

Vision AI–Integrated 4D BIM for Real-Time Construction Progress Monitoring and Anomaly Detection

Kieran Ashcroft*¹, Maeve Donovan²

Email: kashcroft@gmail.com, maeve.dono@gmail.com

¹Department of Civil Engineering, School of Engineering, University of Galway, Galway, Ireland, H91 TK33

²School of Architecture, Planning and Environmental Policy, University College Dublin, Dublin, Ireland, D04 V1W8

*Corresponding Author

Abstract

Construction progress monitoring and quality inspection in high-rise building projects are still largely dependent on manual processes, resulting in inefficient reporting, delayed decision-making, and inconsistent inspection quality. Existing studies generally focus on either Vision-AI-based progress monitoring or BIM-based schedule management, with limited integration of geometric anomaly detection into a unified framework. This study aims to develop and evaluate an integrated Vision-AI and 4D-BIM framework for automated construction progress monitoring and structural geometric anomaly detection. The proposed method combines transformer-based object detection, three-dimensional coordinate transformation, and 4D-BIM schedule synchronization to identify structural components, compare as-built conditions with BIM models, and monitor construction progress in real time. Experimental results demonstrated an average precision of 0.882, an average recall of 0.847, and an inference speed of 41.5 FPS, while successfully detecting structural deviations exceeding predefined tolerance limits and synchronizing construction progress with planned schedules. The novelty of this research lies in the integration of transformer-based Vision-AI, 4D-BIM synchronization, and automated geometric anomaly detection within a single computational framework capable of simultaneously performing progress monitoring and dimensional quality assessment. The proposed framework contributes to digital construction management by improving monitoring efficiency, reducing reliance on manual inspection, and supporting faster identification of schedule deviations and quality issues. These findings demonstrate that the proposed framework provides an effective and practical solution for intelligent construction monitoring in high-rise building projects.

Keywords: Building Information Modeling, Computer Vision, Construction Progress Monitoring, Geometric Anomaly Detection, High-Rise Construction.

I. INTRODUCTION

The rapid expansion of urban populations and the limited availability of land resources have accelerated the development of high-rise buildings in metropolitan areas worldwide. The increasing complexity of modern vertical construction projects requires more sophisticated approaches to monitor project progress and ensure compliance with design specifications. Traditional construction monitoring methods continue to rely heavily on manual inspections, visual observations, and periodic site reports, which are often labor-intensive and susceptible to human subjectivity. Delays in identifying discrepancies between planned schedules and actual field conditions can result in schedule overruns, cost escalation, productivity losses, and extensive rework activities. The challenge becomes more critical in high-rise buildings, where deviations in structural elements at early construction stages may propagate into significant safety and quality issues during subsequent phases. Advances in digital construction technologies have created opportunities for automated monitoring systems capable of providing objective and

continuous assessments of both construction progress and structural conditions. The integration of artificial intelligence and digital building models has attracted growing attention as a potential solution for improving construction project control in complex vertical developments.

Building Information Modeling has evolved into a powerful platform for integrating geometric information with project schedules through 4D visualization, while computer vision technologies have demonstrated considerable capability in recognizing and classifying structural components from site images and videos. (Setyadi & Purwanto, 2025) demonstrated that image-based reconstruction techniques can support automated construction progress monitoring by converting visual information into measurable project indicators. Similarly, (Carizo, 2026) developed methods for integrating visual sensing technologies with BIM environments to estimate project performance and compare actual progress against planned schedules. A representative example can be observed in high-rise construction projects where delays in beam installation or slab formwork completion are frequently identified only after weekly reporting cycles, causing cascading impacts on subsequent structural activities and project milestones.

Although recent advancements in digital construction technologies have significantly improved automated site monitoring capabilities, several important research gaps remain unresolved. (Chong et al., 2025; Park & Kang, 2026) demonstrated the potential of integrating image-based monitoring and BIM environments for construction progress assessment, yet their approaches did not simultaneously incorporate schedule synchronization and geometric quality evaluation. Similarly, (Amirgaliyev et al., 2025; Rathee et al., 2023) confirmed the effectiveness of deep learning for object recognition and automated progress estimation, but the integration of temporal progress tracking with spatial anomaly detection remained largely unexplored.

Photogrammetry-based quality control methods reported by (Seid Ahmed & Amorim, 2025) achieved high geometric accuracy; however, progress monitoring and dimensional verification were still treated as separate processes rather than components of a unified computational framework. Furthermore, research focusing specifically on high-rise construction environments remains limited despite the additional challenges posed by severe occlusion, dynamic lighting conditions, dense structural configurations, and vertical logistics constraints. Therefore, this study aims to develop and validate an integrated Vision-AI and 4D-BIM framework capable of simultaneously performing automated construction progress tracking and geometric anomaly detection in high-rise building projects.

The research seeks to answer the following question: *How can the integration of Vision-AI and 4D-BIM improve real-time construction progress monitoring and anomaly detection in high-rise buildings?* The novelty of this study lies in the development of a unified computational workflow that combines object detection, schedule synchronization, spatial coordinate transformation, and

structural anomaly identification within a single framework. The proposed research hypothesizes that the integration of Vision-AI and 4D-BIM can improve monitoring accuracy, reduce inspection time, and enable earlier detection of structural deviations compared with conventional inspection methods. The expected outcome of this study is the establishment of a reliable digital quality control mechanism that supports more efficient and data-driven decision-making in modern construction management environments.

II. LITERATURE REVIEW

A. *Standardization of Multi-Platform Data Interoperability in Modern Civil Engineering*

The growing complexity of digital workflows in civil engineering has increased the importance of seamless information exchange among multiple software platforms. According to (Berlato et al., 2025), the evolution of Building Information Modeling has shifted engineering information management from isolated datasets toward integrated digital ecosystems involving architectural, structural, and construction stakeholders. Building upon this perspective, (Jia et al., 2024) emphasized the role of Industry Foundation Classes (IFC) in preserving semantic relationships and object properties during information exchange between heterogeneous software environments. (Souza et al., 2025) further observed that interoperability standards facilitate the continuous transfer of both geometric and non-geometric information throughout the entire building lifecycle, particularly in multidisciplinary projects. (Omrany et al., 2023) added that interoperability frameworks support consistency among design models, scheduling information, and construction documentation generated from different digital platforms. Similarly, (Fattore et al., 2025) reported that standardized data environments maintain object identifiers and metadata integrity while enabling synchronization between BIM repositories, planning systems, and project scheduling applications.

B. *Transformation of Deep Learning Architectures in Structural Element Detection*

Recent advances in artificial intelligence have significantly influenced the interpretation of visual information in construction environments. According to (Ullah et al., 2023), convolutional neural networks introduced hierarchical feature extraction capabilities that substantially improved image recognition performance compared with traditional handcrafted features. (Liu et al., 2024) subsequently reported that region proposal mechanisms enhanced object localization accuracy in complex visual scenes containing multiple overlapping elements. (Habash et al., 2025) later observed that single-stage object detectors provided substantial computational advantages while preserving competitive detection performance for real-time applications. (Li & Singh, 2026) added that focal loss strategies improved recognition performance for small and partially occluded objects frequently encountered in dense environments. More recently, (Yu et al., 2025) highlighted that transformer-based detection architectures strengthened contextual understanding

among neighboring objects and their surroundings, leading to increased interest in recognizing structural elements such as columns, beams, reinforcement bars, and formwork systems from construction imagery.

C. Implementation of Machine Learning for Temporal Progress Monitoring Optimization

The increasing availability of visual construction data has encouraged the adoption of machine learning approaches for monitoring project progress over time. According to (Chen et al., 2025), image-based reconstruction techniques enabled visual observations from construction sites to be associated with measurable progress indicators during project execution. (Doan et al., 2025) subsequently reported that the integration of visual sensing technologies with BIM environments provided richer representations of actual site conditions than conventional reporting approaches. (Manakitsa et al., 2024) further observed that machine learning algorithms improved the interpretation of repetitive construction activities by identifying patterns embedded within temporal image sequences. (In-Hwa et al., 2025) added that deep learning models trained on construction datasets were capable of distinguishing completed, ongoing, and pending activities across multiple project environments. Similarly, (Yoon & Kim, 2025) noted that automated interpretation of visual information enabled the association of recognized structural components with predefined scheduling activities and construction sequences.

D. Integration of Photogrammetric Metrology in Geometric Quality Control of Buildings

The advancement of image-based measurement technologies has expanded the application of photogrammetric methods in construction quality assessment. According to (Dong et al., 2025), dense image matching techniques generated detailed point clouds capable of representing subtle geometric variations on structural surfaces. (Tan et al., 2024) reported that photogrammetric point cloud reconstruction achieved dimensional measurements comparable to terrestrial laser scanning under controlled acquisition conditions. (Qi et al., 2024) further observed that close-range photogrammetry enabled highly precise geometric measurements suitable for engineering applications requiring millimeter-level accuracy.

Table 1. Comparison of Previous Photogrammetric Studies

Author(s) and Year	Measurement Technique	Application Context	Reported Accuracy
(Dong et al., 2025)	Dense Image Matching	Surface representation of structural elements	Centimeter level
(Tan et al., 2024)	Photogrammetric Point Cloud Reconstruction	Dimensional comparison with LiDAR measurements	Sub-centimeter level
(Qi et al., 2024)	Close-Range Photogrammetry	Engineering geometric measurements	Millimeter level
(Luo et al., 2024)	Multi-View Stereo Reconstruction	Spatial reconstruction of built environments	Millimeter level
(Payawal & Kim, 2023)	Image-Based Structural Measurement	Geometric documentation in construction projects	Millimeter level

(Luo et al., 2024) added that multi-view image acquisition improved spatial reconstruction quality through increased redundancy among image correspondences and camera positions. Similarly, (Payawal & Kim, 2023) highlighted the growing adoption of image-based measurements for documenting geometric characteristics of structural components in civil engineering environments. Comparative characteristics of previous studies concerning photogrammetric metrology applications in geometric quality control are summarized in Table 1.

III. RESEARCH METHOD

A. Proposed Integrated System Framework

The methodological design of this study was developed to establish a unified computational workflow capable of connecting visual observations from construction sites with digital information contained in 4D-BIM environments. The proposed framework combines computer vision, deep learning, spatial metrology, and BIM interoperability within a single processing pipeline to improve consistency between planned and actual construction conditions. Visual information acquired from multiple sensing platforms is processed through a sequence of detection, synchronization, and verification procedures that operate continuously throughout the project lifecycle. The framework also accommodates both temporal monitoring and geometric verification to represent construction progress and dimensional quality simultaneously. Such integration enables construction information originating from heterogeneous data sources to be transformed into measurable indicators associated with scheduling performance and structural conditions.

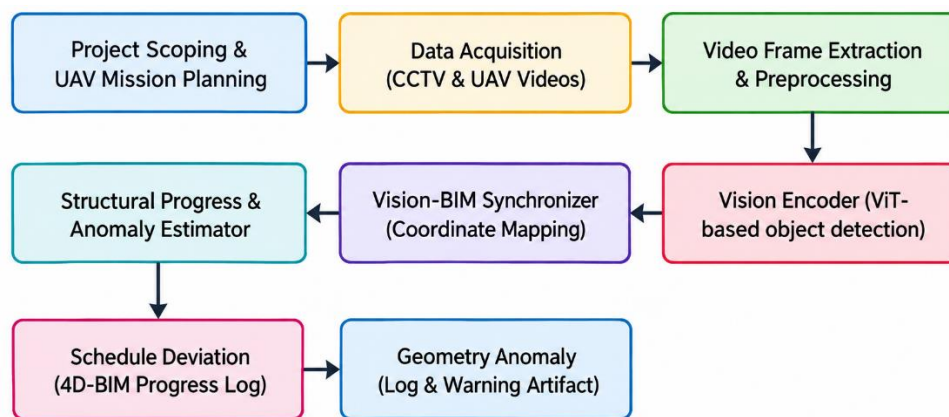


Figure 1. Proposed Integrated Vision-AI and 4D-BIM Framework

The overall workflow of the proposed system architecture is illustrated in Figure 1. The framework begins with project scoping activities and UAV mission planning, followed by visual data acquisition through fixed CCTV cameras and periodic UAV inspections. The collected visual data are subsequently converted into image sequences through frame extraction and preprocessing procedures prior to feature learning by the Vision-AI encoder. The extracted visual information

is then synchronized with the 4D-BIM environment using coordinate mapping and object association mechanisms. Finally, the integrated system produces two independent outputs consisting of schedule deviation information and geometric anomaly records for quality control purposes.

Figure 1 illustrates the relationship between the visual intelligence module and the BIM synchronization module within the integrated computational framework. The Vision Transformer-based encoder functions as the primary feature extraction component responsible for recognizing structural elements from construction imagery. The synchronization layer subsequently associates the detected objects with corresponding BIM components using spatial coordinates and object identifiers. Progress estimation modules evaluate the consistency between actual construction conditions and planned schedules stored within the BIM environment. In parallel, anomaly estimation modules calculate geometric deviations and generate warning artifacts whenever dimensional discrepancies exceed predefined tolerance thresholds.

B. Dataset Characteristics and Visual Data Acquisition Procedure

The experimental dataset was constructed using visual observations collected from an ongoing high-rise building construction project through multiple sensing platforms. The use of heterogeneous visual sources was intended to capture different structural elements under varying environmental conditions and observation perspectives. Fixed CCTV cameras were utilized to continuously monitor repetitive construction activities occurring in predefined locations, whereas UAV surveys provided additional observations for upper floors and inaccessible zones. Handheld cameras were further employed to document indoor activities and enclosed environments where aerial visibility was limited. The resulting dataset contained a wide range of illumination conditions, obstruction levels, and viewing geometries representative of realistic construction environments.

The characteristics of the dataset used for training and validating the proposed system are summarized in Table 2. The dataset was categorized according to structural element type, target component, image source, and environmental conditions observed during acquisition activities. Particular attention was devoted to the representation of occlusion conditions because construction sites frequently contain scaffolding systems, safety nets, and equipment that partially obstruct visual observations. Separate dataset partitions were established for training, validation, and testing purposes to ensure unbiased performance evaluation. The inclusion of diverse image sources and environmental scenarios allows the trained model to learn robust representations of structural components under realistic operational conditions.

The dataset statistics presented in Table 2 indicate considerable variation in image characteristics and scene complexity among structural classes. Column elements exhibit relatively low occlusion

levels because they are generally located in open environments during construction activities. Beam and formwork components experience greater visual obstruction due to the presence of temporary supporting systems and surrounding equipment. Variations in illumination conditions were intentionally preserved to improve model generalization under realistic site environments. The resulting dataset composition reflects the operational conditions typically encountered in high-rise construction projects.

Table 2. Characteristics of the Visual Construction Dataset

Structural Element Type	Target Component	Total Frames	Resolution / Source	Lighting & Environment Variation	Occlusion Level & Obstruction Type	Dataset Split (Train/Val/Test)
C-1: Reinforced Concrete Column	Main Core Column	2,500	3840×2160 (Fixed CCTV)	Direct Sunlight (High Contrast)	Low (<10%) – Clear View	70 / 20 / 10
C-2: Reinforced Concrete Column	Perimeter Column	2,000	3840×2160 (Fixed CCTV)	Dusk / Low Light Condition	Medium (15%) – Safety Nets	70 / 20 / 10
B-1: Primary Structural Beams	Longitudinal Beam	1,800	1920×1080 (UAV Drone)	Overcast / Cloudy Condition	Medium (25%) – Scaffolding Poles	80 / 10 / 10
B-2: Secondary Structural Beams	Transverse Beam	1,400	1920×1080 (UAV Drone)	Mid-Day Shadows	High (35%) – Crane Shadow	80 / 10 / 10
F-1: Slab Formwork and Decking	Conventional Plywood	1,500	1920×1080 (Handheld Camera)	Artificial LED Lighting	Low (<5%) – Clear Surface	75 / 15 / 10
F-2: System Formwork and Shoring	Aluminium Formwork	1,300	1920×1080 (Handheld Camera)	Semi-Enclosed Dark Zones	High (>40%) – Reinforcement Congestion	75 / 15 / 10

C. Vision-AI Architecture Configuration and Spatial Synchronization

The Vision-AI module was designed as a single-stage object detection architecture optimized for edge deployment within construction environments. The model architecture prioritizes computational efficiency while maintaining the ability to distinguish subtle texture differences among structural materials and construction states. A Vision Transformer backbone was selected to improve contextual understanding between neighboring structural components and surrounding environments. Multiple loss functions were combined to balance localization accuracy and classification consistency during training. Additional augmentation strategies were introduced to improve robustness under adverse weather conditions and severe visual obstruction.

The configuration of the proposed Vision-AI model is summarized in Table 3. The selected hyperparameters were determined through iterative experimentation involving accuracy, inference speed, and computational efficiency considerations. Optimization strategies focused on reducing overfitting caused by repetitive visual patterns frequently observed in reinforced concrete structures. Learning rate scheduling mechanisms were introduced to stabilize parameter convergence throughout the training process. Loss scaling parameters were adjusted to improve localization performance near partially occluded structural boundaries.

Table 3. Hyperparameter Configuration of the Proposed Vision-AI Model

Architecture Component	Hyperparameter	Configured Value	Algorithm	Scientific Justification
Network Input	Dynamic Resolution	640×640×3	RGB Tensor	Maintains real-time inference performance
Optimization	Weight Decay	1×10^{-2}	AdamW	Reduces overfitting on repetitive textures
Learning Rate	Base Learning Rate	1×10^{-3}	Cosine Annealing	Stabilizes optimization process
Regression Loss	λ_{box}	7.5	CIoU Loss	Improves boundary localization
Classification Loss	λ_{cls}	0.5	BCE Loss	Supports multi-label classification
Distribution Loss	λ_{dfl}	1.5	Focal Loss Variant	Enhances performance under occlusion
Data Augmentation	Mosaic / MixUp	0.8 / 0.15	Random Scale and Blur	Improves robustness under environmental variability

The hyperparameter configuration shown in Table 3 reflects the balance between computational efficiency and detection reliability required in construction monitoring applications. The use of CIoU loss prioritizes spatial alignment between predicted and actual object boundaries during localization tasks. Focal loss mechanisms improve sensitivity toward partially visible components frequently encountered in dense construction environments. Data augmentation methods expose the model to diverse environmental conditions during training to improve generalization performance. Collectively, these configurations support stable inference performance under realistic site conditions.

The computational workflow responsible for geometric synchronization between image observations and BIM coordinates is illustrated in Figure 2. The synchronization process begins with object detection and camera pose estimation to establish the relationship between image coordinates and the physical environment. Spatial transformation matrices are subsequently employed to project detected objects into the BIM coordinate system. Iterative Closest Point alignment procedures further refine the registration accuracy between as-planned and as-built geometries. The resulting spatial representation enables quantitative evaluation of schedule deviations and geometric anomalies.

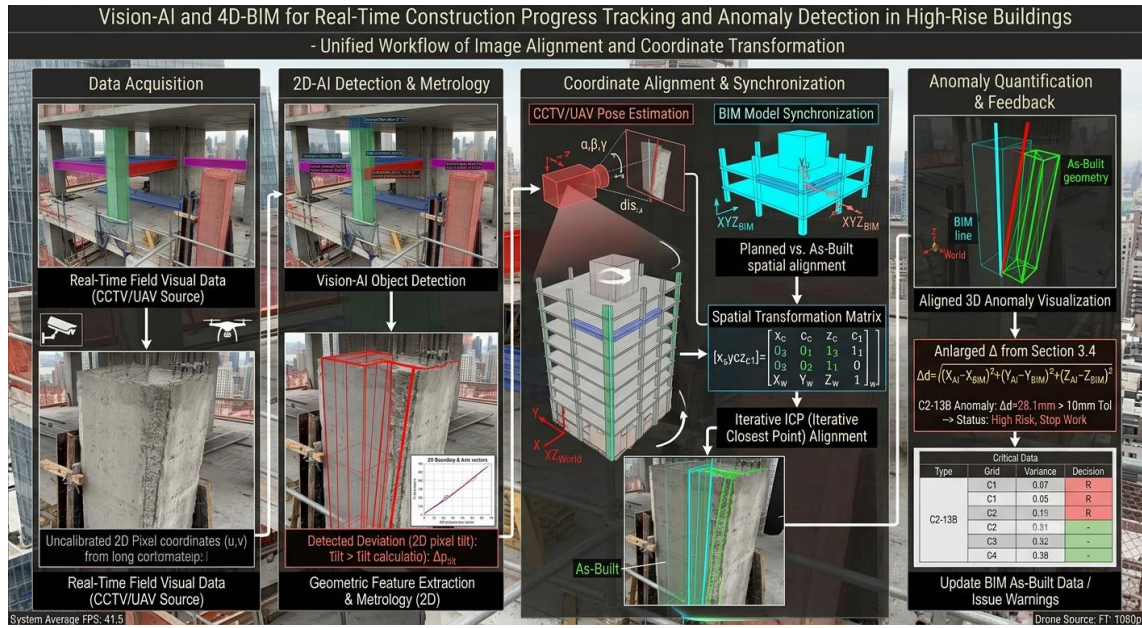


Figure 2. Spatial Alignment and Coordinate Transformation Workflow

Figure 2 illustrates the sequence of operations required to transform image observations into measurable BIM information. Visual features extracted from CCTV and UAV imagery are first associated with structural components detected by the Vision-AI model. Camera pose estimation algorithms subsequently estimate the relative position and orientation of the sensing platform with respect to the building coordinate system. Spatial transformation operations convert two-dimensional image observations into three-dimensional BIM coordinates. The synchronized spatial representation then becomes the basis for anomaly quantification and progress assessment procedures.

D. Mathematical Formulation and Computational Algorithms

The integrated computational framework combines visual interpretation, spatial synchronization, and geometric verification within a unified analytical environment. Information generated by the Vision-AI module and the 4D-BIM platform is represented numerically to enable quantitative comparison between planned and actual construction conditions. The computational procedures involve object localization, optimization during model training, coordinate transformation, and spatial deviation assessment. Each analytical component contributes to the interpretation of construction progress and structural geometry from different perspectives. The resulting mathematical representation forms the basis for anomaly identification and progress monitoring within the proposed system.

The quality of object localization produced by the Vision-AI detector is evaluated using the Intersection over Union metric, which measures the degree of overlap between predicted and reference bounding boxes. Higher overlap values indicate better agreement between the detected structural element and the manually annotated ground truth data, as expressed in Equation (1).

$$IoU = \frac{Area_{Union}}{Area_{Overlap}} \quad (1)$$

where IoU represents the Intersection over Union score, $Area_{Overlap}$ denotes the intersection area shared by the predicted and reference bounding boxes, and $Area_{Union}$ represents the total area covered by both regions. The relationship among these variables is used to quantify the localization performance of the object detection model during training and validation procedures.

The optimization of the Vision-AI model is performed using a weighted combination of localization, classification, and distribution losses to ensure stable convergence during the training process. The contribution of each loss component is regulated through balancing coefficients as formulated in Equation (2).

$$L_{total} = \lambda_{box}L_{box} + \lambda_{cls}L_{cls} + \lambda_{dfl}L_{dfl} \quad (2)$$

where L_{total} denotes the total optimization loss, L_{box} represents the localization loss associated with bounding box regression, L_{cls} corresponds to the classification loss, and L_{dfl} indicates the distribution focal loss used for boundary refinement. The coefficients λ_{box} , λ_{cls} , and λ_{dfl} regulate the relative contribution of each loss component during parameter optimization.

The synchronization process between visual observations and BIM information requires the transformation of image coordinates into the three-dimensional coordinate system of the building model. This projection process is performed using a homogeneous transformation matrix that represents both camera orientation and spatial translation, as shown in Equation (3).

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3)$$

where x_c , y_c , and z_c represent the coordinates of detected objects in the camera coordinate system, while X_w , Y_w , and Z_w denote the corresponding coordinates in the BIM reference system. The matrix $R_{3 \times 3}$ describes rotational transformations, whereas $T_{3 \times 1}$ represents translational displacement between the camera and BIM coordinate frames.

The geometric consistency between the detected construction components and the corresponding BIM elements is quantified using a three-dimensional Euclidean distance metric. The magnitude of spatial deviation is computed according to Equation (4).

$$\Delta d = \sqrt{(X_{AI} - X_{BIM})^2 + (Y_{AI} - Y_{BIM})^2 + (Z_{AI} - Z_{BIM})^2} \quad (4)$$

where Δd represents the spatial deviation distance, while X_{AI} , Y_{AI} , and Z_{AI} correspond to the coordinates estimated by the Vision-AI model and X_{BIM} , Y_{BIM} , and Z_{BIM} denote the reference coordinates extracted from the BIM model. Deviations exceeding the predefined tolerance

threshold of 20 mm are recorded as geometric anomalies and subsequently stored within the warning log artifact generated by the integrated monitoring framework.

IV. RESULT

A. Quantitative Evaluation of Vision-AI Model Performance

The performance evaluation was conducted to assess the capability of the Vision-AI framework in recognizing structural components under realistic construction site conditions. The experimental scenarios included variations in illumination, object scale, camera perspective, and visual obstruction caused by temporary construction facilities. The evaluation process considered both detection accuracy and computational efficiency to represent operational requirements commonly encountered in high-rise projects. Performance measurements were obtained from testing datasets that were isolated from the training and validation subsets to preserve objectivity during assessment. The resulting measurements provide a detailed representation of model behavior under diverse environmental conditions.

The quantitative performance metrics obtained during model inference are summarized in Table 4. The evaluation includes object localization performance, class recognition capability, and computational responsiveness across multiple structural categories. Reinforced concrete columns, structural beams, slab formwork, and exposed reinforcement meshes were selected because they represent the dominant components observed during the structural construction phase. Several complementary evaluation metrics were employed to capture both coarse and fine localization behavior. Together, these indicators provide a detailed representation of detection performance under realistic construction environments.

Table 4. Detailed Confusion Matrix and Performance Metrics of Object Detection

Structural Class Target	TP	FP	Precision	Recall	F1-Score	mAP@0.5	mAP@0.5:0.95	Inference Speed
Concrete Column	2,245	182	0.925	0.891	0.908	0.912	0.745	45.2 FPS
Structural Beam	1,512	204	0.881	0.854	0.867	0.879	0.692	42.1 FPS
Slab Formwork	1,221	208	0.854	0.812	0.832	0.841	0.618	38.5 FPS
Exposed Rebar Mesh	985	145	0.871	0.834	0.852	0.863	0.635	40.2 FPS
System Average	1,491	184	0.882	0.847	0.864	0.873	0.672	41.5 FPS

The values reported in Table 4 show that reinforced concrete columns achieved the highest precision and recall values among all structural categories. Beam elements exhibited slightly lower performance due to the presence of scaffolding systems and temporary supports that partially obstructed their boundaries. Slab formwork generated the lowest localization accuracy

because repetitive textures and weak visual contrast reduced feature separability during inference. The average inference speed remained above 40 FPS across all experiments, satisfying real-time monitoring requirements in dynamic construction environments. The difference between $mAP@0.5$ and $mAP@0.5:0.95$ reflects the increased sensitivity of stricter localization criteria during evaluation.

Representative examples of object detection outputs generated by the Vision-AI framework are presented in Figure 3. The selected visualization corresponds to structural activities observed during active construction operations on upper floors of the building. Multiple structural categories were detected simultaneously within the same scene despite the presence of temporary facilities and background clutter. Confidence values generated during inference are displayed together with object labels and detection boundaries. The visual examples provide direct observations regarding model behavior under unconstrained field conditions.



Figure 3. Vision-AI Bounding Box Predictions on Construction Site Visuals

The visual output shown in Figure 3 demonstrates successful localization of reinforced concrete columns, structural beams, and temporary formwork elements within active construction environments. Structural columns were identified with confidence levels exceeding 90% in most observations due to their distinctive geometry and appearance. Beam elements remained detectable despite partial visual obstruction caused by scaffolding systems and construction equipment. The generated bounding boxes preserved spatial consistency across neighboring structural components and adjacent image regions. Detection performance remained stable under variations in perspective, illumination intensity, and scene complexity.

B. Temporal Progress Synchronization Analysis Using 4D-BIM Data

The temporal synchronization experiment was performed to evaluate the consistency between construction activities observed by Vision-AI and schedule information contained within the 4D-BIM environment. The analysis focused on Floors 12 to 14 because these levels contained multiple concurrent structural activities during the observation period. Progress information extracted from visual observations was associated with corresponding BIM components using object identifiers and spatial coordinates. Schedule variance calculations were subsequently performed at the component level to preserve spatial granularity during analysis. The resulting information provides a detailed representation of progress behavior within the monitored construction zones.

The comparison between planned schedules and actual construction observations is summarized in Table 5. The evaluation includes baseline completion dates, detected completion dates, schedule variance, and dominant field factors associated with deviations. Individual structural components were analyzed independently to preserve object-level progress information within the BIM environment. Both accelerated and delayed activities were observed during the monitoring period. The collected records describe the temporal relationship between planned and actual project execution.

Table 5. Granular Comparison of Actual Progress and 4D-BIM Baseline Schedule

Element ID	Structural Element	Floor and Zone	Baseline Finish	Actual Detection	Variance	Schedule Status	Dominant Field Factor
3B\$xFd9_ExA109	Core Column C1	Floor 12 - Zone A	12-May-2026	11-May-2026	-1 Day	Ahead of Schedule	High worker productivity and efficient curing
7K@zWp2_LxB214	Primary Beam B1	Floor 12 - Zone B	14-May-2026	19-May-2026	+5 Days	Critical Delay	Steel delivery disruption
1M#qYt5_KxC302	Slab Decking F1	Floor 12 - Zone B	18-May-2026	21-May-2026	+3 Days	Delayed	Dependency on beam installation
5N*vRp1_MxA201	Core Column C1	Floor 13 - Zone A	21-May-2026	21-May-2026	0 Day	On Schedule	Standard construction cycle
9P&wTs4_LxB105	Perimeter Column C2	Floor 13 - Zone B	23-May-2026	25-May-2026	+2 Days	Delayed	Wind-related crane restrictions
2L\$zQv8_KxC110	Primary Beam B2	Floor 14 - Zone A	29-May-2026	27-May-2026	-2 Days	Ahead of Schedule	Early-strength concrete implementation

The values summarized in Table 5 reveal considerable variation in schedule performance among structural components and floor locations. Several components progressed faster than the planned

schedule due to improved productivity and accelerated construction methods. Delays were predominantly associated with material logistics, weather restrictions, and dependencies between predecessor activities. Components associated with beam construction exhibited the largest schedule deviations because downstream activities depended directly on their completion status. The component-level representation preserves temporal information that is often lost in conventional project reporting systems.

The visual comparison between planned construction states and field observations is presented in Figure 4. The visualization displays baseline conditions and actual site progress within the same reporting period. Structural components exhibiting schedule discrepancies are highlighted to improve visibility during interpretation. The side-by-side representation preserves the spatial relationship between construction activities and project schedules. The visualization provides a direct comparison between digital planning information and physical site conditions.



Figure 4. Side-by-Side Comparison of 4D-BIM Planned Model and Actual Field Conditions

The side-by-side comparison presented in Figure 4 shows visible differences between planned construction states and actual field observations in several monitored areas. Components highlighted in red correspond to structural activities that remained incomplete at the time of inspection despite being scheduled for completion within the review period. Spatial localization of delayed components remains consistent between the BIM model and field imagery. The synchronized visualization preserves contextual information regarding neighboring structural elements and ongoing construction operations. Schedule discrepancies become immediately observable through the combined representation of temporal and spatial information.

C. Metrological Precision Analysis and Geometric Anomaly Detection

The geometric verification analysis was conducted to evaluate the capability of the proposed framework in identifying dimensional inconsistencies between field observations and BIM geometry. Spatial comparisons were performed using three-dimensional coordinates obtained from the image registration and alignment procedures described in the methodology section. The evaluation focused on structural elements that influence vertical continuity and subsequent construction activities. Detected deviations were compared against tolerance limits commonly adopted for reinforced concrete high-rise buildings. The resulting measurements provide a quantitative representation of dimensional accuracy within the monitored project area.

The detected anomalies and dimensional deviations identified during monitoring activities are summarized in Table 6. The analysis includes information regarding anomaly type, deviation magnitude, permissible tolerance, and corresponding field actions. Spatial deviations were quantified using Euclidean distance calculations within the BIM coordinate system. Multiple categories of geometric inconsistencies were observed throughout the monitored floors. The collected measurements describe the dimensional relationship between as-built and as-designed conditions.

Table 6. Quantitative Summary of Structural Anomalies and Dimensional Deviations

Anomaly ID	Structural Element	Grid Location	Deviation Type	Measured Δd	Tolerance	Δd/Tol	Safety Action
AN-2026-012	Column C12-A	Floor 12 / Grid C-4	Vertical Tilting	24.5 mm	±15.0 mm	1.63	Immediate realignment before concrete pouring
AN-2026-015	Beam B12-B2	Floor 12 / Grid A-2	Axial Eccentricity	12.2 mm	±20.0 mm	0.61	Accepted and logged in BIM
AN-2026-022	Column C13-B	Floor 13 / Grid D-1	Component Misalignment	28.1 mm	±10.0 mm	2.81	Stop work and re-establish alignment
AN-2026-029	Slab S13-A3	Floor 13 / Grid B-3	Surface Flatness Deviation	18.4 mm	±12.0 mm	1.53	Temporary shoring reinforcement
AN-2026-035	Beam B14-A1	Floor 14 / Grid E-2	Span Deviation	32.0 mm	±25.0 mm	1.28	Formwork revision prior to casting

The anomaly statistics reported in Table 6 show substantial variation in severity among detected geometric inconsistencies. Column misalignment generated the highest deviation ratios relative to allowable tolerances due to the sensitivity of vertical structural alignment in high-rise construction. Beam eccentricity remained within acceptable tolerance limits and was incorporated into the as-built BIM database without corrective intervention. Surface flatness irregularities and

span deviations exhibited intermediate severity levels requiring localized corrective actions. The deviation ratio provides additional information regarding the magnitude of dimensional inconsistency relative to design tolerances.

The graphical representation of geometric anomaly detection results is presented in Figure 5. The visualization overlays BIM reference geometry with field observations extracted from construction imagery. Spatial differences between planned and actual positions are displayed directly on the observed structural elements. Numerical annotations associated with each deviation are included within the visual representation. The resulting visualization provides an intuitive interpretation of dimensional discrepancies within the construction environment.

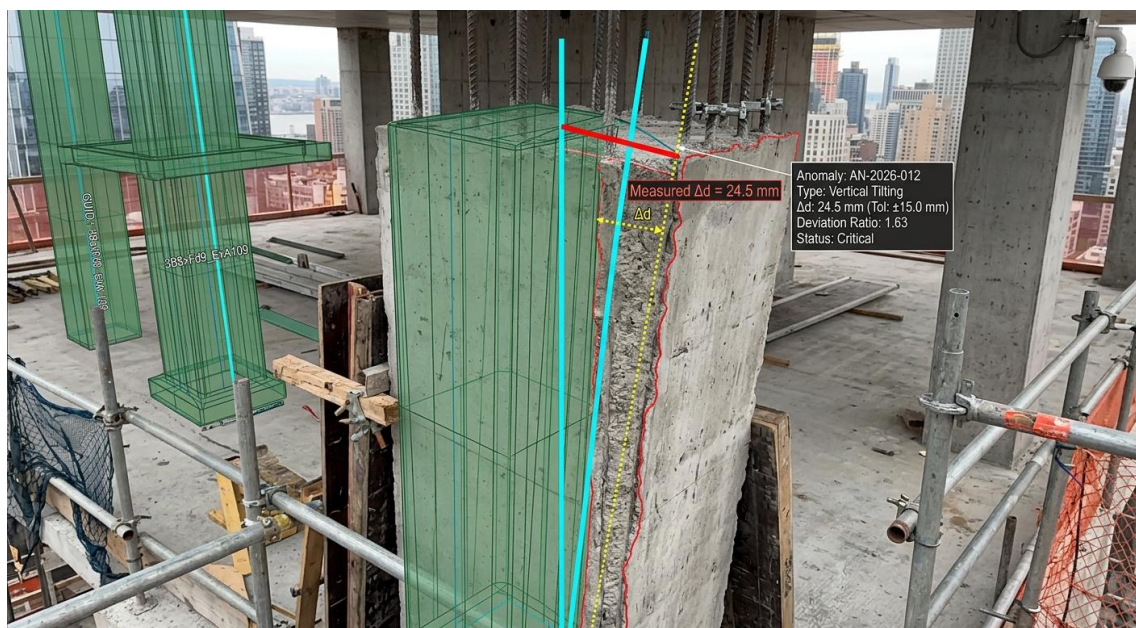


Figure 5. Graphical Visualization of Structural Anomaly Detection

The graphical overlay illustrated in Figure 5 represents the spatial displacement between digital BIM references and the physical position of structural components observed on site. Offset vectors and reference lines are superimposed on the structural surface to represent the magnitude and direction of displacement. The detected column anomaly exhibits visible deviation relative to the reference alignment axis generated from the BIM environment. Quantitative information regarding displacement magnitude and anomaly severity is displayed together with the visual overlay. The combined representation preserves both numerical accuracy and visual interpretability during field inspection activities.

D. Computational Performance and Operational Characteristics

The implementation of the integrated Vision-AI and 4D-BIM framework produced measurable changes in construction monitoring workflows during field deployment in high-rise building environments. Automated interpretation of visual information reduced the dependency on repetitive manual inspection activities that are traditionally performed by field engineers and

project supervisors. Continuous synchronization between construction imagery and BIM schedules enabled progress information to be updated immediately after visual acquisition without requiring additional reporting procedures. Geometric verification processes became available during early construction stages prior to subsequent structural activities such as reinforcement installation and concrete casting. The observed operational behavior reflects the capability of the framework to maintain simultaneous monitoring of temporal progress and dimensional conformity within a unified computational environment.

The computational behavior of the framework during deployment exhibited sensitivity to several environmental and operational conditions commonly encountered in construction sites. Detection performance remained strongly influenced by the availability of unobstructed camera viewpoints and stable visual coverage of structural components throughout the monitoring period. High concentrations of dust particles and unfavorable atmospheric conditions occasionally reduced image clarity and weakened object boundary definition during inference processes. Dense construction environments containing temporary facilities, scaffolding systems, and lifting equipment increased the occurrence of partial occlusion and fragmented observations. Variations in camera calibration quality and perspective alignment also affected the precision of spatial registration between field imagery and BIM coordinate systems.

V. DISCUSSION

The findings demonstrate that the proposed integration of Vision-AI and 4D-BIM effectively addresses the primary research objective by enabling simultaneous real-time construction progress tracking and geometric anomaly detection within a unified computational framework. The object detection model achieved an average precision of 0.882, an average recall of 0.847, and an inference speed of 41.5 FPS, indicating that the framework is capable of supporting continuous monitoring under realistic high-rise construction conditions. The integration of transformer-based visual recognition with BIM-based schedule synchronization provides reliable and timely information for project control while reducing dependence on manual inspections. The superior detection performance observed for reinforced concrete columns compared with beams and slab formwork reflects the influence of object geometry, texture characteristics, and visual complexity on recognition accuracy within construction environments. These performance differences illustrate the capability of the proposed framework to capture both temporal progress and geometric information required for digital construction management.

The present findings are generally consistent with previous studies reporting the effectiveness of image-based monitoring and BIM integration for construction progress assessment, while also extending their capabilities. Similar to the observations of (Chong et al., 2025; Park & Kang, 2026), the proposed framework successfully synchronized construction observations with

planned schedules, enabling direct identification of schedule deviations at the component level. Likewise, the object detection performance supports the conclusions of (Amirgaliyev et al., 2025; Rathee et al., 2023), who demonstrated that deep learning architectures provide reliable recognition performance in complex visual environments. However, unlike earlier approaches that primarily emphasized either schedule monitoring or object recognition independently, the proposed framework simultaneously integrates progress synchronization with geometric anomaly detection, directly addressing the knowledge gap identified in the literature review. The incorporation of three-dimensional coordinate transformation and BIM-based dimensional verification expands the practical application of previous Vision-AI frameworks by linking visual interpretation with engineering quality assessment rather than limiting the analysis to construction progress alone.

Another important contribution of this study is the ability of the integrated framework to identify structural deviations exceeding predefined tolerance limits while maintaining continuous schedule monitoring within the same computational workflow. This finding reinforces previous observations by (Seid Ahmed & Amorim, 2025), who emphasized the importance of photogrammetric techniques for geometric quality evaluation, while demonstrating that quality inspection and progress monitoring can be integrated within a single computational process. Several performance variations were observed across different structural components, particularly for slab formwork and beam elements affected by severe occlusion, repetitive textures, and unfavorable lighting conditions. These variations indicate that environmental complexity continues to influence visual recognition despite the contextual learning capability provided by transformer-based architectures, which is consistent with the observations of (Li & Singh, 2026) regarding partially occluded object detection. The observed performance patterns illustrate the influence of sensing conditions and image quality on the robustness of automated monitoring in dense construction environments.

The findings also provide theoretical and practical insights into digital construction management. From a theoretical perspective, this study contributes to the growing body of knowledge on integrated digital construction systems by demonstrating how Vision-AI, 4D-BIM synchronization, and geometric metrology can operate within a unified computational framework that supports both schedule control and quality assurance. From a practical perspective, the framework enables earlier identification of construction delays and dimensional deviations, allowing project teams to implement corrective actions before subsequent construction activities are affected. Several limitations should nevertheless be acknowledged, including the use of data collected from a single high-rise construction project, sensitivity to occlusion, lighting variability, camera calibration accuracy, and environmental conditions such as dust and temporary site obstructions, all of which may influence detection reliability and spatial registration precision.

These conditions indicate that the reported performance is closely associated with the characteristics of the evaluated project and operational environment, requiring careful interpretation when applying the framework to different construction contexts.

VI. CONCLUSION AND RECOMMENDATION

This study developed and evaluated an integrated Vision-AI and 4D-BIM framework for automated construction progress monitoring and geometric anomaly detection in high-rise building projects. The experimental results demonstrated that the proposed framework achieved reliable object detection performance, accurate synchronization between construction imagery and BIM schedules, and effective identification of structural deviations beyond predefined tolerance limits. The integration of transformer-based computer vision with BIM-based spatial and temporal information enabled simultaneous assessment of construction progress and dimensional conformity within a unified digital environment. The framework also reduced reliance on repetitive manual inspection by providing continuous visual interpretation and automated progress updates throughout the construction process. These outcomes indicate that the proposed approach offers a practical digital solution for improving monitoring efficiency, quality control, and decision-making in modern construction project management.

Future research may extend the proposed framework by validating its performance across different project types, structural systems, and construction environments to evaluate its robustness under more diverse operational conditions. Additional investigations may also incorporate multimodal sensing technologies, such as UAV imagery, terrestrial laser scanning, LiDAR, or IoT-based monitoring systems, to improve geometric accuracy and reduce the influence of visual occlusion and unfavorable environmental conditions. The development of adaptive learning strategies capable of handling varying lighting conditions, temporary site obstructions, and complex construction activities may further enhance object detection consistency throughout project execution. Future studies may also explore tighter integration with digital twin platforms and predictive analytics to support proactive construction management and real-time decision support. These research directions may contribute to the advancement of intelligent construction monitoring systems capable of delivering higher reliability, scalability, and automation for increasingly complex infrastructure and building projects.

AI Use

The authors confirm that artificial intelligence applications, including ChatGPT, were used exclusively to support language editing and manuscript refinement, including improvements to grammar, sentence structure, readability, wording, and overall writing clarity. These AI-based tools did not contribute to any stage of the research process. All scientific activities, including the formulation of the research concept, study design, data acquisition, methodological development,

data analysis, interpretation of the findings, and preparation of the final conclusions, were conducted entirely by the authors without any involvement of AI. The authors accept full responsibility for the manuscript and affirm the authenticity, accuracy, validity, integrity, and overall content of the work.

REFERENCES

- Amirgaliyev, B., Mussabek, M., Rakhimzhanova, T., & Zhumadillayeva, A. (2025). A Review of Machine Learning and Deep Learning Methods for Person Detection, Tracking and Identification, and Face Recognition with Applications. *Sensors*, 25(5), 1410. <https://doi.org/10.3390/s25051410>
- Berlato, M., Binni, L., Durmus, D., Gatto, C., Giusti, L., Massari, A., Toldo, B. M., Cascone, S., & Mirarchi, C. (2025). Digital Platforms for the Built Environment: A Systematic Review Across Sectors and Scales. *Buildings*, 15(14), 2432. <https://doi.org/10.3390/buildings15142432>
- Carizo, G. (2026). Integration of Building Information Modeling (BIM) and Artificial Intelligence for Predicting Construction Project Risks. *Civil Engineering Science and Technology*, 2(1), 36–53. <https://doi.org/10.51903/s8j6ge78>
- Chen, Y., Liu, X., Zhu, B., Zhu, D., Zuo, X., & Li, Q. (2025). UAV Image-Based 3D Reconstruction Technology in Landslide Disasters: A Review. *Remote Sensing*, 17(17), 3117. <https://doi.org/10.3390/rs17173117>
- Chong, H.-Y., Yang, X., Goh, C. S., & Luo, Y. (2025). BIM and AI Integration for Dynamic Schedule Management: A Practical Framework and Case Study. *Buildings*, 15(14), 2451. <https://doi.org/10.3390/buildings15142451>
- Doan, D. T., Atencio, E., Muñoz La Rivera, F., & Alnajjar, O. (2025). A Systematic Literature Review of Building Information Modelling (BIM) and Offsite Construction (OSC) Integration: Emerging Technologies and Future Trends. *Applied Sciences*, 15(18), 9981. <https://doi.org/10.3390/app15189981>
- Dong, S., Wu, D., Kong, W., Liu, W., & Xia, N. (2025). Research on Intelligent Generation of Line Drawings from Point Clouds for Ancient Architectural Heritage. *Buildings*, 15(18), 3341. <https://doi.org/10.3390/buildings15183341>
- Fattore, C., Buldo, M., Priore, A., Porcari, S., Porcari, V. D., & De Fino, M. (2025). A Comprehensive Overview of Heritage BIM Frameworks: Platforms and Technologies Integrating Multi-Scale Analyses, Data Repositories, and Sensor Systems. *Heritage*, 8(7), 247. <https://doi.org/10.3390/heritage8070247>
- Habash, N., Alqumsan, A. A., & Zhou, T. (2025). Recent Real-Time Aerial Object Detection Approaches, Performance, Optimization, and Efficient Design Trends for Onboard Performance: A Survey. *Sensors*, 25(24), 7563. <https://doi.org/10.3390/s25247563>
- In-Hwa, K., Song-Il, P., Na, I. H., Young, K. A., & Chul, K. Y. (2025). Ethical AI in the Metaverse: A Mixed-Methods Study on Design Innovation, Social Implications, and Fairness Challenges. *International Journal of Graphic Design*, 3(1), 53–65.

<https://doi.org/10.51903/ijgd.v3i1.2791>

- Jia, J., Ma, H., & Zhang, Z. (2024). Integration of Industry Foundation Classes and Ontology: Data, Applications, Modes, Challenges, and Opportunities. *Buildings*, 14(4), 911. <https://doi.org/10.3390/buildings14040911>
- Li, Z., & Singh, B. (2026). Robust Occluded Object Detection in Multimodal Autonomous Driving: A Fusion-Aware Learning Framework. *Electronics*, 15(1), 245. <https://doi.org/10.3390/electronics15010245>
- Liu, X., He, Y., Li, J., Yan, R., Li, X., & Huang, H. (2024). A Comparative Review on Enhancing Visual Simultaneous Localization and Mapping with Deep Semantic Segmentation. *Sensors*, 24(11), 3388. <https://doi.org/10.3390/s24113388>
- Luo, H., Zhang, J., Liu, X., Zhang, L., & Liu, J. (2024). Large-Scale 3D Reconstruction from Multi-View Imagery: A Comprehensive Review. *Remote Sensing*, 16(5), 773. <https://doi.org/10.3390/rs16050773>
- Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision. *Technologies*, 12(2), 15. <https://doi.org/10.3390/technologies12020015>
- Omrany, H., Al-Obaidi, K. M., Husain, A., & Ghaffarianhoseini, A. (2023). Digital Twins in the Construction Industry: A Comprehensive Review of Current Implementations, Enabling Technologies, and Future Directions. *Sustainability*, 15(14), 10908. <https://doi.org/10.3390/su151410908>
- Park, S.-M., & Kang, L.-S. (2026). Visual Evaluation of Construction Schedule Progress by Linking Photographs and 4D Model. *Buildings*, 16(4), 733. <https://doi.org/10.3390/buildings16040733>
- Payawal, J. M., & Kim, D.-K. (2023). Image-Based Structural Health Monitoring: A Systematic Review. *Applied Sciences*, 13(2), 968. <https://doi.org/10.3390/app13020968>
- Qi, L., Liu, Y., Yu, Y., Chen, L., & Chen, R. (2024). Current Status and Future Trends of Meter-Level Indoor Positioning Technology: A Review. *Remote Sensing*, 16(2), 398. <https://doi.org/10.3390/rs16020398>
- Rathee, M., Bačić, B., & Doborjeh, M. (2023). Automated Road Defect and Anomaly Detection for Traffic Safety: A Systematic Review. *Sensors*, 23(12), 5656. <https://doi.org/10.3390/s23125656>
- Seid Ahmed, Y., & Amorim, F. L. (2025). Advances in Computer Numerical Control Geometric Error Compensation: Integrating AI and On-Machine Technologies for Ultra-Precision Manufacturing. *Machines*, 13(2), 140. <https://doi.org/10.3390/machines13020140>
- Setyadi, A., & Purwanto, P. (2025). Deteksi Kerusakan Beton Bertulang Jangka Panjang via IoT dan Model Machine Learning. *Jurnal Rekayasa Sipil Dan Arsitektur*, 1(2), 121–132. <https://doi.org/10.51903/xak5p815>

- Souza, C. E., Chinelli, C. K., Soares, C. A., & Longo, O. C. (2025). BIM as a Tool for Developing Smart Buildings in Smart Cities: Potentialities and Challenges. *Architecture*, 5(4), 103. <https://doi.org/10.3390/architecture5040103>
- Tan, Y., Liu, X., Jin, S., Wang, Q., Wang, D., & Xie, X. (2024). A Terrestrial Laser Scanning-Based Method for Indoor Geometric Quality Measurement. *Remote Sensing*, 16(1), 59. <https://doi.org/10.3390/rs16010059>
- Ullah, F., Nadeem, M., Abrar, M., Al-Razgan, M., Alfakih, T., Amin, F., & Salam, A. (2023). Brain Tumor Segmentation from MRI Images Using Handcrafted Convolutional Neural Network. *Diagnostics*, 13(16), 2650. <https://doi.org/10.3390/diagnostics13162650>
- Yoon, S., & Kim, H. (2025). Time-Series Image-Based Automated Monitoring Framework for Visible Facilities: Focusing on Installation and Retention Period. *Sensors*, 25(2), 574. <https://doi.org/10.3390/s25020574>
- Yu, L., Tang, L., & Mu, L. (2025). A Review of DEtection TRansformer: From Basic Architecture to Advanced Developments and Visual Perception Applications. *Sensors*, 25(13), 3952. <https://doi.org/10.3390/s25133952>