Evidence-Based Inference and Quantification of Urban Expansion Using YOLOv8 and High-Resolution Satellite Imagery

Deepika Pahuja
Amity Institute of Information Technology, Amity University, India
deepika.pahuja@student.amity.edu (corresponding author)

Sarika Jain
Amity Institute of Information Technology, Amity University, India
sjain@amity.edu

Shishir Kumar
School of Information Science and Technology, Babasaheb Bhimrao Ambedkar University, India
shishir.cs@bbau.ac.in

Received: 19 May 2025 | Revised: 26 June 2025 and 9 July 2025 | Accepted: 11 July 2025

 DOI: https://doi.org/10.48084/etasr.12235

ABSTRACT

Urban expansion significantly impacts land use and ecological balance, necessitating efficient monitoring tools. This study introduces a deep learning-based approach for detecting urban growth in Greater Faridabad, Haryana, India, using high-resolution satellite imagery and the YOLOv8 object detection model. A custom dataset was created from Google Earth Pro imagery (2014-2024) and manually annotated to include key urban features such as buildings, roads, and construction areas. Images from 2015, 2020, and 2025 supported a temporal assessment of urbanisation trends. The YOLOv8m model exhibited strong performance, achieving a mean Average Precision (mAP) of 0.983 at IoU 0.5, 0.808 for mAP50–95, with 94.6% precision and 97.9% recall, indicating high detection accuracy with minimal false positives and negatives. Urban area grew from 26,837 px² in 2014 to 106,330 px² in 2020 (a 296% increase), and to 147,450 px² by 2025 - an additional 39%, resulting in a cumulative increase of 450%. These results reflect rapid urbanisation in the study area. The proposed method offers a reliable, scalable solution for automated urban change detection, supporting data-driven urban planning and sustainable development.

Keywords-urban expansion; LULC; Yolov8; Google Earth Pro; satellite imagery; deep learning

I. INTRODUCTION

Object detection plays a fundamental role in computer vision, enabling the automatic identification and localization of specific features within images. This technique has found applications across various fields, including precision agriculture [1], traffic surveillance [2], wildlife monitoring [3], and infrastructure assessment. However, when applied to satellite imagery, object detection faces distinct challenges, such as a top-down perspective, a large spatial coverage, and the complexity of urban structures. These challenges are further compounded in urban environments, where high object density, scale variability, and frequent occlusions make accurate detection more difficult [4-5].

Urbanization is one of the most transformative global trends of the 21st century. With more than half of the world's population now living in cities, the need to monitor urban growth has become increasingly urgent for sustainable development and effective planning. In India, this trend is particularly pronounced in peri-urban areas such as Greater Faridabad, located within the National Capital Region (NCR). During the last decade, this region has undergone rapid transformation from agricultural land to a densely built urban landscape, driven by infrastructure development, population growth, and its proximity to Delhi. Rapid urban growth in such areas presents critical challenges in spatial planning. Existing monitoring approaches often lack spatiotemporal depth or are too resource-intensive for frequent use. This study addresses these limitations by introducing a lightweight, scalable framework for accurate, long-term urban analysis using multi-temporal satellite imagery and the YOLOv8 model.

High-resolution satellite imagery from platforms such as Bhuvan, Sentinel-2, and Google Earth Pro has significantly improved the ability to observe and analyze such transformations. These sources make it possible to identify urban features such as roads, buildings, and construction sites

with considerable accuracy. However, the volume and frequency of data acquisition have made manual interpretation increasingly impractical, necessitating the development of automated intelligent systems capable of extracting information from satellite imagery efficiently and at scale.

Traditional classification approaches for satellite imagery, including both pixel-based and object-based methods, often rely on extensive manual feature engineering. These methods, while suitable for simpler or smaller-scale applications, typically struggle in complex urban settings. Early machine learning techniques such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (k-NN) were constrained by their dependence on handcrafted features and limited scalability [7, 8, 9, 10]. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), such as AlexNet, VGGNet, and ResNet, introduced a paradigm shift by enabling automated feature extraction, significantly improving classification performance [11]. However, these CNNs are more suitable for scene-level classification and less effective for detailed object localization.

To address this gap, object detection models such as Faster R-CNN and SSD were introduced, offering improved localization capabilities ideal for high-resolution imagery, although at the expense of speed and computational efficiency [12-13]. YOLO (You Only Look Once) models emerged as a compelling alternative by combining high detection speed with satisfactory accuracy. For example, YOLOv2 was effectively applied for detecting urban vehicle fleets from aerial imagery [14], and YOLOv3 and YOLOv4 gained popularity for real-time detection tasks involving UAV and satellite data [6, 15-17]. YOLOv5 has been used to detect urban trees in multi-temporal UAV imagery, while YOLT (You Only Look Twice) introduced object-level tracking of buildings over time from satellite images [18].

Despite these advances, earlier YOLO versions have shown limitations in detecting small or overlapping objects, especially in densely built areas. YOLOv8 overcomes many of these challenges through its anchor-free design and its enhanced ability to detect small-scale features. These characteristics make it especially suitable for analyzing temporal changes in complex urban environments. However, its application to multi-temporal, high-resolution satellite imagery remains relatively underexplored. This study aimed to address this gap by applying YOLOv8 to assess urban expansion in Greater Faridabad, India. Using high-resolution Google Earth Pro imagery from 2014 to 2025, the model was trained and validated to detect impervious surfaces, primarily residential and commercial structures. The detected features were translated into spatial units (in px²) and converted to georeferenced formats such as GeoJSON and shapefiles. These outputs were integrated into QGIS to enable spatial analysis and temporal comparisons.

The practical implications of this approach are substantial: it provides planners, environmental analysts, and policymakers with a scalable, automated framework to track urban sprawl, assess infrastructure development, and support data-driven land-use decisions in rapidly changing regions. By bridging deep learning detection outputs with geospatial analysis tools, this study aligns with recent efforts to develop integrated systems for urban monitoring using commercial satellites such as KhalifaSAT [9] and platforms such as Google Earth Engine [10]. Thus, this study not only advances technical innovation in urban analytics but also offers a transferable method for policy-driven spatial planning and sustainable development.

II. STUDY AREA

This study focuses on monitoring urban growth in Greater Faridabad, located approximately at 28°23'30.08" N and 77°20'48.48" E in India. Over the past 15 years, this region has undergone significant changes in land use and cover, largely due to substantial increases in construction activity and urban sprawl. This makes it a relevant and insightful case for examining the pattern of urban growth. The study area covers 2.80 Km^2, including both densely populated urban centers and rapidly transforming peri-urban zones. Figure 1 illustrates the spatial extent of the study area, highlighting its administrative limits and major landmarks.

Fig. 1. Satellite Image of Sec-81, Greater Faridabad, India (Google Earth Pro, 2025).

III. METHODOLOGY

This study leverages the YOLOv8 model, a deep learning-based object detection framework, to analyze urbanisation trends from high-resolution multi-temporal satellite imagery. Figure 2 illustrates the method for automated detection of urban features, which consists of five distinct stages: i) collection of satellite images, ii) Data annotation, iii) model training, iv) model evaluation, and v) inference and urban detection. A custom dataset of 50 satellite images (1024×768 px) was curated from Google Earth Pro for the period 2014-2024. The images were selected based on cloud-free and seasonally consistent conditions, with an intended five images per year, although some years had fewer due to availability. All images were resized to 640×640 px to meet the YOLOv8m input requirements.

A. Satellite Imagery Collection

To maintain temporal consistency, images showing seasonal or vegetation anomalies were removed from the dataset. All satellite imagery was captured at a spatial resolution of approximately 0.5 meters per pixel, enabling high-precision localization and delineation of urban features, including built infrastructures, road networks, and ancillary urban systems.

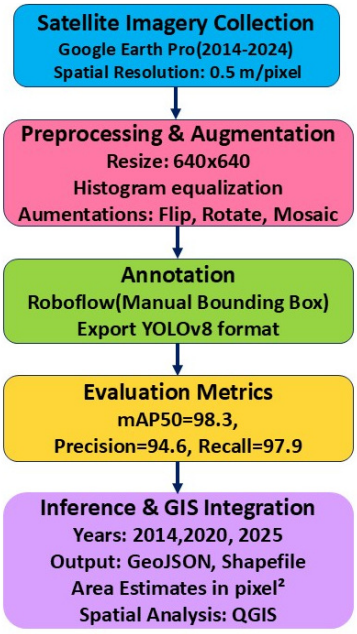

Fig. 2. Framework for urban feature extraction using YOLOv8 and satellite imagery.

B. Data Annotation

Before the training phase, the dataset was refined through a structured preprocessing pipeline and an annotation workflow. First, the images were rescaled to a consistent resolution of 640×640 px, which corresponds to the input requirements of the YOLOv8m model. Histogram equalization and contrast normalization were applied to improve image quality and optimize the clarity of urban features. To enrich the dataset and improve the robustness of the model, each original image was augmented 10 times using a variety of transformations, including horizontal flips to handle lateral orientation changes, random cropping with up to 20% zoom to simulate variations in spatial scale, and random rotations ranging from -37^0 to $+37^0$ to increase tolerance to different viewing angles. Additionally, mosaic augmentation was used, blending multiple images into a single image frame to add contextual and spatial diversity, helping the model learn from a broader variety of urban scenarios.

Annotation was performed manually using the Roboflow platform, which provided an intuitive interface for labelling and dataset organization. Residential buildings were manually annotated by drawing bounding boxes around each visible instance. Annotations were saved in a format compatible with YOLOv8, ensuring seamless integration into the object detection models. Figure 3 shows an example of the manual annotation process, where bounding boxes highlight residential structures. Roboflow's version control and preprocessing tools were also used to standardize image sizes and streamline augmentation. After annotation, the dataset was stratified into a ratio of 9:1 for training and validation splits, helping to maintain a diverse representation of spatial features and reduce sampling bias.

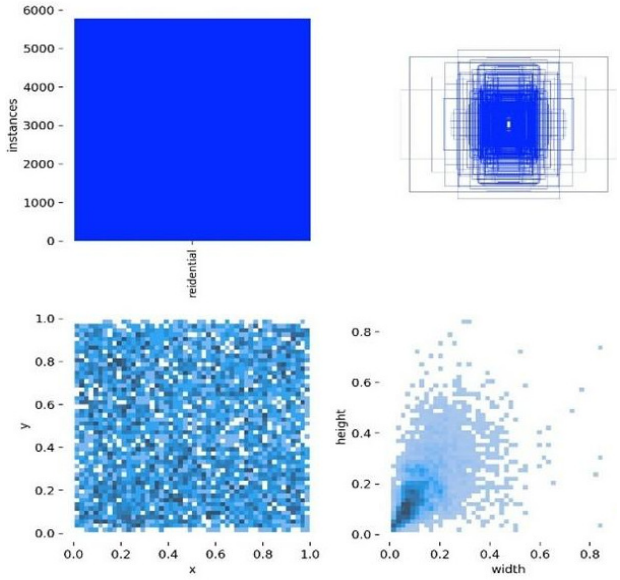

Fig. 3. Manual annotation specifications.

C. Model Training

The medium-scale YOLOv8m model was trained on annotated satellite imagery, targeting urban features, particularly residential areas. Training was conducted on a GPU-accelerated environment, with a NVIDIA Tesla T4 GPU (40 GB VRAM). The setup utilized the Ultralytics YOLOv8 framework (version 8.3.78) within a Python 3.11.11 environment, supported by PyTorch, torch vision 2.5.1, CUDA 11.2, and cuDNN 8.1.0. The model was trained for 30 epochs using a batch size of 16 and the Stochastic Gradient Descent (SGD) optimizer. The initial learning rate was set at 0.001.

Figure 4 presents a representative output from the validation phase, in which urban residential structures are accurately detected and highlighted using blue bounding boxes. These predictions demonstrate the model's ability to localize densely built-up areas within high-resolution imagery.

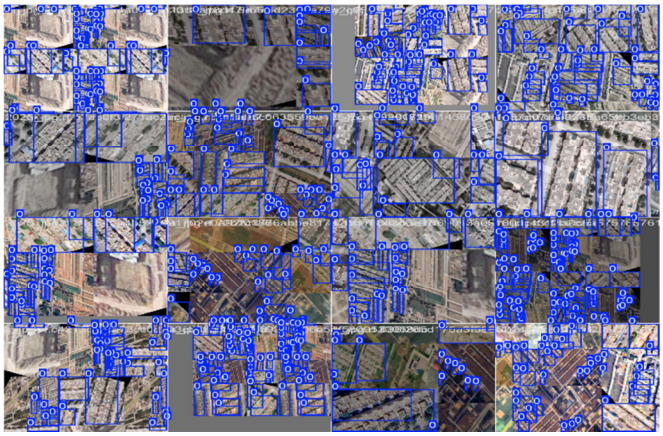

Fig. 4. Batch visualization of YOLOv8 model validation results on urban residential structures.

D. Model Evaluation

Figure 5 illustrates that the training and validation losses exhibited a consistent decline throughout training, while evaluation metrics such as mAP and recall improved steadily, indicating effective learning and strong generalization. The performance of the YOLOv8m model, evaluated on the validation dataset, demonstrated high detection accuracy across all metrics. Specifically, the model achieved a mean Average Precision (mAP) of 0.983 at an IoU threshold of 0.5 (mAP50) and 0.808 across the 0.5 to 0.95 IoU range (mAP50-95). It also attained a precision of 0.946 and a recall of 0.979, reflecting a high level of reliability in detecting urban features with minimal false positives and false negatives. These quantitative results highlight the robustness and applicability of the model for fine-grained urban feature extraction from high-resolution satellite imagery.

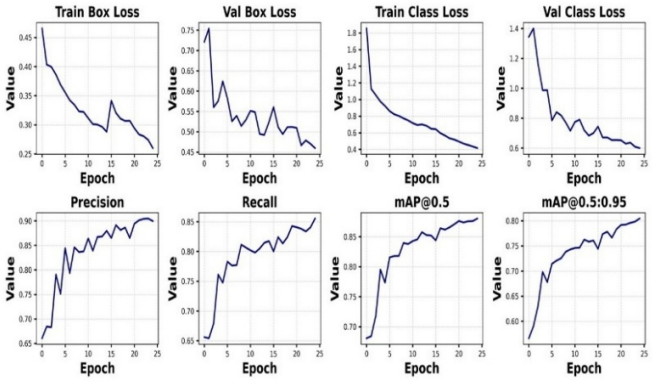

Fig. 5. Training and validation performance of YOLOv8 across 30 epochs.

Fig. 6. Validation results: Detected urban residential structures with YOLOv8.

To qualitatively assess the model's performance, sample validation images were reviewed with annotated bounding boxes. Figure 6 illustrates a sample validation image used in the evaluation phase of the object detection model. The annotated bounding boxes indicate residential areas in high-resolution satellite imagery. Manually reviewed annotations ensured accuracy by correcting minor labeling errors. The image showcases the model's precise localization of densely built-up residential zones.

E. Inference and Detection

The YOLOv8m model was deployed on satellite imagery from 2014, 2020, and 2025 to identify impervious surfaces, with a focus on residential structures. Inference was accelerated using GPU processing. Detections with confidence scores below 0.5 were discarded and Non-Maximum Suppression (NMS) was applied to eliminate redundant overlapping boxes. The resulting bounding boxes were processed using Python to calculate the spatial footprint of each structure in pixel square units.

For geospatial analysis, the detections were georeferenced using spatial metadata extracted from Google Earth Pro, including image extent, resolution, and projection information. Bounding boxes were first converted to polygon features using the geopandas and shapely libraries and exported in GeoJSON and ESRI shapefile formats. Area estimates (in px²) were calculated from bounding box dimensions and converted to m^2 using the image resolution. These geospatial files were then imported into QGIS, where they were spatially aligned and used for overlay analysis over the three study years.

This detection-to-GIS pipeline enables both visual and quantitative assessment of urban growth. By producing GIS-ready layers, the workflow supports practical applications, such as monitoring construction trends, identifying high-growth areas, and informing infrastructure and zoning decisions. Figure 7 illustrates the results, showing the geographic distribution of the detected features in Greater Faridabad.

Fig. 7. Georeferenced detection outputs over Greater Faridabad, visualized in Google Earth Pro and exported for spatial analysis in QGIS.

IV. RESULTS

A. Urban Area Extraction

Quantitative measurement of residential urban expansion was performed using high-resolution satellite imagery from 2014, 2020, and 2025, sourced via Google Earth Pro. The detected built-up areas, calculated in pixel square units (px²), showed significant growth over the 11 years. Table I presents

the temporal progression of urban expansion in sec-81, Greater Faridabad, detailing both interval and cumulative percentage growth from 2014 to 2025. This substantial increase in residential area indicates intense urban development, likely driven by population growth and infrastructure expansion.

TABLE I. URBAN AREA EXPANSION WITH PERIODIC AND CUMULATIVE GROWTH RATES (2014–2025)

Year	Urban area	Increment (%)	Cumulative %
2014	26,837	–	–
2020	106,330	296%	296%
2025	147,450	39%	450%

B. *Urban Growth Analysis*

1) *Spatial Distribution and Land Use Change*

Figure 8 illustrates the spatial and temporal growth of residential development. In 2014, built-up areas were sparse, totaling ~26,837 px². By 2020, urban clusters expanded significantly to ~106,330 px². By 2025, widespread sprawl covered most of the region, reaching ~147,450 px² with minimal undeveloped land remaining. These trends indicate three patterns: infill development, outward expansion of rural areas, and increasing housing density, indicating a blend of horizontal sprawl and internal densification.

The geospatial outputs derived from the YOLOv8 detection results were imported into QGIS, enabling spatial overlays across the three years. This allowed direct visual and quantitative comparison of expansion patterns, revealing hotspot zones of rapid development. Such GIS-integrated analysis enhances the applicability of the model outputs, providing urban planners with precise, time-stamped spatial data to support infrastructure planning, land-use regulation, and sustainable development strategies in rapidly growing regions, such as Greater Faridabad.

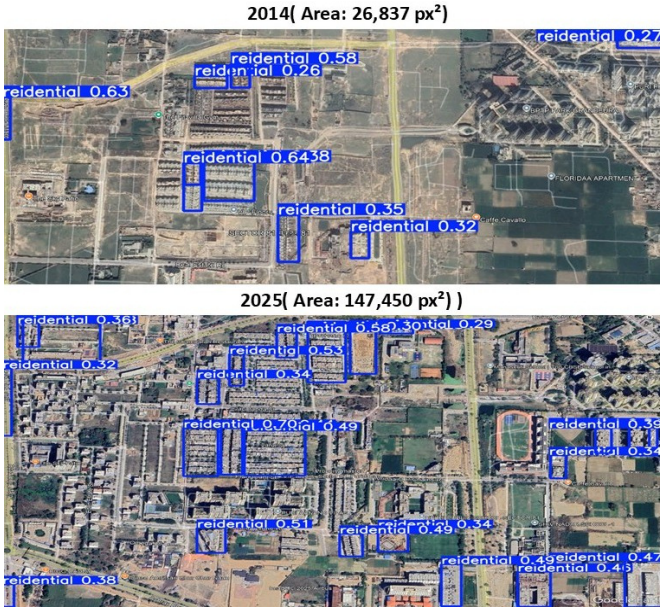

Fig. 8. Urban crawl spans over the decade 2014-2025.

2) *Visualization of Urban Growth Trends*

Figure 9 illustrates a continuous increase in the urban areas detected between 2014 and 2025, measured in px². The built-up area increased from 30,000 px² in 2014 to 140,000 px² by 2025, indicating a substantial expansion of impervious surfaces in Greater Faridabad. The most rapid expansion occurred during the earlier years of the study period, while the rate of increase appears to slow in the later years. This trend change may indicate a shift toward more compact urban development, possibly influenced by planning policies, spatial constraints, or the limited availability of open land. These temporal patterns, when visualized through GIS, provide valuable input to anticipate future land use dynamics and guide data-driven urban policy.

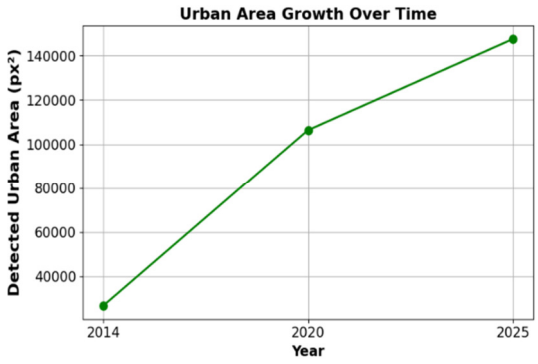

Fig. 9. Temporal trend in urban residential area (2014–2025), in px².

V. V. COMPARATIVE EVALUATION WITH EXISTING APPROACHES

To assess the effectiveness of the proposed YOLOv8-based approach, Table II presents a comparison with recent methods. This table summarizes key studies for 2019-2025, highlighting their datasets, methods, accuracy, and temporal capabilities.

TABLE II. COMPARATIVE ANALYSIS OF RECENT REMOTE SENSING APPROACHES (2019-2025)

Ref.	Region/ source	Years	Model	Target	Scores (%)
[4]	UC Merced, AID (Optical)	Single-date	Fine-tuned VGG Net	Scene classification	90.36 (OA)
[8]	Senitel-2 (Multispectral)	Single-date	SVM, RF, KNN	Land cover classification	93.85 (SVM - OA)
[9]	Dubai / KhalifaSAT, HRsat)	2021-2022	U2-Net	Buildings. Vegetation	82 (OA)
[10]	Kaduna Basin (Landsat+GEE)	1987-2020	RF	Urban Encroachment	80 (OA)
[18]	SpaceNet MUDS (Satellite)	2019-2020	YOLT	Buildings	F1 = 0.45
This work	Greater Faridabad, India (Google Earth)	2014–2025	YOLOv8	Residential Structures	94.0 (mAP@0.5)

Although classical models such as SVM and RF perform well on limited data, they lack scalability. YOLO-based methods offer speed and precision but are often applied to

single-date imagery. The proposed method improves on these by enabling multi-year analysis, achieving 94% mAP@0.5, and producing GIS-compatible outputs for spatial planning. This comparative evaluation underscores the robustness of the proposed YOLOv8 framework for multi-temporal urban analysis, offering superior accuracy and operational scalability over both classical and contemporary object detection approaches.

VI. CONCLUSIONS

This study presents an automated YOLOv8m-based framework to detect residential structures in Greater Faridabad using high-resolution Google Earth imagery from 2014, 2020, and 2025. The pipeline produced GIS-ready output, enabling efficient spatiotemporal monitoring of urban growth. Comparative evaluation with recent methods (2019-2025) demonstrated that although classical approaches such as SVM and RF achieve good accuracy on limited datasets, they lack scalability and temporal depth. YOLO-based methods, such as YOLOv5 and YOLT, offer real-time performance but are often applied to UAV or single-date imagery. In contrast, the proposed method achieved superior accuracy (98.3% mAP@0.5), supported multi-year analysis, and delivered spatial outputs suitable for urban planning applications. While the current implementation focuses solely on residential features and optical imagery, future work will extend to multi-class detection (e.g., roads, vegetation), integration of multispectral or SAR data for improved robustness [19], and deployment via cloud platforms for scalable, near-real-time urban monitoring [20].

REFERENCES

[1] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018, https://doi.org/10.1016/j.compag.2018.02.016.

[2] Y. Hou, X. Zheng, C. Han, W. Wei, R. Scherer, and D. Połap, "Deep Learning Methods in Short-Term Traffic Prediction: A Survey," *Information Technology and Control*, vol. 51, no. 1, pp. 139–157, Mar. 2022, https://doi.org/10.5755/j01.itc.51.1.29947.

[3] M. S. Norouzzadeh *et al.*, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716–E5725, Jun. 2018, https://doi.org/10.1073/pnas.1719367115.

[4] G. Cheng, J. Han, and X. Lu, "Remote Sensing Image Scene Classification: Benchmark and State of the Art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017, https://doi.org/10.1109/jproc.2017.2675998.

[5] G. S. Xia *et al.*, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, Jun. 2018, https://doi.org/10.1109/cvpr.2018.00418.

[6] G. Jocher, J. Qiu, and A. Chaurasia, "Ultralytics YOLO." Jan. 2023, [Online]. Available: https://github.com/ultralytics/ultralytics.

[7] P. T. Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors*, vol. 18, no. 1, Jan. 2018, Art. no. 18, https://doi.org/10.3390/s18010018.

[8] Atijosan, Abimbola, and Muibi, "A Performance Comparison of Three Machine Learning Algorithms for Urban Land Cover Classification Using High- Resolution Imagery," *Covenant Journal of Engineering Technology*, vol. 8, no. 2, Dec. 2024.

[9] M. A. A. Nuaimi *et al.*, "Deep Learning-Based Assessment of Urban and Vegetation Changes Using High-Resolution Khalifasat Satellite Imagery over Dubai," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. X-4/W4-2024, pp. 137–143, May 2024, https://doi.org/10.5194/isprs-annals-x-4-w4-2024-137-2024.

[10] O. S. Durowoju, R. O. Obateru, S. Adelabu, and A. Olusola, "Urban change detection: assessing biophysical drivers using machine learning and Google Earth Engine," *Environmental Monitoring and Assessment*, vol. 197, no. 4, Mar. 2025, Art. no. 441, https://doi.org/10.1007/s10661-025-13863-4.

[11] S. Zhao, K. Tu, S. Ye, H. Tang, Y. Hu, and C. Xie, "Land Use and Land Cover Classification Meets Deep Learning: A Review," *Sensors*, vol. 23, no. 21, Nov. 2023, Art. no. 8966, https://doi.org/10.3390/s23218966.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, https://doi.org/10.1109/tpami.2016.2577031.

[13] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, 2016, pp. 21–37, https://doi.org/10.1007/978-3-319-46448-0_2.

[14] H. Lechgar, H. Bekkar, and H. Rhinane, "Detection of Cities Vehicle Fleet Using Yolo v2 and Aerial Images," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4-W12, pp. 121–126, Feb. 2019, https://doi.org/10.5194/isprs-archives-XLII-4-W12-121-2019.

[15] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." arXiv, Apr. 08, 2018, https://doi.org/10.48550/arXiv.1804.02767.

[16] D. Sankhe and S. Bhosale, "Vessel Detection in Satellite Images using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18357–18362, Dec. 2024, https://doi.org/10.48084/etasr.8755.

[17] "Real-time global intelligence using satellite-AI | Granular.ai." https://www.granular.ai/.

[18] A. Van Etten, D. Hogan, J. Martinez-Manso, J. Shermeyer, N. Weir, and R. Lewis, "The Multi-Temporal Urban Development SpaceNet Dataset." arXiv, 2021, https://doi.org/10.48550/ARXIV.2102.04420.

[19] A. Irfan, Y. Li, X. E, and G. Sun, "Land Use and Land Cover Classification with Deep Learning-Based Fusion of SAR and Optical Data," *Remote Sensing*, vol. 17, no. 7, Apr. 2025, Art. no. 1298, https://doi.org/10.3390/rs17071298.

[20] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1–20, Jan. 2020, https://doi.org/10.1080/15481603.2019.1650447.