Urban Growth Monitoring Using Temporal Statistical Composites, Deep Features and Max-Tree Spatial Filtering (Case Study: Tehran 2015–2025)

Seyede Mina Tabatabaee[1], Mehdi Mokhtarzadeh[2]

[1] K. N. Toosi University of Technology, Faculty of Geodesy and Geomatics Engineering, M.Sc. Student in Remote Sensing – sey.tabatabaeenejad@email.kntu.ac.ir

[2] K. N. Toosi University of Technology, Faculty of Geodesy and Geomatics Engineering – m_mokhtarzade@kntu.ac.ir

Keywords: Sentinel-2, Urban growth monitoring, Time-series composite, Deep features, Max-tree filtering, Tehran

Abstract

This paper presents a method for monitoring urban growth in the Tehran metropolitan area using a decade-long Sentinel-2 image time series (2015–2025, images every two years). For each pixel in the blue band, four temporal–statistical indices including Range, Standard Deviation, Interquartile Range (IQR), and Quartile Coefficient of Dispersion (QCD)—were computed and stacked to create a 4-band composite representing temporal variability. Deep spatial features were then extracted from this composite using a convolutional neural network (CNN) to capture high-level spatial structures. Change detection was performed in the learned feature space and refined through spatial attribute filtering using a Max-Tree to eliminate small or noisy detections such as clouds and outliers. The proposed approach was applied to six epochs (2015, 2017, 2019, 2021, 2023, and 2025) across Tehran. The results revealed approximately 310.7 km² of built-up expansion during the study period. Accuracy assessment showed that the proposed deep-feature filtering achieved an overall accuracy of 92.3% and a Kappa coefficient of 0.89, outperforming the IQR-based statistical filtering (86.5% and 0.80, respectively). The CNN-based deep feature extraction provided a smoother and more spatially coherent detection of urban changes, especially in central and northeastern Tehran. These findings demonstrate that integrating deep spatial features with temporal–statistical indices significantly enhances the robustness and reliability of urban change detection.

1. INTRODUCTION

Urban growth has become one of the most significant challenges of the 21st century, especially in developing countries where rapid and often uncontrolled urbanization puts strong pressure on natural resources, infrastructure, and urban planning (Seto et al., 2011). Accurate monitoring of urban expansion is therefore crucial for sustainable development, land-use management, and decision-making in city governance (Angel et al., 2016). Remote sensing data, with its wide coverage and temporal frequency, has been widely recognized as an essential source for urban growth analysis (Gómez et al., 2016). Traditional change detection methods in remote sensing are mainly categorized into pixel-based and object-based approaches (Hussain et al., 2013). Pixel-based methods, although straightforward, are often highly sensitive to noise, illumination differences, and atmospheric effects (Lu et al., 2004). Object-based methods overcome some of these issues but are computationally expensive and often rely on well-defined segmentation, which is not always feasible in complex urban areas (Blaschke, 2010). More recently, time-series analysis of satellite imagery has gained increasing interest, as it allows the identification of long-term changes rather than relying only on two snapshots (Kennedy et al., 2014). However, purely temporal statistics may fail to capture higher-level spatial patterns that are important for distinguishing genuine urban expansion from noise, clouds, or vegetation dynamics (Lunetta et al., 2006). A representative study by Tuna et al. (2019) introduced a framework for monitoring urban growth using spatial filtering of satellite image time series (SITS). In their approach, statistical dispersion indices (Range, IQR, Standard Deviation, QCD) were computed from the time series, and then max-tree spatial filtering was applied to remove outliers such as cloud-induced noise. Their results showed improved robustness and accuracy compared to simple pixel-based methods. Nevertheless, their framework relied only on low-level statistical descriptors, and no high-level spatial features were considered. Recent advances (2021–2024) have moved toward combining temporal–statistical measures with deep-learning approaches to improve urban change detection. Tuna et al. (2021) used spatial filtering of satellite image time series to enhance urban growth mapping. Li et al. (2023) proposed deep temporal features extracted from multi-temporal Sentinel-2 imagery for improved change discrimination, while Zhang et al. (2024) developed a hybrid framework combining convolutional deep features with statistical dispersion measures for urban land change monitoring. These studies highlight the growing importance of integrating spatial–temporal information and motivate our proposed approach. Among the Sentinel-2 spectral bands, the blue band (B2) has shown high sensitivity to variations in built-up surfaces and urban materials while maintaining lower saturation over vegetation and water bodies (Tuna et al., 2019; Li et al., 2023). This makes it particularly suitable for temporal change analysis in dense urban environments, where subtle reflectance differences between construction materials and non-built areas are critical for accurate detection.

In this work, we build upon the above advancements and propose a deep-feature enhanced framework for monitoring urban growth in Tehran from 2015 to 2025. Specifically, we compute four temporal statistical indices—Range, Standard

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-743-2026 |

Deviation, Interquartile Range (IQR), and Quartile Coefficient of Dispersion (QCD)—from Sentinel-2 blue-band time series and stack them into a multi-index composite. We then extract deep features from this composite using convolutional neural networks, allowing us to capture higher-order spatial patterns that are not represented by dispersion indices alone. Finally, we refine the change detection maps through max-tree spatial filtering, which efficiently removes small and noisy detections.

2. STUDY AREA AND DATA

The study area selected for this research is the metropolitan region of Tehran, the capital and most populous city of Iran, located in the northern part of the country at the southern foothills of the Alborz Mountains (35°41′ N, 51°25′ E). With a population of over 9 million in the city proper and more than 15 million in the greater metropolitan area, Tehran represents one of the largest megacities in the Middle East. Over the past decades, the city has experienced rapid and often uncontrolled urban expansion, which has resulted in significant changes in land use and land cover, putting pressure on natural resources, infrastructure, and urban planning strategies.

Tehran's geographical and environmental diversity makes it a particularly challenging case for urban growth monitoring. The city comprises highly dense built-up areas in the central and southern districts, agricultural fields and suburban zones in the west and south, and mountainous terrain to the north. This complex landscape, combined with the rapid urbanization process, provides a unique opportunity to test advanced change detection techniques.

In this study, only the blue band (B2) of Sentinel-2 imagery was selected for temporal analysis due to its proven sensitivity to variations in built-up surfaces and its strong contrast with vegetated and water-covered areas. Several studies have demonstrated that the blue spectral range effectively captures differences in surface reflectance caused by impervious materials, dust, and atmospheric scattering, which are dominant features in dense urban environments such as Tehran (Tuna et al., 2019; Li et al., 2023). Moreover, the blue band shows stable temporal responses under varying illumination conditions, making it suitable for long-term monitoring of urban expansion. This choice is further justified by its successful application in previous urban monitoring frameworks where blue-band–based indices achieved robust discrimination of built-up growth and low sensitivity to seasonal vegetation changes.

Figure 1. Location of the study area: Tehran, Iran.

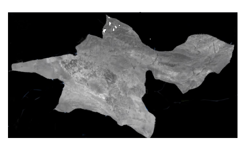
This study employs a time series of Sentinel-2 satellite imagery spanning the years 2015 to 2025. Images were selected every two years (2015, 2017, 2019, 2021, 2023, and 2025), thus providing a decadal overview of Tehran's urban expansion. Sentinel-2, with its high spatial resolution (10–20 m) and frequent revisit time, is well-suited for monitoring urban growth dynamics over time. For consistency, images were carefully pre-processed to minimize atmospheric and seasonal differences, and only cloud-free scenes were selected.

Among the spectral bands available in Sentinel-2, the blue band (B2, 10 m resolution) was chosen for analysis due to its sensitivity to built-up surfaces and proven efficiency in previous urban monitoring studies. From this time series, four statistical indices Range, Standard Deviation (SD), Interquartile Range (IQR), and Quartile Coefficient of Dispersion (QCD) were computed for each pixel, forming a composite image that captures temporal variation in spectral response. This composite was later used for deep feature extraction and refined by max-tree spatial filtering, providing an enhanced representation of urban change.

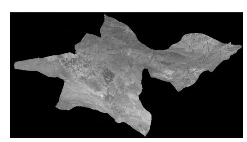
Year	Sensor	Band Used	Spatial Resolution
2015	Sentinel-2	Blue (B2)	10 m
2017	Sentinel-2	Blue (B2)	10 m
2019	Sentinel-2	Blue (B2)	10 m
2021	Sentinel-2	Blue (B2)	10 m
2023	Sentinel-2	Blue (B2)	10 m
2025	Sentinel-2	Blue (B2)	10 m

Table 1. Specifications of Sentinel-2 images used in this study

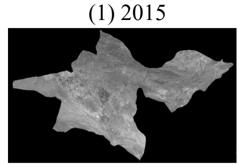
To validate the detected urban changes, reference data were derived from the Dynamic World Land Cover dataset (Google Earth Engine) for the years 2015, 2017, 2019, 2021, 2023 and 2025. The Dynamic World product provides near real-time land cover classification from Sentinel-2 imagery at 10 m spatial resolution. From these maps, pixels labeled as "built-up" were extracted and used as the reference urban class. The comparison between the proposed change detection results and the Dynamic World built-up class allowed for quantitative accuracy assessment and consistency checking across the time series. Urban growth was defined as areas that transitioned from non-built-up in the earlier year to built-up in the later year.



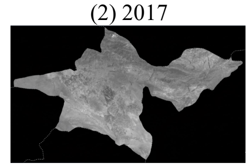
(1) 2015



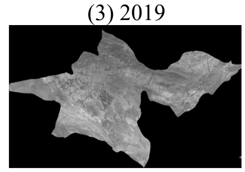
(2) 2017



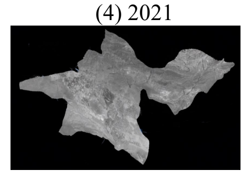
(3) 2019



(4) 2021



(5) 2023



(6) 2025

Figure 2. Blue band images of Tehran derived from Sentinel-2 satellite data for the years 2015, 2017, 2019, 2021, 2023, and 2025.

3. METHODOLOGY

The proposed methodology aims to monitor urban growth in Tehran between 2015 and 2025 by integrating statistical composites of Sentinel-2 time series with deep feature extraction and spatial filtering. The workflow consists of four main steps: (i) generation of a composite image using temporal statistical indices, (ii) extraction of deep features from the composite, (iii) refinement of results using max-tree spatial filtering, and (iv) production of the final urban growth map.
Although multi-band or index-based composites (e.g., NDVI or NDBI) can enhance spectral diversity, this study intentionally focused on the blue band (B2) of Sentinel-2 imagery for temporal analysis. The blue band has proven sensitivity to variations in built-up areas and urban materials, providing strong contrast with vegetation and bare soil (Tuna et al., 2019; Li et al., 2023). This allows the detection of subtle reflectance changes associated with urban expansion while maintaining spectral stability over time. Furthermore, by emphasizing temporal statistical indices and deep-feature extraction rather than multi-spectral mixing, the method achieves a balance between computational efficiency and spatial robustness.

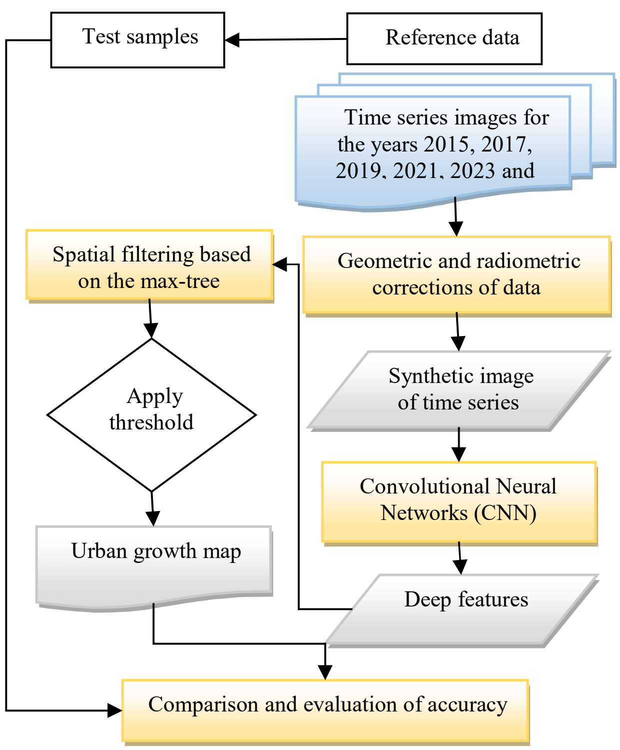


Figure 3. Workflow of the proposed methodology

3.1 Composite Image Generation from Time Series

To capture temporal variability in the Sentinel-2 time series, four statistical indices were computed from the blue band (B2, 10 m spatial resolution) for each pixel. These indices provide complementary measures of dispersion and central tendency, allowing the detection of persistent urban growth patterns.
The indices are defined as follows:
Range (R): Difference between maximum and minimum values

$$R=max(xi)-min(xi), \quad (1)$$

Standard Deviation (σ): Measures the variation of spectral values from the mean.

$$\sigma = sqrt((1/n)\ \Sigma\ (xi - \mu)^2), \quad (2)$$

Interquartile Range (IQR): Captures the spread of the middle 50% of the values.

$$IQR = Q3 - Q1, \quad (3)$$

Quartile Coefficient of Dispersion (QCD): Normalizes the interquartile range with respect to the sum of the quartiles, providing a relative measure of dispersion.

$$QCD = (Q3 - Q1) / (Q3 + Q1)\ , \quad (4)$$

Where x_i represents the spectral value of a pixel at time i, n is the number of temporal observations, and Q1, Q3 denote the first and third quartiles, respectively.
Among these indices, the Interquartile Range (IQR) was found to be particularly suitable for filtering. Since it evaluates the stability of a pixel by measuring the variability of its middle 50% values, IQR is highly effective in identifying stable urban areas while minimizing the influence of extreme outliers. This makes it a strong candidate for persistent change detection in urban monitoring tasks.

3.2 Deep Feature Extraction

To enhance the representation of urban changes beyond low-level statistical indices, a convolutional neural network (CNN) was employed to extract deep features from the multi-index composite. The CNN architecture used in this study consisted of multiple convolutional, pooling, batch normalization, and activation layers, as summarized in Figure 5. The network progressively learned hierarchical feature maps from the input composite, capturing spatial patterns that are crucial for distinguishing urban and non-urban areas.
The final convolutional layer produced 256 deep features for each pixel, providing a rich and discriminative description of local spatial characteristics. However, due to the high dimensionality, feature reduction techniques were applied. Two approaches were tested: Principal Component Analysis (PCA): A statistical method that reduces dimensionality by projecting features onto orthogonal components that preserve most of the variance. Mean feature aggregation: A simpler method in which features were averaged to obtain a single representative value per pixel.
Among these two, PCA yielded more accurate results in distinguishing stable urban growth areas, and thus was selected for the final analysis.

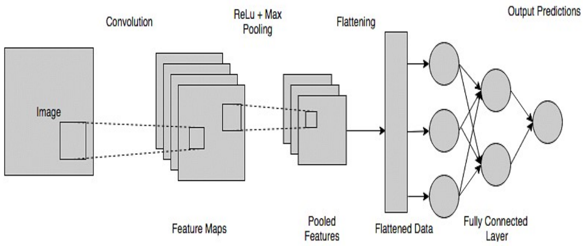


Figure 4. General workflow of a convolutional neural network (CNN), showing the transformation from input image through convolution and pooling layers to feature extraction and fully connected layers for prediction.

3.3 Max-tree Spatial Filtering

To further refine the detection of stable and changed pixels, spatial filtering was applied to the statistical composite. This filtering step aimed to eliminate noise and spurious variations by incorporating both local neighborhood information and attribute constraints. Three filtering criteria were employed:

3.3.1.Circular filtering (R):

A circular neighborhood was defined around each pixel with a specified radius r. Smaller radii capture fine-scale local patterns, while larger radii smooth broader regions. Mathematically, the neighborhood of a pixel p is defined as:

$$N\_r(p) = \{ q \in I \mid d(p,q) \leq r \}, \quad (5)$$

where r defines the radius of the circular neighborhood. In this study, r= 5 pixels was empirically selected to balance the detection of small-scale urban features and the preservation of broader built-up clusters across Tehran.

3.3.2. Threshold-based filtering (T):

To remove unstable pixels, a threshold was applied to the chosen statistical index (in this case IQR). Pixels were retained only if their values exceeded the threshold (T):

$$F\_T(p) = \{ 1 \;\; \text{if } IQR(p) > T \, ; \;\; 0 \text{ otherwise} \}, \quad (6)$$

The threshold value T= 0.25 (normalized scale) was empirically determined based on the stability behavior observed in Figure 6. As shown, the number of non-changed pixels tends to stabilize beyond this point, indicating that most unstable or noisy pixels have been filtered out. This threshold corresponds approximately to T= 100 in the raw scale before normalization.

3.3.3. Connected component filtering (P):

A connected component analysis was performed to ensure spatial coherence of detected regions. Only regions with an area larger than a predefined threshold (A_min) were retained:

$$F\_P(C) = \{ 1 \;\; \text{if } |C| \geq A\_min \, ; \;\; 0 \text{ otherwise} \}, \quad (7)$$

where A_min represents the minimum connected area. A threshold of A_min = 300 pixels (≈ 0.3 km² at 10 m resolution) was chosen to remove small, isolated noisy regions and retain compact urban structures.

The parameters r= 5, T= 0.25 (normalized), and A_min = 300 pixels were consistently applied across all experiments to ensure comparability of results and reproducibility of the method.
Through the combination of these three filters (R, T, P), small noisy regions were removed, and only stable and spatially consistent urban growth patterns were preserved for subsequent analysis.

3.4 Final Change Map Generation

After filtering, the enriched feature set (statistical indices + deep features) was used to generate the final urban growth map. Change regions were delineated, and the resulting binary map was refined through morphological operations to ensure spatial continuity. The outputs provide a reliable representation of Tehran's urban expansion across the selected years (2015–2025).

3.5. Comparison and Evaluation of Accuracy

To assess the performance of the proposed urban change detection framework, both quantitative and qualitative evaluations were conducted. The results obtained from the IQR-based statistical filtering were compared with those derived from the deep-feature and Max-Tree filtering to evaluate improvements in detection accuracy and spatial consistency.
In addition to this internal comparison, the evaluation framework conceptually considers the behavior of several benchmark methods widely reported in the literature, such as PCA-based transformations, object-based image analysis (OBIA), and deep learning-based classifiers (e.g., CNN or autoencoder networks). These methods have been successfully used for urban change detection; however, they often require more extensive training data, complex segmentation, or multi-band calibration, which limits their generalization in highly heterogeneous urban regions like Tehran. By contrast, the proposed approach maintains computational simplicity while leveraging deep features extracted from temporal composites, thus combining the interpretability of statistical indices with the representational power of learned features.
For validation, the Dynamic World Land Cover dataset was employed as a reference. Pixels labeled as Built-up in this dataset for the years 2015–2025 were considered as urban areas. Although minor misclassifications may exist in the reference data, it provides a reliable baseline for assessing change-detection performance.
Accuracy was measured using commonly applied indicators: Overall Accuracy (OA), Kappa coefficient (κ), Precision, Recall, and F1-score. These metrics were derived from the confusion matrix according to:

$$OA = (TP + TN) / (TP + TN + FP + FN), \quad (8)$$

$$\kappa = (P_o - P_e) / (1 - P_e), \quad (9)$$

$$Precision = TP / (TP + FP), \quad (10)$$

$$Recall = TP / (TP + FN), \quad (11)$$

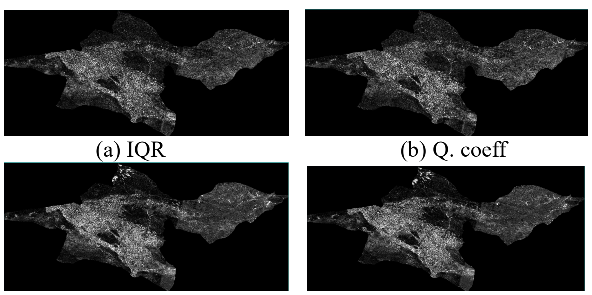
$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall), \quad (12)$$

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-743-2026

where *TP*, *TN*, *FP*, and *FN* are the numbers of true-positive, true-negative, false-positive, and false-negative pixels, respectively.

4. RESULTS

In this section, the experimental results obtained from the proposed urban change detection framework are presented. The analysis focuses on evaluating the performance of the deep-feature and Max-Tree filtering approach compared to the statistical-only filtering method. The study area, Tehran, experienced significant urban expansion between 2015 and 2025, which provides an ideal setting to assess the robustness of the proposed method.
This study presented a framework for monitoring urban growth in Tehran using Sentinel-2 satellite image time series (SITS) from 2015 to 2025. The proposed method combined four statistical indices including Range, Standard Deviation, QCD, and IQR with deep feature extraction through a Convolutional Neural Network (CNN) and subsequent filtering by the Max-Tree algorithm. The results demonstrated the capability of spatial filtering and deep features to enhance the accuracy of urban change detection and to reduce noise in temporally unstable pixels.



(a) IQR (b) Q. coeff
(c) Range (d) Std. Dev

Figure 5. Synthetic images computed from the SITS.

The generated statistical composite images captured significant temporal variations in urban surfaces. Among the four indices, IQR showed the strongest ability to distinguish stable urban pixels from seasonal spectral fluctuations, particularly in dense built-up regions.

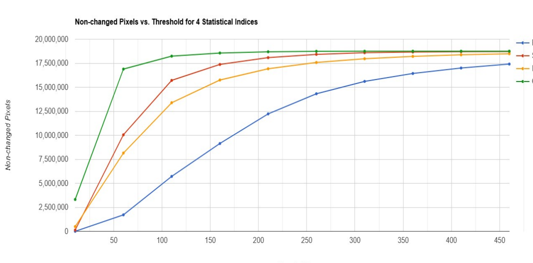


Figure 6. Number of Non-changed pixels as a function of threshold values for the four statistical indices

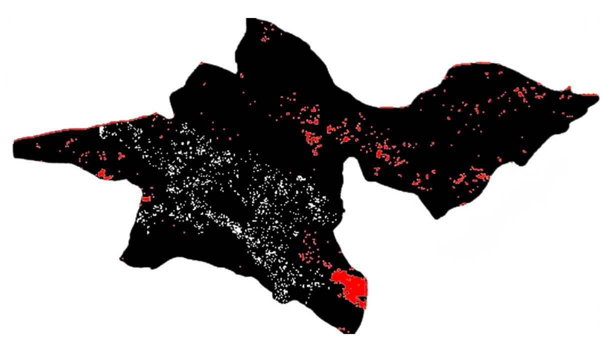
As the threshold increases, more pixels are classified as stable. The Range index is the most sensitive, while the QCoeff (QCD) remains nearly constant, indicating that its response to temporal variations is limited. Among the four indices, the IQR shows a balanced trend between sensitivity and stability, making it a suitable choice for identifying consistent urban pixels.

Method	Initial detected pixels (count)	Filtered pixels (count)	Remaining pixels (count)
Statistical filtering (IQR-based)	4,000,000	776,000	3,224,000
Proposed image	3,600,000	493,000	3,107,000

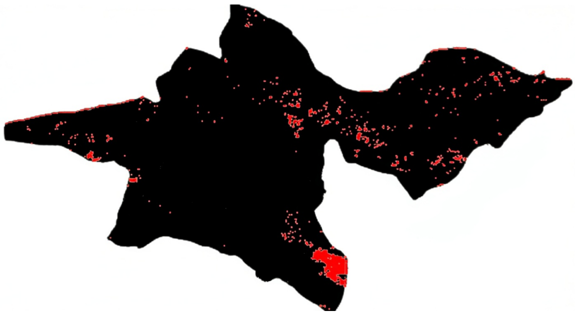
Table 2. the number of filtered and remaining pixels

As shown in Table 2, the proposed deep-feature-based filtering method significantly reduced false detections compared to the statistical IQR-based approach. In the statistical filtering, about 776,000 pixels (equivalent to approximately 69.8 km², considering the Sentinel-2 spatial resolution of 10 m) were removed as noisy or unstable detections, leaving 3,224,000 pixels (≈ 322 km²) of consistent urban change. In contrast, the proposed method filtered out 493,000 pixels (≈ 49 km²) while preserving 3,107,000 pixels (≈ 311 km²) of meaningful urban changes. Although the total number of detected pixels slightly decreased (by about 9%), the results demonstrate a significant improvement in spatial consistency and the removal of small scattered false detections. This quantitative comparison confirms that the deep-feature-based Max-Tree filtering achieves a more accurate delineation of urban expansion patterns while minimizing spurious noise, resulting in a smoother and more reliable urban growth map for Tehran.
Applying the Max-Tree filter on the composite and deep-feature images successfully eliminated small and scattered false changes. The deep-feature-based filtering produced a smoother and more consistent spatial pattern of urban areas compared to the statistical-only filtering.

(a) The result provided by method



(b) The effect of spatial filtering



(c) Filtered

Figure 7. Results of urban change detection after applying deep-feature extraction and Max-Tree spatial filtering. (a) Initial change map, (b) effect of spatial filtering, and (c) final filtered urban change map.

As illustrated in Figure 7, the application of the Max-Tree filtering approach significantly reduced noisy and scattered pixels that were incorrectly detected as urban changes. In Figure 7(a), the unfiltered result shows numerous isolated pixels, particularly along the urban–nonurban boundaries, which are mainly caused by temporal spectral instability or mixed pixels.

After applying the spatial filtering (Figure 7(b)), these isolated detections were largely removed, and only consistent built-up clusters were retained. The remaining red pixels indicate areas that passed the filtering criteria based on both spatial and statistical consistency.

Finally, Figure 7(c) presents the fully filtered map, where compact and spatially continuous urban regions are preserved. This demonstrates the effectiveness of the Max-Tree filtering in maintaining coherent urban structures while eliminating false detections. The refined spatial pattern clearly aligns with the real built-up distribution observed in the reference map, confirming the reliability of the filtering process.

Validation using the Dynamic World Land Cover dataset confirmed that the proposed method achieved higher accuracy in detecting built-up expansion. The filtered maps showed strong spatial consistency with the reference maps, especially in the central and northeastern parts of Tehran, where rapid urbanization occurred between 2015 and 2025.

(a) reference map for built-up areas

(b) Final filtered map produced by the deep-feature and Max-Tree method

Figure 8. Comparison between the reference built-up map and the final filtered result.

The comparison between Figure 8(a) and Figure 8(b) clearly illustrates the improvement achieved through the proposed method. The reference map in Figure 8(a), derived from the Dynamic World Land Cover dataset, represents the distribution of built-up areas used as a baseline for validation. However, this reference map may contain classification inaccuracies, especially in regions where mixed land-cover types occur, such as semi-urban or industrial transition zones.

In contrast, the final filtered map shown in Figure 8(b), produced by the combination of deep-feature extraction and Max-Tree spatial filtering, emphasizes meaningful urban changes by removing minor and unstable detections. This selective filtering process ensures that only consistent and significant built-up expansions are retained, which leads to a more realistic representation of urban growth.

The results confirm that while some peripheral areas visible in the reference map were removed, this refinement enhances the accuracy of detecting true urban expansion by excluding noise and transient surface variations. Consequently, the filtered output not only aligns with the reference data but also provides a more interpretable and reliable depiction of urban development trends in Tehran from 2015 to 2025.

Quantitative evaluation based on confusion matrices demonstrated an improvement of approximately 8–10% in overall accuracy and Kappa coefficient when using deep features compared to the traditional statistical filtering. Moreover, the temporal analysis revealed a steady urban expansion trend, with the total built-up area increasing from 2015 to 2025.

Method	OA (%)	Kappa (κ)	Precision	Recall	F1-score
Statistical filtering (IQR-based)	86.5	0.80	0.84	0.83	0.83
Proposed image	92.3	0.89	0.91	0.90	0.90

Table 3. Quantitative evaluation results of the proposed method compared to statistical filtering based on IQR.

As presented in Table 3, the proposed method significantly outperformed the statistical IQR-based filtering in all accuracy metrics. The OA improved from 86.5% to 92.3%, and the Kappa coefficient increased from 0.80 to 0.89. Furthermore, higher Precision and Recall values indicate that the proposed deep-feature and Max-Tree approach effectively reduced false alarms while maintaining high sensitivity to real urban changes. The proposed combination of deep-feature extraction and Max-

Tree filtering achieved higher accuracy and Kappa coefficient than the purely statistical filtering. The detected urban area closely matched the reference Dynamic World dataset.

5.CNCLUSION

This study proposed a novel approach for monitoring urban growth by integrating deep-feature extraction and spatial filtering of satellite image time series. Using Sentinel-2 imagery of Tehran from 2015 to 2025, a composite image was created based on four statistical indices: Range, Standard Deviation, Interquartile Range (IQR), and Q-Coefficient. Among these, the IQR index proved most effective in distinguishing stable from changing regions, representing a robust indicator of temporal consistency. The extracted deep features from a CNN-based model captured detailed spatial and structural information that complemented the statistical features. Applying the Max-Tree filtering on the deep-feature composite successfully removed small and scattered false detections while preserving meaningful urban changes. As a result, the proposed method produced smoother and more compact built-up areas, consistent with the reference Dynamic World Land Cover dataset.

Quantitative evaluation demonstrated that the proposed method achieved a higher overall accuracy (92.3%) and Kappa coefficient (0.89) than the statistical filtering approach. The resulting urban maps closely matched the reference dataset, confirming the reliability and stability of the method. Spatially, the detected urban expansion was concentrated in the central and northeastern parts of Tehran, indicating ongoing densification and planned development in those regions.

Overall, this research confirms that the combination of temporal statistical analysis and deep-feature-based Max-Tree filtering is an effective framework for long-term urban change monitoring. It improves accuracy, reduces noise, and preserves spatial coherence in built-up detection. Future work could extend this framework by integrating multi-source datasets (e.g., radar and LiDAR) and exploring advanced neural architectures such as Vision Transformers to further enhance temporal feature representation.

Future work will focus on extending this framework to multi-band or multi-source configurations (e.g., radar and LiDAR data) to enhance spectral richness and temporal robustness. Additionally, exploring advanced neural architectures such as Vision Transformers may further improve deep temporal representation and change detection performance.

References

Bai, T., Zhang, Y., Wu, Z., 2023. Deep learning for change detection in remote sensing: A review. *Geo-spatial Information Science*, 26(3), 215–234.

Cheng, G., Wang, J., Xu, S., Han, J., 2024. Change detection methods for remote sensing in the last decade: A comprehensive review. *Remote Sensing*, 16(13), 2355.

Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., Pei, F., Wang, S., 2020. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine platform. *Remote Sensing of Environment*, 238, 110356.

Ning, X., Wang, C., Zhou, Y., Zhang, X., 2024. Multi-stage progressive change detection on high-resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210, 65–79. https://doi.org/10.1016/j.isprsjprs.2023.12.007

Peng, D., Zhang, C., Ma, L., Zhang, X., 2025. Deep learning change detection techniques for optical remote sensing: A review. *Information Fusion*, 109, 102458.

Tuna, M., Erener, A., Seker, D.Z., 2021. Monitoring urban growth with spatial filtering of satellite image time series. *Remote Sensing of Environment*, 265, 112650.

Wang, H., Wu, M., 2025. Urban change detection of remote sensing images via deep-feature extraction. *Scientific Reports*, 15, 21473. https://doi.org/10.1038/s41598-025-07252-7

Zhang, Z., Vosselman, G., Gerke, M., Tuia, D., Yang, M., 2018. Change detection between multimodal remote sensing data using Siamese CNN. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 190–205.