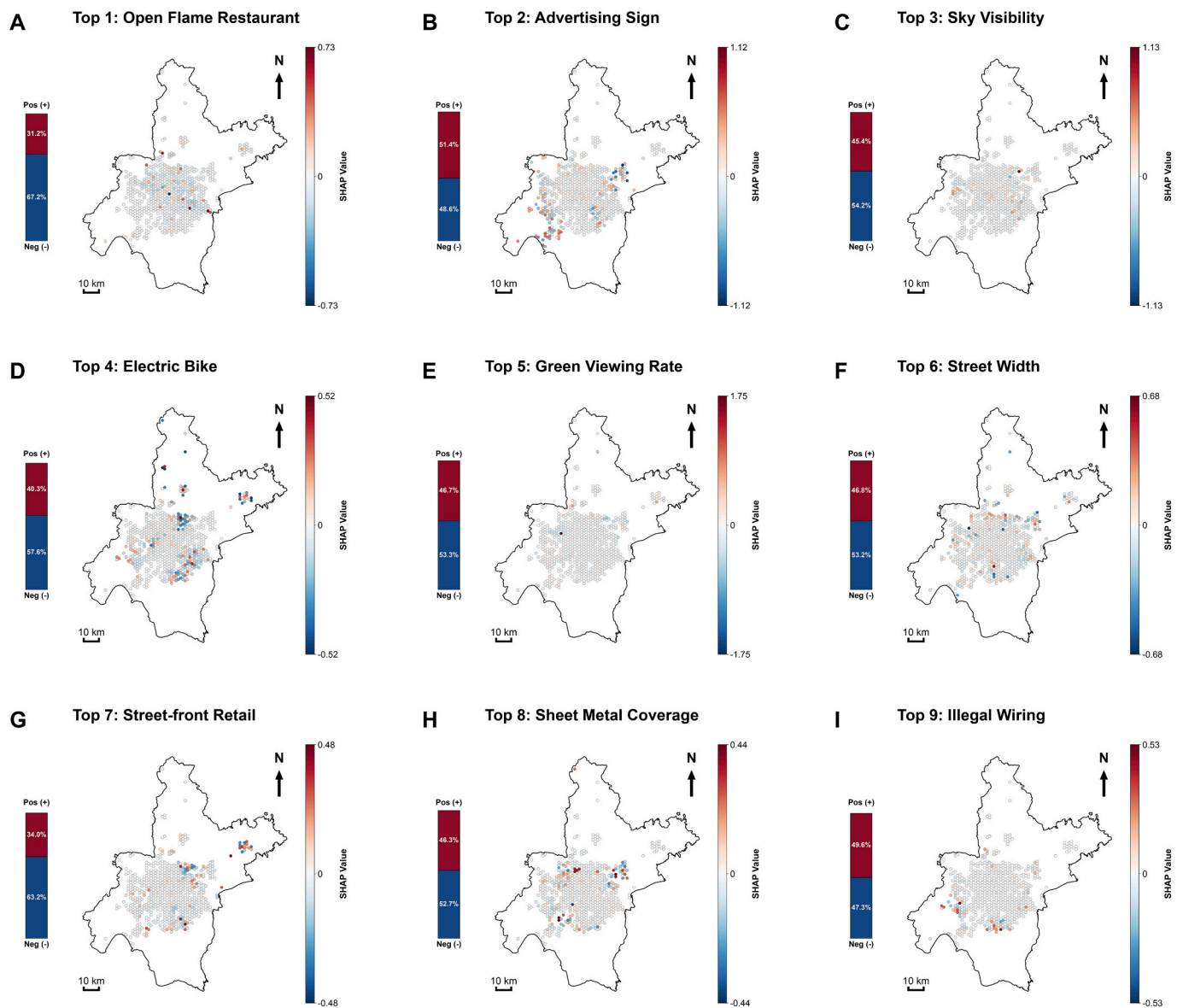


Fig. 10. Spatial distributions of local SHAP values for the nine highest-ranking visual features in Model M4: (A) Open Flame Restaurant; (B) Advertising Sign; (C) Sky Visibility; (D) Electric Bike; (E) Green Viewing Rate; (F) Street Width; (G) Street-front Retail; (H) Sheet Metal Coverage; and (I) Illegal Wiring.

center, specifically covering the transitional zones (2nd to 3rd Ring Road) densely populated by urban villages and aging communities. This pattern reflects the structural mismatch between housing and jobs (“last-mile” commuting) in high-density cities, positioning this transitional belt as the “main battlefield” where infrastructure upgrades will yield the highest marginal safety returns.

Finally, the potential map for Illegal Wiring displays a reticulated and fragmented texture, infiltrating the capillary streets of historic districts and the aging centers of distal districts like Xinzhou and Jiangxia. Although the global statistical gain is lowest, the localized importance of preserving these heritage sites is irreplaceable. Synthesizing these findings confirms a phenomenon of spatial decoupling in urban fire risk governance: different visual proxies exhibit varying predictive sensitivities across the urban fabric. This scientific evidence advocates for a differentiated spatial intervention strategy: prioritizing operational regulation (restaurants) in commercial cores, infrastructure retrofitting (e-bikes) in transitional residential belts, and micro-renewal (wiring) in historic preservation zones.

5. Discussion

5.1. Advancing urban fire risk assessment through multi-modal integration

The superior performance of the full-fusion model (M5) over single-modal baselines underscores a fundamental paradigm shift in urban risk assessment: the evolution from reliance on static “statistical proxies” to dynamic “semantic perception.” Traditional urban analytics have long depended on structured socioeconomic metrics (e.g., POI density, building age) to approximate risk states. While these indicators effectively establish a macro-baseline, they function inherently as a coarse filter, limited by a profound semantic gap—they quantify the intensity of urban functions but fail to capture the quality and condition of the micro-environment. For instance, a well-managed high-rise complex and a congested Urban Village may exhibit identical density profiles yet possess diametrically opposed risk profiles. Relying solely on such numerical proxies inevitably leads to false positives in dense but safe districts, or false negatives in sparse but hazardous informal settlements. This lack of granularity explains the bottleneck of the static-only model

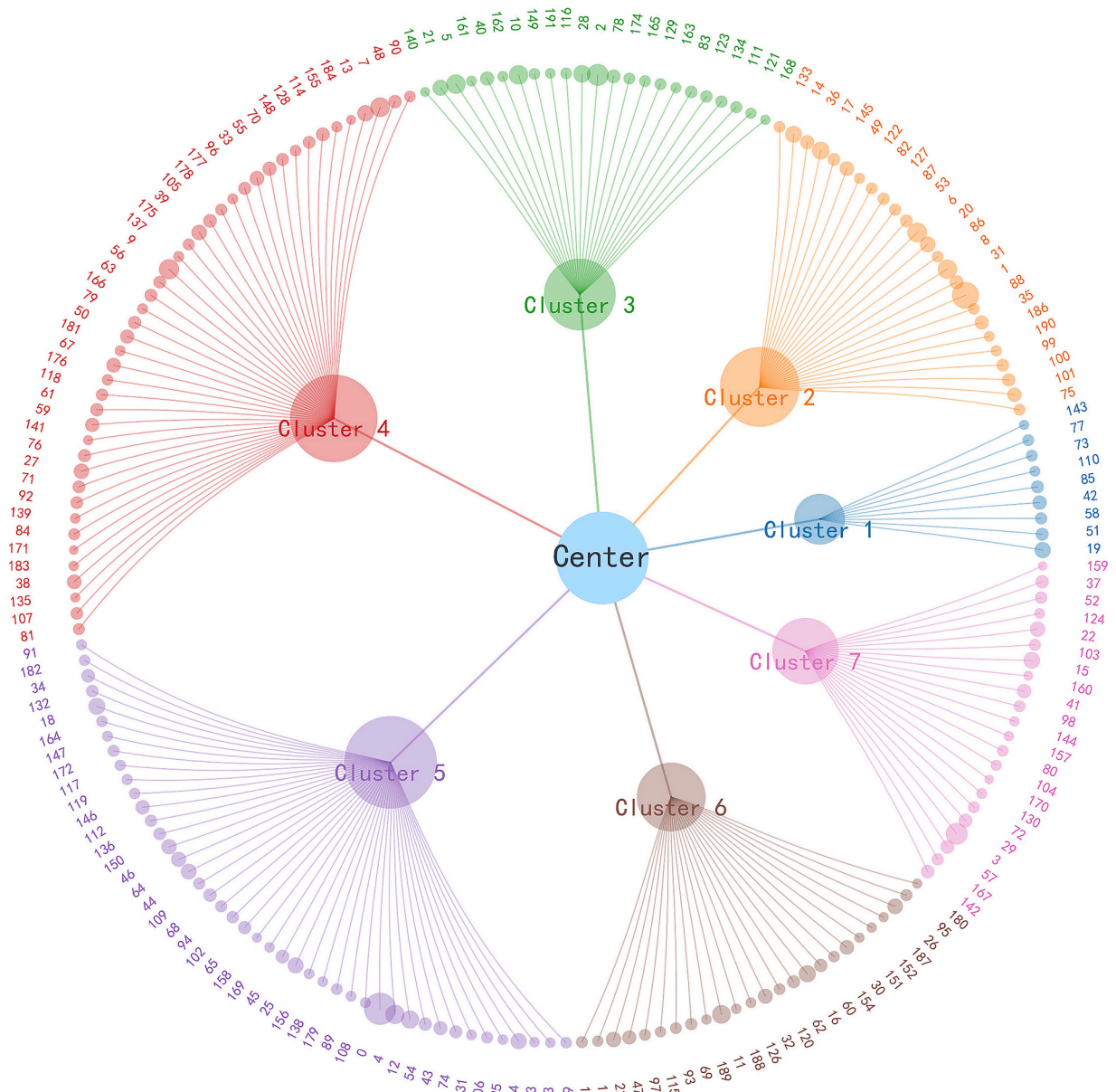


Fig. 11. The relationship between the seven identified clusters and their constituent micro-topics.

(M1), which lacks the cognitive resolution to distinguish macro-level conditions from micro-scale hazards.

The integration of MLLM-derived visual semantics breaks this bottleneck by decoupling specific physical hazards from general urban activities. Unlike indirect statistical inferences, visual features—such as Open Flame Restaurant, Illegal Wiring, and Outdoor Clutter—serve as direct semantic indicators of potential hazards. This multimodal integration empowers the model to identify latent hazards that structured data overlooks; for example, while POI data merely records the existence of a restaurant, the visual semantic layer acutely distinguishes between a benign modern kitchen and a hazardous street-front operation. The high SHAP rankings of e-bikes and billboards further validate that the model is learning the specific manifestations of spatial disorder and combustible accumulation. By fusing structured priors with unstructured visual reasoning, the proposed framework achieves a holistic risk profile, precisely locking onto areas that are not only structurally vulnerable but also functionally compromised.

5.2. Deciphering the spatial non-stationarity of fire predictors

The spatially explicit results from GW-XGBoost powerfully refute the “spatial stationarity” assumption implicit in traditional risk assessments. The potential mechanisms and transmission pathways of risk are not uniformly distributed but exhibit profound heterogeneity, which we conceptualize as the safety dimension projection of a “function-form mismatch” during urban evolution. To give an example, in the historic urban core (e.g., Hankou and Wuchang), risk is structurally associated with functional overload. Although these historic districts in the city centre now enjoy excellent commercial accessibility, the character of their built environment was established decades ago, and in some cases nearly a century ago. This temporal lag forces modern, high-intensity commercial activities (retail, dining) into obsolete, narrow physical spaces, creating a hazardous “vertical mix”—high-fire-load commercial uses on the ground floor with aging residential units above. Large-scale advertising signage, when situated in close proximity to key fire hazards such as restaurants with open flames, is bound to create more complex fire risks. The Sankey diagram further corroborates this, showing a

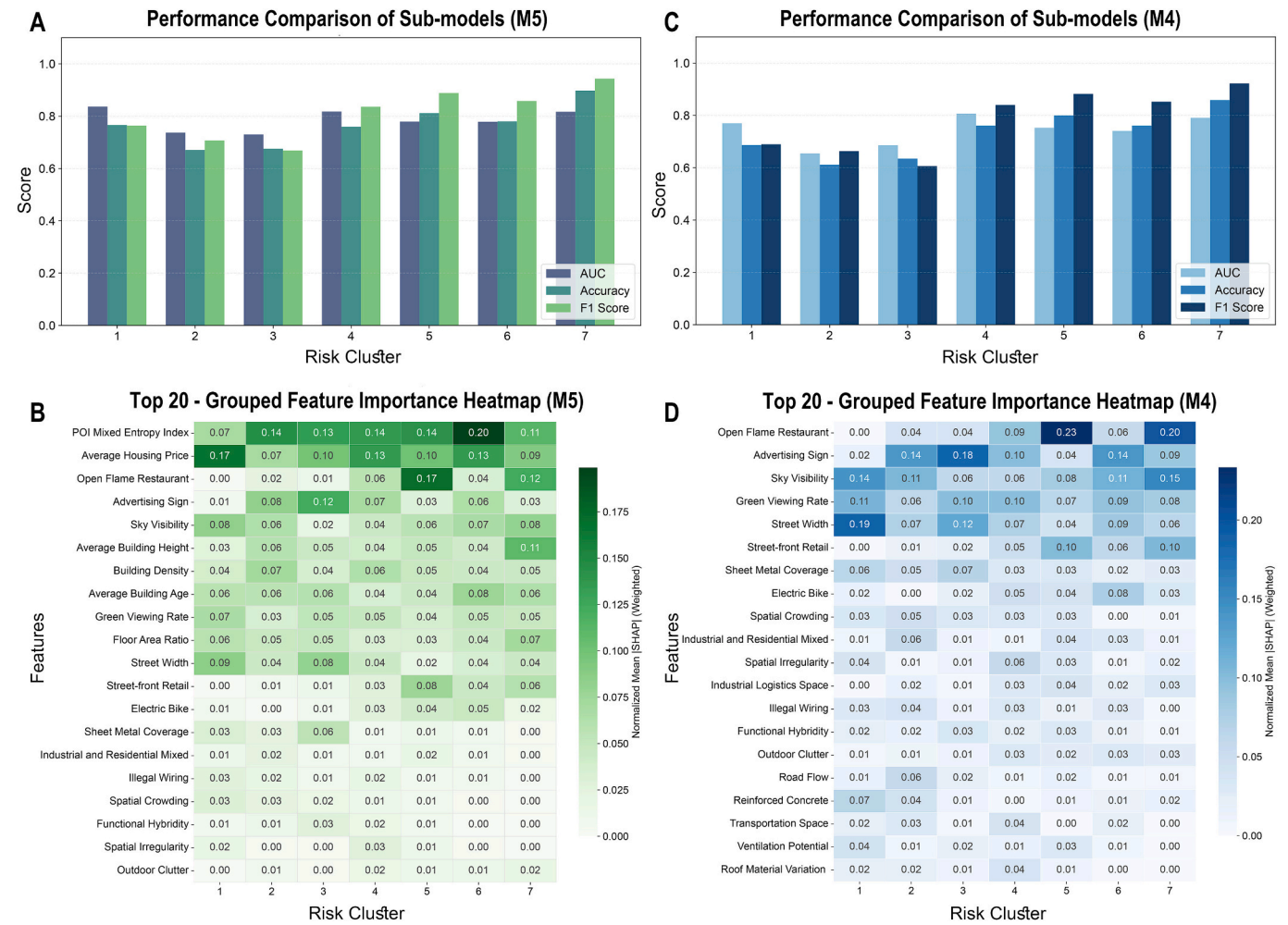


Fig. 12. Cluster-specific performance and feature importance of the GW-XGBoost submodels: (A) performance metrics for Model M5 submodels; (B) top-20 feature-importance heatmap for Model M5; (C) performance metrics for Model M4 submodels; and (D) top-20 feature-importance heatmap for Model M4.

predominant flow from central districts to Cluster 5, suggesting that a primary vulnerability in the core lies in the structural friction where urban renewal speeds lag behind commercial iteration.

Moving outward to the Transitional Zone (between the 2nd and 3rd Ring Roads), the dominant predictors undergo a sharp phase transition, with Electric Bike and Illegal Wiring becoming the prominent visual predictors of risk. This zone corresponds to the “informal development stage” of the urban lifecycle (e.g., Urban Villages), characterized by a severe infrastructure deficit. Here, the rapid influx of a job-housing separated population outpaces the supply of formal amenities. Consequently, residents adopt adaptive strategies—such as illicit wiring and corridor parking—to meet basic commuting and energy needs. The high SHAP values captured by the model are essentially quantitative mappings of this lifestyle-induced risk. Unlike the commercial prosperity of the core, predicted risk here is fundamentally linked to the spatiotemporal mismatch between public service provision and population agglomeration.

Finally, in the urban periphery (new development zones), the risk mechanism inverts again. Environmental factors like Green Viewing Rate and Sky Visibility exhibit significant mitigating predictive associations, reflecting the dividends of modern planning or resilience by design. As the nascent phase of the urban lifecycle, these areas adhere to standards of low density and wide spacing, which naturally sever fire propagation paths and provide ample buffering. The sporadic appearance of industrial logistics risks aligns with functional zoning principles that sequester hazardous facilities away from population centers. Thus,

the periphery exists in a state of “low entropy” (ordered control), forming a stark contrast to the “high entropy” (chaotic mixture) of the urban core.

5.3. Precision governance: from macro-strategy to micro-intervention

Synthesizing the spatial and semantic diagnostics, we propose a two-tier precision governance framework that cascades from macro-zonal strategies to micro-operational levers. At the strategic level, the seven semantic archetypes naturally differentiate into four quadrants based on built environment formalization and risk intensity, informing tailored spatial policies (Fig. 16). For low-formalization sectors (e.g., urban villages and dense mixed-use areas), which face severe structural deficits, mass demolition is often economically unviable. Therefore, the strategy for high-risk residential enclaves points toward micro-invasive renewal, involving drone oblique photography-guided clearance of facade clutter to reopen evacuation channels and the strategic insertion of safety facilities at community peripheries. In mid-risk mixed-use zones, governance should prioritize physical buffering, enforcing strict zoning redlines and green belts to mitigate potential contagion paths between industrial and residential functions, alongside a material substitution plan to cap environmental combustibility in informal structures. This governance model will enhance the performance of cross-departmental fire management supported by urban big data, thereby improving the level of urban safety resilience (Song & Liu, 2024).

Conversely, within high-formalization sectors, governance pivots

Multi-stage Risk Transmission Path

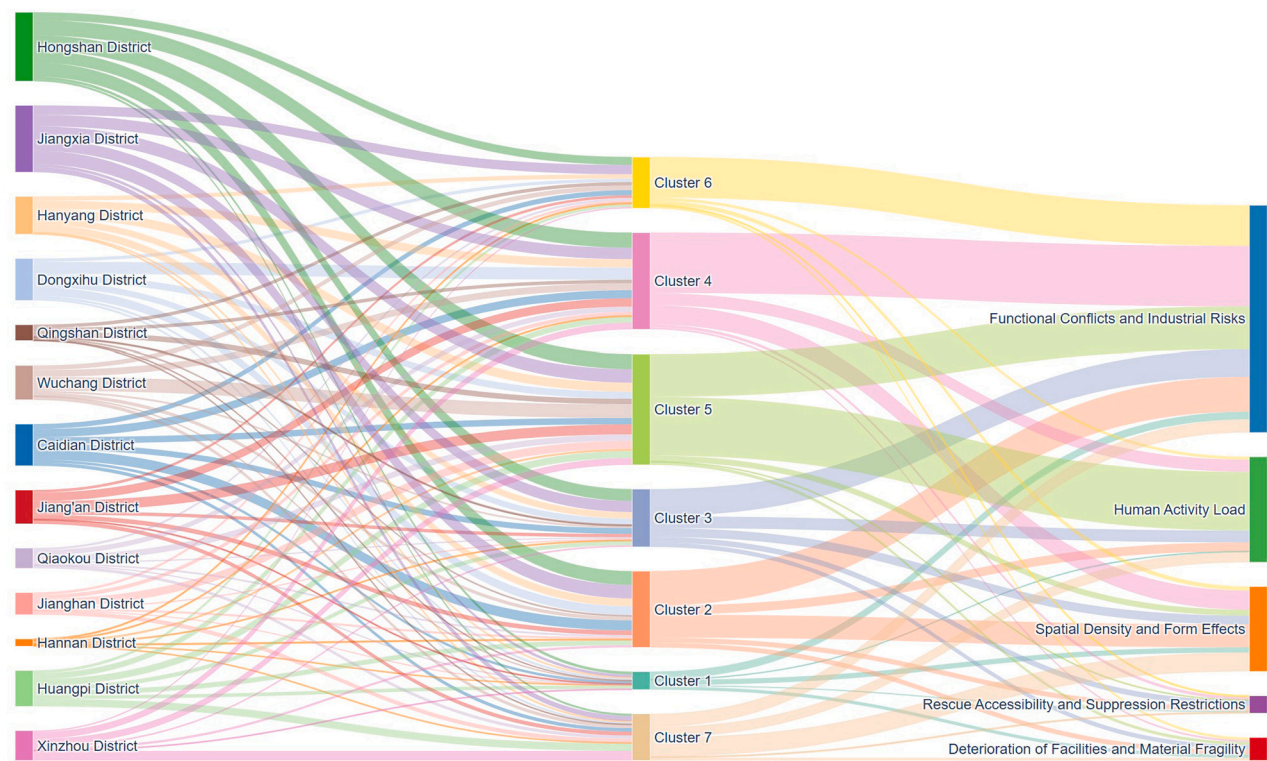


Fig. 13. Sankey diagram linking Wuhan's administrative districts, seven semantic fire-risk clusters, and five aggregated predictor categories.

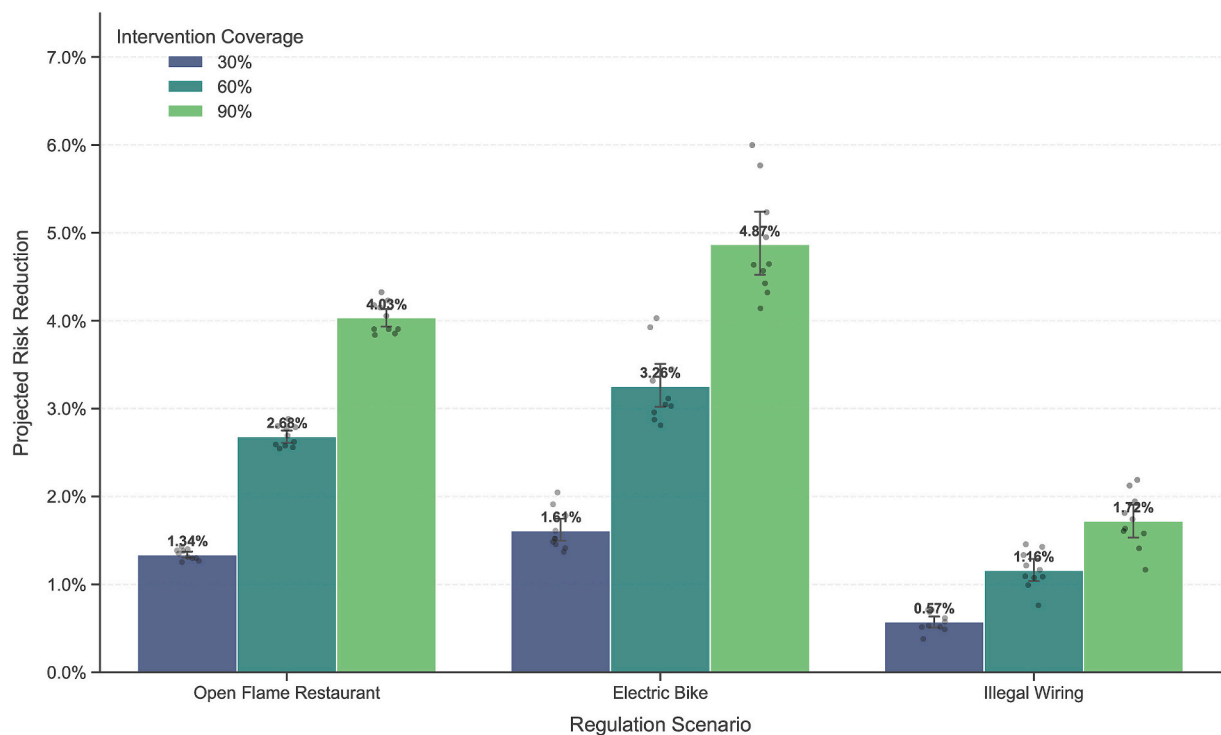


Fig. 14. Monte Carlo simulations of citywide predicted risk reduction under 30%, 60%, and 90% intervention scenarios for Open Flame Restaurant, Electric Bike, and Illegal Wiring.

toward lifecycle management and preservation. For high-risk new development zones, despite modern infrastructure, transient construction risks and material vulnerabilities require a lifecycle management approach, tightening supervision from design approval to

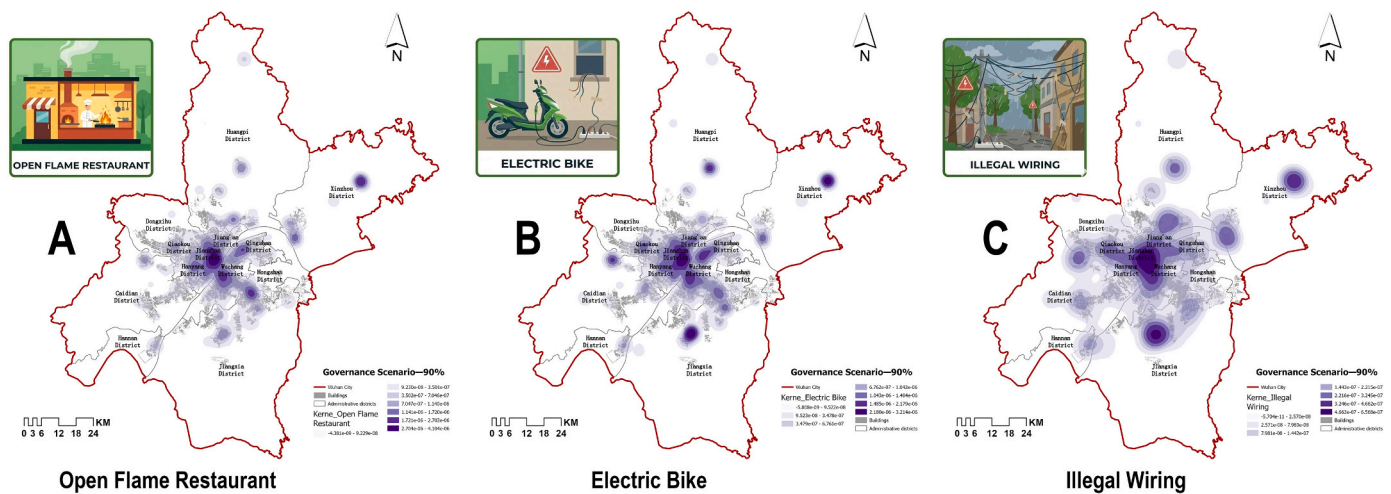


Fig. 15. Spatial distributions of risk-reduction potential under the 90% intervention scenario for (A) Open Flame Restaurant, (B) Electric Bike, and (C) Illegal Wiring.

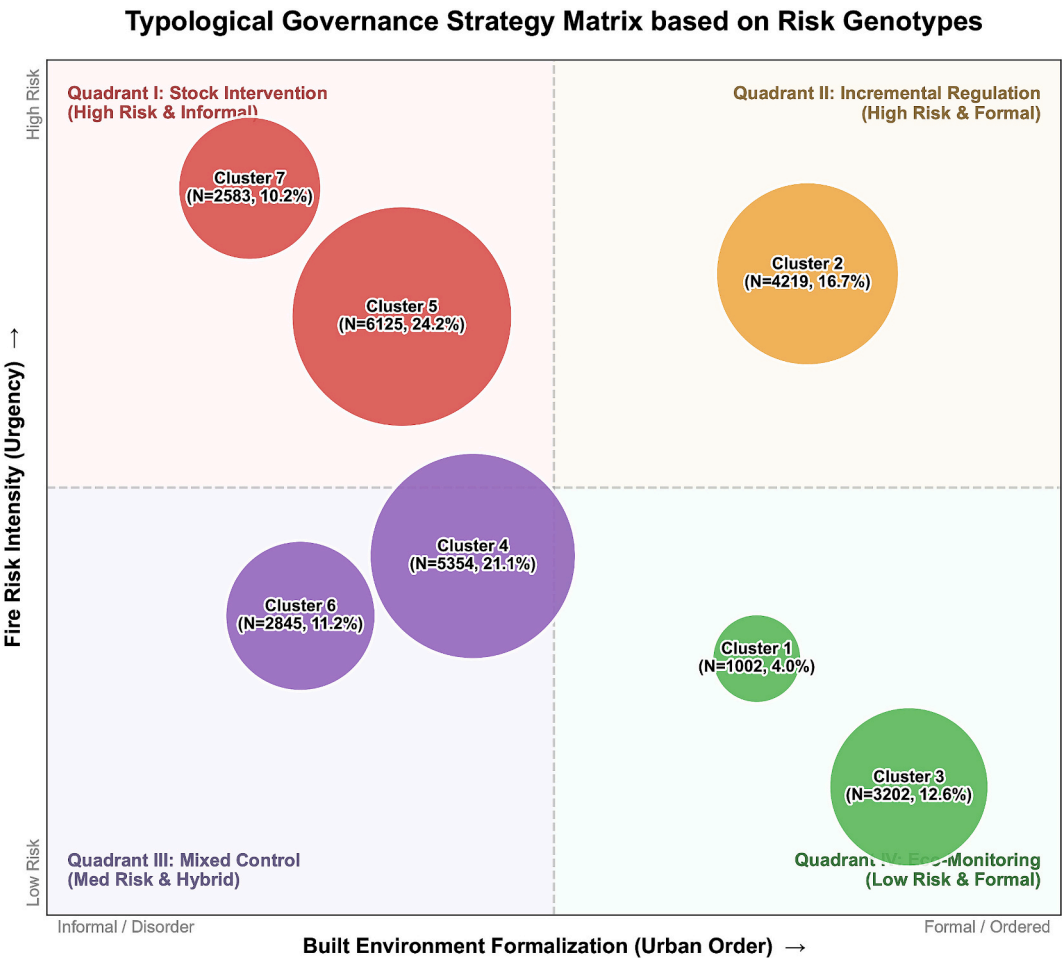


Fig. 16. Typological governance strategy matrix for the seven semantic fire-risk clusters based on fire-risk intensity and built-environment formalization.

operational monitoring to ensure continued compliance with hardware standards (Tian & Comm, 2011). For low-risk ecological and open spaces, which function as the city's safety baseline, the primary goal is entropy control. This involves strict land-use regulation to prevent the encroachment of high-risk functions and illegal structures, employing seasonal monitoring to preserve the integrity of the urban resilience system. This macro-zonal differentiation transitions urban fire

governance from homogeneous regulation to an evidence-based, spatial intervention paradigm. As a concrete extension of the macro-level strategy, the methodology proposed in this study enables the identification of visual features strongly correlated with urban fire risk and their translation into intervention targets offering a high return on investment, thereby providing a highly precise operational lever for urban safety governance. Taking the

three effective governance measures outlined above as examples, each has its own distinct implementation logic and approach to addressing fire risks. For open flame restaurants, where urban vitality conflicts with ignition risk, simple eradication is unviable. Instead, identified hotspots define priority zones for digital fire prevention, deploying IoT-based monitoring to achieve closed-loop control without disrupting commercial ecosystems (Qian, 2023). For electric bikes, the diffuse risk distribution signals a mismatch between infrastructure supply and mobility demand, necessitating supply-side redirection. Visual data on illicit parking should serve as a site-selection guide, channeling informal mobility into regulated systems via embedded charging facilities. Finally, illegal wiring serves as the most actionable entry point for visual rehabilitation. Model-generated maps can guide targeted micro-renewal interventions—such as bundling or burying lines—to disrupt potential electrical fire pathways and visibly enhance community safety perception with minimal coordination costs.

5.4. Limitations and future prospects

This study establishes a pioneering framework for urban fire risk identification by integrating MLLMs with street-view and remote sensing imagery. Unlike traditional methods reliant on static statistical proxies, this dual-perspective approach not only enhances predictive accuracy but also captures subtle, high-risk visual features that are typically overlooked. By synthesizing sub-group semantic analysis with scenario-based governance simulations, the research successfully decodes the environmental predictive patterns and transmission pathways of fire risks in complex urban environments, providing a granular, evidence-based foundation for precision governance.

However, certain limitations warrant consideration to properly contextualize these findings. First, the spatial discontinuity of current street-view coverage and the resolution limits of satellite imagery may obscure micro-scale ground details, particularly in narrow alleyways. Second, the model currently relies on static visual snapshots, potentially missing highly dynamic behavioral risks such as evening rush-hour traffic congestion or temporary material stacking. Furthermore, while the application of spatial SHAP greatly enhances interpretability, the inherent opacity of black-box machine learning models remains an ongoing challenge. Crucially, it is important to note that the relationships identified by the GW-XGBoost and SHAP frameworks represent statistical correlations and predictive associations rather than strict causal mechanisms. Future studies incorporating formal causal inference designs are necessary to isolate the definitive causal drivers of urban fire occurrences.

Future research should conduct cross-model benchmarking to systematically evaluate the comparative advantages of various MLLMs across heterogeneous research contexts, with a particular focus on their generalizability and adaptability in processing dual-dimensional spatiotemporal features. Temporally, utilizing Generative AI to simulate time-specific contexts (e.g., daytime vs. nighttime streetscapes) alongside high-frequency data—such as hourly mobile signaling and timestamped fire records—can capture dynamic behavioral risks. Spatially, incorporating low-altitude UAVs can provide flexible, high-resolution imagery to eliminate visual blind spots in narrow alleyways inaccessible to street-view cars. Ultimately, using causal inference models to establish the relationship between urban multimodal data and fire risk will enable us to gain a deeper understanding of how human-environment interactions drive fire risk around the clock.

6. Conclusion

By integrating MLLMs with geographically weighted machine learning, this study established a comprehensive framework for urban fire risk assessment. Through rigorous controlled experiments across diverse data sources and semantic scenarios, we successfully identified the environmental predictive patterns and associations of fire risk in

Wuhan. The study systematically realized the workflow from semantic feature extraction and dominant factor identification to scenario-based governance simulation. The key conclusions are as follows:

- (1) The GW-XGBoost model demonstrated superior predictive capability, outperforming baseline models by 3.1% to 15.7%. While the holistic model (M5) achieved the highest accuracy driven by macro-metrics (e.g., POI Mixed Entropy, Building Age), the Visual-Only Model (M4) proved more practically significant for governance. It effectively identified micro-scale visual proxies—specifically Open Flame Restaurant, Advertising Sign, and Sky Visibility—as the three highest-intensity local visual predictors, proving that visual semantics can independently capture risk variations.
- (2) Spatial SHAP analysis confirmed a distinct core-periphery divergence in risk mechanisms. The urban core is dominated by commercial load features (Advertising Sign, Open Flame Restaurant), whereas the transitional zone is governed by informal mobility risks (Electric Bike). Conversely, environmental factors like Sky Visibility and Green Viewing Rate act as risk suppressors in the periphery or waterfront areas, contrasting sharply with the cumulative risk accumulation in the city center.
- (3) The comparative study of the seven risk clusters revealed the structural logic of urban vulnerability. Central districts are characterized by compound risks in high-density spaces (Clusters 4 & 5), posing the severest threats due to obstructed evacuation potential. In contrast, peripheral risks (Clusters 2 & 7) stem from the structural friction between new urban expansion and legacy renovations, highlighting the vulnerability of transitional interfaces during urban renewal.
- (4) Scenario simulations quantified the marginal benefits of intervention. Regulating Electric Bike yielded the highest city-wide risk reduction (4.87%), followed by Open Flame Restaurant (4.03%) and Illegal Wiring (1.72%). Crucially, the high-yield zones for these interventions are spatially decoupled: restaurant regulation is most effective in core commercial hubs, e-bike infrastructure upgrades in transitional urban villages, and wiring rehabilitation in historic districts.
- (5) Unstructured visual data processed via MLLMs can effectively bridge the semantic gap inherent in traditional static data (e.g., POI or census data). By decoupling specific physical hazards from general urban intensity, the proposed M4 framework offers a scalable, cost-effective solution for precision risk sensing. This implies that even in the absence of granular socioeconomic data, cities can leverage accessible street-view imagery to construct high-fidelity risk profiles, offering significant practical value for rapidly urbanizing or data-scarce regions.

CRedit authorship contribution statement

Dong Xie: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Senhong Cai:** Writing – original draft, Validation, Formal analysis, Conceptualization. **Ke Wang:** Visualization, Validation, Software, Methodology. **Ce Hou:** Writing – review & editing, Software, Resources, Data curation, Conceptualization. **Shaoqing Dai:** Software, Resources, Methodology, Conceptualization. **Jingnan Huang:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgments

This research was supported in part by China Association for Science

and Technology Youth Science and Technology Talent Cultivation Project: Doctoral Student Scheme. Additionally, we extend our gratitude to the Fire and Rescue Department of Hubei Province for their research support.

Appendix A. MLLM prompt engineering framework for urban fire risk extraction

To ensure reproducibility and enforce strict semantic alignment between visual features and fire risk mechanisms, we designed a structured prompt engineering framework. This section provides a detailed overview, illustrated with examples, of prompt design for micro-scale feature extraction using street view and remote sensing imagery.

A.1. Micro-scale hazard extraction (street view imagery)

The prompt for SVI analysis is designed to simulate a professional inspection by a fire safety expert, strictly adhering to the Code for Fire Protection Design of Buildings (GB 50016). The prompt structure consists of three core components: Role Definition, Task Specification, and Evaluation Criteria.

Prompt component	Content/instruction design
System Prompt (Role Definition)	Role: You are a professional Urban Fire Safety Assessment Expert. Objective: Conduct a detailed fire risk assessment of the provided street-view image. You must visually correct for panoramic distortion and focus on direct fire hazards. Constraints: Output the result strictly in JSON format without markdown or additional text.
User Instruction (Task Specification)	1. Scene & Structure Analysis: • Classify the Scene Category (e.g., Core Commercial, Old Residential, Urban Village). • Identify Primary Building Materials (e.g., Concrete, Brick-Concrete, Steel Structure, Temporary Sheet Metal). 2. Ignition Source Identification: • Open Flame Restaurants: Detect street-front cooking equipment or exposed gas canisters. • E-mobility Risks: Identify Electric Bikes involved in illegal parking (blocking paths) or illicit charging (cables extending from windows). • Electrical Infrastructure: Detect “Flying Wires” (chaotic overhead cables) and irregular connections. 3. Vulnerability & Load Assessment: • Facade Flammability: Assess the combustibility of the facade and insulation materials. • Outdoor Clutter: Evaluate the accumulation of combustible debris (e.g., cardboard, waste). • Ad Signs: Assess the density and material of attached advertising structures. 4. Safety & Environment: • Safety Measures: Check for visible fire extinguishers or hydrants; assess if evacuation routes are obstructed. • Visual Indicators: Quantify Green View Rate and Sky Visibility as continuous variables.
Evaluation Criteria (Standardization)	To ensure consistency, specific grading criteria were embedded in the prompt: • Open Flame Risk: None (0); Few (1–2 stores); Dense (≥ 3 stores, high risk). • Outdoor Clutter: None; Light; Medium (Visible load); Severe (High fire load blocking paths). • Facade Flammability: Low (Concrete/Brick); Medium (Partial combustibles); High (Extensive wood/insulation/sheet metal).
Output Requirement	Format: Strict JSON structure. Reasoning: Include a risk_assessment_basis field to explain the visual evidence supporting the final fire_risk_score (0–10).

A.2. Meso-scale context extraction (remote sensing imagery)

The prompt for RS analysis focuses on extracting urban morphological features and broad-scale risk patterns from 1.2 m resolution satellite imagery tiles (512 × 512 pixels), simulating an expert interpretation of urban texture and spatial configuration.

Prompt component	Content/instruction design
System Prompt (Role Definition)	Role: You are a specialist in Urban Disaster Prevention Planning and Remote Sensing Interpretation with 20 years of experience. Objective: Analyze the provided satellite image tile (Wuhan, China context) to infer potential fire spread risks based on urban texture. Constraints: Output the result strictly in JSON format without markdown.
User Instruction (Task Specification)	1. Spatial Pattern & Morphology: • Building Layout: Assess Density (Coverage ratio) and Regularity (Orderliness of arrangement). • Height Variation: Infer vertical heterogeneity based on shadow distribution patterns (Uniform vs. Chaotic shadows). 2. Material & Interface Risks: • Roof Composition: Quantify the ratio of Blue Sheet Metal (indicating temporary/informal structures) and roof material consistency. • Functional Interface: Detect Industrial-Residential Mixing (large-span factories adjacent to dense housing). 3. Rescue & Spread Potential: • Road Effectiveness: Estimate internal road width and network connectivity (Grid vs. Dead-ends). • Ventilation: Assess the continuity of open spaces or corridors for smoke dissipation. 4. Comprehensive Assessment: • Disaster Chain: Estimate the potential for cascading disasters based on proximity to hazard sources. • Scoring: Provide a holistic fire_risk_score (0–10) with a key_risk_description.
Evaluation Criteria (Standardization)	To ensure consistent quantization of visual features, the following standards were applied: • Building Density: Low (<20%); Medium (20–40%); High (40–60%); Extreme (>60% or contiguous).

(continued on next page)

(continued)

Prompt component	Content/instruction design
Output Requirement	• Road Effectiveness: Effective (>6 m width); Restricted (3-6 m); Severely Restricted (<3 m). • Sheet Metal Coverage: None; Low (<10%); Medium; High (>30%). • Risk Scoring: Low (1–3); Medium (4–6); High (7–8); Severe (9–10, e.g., Urban Villages with restricted access). Format: Strict JSON structure. Reasoning: Include a qualitative description explaining the morphological evidence supporting the risk score.

Appendix B

To ensure the methodological rigor, transparency, and reproducibility of the proposed urban fire risk governance framework, this appendix presents a comprehensive consistency assessment between the MLLM (Qwen3-plus) and human planning experts. Fig. B.1 shows the results of tests for binary, ordinal and transformed continuous variables. Fig. B.2 and Fig. B.3 show the confusion matrix results for multi-class variables (one-hot encoding).

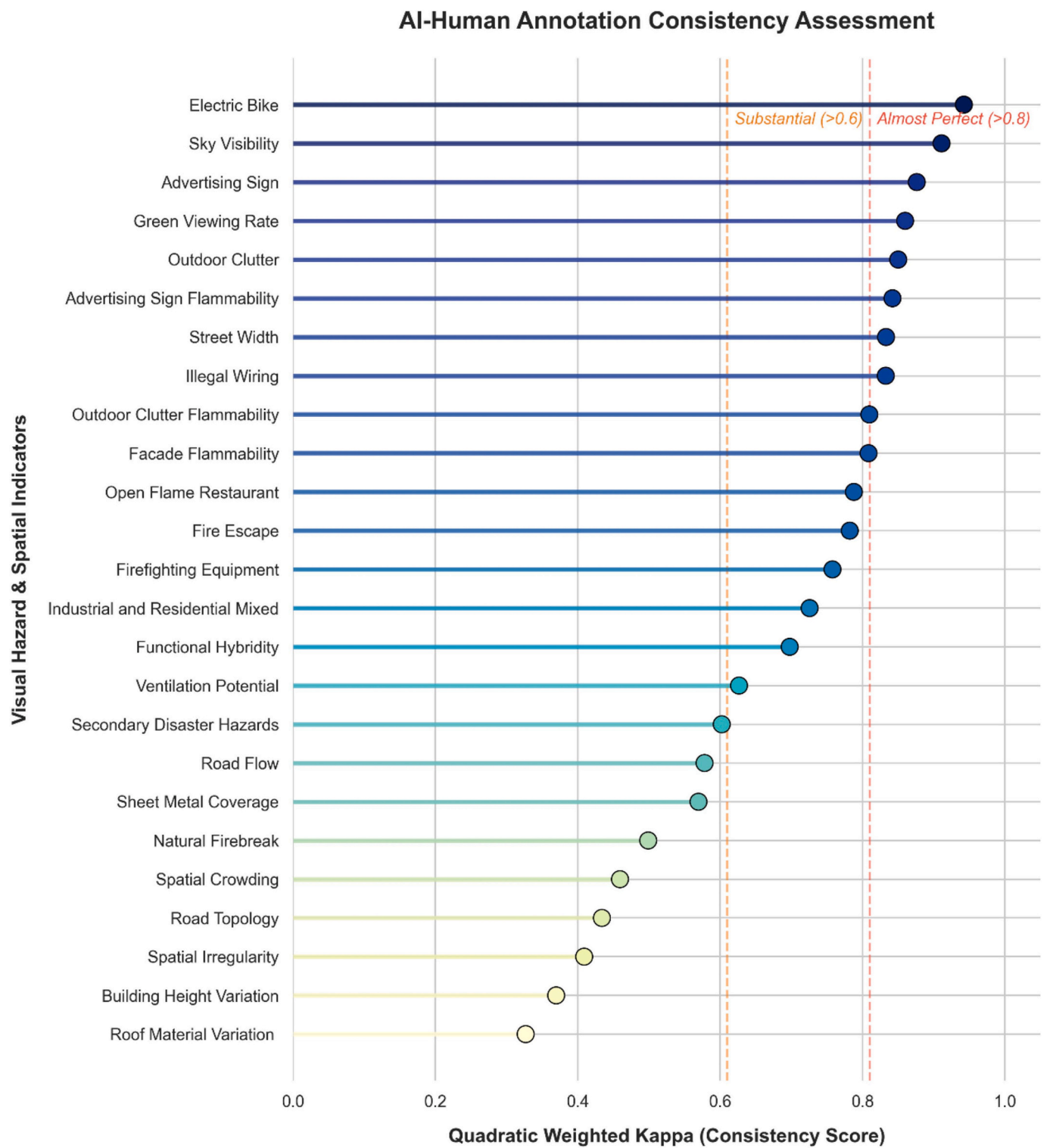


Fig. B.1. Consistency between human annotations and Qwen3-plus outputs for binary, ordinal, and discretized continuous indicators, measured using quadratic weighted kappa.

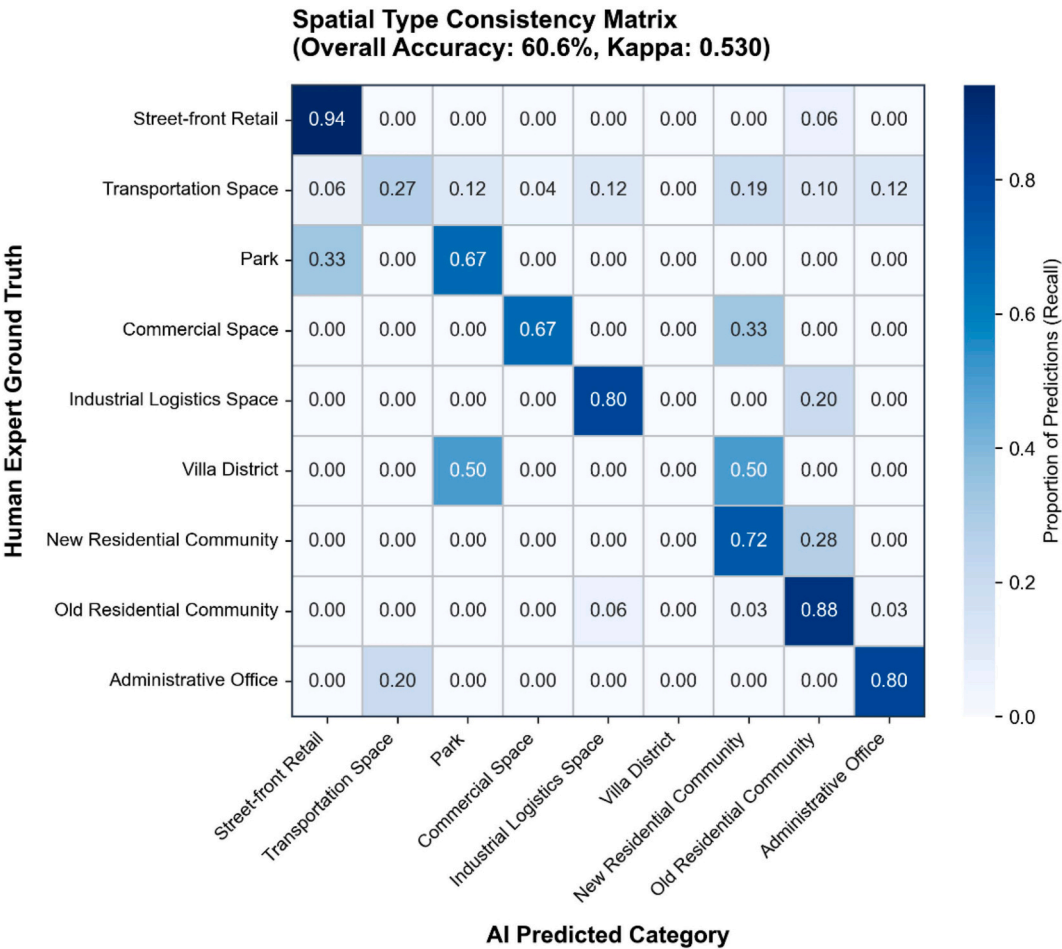


Fig. B.2. Confusion matrix comparing human annotations and Qwen3-plus predictions for spatial-type classification.

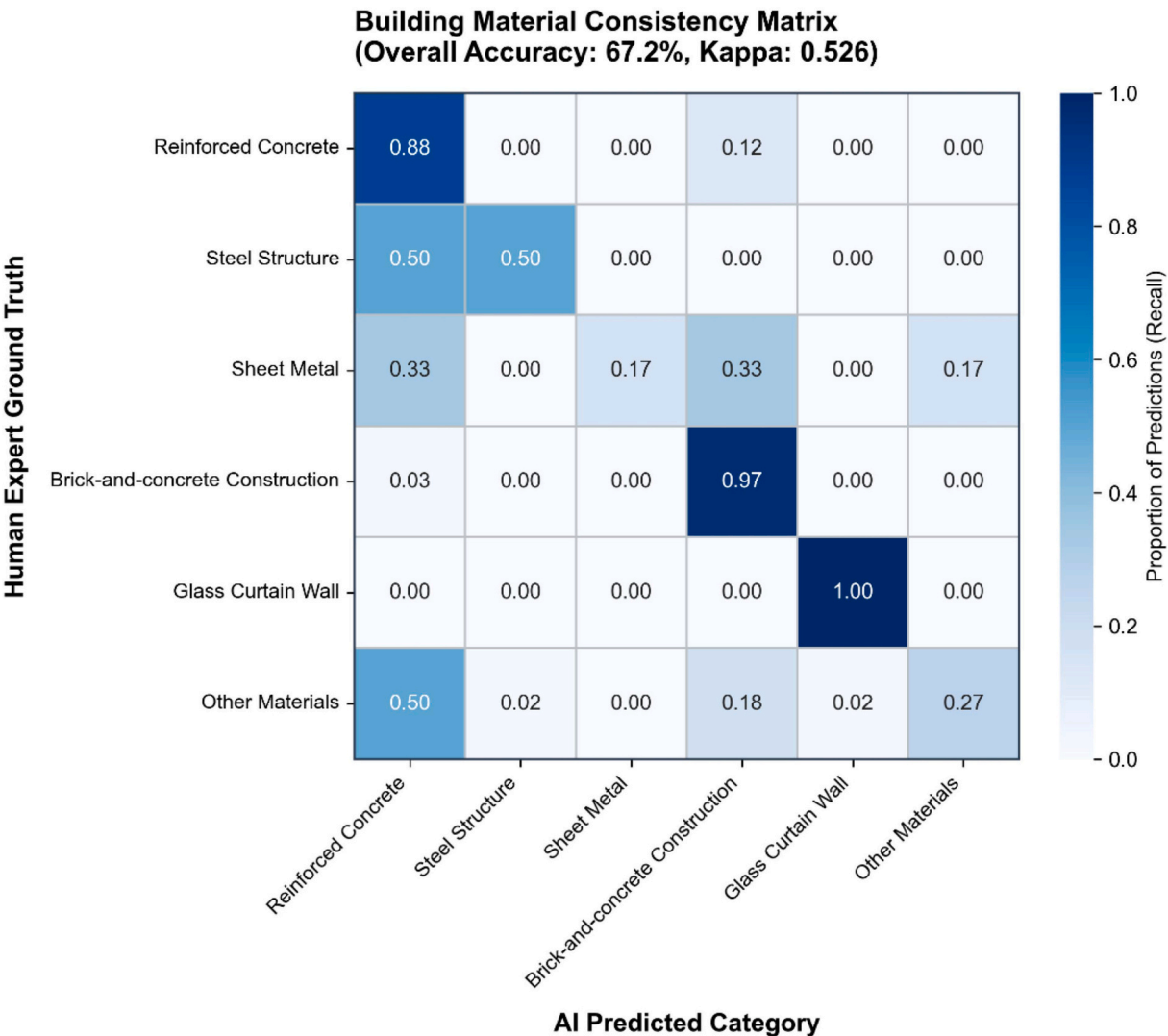


Fig. B.3. Confusion matrix comparing human annotations and Qwen3-plus predictions for building-material classification.

The multi-dimensional consistency evaluation confirms that the MLLM is a highly reliable and robust surrogate for human experts in micro-scale urban sensing. The validation reveals a clear, two-tiered performance paradigm that closely mirrors human cognitive patterns:

For high-risk, concrete physical elements (e.g., Electric Bikes, Illegal Wiring), the model achieved “Almost Perfect” agreement (Kappa >0.80). This demonstrates the MLLM’s remarkable zero-shot capability in identifying explicit visual hazard triggers.

For highly abstract spatial metrics (e.g., Spatial Crowding, Spatial Irregularity) and transitional functional zones (e.g., Transportation Spaces), the consistency was moderate. However, as illustrated in the confusion matrices, these deviations were not random hallucinations; rather, they reflected meaningful visual-semantic overlaps (e.g., visually confusing New Residential Communities with Villa Districts due to similar low-density, high-greenery traits). These discrepancies accurately capture the inherent subjective ambiguity of complex urban environments, a challenge similarly faced by human annotators.

Ultimately, this rigorous assessment substantiates that the extracted micro-scale visual features are statistically sound, laying a trustworthy empirical foundation for the subsequent GW-XGBoost spatial modeling.

Appendix C. Summary of identified fire risk clusters and representative textual keywords

Clusters	Risk archetype (topic name)	Representative keywords
Cluster 1	Combustible Exposure in Low-Density Open Spaces	Flammables, open space, vegetation, environment, low density, green belt, park parking, open flame source
Cluster 2	Material Flammability & Construction Risks in New Developments	Concrete, newly built, commercial housing, greening rate, corrugated steel, steel structure, material, ignition, roof, factory building
Cluster 3	Controllable Risk in Well-Ventilated & Structurally Intact Zones	Hazard source, unobstructed, controllable, natural ventilation, tidy, non-combustible, concrete, greening rate, hidden danger

(continued on next page)

(continued)

Clusters	Risk archetype (topic name)	Representative keywords
Cluster 4	Localized Compound Risks in Confined Non-Combustible Structures	Confined space, structure, non-combustible, fire prevention, parking, combustibles, corrugated steel, risk, adjacent, brick-concrete, open flame, density
Cluster 5	Fire Spread & Evacuation Obstruction in High-Density Residential Areas	Obstructed, fire incident, residential compound, evacuation, corrugated steel, roof, illegal wiring, dense, fire intensity, high density, spread, billboards, flammable
Cluster 6	Fire Propagation Risks at Industrial-Residential Interfaces	Disaster, high density, potential, firefighting, corrugated steel, adjacent, restricted, narrow, industrial zone, combustibles, non-combustible
Cluster 7	Fire Access Blockage & Rapid Spread in Typical Urban Villages	Typical Urban Village, chaotic, fire truck, narrow, occluded, evacuation, dense, high risk, obstructed, street-front, extremely difficult

Appendix D. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cities.2026.107449>.

Data availability

Data will be made available on request.

References

Ahn, S., Won, J., Lee, J., & Choi, C. (2024). Comprehensive building fire risk prediction using machine learning and stacking ensemble methods. *Fire*, 7(10), 336.

Alshawawreh, L., & Rush, D. (2024). Initial fire safety analysis of shelters in Zaatari and Azraq refugee camps. *International Journal of Disaster Risk Reduction*, 108, Article 104557.

Ardianto, R., & Chhetri, P. (2019). Modeling spatial-temporal dynamics of urban residential fire risk using a markov chain technique. *International Journal of Disaster Risk Science*, 10, 57–73.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, Article 104217.

Cao, Y., Jiang, J., Sun, H., Chen, W., & Ye, J. (2025). Modeling of urban fire spread with integrated cellular automata (CA)-unity framework: Case study on 2025 wildfire in Altadena. *Sustainable Cities and Society*, 134, Article 106961.

Chacra, D. B. A., & Zelek, J. S. (2017). Fully automated road defect detection using street view images. In *2017 14th conference on computer and robot vision (CRV)* (pp. 353–360).

Chen, T. (2016). *XGBoost: A scalable tree boosting system*. Cornell University.

Chen, J., Chen, L., Li, Y., Zhang, W., & Long, Y. (2023). Measuring physical disorder in urban street spaces: A large-scale analysis using street view images and deep learning. *Annals of the American Association of Geographers*, 113, 469–487.

Chen, M., Tao, Z., Tang, W., Qin, T., Yang, R., & Zhu, C. (2024). Enhancing emergency decision-making with knowledge graphs and large language models. *International Journal of Disaster Risk Reduction*, 113, Article 104804.

Chhetri, P., Corcoran, J., Stimson, R. J., & Inbakaran, R. (2010). Modelling potential socio-economic determinants of building fires in south east Queensland. *Geographical Research*, 48, 75–85.

Corcoran, J., Higgs, G., & Higginson, A. (2011). Fire incidence in metropolitan areas: A comparative study of Brisbane (Australia) and Cardiff (United Kingdom). *Applied Geography*, 31, 65–75.

Cui, S., Wang, N., Zhao, E., Zhang, J., & Zhang, C. (2024). Assessing urban fire risk: An ensemble learning approach based on scenarios and cases. *International Journal of Disaster Risk Reduction*, 114, Article 104941.

Dong, X.-m., Li, Y., Pan, Y.-l., Huang, Y.-j., & Cheng, X.-d. (2018). Study on urban fire station planning based on fire risk assessment and GIS technology. *Procedia Engineering*, 211, 124–130.

Draganov, A., Jorgensen, J., Scheel, K., Mottin, D., Assent, I., Berry, T., & Aslay, C. (2023). *ActUp: Analyzing and consolidating tSNE & UMAP*, 32nd International Joint Conference on Artificial Intelligence (IJCAI), Macao, PEOPLES R CHINA (pp. 3651–3658).

El-Niss, A., Alzu'Bi, A., & Abuarqoub, A. (2024). *Multimodal fusion for disaster event classification on social media: A deep federated learning approach, proceedings of the 7th international conference on future networks and distributed systems* (pp. 758–763). Dubai, United Arab Emirates: Association for Computing Machinery.

Elo, I. T., Mykyta, L., Margolis, R., & Culhane, J. F. (2009). Perceptions of neighborhood disorder: The role of individual and neighborhood characteristics. 90, 1298–1320.

Fan, C., Jiang, Y., & Mostafavi, A. (2020). Social sensing in disaster city digital twin: Integrated textual-visual-geo framework for situational awareness during built environment disruptions. 36, Article 04020002.

Flannigan, M. D., & Harrington, J. B. (1988). A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). *Journal of Applied Meteorology and Climatology*, 27, 441–452.

Grekosis, G. (2025). Geographical-XGBoost: A new ensemble model for spatially local regression based on gradient-boosted trees. *Journal of Geographical Systems*, 27, 169–195.

Hossain, M. R., & Smirnov, O. (2023). Analyzing the risk factors of residential fires in urban and rural census tracts of Ohio using panel data analysis. *Applied Geography*, 151, Article 102863.

Hou, C., Zhang, F., Li, Y., Li, H. F., Mai, G. C., Kang, Y. H., ... Liu, Y. (2025). Urban sensing in the era of large language models. *Innovation (Cambridge, Mass.)*, 6.

Hu, J., Shu, X., Xie, S., Tang, S., Wu, J., & Deng, B. (2019). Socioeconomic determinants of urban fire risk: A city-wide analysis of 283 Chinese cities from 2013 to 2016. *Fire Safety Journal*, 110, Article 102890.

Iban, M. C., & Sekertekin, A. (2022). Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics*, 69, Article 101647.

Ibrahim, M. R., Haworth, J., & Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, 96, Article 102481.

Imbe, T., Maruyama, K., & Acm. (2021). *Finding the noteworthy learning material about a keyword through courses already learned and to be learned*, 52nd Technical Symposium on Computer Science Education, null, NY (p. 1268).

Jin, G., Wang, Q., Zhu, C., Feng, Y., Huang, J., & Hu, X. (2020). Urban fire situation forecasting: Deep sequence learning with spatio-temporal dynamics. *Applied Soft Computing*, 97, Article 106730.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2012). Using Google Street View to audit the built environment: Inter-rater reliability results. *Annals of Behavioral Medicine*, 45, S108–S112.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363–374.

Li, L., Liu, B., Zheng, W., Wu, X., Song, L., & Dong, W. (2022). Investigation and numerical reconstruction of a full-scale electric bicycle fire experiment in high-rise residential building. *Case Studies in Thermal Engineering*, 37, Article 102304.

Li, T., Chen, H., Fung, J. C. H., Chan, D. H. L., Yu, A. L. C., Leung, K. K. M., & Yu, J. Z. (2023). Large presence of bromine and toxic metals in ambient fine particles from urban fires. *Atmospheric Environment*, 295, Article 119554.

Liu, F., Zhao, S., Weng, M., & Liu, Y. (2017). Fire risk assessment for large-scale commercial buildings based on structure entropy weight method. *Safety Science*, 94, 26–40.

Liu, Z., Xu, Y., Li, Z., Zhai, M., Yang, W., Lin, J., & Sun, Y. (2025). Towards evidence-based fire prevention policy: Uncovering drivers of urban residential fire spread via explainable machine learning. *Developments in the Built Environment*, 24, Article 100761.

Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*, 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA.

Ma, H., Sun, X., & Zhao, J. (2025). City-scale fire risk modeling based on spatial regression methods. *Safety and Emergency Science*, 1, Article 9590002.

Masoumi, Z., van L. Genderen, J., & Maleki, J. (2019). Fire risk assessment in dense urban areas using information fusion techniques. *ISPRS International Journal of Geo-Information*, 579.

Neto, A. C. A., Sander, J., Campello, R., & Nascimento, M. A. (2017). *Efficient computation of multiple density-based clustering hierarchies*, 17th IEEE International Conference on Data Mining (ICDMW), New Orleans, LA (pp. 991–996).

Parente, J., Tonini, M., Amraoui, M., & Pareira, M. (2024). Socioeconomic impacts and regional drivers of fire management: The case of Portugal. In J. Rodrigo-Comino, & L. Salvati (Eds.), *Fire hazards: Socio-economic and regional issues* (pp. 181–194). Cham: Springer International Publishing.

Qian, X. D. (2023). Evaluation on sustainable development of fire safety management policies in smart cities based on big data. *Mathematical Biosciences and Engineering*, 20, 17003–17017.

Qian, J., Du, Y., Liang, F., Yi, J., Wang, N., Tu, W., ... Ma, T. (2024). Quantifying urban linguistic diversity related to rainfall and flood across China with social media data. *ISPRS International Journal of Geo-Information*, 13(3), 92.

Reilley, C., Crandall, M. S., Kline, J. D., Kim, J. B., & de Diego, J. (2023). The influence of socioeconomic factors on human wildfire ignitions in the Pacific Northwest, USA. *Fire*, 6(8), 300.

Rohde, D., Corcoran, J., & Chhetri, P. (2010). Spatial forecasting of residential urban fires: A Bayesian approach. *Computers, Environment and Urban Systems*, 34, 58–69.

- Shaham, Y., & Benenson, I. (2018). Modeling fire spread in cities with non-flammable construction. *International Journal of Disaster Risk Reduction*, 31, 1337–1353.
- Sijabat, K. R., Aufaristama, M., Wirawan Arief, M. C., & Dharmawan, I. A. (2025). Integrating remote sensing and street view imagery with deep learning for urban slum mapping: A case study from Bandung City. *Applied Sciences*, 15(14), 8044.
- Simons, M., Ney, H., Martin, S. C., & Ieee Comp, S.O.C. (1997). *Distant bigram language modelling using maximum entropy*, 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 97), Munich, Germany (pp. 787–790).
- Song, W., & Liu, Z. (2026). Urban fire risk prediction and spatiotemporal analysis based on machine learning. In Y. Liu, L. Chen, S. Wang, M. Deng, B. Zheng, & Q. Mei (Eds.), *Spatial data and intelligence* (pp. 20–34). Singapore: Springer Nature Singapore.
- Song, Y. W., & Liu, J. D. (2024). Exploring the cross-sectoral joint fire management mode driven by fire information in China: From the perspective of organizational interaction. *Fire*, 7.
- Sufianto, H., & Green, A. R. (2012). Urban fire situation in Indonesia. *Fire Technology*, 48, 367–387.
- Sun, B. (2026). Geographic information system-based urban community fire danger evaluation method fusing evidential reasoning and intelligent optimization. *Reliability Engineering & System Safety*, 268, Article 112039.
- Sun, B., Chen, L., & Guo, T. (2025). GIS-based large-scale urban fire situation spread probability model and simulation technique. *International Journal of Disaster Risk Reduction*, 129, Article 105779.
- Sun, Y., Wang, C., & Peng, Y. (2024). *Reducing hallucinations: Enhancing VQA for flood disaster damage assessment with visual contexts, proceedings of the international conference on algorithms, software engineering, and network security* (pp. 587–590). Nanchang, China: Association for Computing Machinery.
- Thakare, K. V., & Tajne, K. M. (2025). A comprehensive review of geographic information systems (GIS)-based methodologies for urban fire risk assessment. *Cureus*, 2, Article e7759/s44388-44024-02916-y.
- The New York Times. (2025). *Death toll rises to 128 in Hong Kong high-rise fire*.
- Thi Hang, H., Mallick, J., Alqadhi, S., Bindajam, A. A., & Abdo, H. G. (2024). Exploring forest fire susceptibility and management strategies in Western Himalaya: Integrating ensemble machine learning and explainable AI for accurate prediction and comprehensive analysis. *Environmental Technology & Innovation*, 35, Article 103655.
- Tian, Y. M., & Comm, E. B. M. O. (2011). *Study on fire risk management in the process of building construction*, 2nd International Conference on Engineering and Business Management, Wuhan, PEOPLES R CHINA (pp. 1394–1397).
- United Nations Office for Disaster Risk Reduction. (2019). *Fire risk reduction on the margins of an urbanizing world*.
- Wang, C., Engler, D., Li, X., Hou, J., Wald, D. J., Jaiswal, K., & Xu, S. (2024). Near-real-time earthquake-induced fatality estimation using crowdsourced data and large-language models. *International Journal of Disaster Risk Reduction*, 111, Article 104680.
- Wang, K., Yuan, Y., Chen, M., & Wang, D. (2021). A POIs based method for determining spatial distribution of urban fire risk. *Process Safety and Environmental Protection*, 154, 447–457.
- Wang, N., Xu, Y., & Wang, S. (2022). Interpretable boosting tree ensemble method for multisource building fire loss prediction. *Reliability Engineering & System Safety*, 225, Article 108587.
- Wang, N., Zhao, S., Cui, S., & Fan, W. (2021). A hybrid ensemble learning method for the identification of gang-related arson cases. *Knowledge-Based Systems*, 218, Article 106875.
- Xia, J., Tong, Y., & Long, Y. (2025). Advancements in the application of large language models in urban studies: A systematic review. *Cities*, 165, Article 106142.
- Xia, Z., Zhang, X., Zhai, G., & Zhang, Y. (2025). Integrating visual spatial vulnerability to quantify fire-prone neighborhoods in cities: A case study of Nanjing, China. *International Journal of Disaster Risk Reduction*, 128, Article 105758.
- Xu, F., Ma, J., Li, N., & Cheng, J. C. P. (2025). Large language model applications in disaster management: An interdisciplinary review. *International Journal of Disaster Risk Reduction*, 127, Article 105642.
- Xu, Z., Li, J., Cheng, S., Rui, X., Zhao, Y., He, H., Guan, H., Sharma, A., Erxleben, M., Chang, R., & Xu, L. L. (2025). Deep learning for wildfire risk prediction: Integrating remote sensing and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 227, 632–677.
- Yang, X., Hao, Y., Ding, H., Yu, C., Liu, J., Li, L., & Chen, J. (2025). Explainable Artificial Intelligence (XAI) framework using XGBoost and SHAP for assessing urban fire risk based on spatial distribution features. *International Journal of Disaster Risk Reduction*, 129, Article 105798.
- Yao, Q., Jiang, D., Zheng, B., Wang, X., Zhu, X., Fang, K., ... Trouet, V. (2024). Anthropogenic warming is a key climate indicator of rising urban fire activity in China. *National Science Review*, 11, Article nwae163.
- Yu, H., Wang, J., Bai, Y., Yang, W., & Xia, G. (2018). *Analysis of large-scale UAV images using a multi-scale hierarchical representation*. 21 pp. 33–44).
- Yu, Z., Shen, X., Xu, R., Wang, Z., Wan, Z., Chen, M., Cui, Y., Fang, Y., & Ma, X. (2024). Understanding the combustion behavior of electric bicycle batteries and unveiling its relationship with fire extinguishing. *Journal of Energy Chemistry*, 91, 609–618.
- Zhang, S., Song, Y., Dong, Q., Yang, H., & Shi, L. (2025). Developing an indicator system for urban fire risk assessment. *Fire Safety Journal*, 158, Article 104536.
- Zhang, Y., Deng, Q., Guo, M., Li, Y., Lu, F., Chen, J., ... Long, Y. (2023). Using street view imagery to examine the association between urban neighborhood disorder and the long-term recurrence risk of patients discharged with acute myocardial infarction in central Beijing, China. *Cities*, 138, Article 104366.
- Zhang, Y., Zhao, H., & Long, Y. (2025). CMAB: A multi-attribute building dataset of China. *Scientific Data*, 12, 430.
- Zhou, B.-x. (2013). Analysis of fire hazards of billboards on exterior walls of buildings and fire control safety countermeasures. *Procedia Engineering*, 52, 693–696.