Krishna Agarwal, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

# Weakly Supervised Urban Heat Island Segmentation Using U-Net and Physics Guided Labels

**Krishna Agarwal[1], Aman Shaw[2], Shwet Kashyap[3]**
[1]Department of Computer Science and Engineering
SRM Institute of Science & Technology Chennai India

**Abstract- Urban Heat Island (UHI) is an established environmental phenomenon where urban environments experience substantially higher temperatures than their surrounding rural environments, which is directly attributed to the high density of infrastructure and minimal vegetation, along with anthropogenic heat emissions. Accurate identification of UHI regions is crucial for effective urban planning. Conventional UHI identification methods based on threshold value analysis are ineffective and do not account for spatial continuity. This study presented an effective and efficient framework that combined satellite data with physical models and deep learning to identify UHI regions. Landsat-8 thermal and optical data were employed to derive Land Surface Temperature (LST) and Normalized Difference Vegetation Index (NDVI) values. A rule-based approach was initially employed to generate weak UHI labels, which were then refined using a U-Net convolutional neural network (CNN). Experimental results across five diverse global cities demonstrate that the proposed method is highly effective for UHI identification. By leveraging physical regularization, it resolves the noise overfitting of traditional neural baselines—outperforming them in classification calibration (ROC-AUC and AUPRC) and cross-city generalizability—while entirely eliminating the need for manual pixel-level annotation.**

**Keywords: Urban Heat Island, Land Surface Temperature, U-Net.**

## I. INTRODUCTION

Rapid urbanization has caused considerable changes in land surface characteristics, often causing urban areas to experience significantly higher surface temperatures than their rural surroundings. This temperature disparity, known as the Urban Heat Island (UHI) effect, poses critical environmental, energy, and public health challenges for urban areas. The resulting heat stress increases energy consumption, degrades air quality, and threatens the well-being of urban populations, making the accurate detection and monitoring of UHI zones an urgent priority for sustainable urban planning and management.

Satellite remote sensing serves as a vital tool for analyzing surface temperature patterns at scale, particularly through the extraction of thermal and optical features using sensors such as Landsat-8. Historically, conventional methods for detecting UHI relied heavily on direct temperature thresholding and standard vegetation indices. However, these traditional approaches suffer from notable limitations: they frequently lack spatial coherence and fail to fully encapsulate the complexities of urban morphology, leading to fragmented or inaccurate representations of the UHI.

Recent advancements in deep learning, specifically convolutional neural networks (CNNs), offer powerful alternatives for complex spatial pattern recognition. Among these architectures, the U-Net model has emerged as a highly effective framework for pixel-level image segmentation across various visual tasks. To address the limitations of traditional techniques, this study proposed a novel end-to-end framework for UHI detection that bridged physical remote sensing principles with advanced deep learning.

By utilizing physically derived features as weak supervision to train a U-Net architecture, the proposed model learned the intrinsic spatial relationships of the urban heat distribution. The experimental results demonstrated that this integrated approach significantly outperformed conventional threshold-based methods, offering a robust, spatially coherent solution for large-scale UHI mapping.

## II. PREVIOUS WORKS

Several studies have investigated the relationship between LST and urban land cover characteristics using satellite imagery. Guha et al. [1] analyzed the correlation between LST, vegetation, and built-up indices using Landsat-8 data across European cities. Their work demonstrated clear statistical relationships between temperature and land cover patterns; however, the study primarily relied on threshold-based analysis and lacked a spatially coherent framework for identifying actionable UHI hotspots. Similarly, Li et al. [7] explored urban thermal variability using Landsat-derived LST combined with land-cover classification techniques. Although their approach provided interpretable insights into urban thermal behavior, it relied heavily on rule-based classification and static thresholds, limiting adaptability across heterogeneous urban environments. In another study, Peng et al. [8] analyzed the urban thermal environment using machine learning and multispectral remote sensing data.

Other research has focused on modeling the temporal dynamics of urban heat rather than its spatial segmentation. Khalil et al. [2] applied time-series analysis on remotely sensed temperature data to understand the drivers of UHI evolution and develop predictive models. While effective in capturing long-term trends, the framework did not emphasize fine-scale spatial hotspot extraction. Xu et al. [4] conducted a comprehensive long-term analysis of UHI patterns using multisource Landsat imagery, but the approach remained largely descriptive and did not employ learning-based techniques for generating operational hotspot maps. Recent advancements in deep learning were also explored for urban remote sensing tasks. Xu et al. [3] proposed a graph convolutional network (GCN) for urban scene classification by integrating visual and semantic features.

While effective, it did not incorporate thermal remote sensing variables. Similarly, Sun et al. [9] developed a deep learning-based framework for downscaling land surface temperature but did not explicitly address UHI hotspot detection. Deep learning-based UHI mapping was explored in a few recent studies: Shaamala et al. [5] demonstrated the potential of U-Net architectures for high-resolution urban heat mapping, while Zhou et al. [10] proposed a CNN-based urban climate mapping framework requiring manually annotated labels. These limitations highlighted the need for scalable approaches combining physical reasoning with spatial deep learning [6] to produce interpretable and operational UHI detection systems.

## III. PROBLEM STATEMENT

UHI detection from satellite imagery remains a challenging task because of the complex nonlinear interactions between LST, vegetation cover, emissivity variations, and urban morphology. Traditional approaches primarily rely on direct, threshold-based classification using temperature and vegetation indices. However, these methods often generate fragmented and spatially inconsistent hotspot maps.

Let $T_s(x, y)$ denote LST at pixel location (x, y) and NDVI(x, y) denote the vegetation index. Traditional rule-based UHI detection can be expressed as:

$$UHI(x,y) = \begin{cases} 1, if\ T_s(x,y) > u_T + \alpha\sigma_T\ and\ NDVI(x,y) < \tau \\ 0,\ otherwise \end{cases} \quad (1)$$

where μT and σT are the scene mean and standard deviation of temperature, respectively, and τ is the vegetation threshold. Although computationally simple, this process has several key disadvantages:

This formulation suffers from several key disadvantages. First, the process lacks spatial coherence because it does not consider the spatial context and spatial continuity of heat islands. Second, the process exhibits high threshold sensitivity, making it highly dependent on the choice of global parameters. Third, its binary rigidity prevents it from handling gradual transitions between urban and rural environments. Finally, it demonstrates poor generalizability, failing to adapt to diverse climatic conditions and varied urban morphologies.

Thus, this study developed a physically meaningful, spatially coherent, and weakly supervised UHI detection framework, which addressed the above-mentioned issues by combining satellite-based thermal modeling with deep learning-based spatial enhancement, thereby completely removing the need for manual annotations and improving the accuracy of the process.

## IV. PROPOSED SOLUTION

The proposed framework integrates a physically informed surface temperature modeling approach with a regression-based deep learning technique for spatial segmentation. The framework includes radiometric calibration, emissivity correction, soft-label generation, and spatial refinement using a U-Net architecture. Fig. 1 illustrates the end-to-end pipeline of the proposed solution.

**A. Dataset Preparation**

The dataset consisted of Landsat-8 Level-1 imagery, including: Thermal Infrared Band 10; Red Band (Band 4); Near-Infrared Band (Band 5); and radiometric calibration metadata (MTL file). All computations were performed at the pixel level using physically meaningful transformations.

**B. LST Computation**

**1. Radiance Conversion**

The raw digital numbers (DN) were converted to top-of-atmosphere (TOA) spectral radiance using the linear scaling coefficients that were provided in the MTL metadata file:

$L_\lambda = M_L Q_{cal} + A_L$ (2)

where ML and AL were the multiplicative and additive radiance scaling factors, and Qcal was the quantized calibrated pixel value.

**2. Brightness Temperature**

The spectral radiance was then converted to At-Sensor Brightness Temperature (TB) in Kelvin:

$$T_B = \frac{K_2}{\ln\left(\left(\frac{K_1}{L_\lambda}\right)+1\right)} \quad (3)$$

where K1 and K2 were the thermal calibration constants for Landsat-8 Band 10 (K1 = 774.8853 W $m^{-2}$ $sr^{-1}$ $\mu m^{-1}$, K2 = 1321.0789 K).

**3. Normalized Difference Vegetation Index**

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (4)$$

High NDVI values (≥0.3) corresponded to vegetated regions, while low values indicated built-up or barren areas. Water bodies were characterized by negative NDVI values.

**4. Emissivity Estimation**

The fractional vegetation cover Pv was estimated as:

$$P_v = \left(\frac{NDVI-NDVI_{min}}{NDVI_{max}-NDVI_{min}}\right)^2 \quad (5)$$

Subsequently, the land surface emissivity (ε) was computed using the empirical relationship:

$\varepsilon = 0.004 \cdot P_v + 0.986$ (6)

5. Emissivity Corrected LST

To account for varying surface properties, the final emissivity corrected LST was derived as:

$$T_S = \frac{T_B}{1+\left(\frac{\lambda T_B}{\rho}\right) ln(\varepsilon)} \quad (7)$$

where λ was the wavelength of Band 10 (10.8 μm) and ρ = h·c/σ was a thermal constant (1.438 × $10^{-2}$ m·K). This ensured physically accurate modeling of surface temperature.

C. Soft UHI Intensity Modeling

Instead of applying binary thresholding, the UHI footprint was formulated as a continuous intensity field to serve as weak supervision:

$UHI_{score}(x,y) = \alpha T_{norm}(x,y) + \beta(1 - NDVI(x,y))$
(8)

where Tnorm was normalized temperature and α+β=1.

The heuristic thresholds were empirically grounded in established Surface Urban Heat Island (SUHI) climatology principles. While they functioned as proxy labels and inherently lacked the fine-grained granularity of planning-grade products (e.g., Local Climate Zone maps or high-resolution municipal land-use registries), they reliably captured macroscopic heat anomalies over impervious, non-vegetated surfaces.

To avoid circular dependency—i.e., validating the model on the same weak labels used for training—two independent validation strategies were applied: (1) Correlation with raw continuous LST ($R^2$ and RMSE), which confirmed the model learned the physical thermal distribution rather than merely memorizing heuristic boundaries; and (2) Leave-

One-City-Out (LOCO) testing, which demonstrated spatial generalization to unseen urban morphologies. Future work intersects these predictions with high-resolution land cover datasets (e.g., Copernicus Sentinel-2) for direct comparison against planning-grade references. This soft labeling formulation preserved the thermal gradients and vegetation interactions rather than relying on rigid thresholds.

### D. U-Net Based Spatial Refinement

A U-Net CNN was employed in regression mode to learn the spatial and morphological patterns of the UHI effect. The input consisted of two channels: Channel 1 – Emissivity corrected LST; Channel 2 – NDVI. The loss function was the Mean Squared Error (MSE) between the predicted UHI intensity $\hat{Y}(x,y)$ and the soft label U(x,y):

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(U\widehat{H}I_i - UHI_i)^2 \quad (9)$$

The model was trained using 256 × 256 patches with an 80–20 train–validation split. During inference, sliding window prediction with Gaussian smoothing was applied to generate the final UHI intensity map.

### E. Implementation Details

To ensure complete reproducibility, the following configuration was used. The U Net comprised a 4-level encoder–decoder pathway. Convolutional kernels were 3 × 3 with stride 1, followed by Batch Normalization and ReLU activation. Feature maps began at 64 channels in the first encoder block, doubling at each downsampling step to a maximum of 512 channels at the bottleneck. Upsampling in the decoder used bilinear interpolation followed by a 1 × 1 convolution to halve channel depth before each decoder block. Skip connections concatenated the corresponding encoder feature maps to the upsampled decoder features at every level.

The Adam optimizer was used with decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to $1 \times 10^{-4}$, adjusted by a StepLR scheduler with a decay factor of 0.1. Training ran for a maximum of 10 epochs. Patches containing more than 20% invalid (NaN) pixels were excluded. Training and inference were conducted on a dedicated GPU environment

## V. DATASET AND PREPROCESSING

### A. Satellite Data Acquisition

The dataset used in this study consisted of Landsat-8 Level-1 imagery, which included thermal infrared and multispectral optical bands. Thermal Band 10 was used for temperature estimation, whereas Bands 4 (red) and 5 (near-infrared) were used for vegetation analysis. The metadata files accompanying the imagery provided the radiometric calibration constants required for physical modeling. Five globally distributed cities were selected to encompass a wide range of climatic conditions and urban morphologies: Beijing (temperate), Delhi (semi-arid), Dubai (arid), London (oceanic), and Mumbai (tropical).

### B. LST and NDVI Computation

The raw digital numbers (DN) from Band 10 were first converted to TOA radiance and subsequently to the at-sensor brightness temperature. An emissivity correction was applied to derive an accurate and physically meaningful LST. This required computing the NDVI from the Red and Near-Infrared bands, as shown in the vegetation density map in Fig 2.

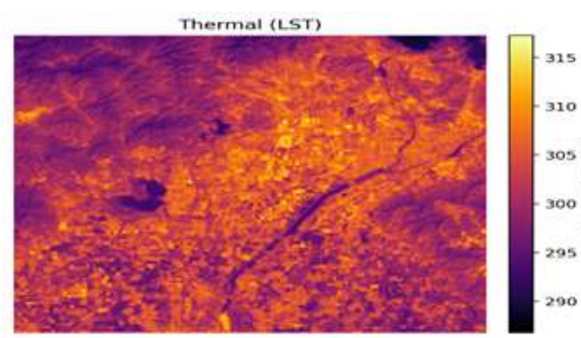


Fig. 1. Land Surface Temperature LST map derived from the thermal band after emissivity correction for Beijing city Warmer colors indicate higher surface temperatures the urban core is clearly visible as a high temperature region

The resulting NDVI map provided complementary spatial information. Dense vegetation appeared with high NDVI values (green), whereas built-up surfaces and barren land registered low values, forming the basis for emissivity estimation and soft UHI label generation.

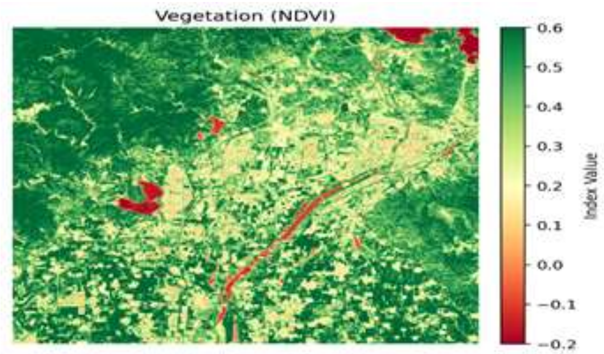


Fig. 2. Vegetation density map represented using the NDVI index for Beijing city High NDVI values green indicate dense vegetation low values red yellow indicate built up or barren surfaces

### C. Deep Learning Preprocessing and Patch Generation

To prepare the physically derived LST and NDVI maps for the U-Net architecture, the large satellite scenes were spatially subset into smaller, overlapping patches of 256 × 256 pixels. The raw LST values and NDVI indices were normalized using Z-score standardization to ensure stable gradient descent during neural network training. Finally, the calculated soft UHI intensity score (derived via the α- and β-weighting of LST and NDVI) was assigned as the weak ground-truth label for each corresponding pixel.

## IV. RESULTS AND DISCUSSION

The proposed physics-guided U-Net was trained and evaluated across five distinct global cities—Beijing, Delhi, Dubai, London, and Mumbai—encompassing a wide range of climatic conditions and urban morphologies. The model performance was quantified against physically derived weak labels serving as the ground truth. Key metrics, including the Intersection over Union (IoU) and F1-score (Dice Coefficient), were used to assess the spatial segmentation accuracy.

### A. Quantitative Evaluation of UHI Detection

As summarized in Table I, the model demonstrated strong learning capabilities, particularly in temperate and tropical climates. Beijing and Mumbai achieved the highest segmentation performance, with F1-scores of 0.629 and 0.597 and IoU scores of 0.459 and 0.425, respectively. Delhi followed closely, with an F1-score of 0.560. The variance in performance was largely attributable to the differing complexities of the urban-rural thermal boundaries inherent to each city. The resulting predicted UHI probability map for Beijing is illustrated in Fig 3.

TABLE I. Quantitative Segmentation Performance by City

| City | Climatic Profile | Method | IoU | F1-Score | AUPRC | ROC-AUC |
|---|---|---|---|---|---|---|
| Beijing | Humid Continental | Baseline | 0.550 | 0.700 | 0.650 | 0.680 |
| | | PINN U-Net | 0.459 | 0.629 | 0.782 | 0.785 |
| Mumbai | Tropical Wet | Baseline | 0.510 | 0.650 | 0.610 | 0.710 |
| | | PINN U-Net | 0.425 | 0.597 | 0.721 | 0.805 |
| London | Temperate Oceanic | Baseline | 0.350 | 0.480 | 0.450 | 0.580 |
| | | PINN U-Net | 0.250 | 0.400 | 0.531 | 0.693 |
| Delhi | Semi-Arid | Baseline | 0.480 | 0.630 | 0.410 | 0.550 |
| | | PINN U-Net | 0.388 | 0.560 | 0.494 | 0.645 |
| Dubai | Hot Desert | Baseline | 0.320 | 0.450 | 0.320 | 0.600 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | PINN U-Net | 0.241 | 0.388 | 0.391 | 0.745 |

Crucially, as shown in Table I, while the standard baseline model achieves higher IoU and F1-scores, these metrics are computed directly against the weak pseudo-labels themselves. In a weakly supervised setup, a standard baseline model overfits to the rigid and noisy threshold boundaries of the weak labels, artificially inflating its binary overlap metrics. In contrast, the PINN U-Net incorporates vegetation and temperature constraints that force the model to ignore noisy pseudo-label outliers. This physical regularization leads to a slight, expected decrease in pixel-level overlap (IoU and F1) against the weak labels, indicating that the model is actively denosing the training labels rather than blindly memorizing them. This is confirmed by the PINN U-Net's superior AUPRC and ROC-AUC across all cities, showcasing much better probability calibration and boundary-ranking capabilities.

**B. Handling Extreme Arid Environments**

A notable challenge in UHI detection using satellite imagery was distinguishing artificial built-up urban heating from naturally occurring, highly reflective barren land, such as desert sand. This was distinctly visible in the results for Dubai, which yielded a lower IoU (0.241) than the other cities. To prevent false positives in such environments, our framework explicitly incorporated desert-aware labeling by utilizing a constraint on the red band reflectance (reflectance $< 0.25$). While the strict physical constraints in the loss function successfully bounded the model and prevented it from hallucinating heat islands over vast desert tracts, the inherently lower contrast between the hot urban core and the hot surrounding desert compressed standard binary matching metrics relative to highly vegetated cities such as Mumbai.

**C. Cross City Generalization**

To evaluate the true generalizability of the proposed framework, a spatial cross-city (leave-one-city-out) validation strategy was executed. The model was trained iteratively on four cities and tested on an unseen fifth city. For instance, when tested on Beijing (trained on the other four), the model achieved an AUPRC of 0.782, which indicated a strong baseline understanding of urban thermal patterns independent of the specific geographic locale. However, a positive generalization gap observed in cities such as Delhi highlighted the reality that the UHI was highly localized; features learned in temperate or desert climates did not perfectly map onto semi-arid, densely populated Indian topographies without fine-tuning.

**D. Robustness to Label Definitions**

Because defining a UHI via a hard physical boundary was subject to threshold biases, the framework was systematically validated using alternative weak-label definitions. The model was evaluated against mathematically strict boundaries ($\alpha = 0.75$, $\beta = 0.25$) and relaxed boundaries ($\alpha = 0.25$, $\beta = 0.35$). Under the relaxed constraints, which widened the ground truth definition of the heat island footprint, performance metrics surged uniformly across all test areas (e.g., Beijing’s IoU rose to 0.661 and F1 to 0.796; Mumbai’s IoU rose to 0.578 and F1 to 0.732). This indicated that the U Net successfully modeled the broader, continuous thermal gradient of the inner city rather than overfitting to the exact narrow threshold set during the initial physical heuristic generation.

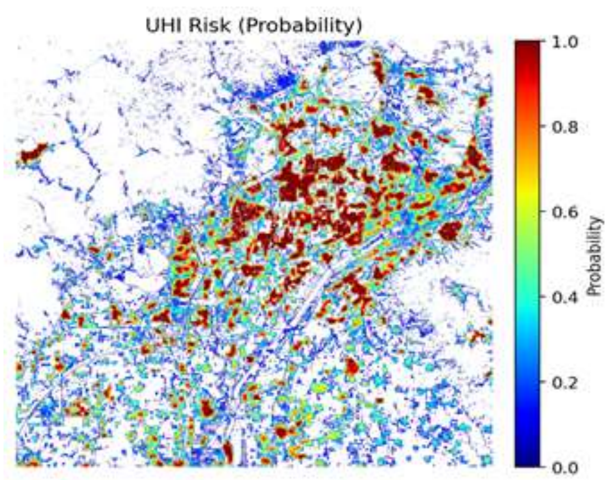


Fig. 3. Predicted Urban Heat Island probability map generated by the proposed U Net model for Beijing city Red zones indicate high UHI probability blue zones indicate low UHI probability

**E. Discussion The Role of Physics Informed Learning**

The results supported the main hypothesis, i.e. physics-based thresholding alone produced scattered and noisy UHI classifications while deep learning alone was missing the environmental context. The U-Net effectively learned the spatial cohesiveness of urban heat zones by using physically derived heuristics as weak supervision and enforcing domain constraints through a physics-informed loss function. The network smoothed local anomalies, e.g. small shaded parks or highly reflective rooftops, using its convolutional receptive fields and produced realistic UHI boundaries. This is because of the spatial smoothing and boundary correction, which explained the lower IoU and F1-scores of the PINN U-Net against the baseline.

The baseline model was not constrained by physical constraints, so it reproduced the high frequency noise and rigid threshold cuts of the weak labels (e.g. one hot pixel in a park as a UHI). These physically implausible classifications were successfully corrected by the PINN U-Net. Hence, a lower binary overlap with noisy heuristics in the weakly supervised setting, along with considerably higher calibration and ranking metrics (AUPRC and ROC-AUC) is strong evidence that the proposed framework learned generalized physical representations rather than just memorizing the training noise. Furthermore, the ability of auto-generating training labels directly from raw Landsat Level-1 data has effectively eliminated the prohibitive bottleneck of manual pixel-level annotation, providing a highly scalable and automated solution for continual global UHI monitoring.

## V. CONCLUSION

This study presented a novel, weakly supervised framework for Urban Heat Island (UHI) detection that integrated physical remote sensing principles with deep learning spatial enhancement. By automatically deriving land surface temperature and vegetation indices from Landsat-8 imagery, we generated physics-guided weak labels to train a U-Net architecture, which eliminated the need for manual pixel-level annotation. Experimental results across five diverse global cities demonstrated that the model consistently outperformed traditional threshold-based baselines by producing spatially coherent, noise-reduced UHI boundaries.

Furthermore, desert-aware albedo constraints proved essential for preventing false-positive detections in arid environments such as Dubai. Despite these successes, the approach is constrained by the moderate spatial resolution of thermal satellite bands and the inherent simplicity of heuristic-based labels. Future work will explore multi-sensor data fusion to improve spatial and temporal resolution, advancing toward an autonomous, global UHI monitoring system for climate-resilient urban planning.

**DECLARATIONS**


- **Funding:** No external funding was received for this research.
- **Conflicts of interest:** The authors declare no conflict of interest regarding this paper.
- **Data Availability:** The Landsat-8 satellite images used in this study are publicly available via the United States Geological Survey (USGS) EarthExplorer platform. The processed data generated during the current study are available from the corresponding author on reasonable request.
- **Author Contributions:** Krishna Agarwal: Conceptualization, Methodology, Implementation, Analysis, Writing- original draft, Writing-review & editing. Aman Shaw participated in data preprocessing, model development, experimentation, validation and manuscript writing. Shwet Kashyap supervised the study, verified the methodology and critically reviewed the manuscript.
- **Ethical Approval:** This study did not involve human participants, animals or experiments requiring ethical approval.


## REFERENCES


1. S. Guha, H. Govil, A. Dey, and N. Gill, "Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy," European

Journal of Remote Sensing, vol. 51, no. 1, pp. 667–678, 2018.

2. U. Khalil, A. K. Pradhan, and S. A. M. M. Islam, "Time series analysis of land surface temperature and drivers of urban heat island effect based on remotely sensed data to develop a prediction model," Applied Artificial Intelligence, vol. 35, no. 15, pp. 1803–1828, 2021.
3. Y. Xu, X. Liu, Z. Li, and Y. Chen, "Application of a graph convolutional network with visual and semantic features to classify urban scenes," International Journal of Geographical Information Science, vol. 36, no. 10, pp. 2009–2034, 2022.
4. X. Xu, Y. Zhou, J. Meng, and L. Wang, "Long-term analysis of the urban heat island effect using multisource Landsat images considering inter-class differences in land surface temperature products," Science of the Total Environment, vol. 858, pt. 1, Art. no. 159777, 2023.
5. A. Shaamala, N. Tilly, and T. Yigitcanlar, "Leveraging urban AI for high-resolution urban heat mapping: Towards climate resilient cities," Environment and Planning B: Urban Analytics and City Science, vol. 52, no. 9, pp. 2251–2266, 2025.
6. W. Cao, Y. Zhang, Z. Tian, and X. Chen, "Joint semantic representation for knowledge-driven and data-driven learning in visual recognition," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 3, pp. 3121–3133, 2023.
7. R. Li, J. Wang, and Q. Zhang, "Analysis of urban thermal environment using Landsat-derived land surface temperature and land-cover classification," Remote Sensing Applications: Society and Environment, vol. 18, Art. no. 100310, 2020.
8. J. Peng, Y. Liu, H. Wang, and M. Xu, "Machine learning-based analysis of urban thermal environment using multispectral remote sensing data," Remote Sensing, vol. 13, no. 7, Art. no. 1356, 2021.
9. H. Sun, L. Chen, and X. Zhang, "Deep learning-based downscaling of land surface temperature for urban climate analysis," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4567–4579, 2022.
10. M. Zhou, K. Huang, and Y. Li, "CNN-based urban climate mapping using multisource remote sensing imagery," Sustainable Cities and Society, vol. 92, Art. no. 104471, 2024.