AI-Driven Construction Quality Management and Intelligent Infrastructure Systems.

Abstract

Quality failures in construction are expensive to put right, and in infrastructure they can linger for decades. Artificial intelligence (AI) is now used both to check quality while assets are built and to monitor their condition in service, yet these two bodies of research are usually studied in isolation. This paper reviews AI for construction quality management and for intelligent infrastructure systems, and proposes a framework that links the two. The corpus holds 30 peer-reviewed journal articles, the earliest from 2020 and the latest from 2025. The search was documented and rerun in OpenAlex on 16 November 2026 with eleven stated search strings, and the selection flow is reported in full. Studies were coded by phase, task, data, technique and type of evidence, and appraised with a short checklist. During construction, deep learning finds surface defects, comparison of scans with building information models measures deviations to within a millimetre, knowledge-based methods automate parts of code checking, and sensing combined with machine learning tracks concrete strength. In operation, machine learning supports damage diagnosis, drones and convolutional networks speed up inspection, and digital twins bring data together for maintenance decisions. In both phases the evidence rests mostly on prototypes and single cases, models often lose accuracy on independent data, and the value of digital investment is seldom shown. The proposed quality continuity framework carries as-designed, as-built and as-operated information forward through shared data, explainable models and value-based decisions. It has not been validated; four propositions are set out, together with a plan for testing them.

Keywords: Artificial Intelligence; Construction Quality Management; Intelligent Infrastructure Systems; Structural Health Monitoring; Building Information Modelling; Digital Twin; Quality Continuity Framework

UNDER PEER REVIEW IN IJAR

1. Introduction

Construction quality tends to be judged twice: once when work is inspected and accepted, and again, often years later, when defects show up in use. Between those two points sit many decisions about materials, tolerances, workmanship and maintenance. A defect that slips past inspection is buried in finished work, and a failure to comply with the design or the regulations may only surface when it is expensive to fix. A systematic review of digital technology for construction quality management called for better handling of defects in concealed work, stronger prevention before construction starts, post-completion testing of how products function, and more research on compliance inspection (Luo et al., 2022).

AI offers new ways of meeting these needs. Machine learning has been described as central to smart construction, with openings in site supervision, automatic detection and intelligent maintenance, although labelled data are hard to collect on complex sites (Xu et al., 2021). Chen et al. (2025) analysed 594 papers quantitatively and 91 qualitatively in a critical review of AI in infrastructure construction. They found that research concentrated on safety monitoring and control and on process management, led by machine learning, computer vision and language processing. Quality was not one of the main themes they identified.

Research on intelligent infrastructure has meanwhile built AI methods for assets in service. Structural health monitoring (SHM) systems use sensors and data-driven models to detect and classify damage, and machine learning for these systems has grown quickly (Gomez-Cabrera & Escamilla-Ambrosio, 2022; Zinno et al., 2022). Drones and imaging sensors are taking over parts of manual bridge and pavement surveys. Digital twins promise to combine monitoring data with a model of the asset to guide maintenance, but in architecture, engineering and construction they remain largely at the prototype stage (Pregnolato et al., 2022).

The two streams share methods, data types and goals, yet they seldom meet. As-built geometry, material test results and inspection records could form the baseline for monitoring in operation, and evidence of deterioration could in turn inform quality requirements for new work. Earlier reviews have mapped AI in construction, SHM and digital twins, but each concentrates on one phase (Section 8, Table 3). This paper makes two contributions. It brings AI for construction-phase quality and for operation-phase condition management into one review with a documented, repeatable search. And it proposes a quality continuity framework, with four

UNDER PEER REVIEW IN IJAR

propositions and a plan for testing them, that treats quality as a thread running from design through to maintenance. The framework is a proposal derived from the literature, not a validated model.

2. Research Aim, Objectives and Questions

The aim is to examine how AI supports quality in construction and in the operation of intelligent infrastructure, and to develop a framework linking quality information across the lifecycle. The objectives are (1) to identify the AI techniques and data sources used for quality tasks during construction; (2) to identify those used for monitoring and managing infrastructure in service; (3) to analyse the conditions, such as data, validation, explainability and organisational value, that hold back or enable adoption; and (4) to propose a framework connecting the two phases and to state how it could be tested.

Three questions follow. RQ1: Which AI techniques and data sources are used for quality management during construction, and what do studies report about their performance? RQ2: How is AI used in intelligent infrastructure systems for monitoring, inspection and asset decisions, and with what results? RQ3: What challenges cut across both phases, and how might construction quality management and infrastructure operation be connected in one framework?

3. Methodology

We used a systematic integrative review, because the question spans technical studies, reviews and conceptual work and the aim is to build theory rather than estimate an effect. The search and screening logs are available on request.

3.1 Sources, Search Strings and Dates

The corpus was first assembled through exploratory, relevance-ranked searching of OpenAlex, an open index of scholarly works, on 13 November 2026. Logs of those searches were not kept. To make the process reproducible, we reran a structured search of OpenAlex on 16 November 2026 and logged every query, hit count and exported record. Crossref was used only to verify bibliographic metadata, publication years and DOIs; it was not used as a search source. Scopus and Web of Science were not searched.

UNDER PEER REVIEW IN IJAR

Every query was limited to English-language articles published from 2020 to 2025 and matched against titles and abstracts. Writing AI for ("artificial intelligence" OR "machine learning" OR "deep learning" OR "computer vision" OR "neural network"), the eight core strings were: S1, AI AND ("construction quality" OR "quality inspection" OR "defect detection") AND (construction OR building OR precast); S2, ("automated compliance checking" OR "code compliance" OR "rule checking") AND ("building information modelling" OR "building information modeling" OR BIM); S3, ("point cloud" OR "laser scanning" OR LiDAR OR "scan-to-BIM") AND ("quality control" OR "quality inspection" OR "dimensional quality" OR deviation) AND (BIM OR "building information"); S4, AI AND ("concrete strength" OR "compressive strength" OR "concrete curing") AND (prediction OR monitoring); S5, AI AND "structural health monitoring" AND (bridge OR building OR infrastructure); S6, (AI terms OR "unmanned aerial vehicle" OR UAV OR drone) AND ("bridge inspection" OR "pavement distress" OR "crack detection") AND (concrete OR bridge OR pavement); S7, "digital twin" AND (infrastructure OR bridge OR "civil engineering" OR construction) AND ("asset management" OR maintenance OR "structural health"); and S8, AI AND (construction OR "construction management" OR "infrastructure construction") AND (review OR overview).

Six of the 30 studies fell outside every core string, because they deal with enabling techniques rather than AI as such. We therefore added three supplementary strings and report this openly: S9, ("Internet of Things" OR IoT OR "wireless sensor") AND concrete AND (curing OR "compressive strength") AND monitoring; S10, "digital twin" AND (definition OR workflow OR concept) AND (construction OR "civil engineering" OR infrastructure); and S11, ("digital technology" OR "artificial intelligence") AND ("quality management" OR "construction management") AND construction. The eleven strings returned 11,838 hits in total (S1 454; S2 92; S3 103; S4 2,937; S5 1,161; S6 1,361; S7 1,206; S8 3,134; S9 37; S10 1,080; S11 273). The 50 most relevant records from each string were exported, or all 37 for S9, giving 537 records.

3.2 Screening and Eligibility

To be included, a record had to be an English-language, peer-reviewed journal article from 2020 to 2025 applying an AI or AI-enabling data technique (such as sensing, scanning, rule-based checking or digital twins) to a quality, inspection, monitoring or asset-condition task in buildings or civil infrastructure. Broad overviews of AI in construction were admitted where they covered

UNDER PEER REVIEW IN IJAR

quality or inspection. Conference papers, editorials, retracted items and records without an obtainable abstract were excluded, as were studies concerned only with worker safety, productivity, cost or scheduling.

Removing duplicates (15 by DOI and 3 by title) left 519 records. Automated checks then excluded 2 retracted items, 9 without a DOI and 10 from non-journal sources. The remaining 498 records were screened on title and abstract, or on title alone where OpenAlex held no abstract. Screening excluded 132 records: 44 fell outside the built environment, 82 did not address a quality, inspection, monitoring or asset-condition task, 5 used no AI or data-driven technique, and 1 was a retraction notice. Of the 366 that passed, 91 had no abstract in OpenAlex; abstracts were sought in Crossref and Semantic Scholar, found for 15, and the other 76 were excluded. That left 290 records assessed for eligibility.

All 290 met the criteria. Twenty-two of them were studies already in the corpus. The other 268 were not added. The corpus was purposive and capped at 30, built so that each quality task had at least two sources and both phases were represented; those 268 mostly repeated tasks already covered. The remaining eight studies came from the earlier exploratory searching. Each was checked on 16 November and found within the hit set of at least one core string, but below the top-50 export. Figure 1a shows this flow.

UNDER PEER REVIEW IN IJAP

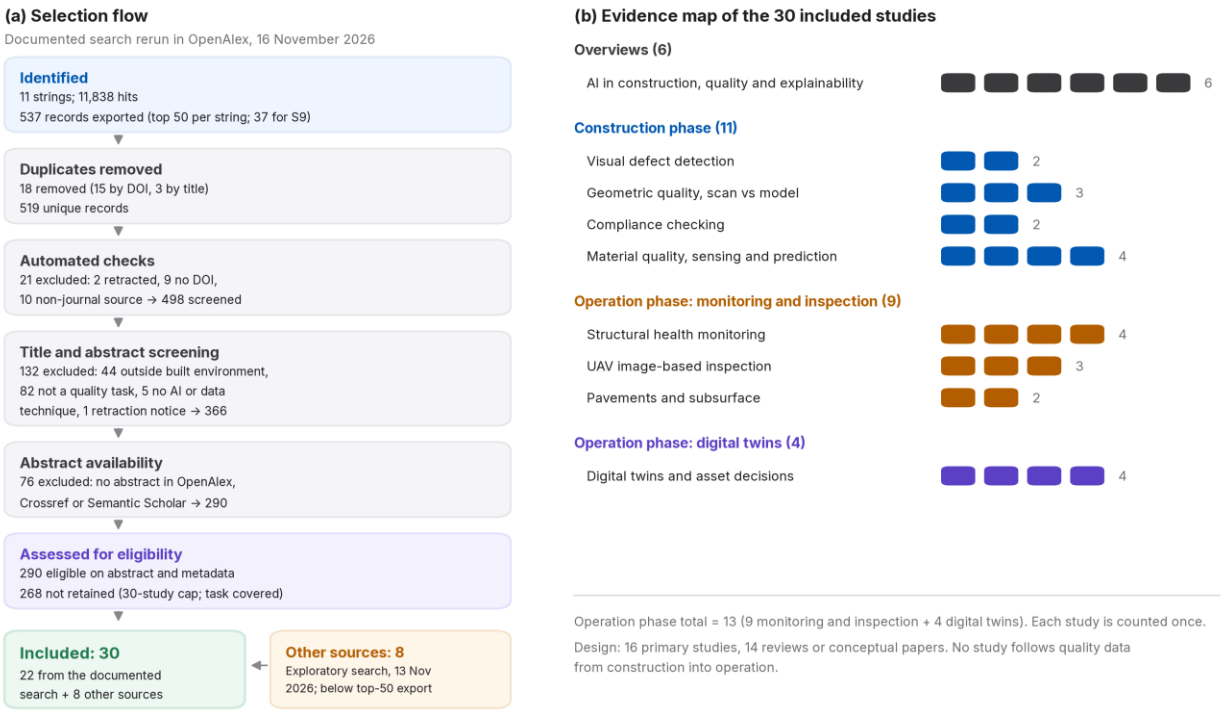


Figure 1. (a) Selection flow (OpenAlex, 16 November 2026); (b) evidence map of the 30 included studies. Operation phase: 9 monitoring and inspection + 4 digital twins = 13.

Eligibility was judged on abstracts and metadata, not full texts. One reviewer, the author, did all the screening, which is a limitation (Section 9). Where a record was ambiguous, for instance a scan-to-BIM study that might concern either modelling or quality, it was judged by its stated purpose: included if it set out to check conformance, monitor condition or support a quality or maintenance decision, and excluded otherwise. Each decision and its reason were recorded in the screening log.

3.3 Classification, Appraisal and Synthesis

An article counted as a review if it synthesised published studies, and as a primary study if it reported new data, a new system or a worked application. Two conceptual papers (Aragón et al., 2025; Eber, 2020) were grouped with the reviews. On this basis the corpus holds 14 reviews or conceptual papers and 16 primary studies. For Figure 1, each article was placed once, by its main phase and task: six overviews of AI in construction, eleven construction-phase studies and thirteen operation-phase studies. The thirteen are nine on monitoring and inspection and four on digital twins.

UNDER PEER REVIEW

We appraised studies proportionately, with a short checklist adapted from mixed-methods and critical-appraisal tools and coded from abstracts. For primary studies it asked whether the aim was clear, whether the data were described, whether performance was measured against a baseline or held-out data, and whether field or site data were used. All 16 stated a clear aim. Ten reported evaluation against a baseline or independent test data, two did so partly, and four offered a demonstration only. Nine used field or site data, five used laboratory data, public databases or worked examples, and for two this was unclear. Only two covered more than one site or case, and only one mentioned limitations in its abstract. Of the 14 reviews, seven described a defined corpus or systematic procedure and seven were narrative or conceptual. These ratings set how much weight each finding carries: a claim resting on one prototype is reported more cautiously than a pattern that recurs.

Reviews and primary studies can overlap. To check, we compared the reference lists recorded in OpenAlex for the 30 articles. Only one included review cites an included primary study: Hakimi et al. (2023) cite Pregnolato et al. (2022). No article is counted twice in Figure 1. Review findings are used to describe patterns in a field, never as extra confirmation of a claim already drawn from a primary study they summarise.

Each article was coded for phase, task, asset type, data source, technique, type of evidence and reported results. Findings were synthesised by theme within each phase and then across phases. Claims were checked against what each study reports, and numbers are given as published, without recalculation or comparison across datasets.

4. AI for Quality Management During Construction

Luo et al. (2022) urged attention to quality culture and behaviour-based assessment as well as technology. Four tasks dominate the construction-phase studies: finding surface defects, measuring geometric deviation, checking compliance with codes, and controlling material quality. Table 1 summarises selected primary studies.

Table 1. Selected primary studies (results as stated by the authors)

Study	Phase and task	Data and technique	Reported result
Lee et al. (2022)	Construction; precast member cracks	Video images; CNN and object detection	13 minutes less total inspection time; records stored in database

UNDER PEER REVIEW IN IJAR

Study	Phase and task	Data and technique	Reported result
Vasquez et al. (2024)	Construction; window frame defects	Phone and robot images; enhanced U-Net	F1 score up 7.43%; IoU up 15.1% versus standard U-Net
Tan et al. (2020)	Construction; prefabricated unit geometry	LiDAR and BIM comparison	0.7 mm (structural) and 0.9 mm (MEP) accuracy
Bassier et al. (2020)	Construction; site positioning errors	Point cloud and BIM matching	Errors computed despite drift, noise and clutter
Ismail et al. (2023)	Design; fire safety compliance	256 clauses; native BIM; visual programming	User validation means of 4.09 to 4.96
Lo et al. (2021)	Construction; concrete curing	IoT moisture sensing and control	Better curing quality and less supervision time
Elshaarawy et al. (2024)	Construction; concrete strength	1,030 records; ensemble learning; SHAP	R^2 0.966; RMSE 3.06 MPa; age most influential
Ayele et al. (2020)	Operation; bridge damage	UAV images; 3D models; deep learning	Methodology demonstrated on a 140 m bridge
Yang et al. (2022)	Operation; pavement subsurface	3D GPR images; detection and classification	87.71% localisation precision; 63.63% five-class accuracy

4.1 Visual Defect Detection

Defects are still mostly found by eye, and inspectors lean heavily on experience. For precast concrete members, Lee et al. (2022) sorted inspection items from existing checklists and specifications, picked out those that could be automated, and built a crack detection system on a convolutional neural network (CNN) and object detection, served through mobile and web servers. Cracks were classed by width and length thresholds. Total work time fell by 13 minutes, and records went straight into a database.

Installed window frames are harder: they are reflective and sit in varied surroundings. Vasquez et al. (2024) combined image enhancement, data augmentation and a U-Net model, and raised the F1 score by 7.43% and intersection over union by 15.1% against a standard U-Net, using five datasets captured by mobile phones and an inspection robot. They noted that the method still needs trying on a wider range of sites. Deep learning can clearly automate parts of visual inspection, but scarce data and variable sites limit how far results travel.

UNDER PEER REVIEW IN IJAR

4.2 Geometric Quality and Scan-versus-Model Comparison

Dimensional errors drive rework, particularly in modular construction. Comparing laser-scanned point clouds with building information models (BIM) gives an objective alternative to tape and total station. For prefabricated bathroom units, Tan et al. (2020) compared light detection and ranging (LiDAR) data with the BIM model for both structural and mechanical, electrical and plumbing elements, reporting accuracies of 0.7 mm and 0.9 mm and large savings in time and labour.

Bassier et al. (2020) worked at the scale of a whole site, matching point clouds to as-designed BIM elements and computing each object's positioning error despite drift, noise, clutter and referencing errors; results were colour-coded inside the BIM software. On a removable floodwall, combining BIM with laser scanning gave a benchmark for tolerance control that was faster and more accurate than manual or contact measurement (Li et al., 2020). Geometric checking can thus move from sampling a few points to assessing every element. Not all of this is AI in the narrow sense, but it yields the structured as-built data that learning-based inspection and digital twins need.

4.3 Automated Compliance Checking

Code checking has traditionally fallen to expert reviewers. Peng and Liu (2023) used natural language processing to turn specification clauses into a knowledge graph, exported BIM data into a resource description framework, and generated review reports through mapping and review rules, showing feasibility on an example. Ismail et al. (2023) translated 256 fire safety clauses from Malaysian regulations in a semi-automated process and built a prototype with native BIM models and visual programming; users rated it between 4.09 and 4.96 on average.

In both studies the translation of regulations into machine-readable rules is the core of the method. Ismail et al. (2023) found that a semi-automated translation drew regulatory experts into the work and made the system easier to use. The studies also mark where automation currently stops. Rule interpretation is still partly manual, and validation has rested on examples or user ratings rather than large comparisons with expert review.

UNDER PEER REVIEW IN IJAR

4.4 Material Quality: Sensing and Prediction

Concrete quality depends on the mix and on curing conditions that are hard to supervise. In site experiments, an Internet of Things (IoT) system that monitored and controlled moisture in hardening concrete beat traditional curing on both curing quality and supervision time (Lo et al., 2021). Iqbal et al. (2024) linked low-cost wireless sensors to a cloud platform and BIM to monitor early-age compressive strength remotely and in real time. The result was a digital twin that could inform decisions such as when to strike formwork or end curing.

Machine learning is also used to predict strength from mix proportions. On 1,030 experimental records, a categorical gradient boosting model was the most accurate of those tested, with a test coefficient of determination of 0.966 and a root-mean-square error of 3.06 MPa; Shapley additive explanations (SHAP) showed concrete age to be the most influential variable (Elshaarawy et al., 2024). Ekanayake et al. (2022) found extreme gradient boosting and Laplacian kernel ridge regression best among their algorithms, and SHAP interpretations agreed with known concrete behaviour.

On RQ1, then: construction-phase quality management uses deep learning on images for defects, point cloud and BIM comparison for geometry, language processing and knowledge graphs for compliance, and IoT sensing with ensemble learning for materials. Results are promising, but most come from prototypes, single components or laboratory data.

5. Intelligent Infrastructure Systems in Operation

Once an asset is in service, quality management becomes condition management: diagnosis from monitoring data, inspection from imaging platforms, and asset decisions supported by digital twins.

5.1 Structural Health Monitoring and Diagnosis

Machine learning is now a central tool for reading SHM data. Across 68 articles, Gomez-Cabrera and Escamilla-Ambrosio (2022) found artificial neural networks, CNNs and support vector machines, combined with preprocessing, used to grade damage in a hierarchical classification, and concluded that they improve the reliability and performance of monitoring systems. Zinno et al. (2022) framed AI-based SHM of bridges as part of intelligent transport systems across

UNDER PEER REVIEW IN IJAR

construction, management and maintenance, with IoT and big data analytics changing how monitoring is done.

Wang et al. (2022) reviewed railway bridge monitoring, covering fibre optic and vision-based sensing, CNNs and transfer learning for condition assessment, and digital twins. Bao et al. (2025) described machine learning in data cleaning, modal parameter estimation, damage identification, digital twins and reliability assessment, and pointed out that diagnosing complex bridges from badly incomplete data is an inherently high-dimensional problem.

5.2 Image-Based Inspection with Unmanned Aerial Vehicles

Visual bridge inspection is slow, subjective and often incomplete, and typically recurs every two years (Feroz & Abu Dabous, 2021). Their review of 65 papers found that unmanned aerial vehicles (UAVs) carrying visual, infrared and LiDAR sensors improve access, cost and safety and avoid closing roads. On a 140 m prestressed concrete bridge in Norway, Ayele et al. (2020) combined UAV data capture, three-dimensional models kept as permanent geometric records and deep learning for damage identification, with the aim of cutting inspection time, cost and the maintenance backlog.

Yu et al. (2022) tuned a deep CNN with an enhanced chicken swarm algorithm to diagnose cracks in concrete image patches and compared it with other techniques on statistical indicators. In our reading, the practical worth of such methods depends on reliable image capture and on linking detected damage to a condition rating, and the primary studies here have not yet shown that at scale.

5.3 Pavements and Subsurface Distress

A random-effects meta-analysis of deep learning for pavement distress detection noted that leading models did poorly on independent data in an object detection competition, and that performance measures were often computed in unclear ways (Zihan et al., 2023). Below the surface, deep learning on horizontal slices of three-dimensional ground penetrating radar data located anomalous waveforms with 87.71% precision, but sorting them into five classes of structural distress reached only 63.63% accuracy (Yang et al., 2022). Finding hidden defects, in other words, is easier than diagnosing them.

UNDER PEER REVIEW IN IJAP

5.4 Digital Twins and Asset Decisions

Digital twins are meant to join data and models so that owners can act on condition information. Pregnolato et al. (2022) proposed a step-by-step workflow for twinning an existing asset and applied it to the Clifton Suspension Bridge, noting competing definitions and unclear processes in the sector. Hakimi et al. (2023) reviewed 105 papers and found that predictive maintenance often relies on single-source data; they named data heterogeneity, interoperability, data quality and algorithm choice as gaps and proposed multilayer data fusion using open BIM, geographic information systems and extended industry foundation classes.

Aragón et al. (2025) argued that a twin is only as useful as its representation of the asset is accurate, and proposed first steps towards a European definition and standard processes. Vieira et al. (2024) observed that the value of digital twins is often claimed but rarely shown, and offered a value-based analysis for weighing twinning opportunities against investment in physical assets, tested on three rail and road cases.

On RQ2: intelligent infrastructure systems apply machine learning to monitoring data, deep learning to images from UAVs and radar, and digital twins to pull information together for maintenance. Reviews dominate this strand, the primary studies are case-based, and evidence of decision value is thin.

6. Cross-Cutting Synthesis

Read together, the two phases share six challenges, set out in Table 2. Data come first. Labelled data are hard to get on site (Xu et al., 2021), defect datasets are small and tied to their surroundings (Vasquez et al., 2024), and monitoring data are heterogeneous and often incomplete (Bao et al., 2025; Hakimi et al., 2023). Jiang and Messner (2023), reviewing 103 journal papers on computer vision across construction and asset management, likewise found practical limits on vision-based methods in the industry.

Table 2. Six cross-cutting challenges and the responses built into the framework

Challenge	Evidence from the corpus	Response in the framework	Key sources
1. Scarce and heterogeneous data	Labelled site data hard to obtain; monitoring data incomplete and multi-source	Shared data layer; as-built data structured for reuse	Xu et al. (2021); Hakimi et al. (2023); Bao et al. (2025)

UNDER PEER REVIEW IN IJAR

Challenge	Evidence from the corpus	Response in the framework	Key sources
2. Weak generalisation	Models underperform on independent datasets; single-site tests	Independent and cross-site validation rule	Zihan et al. (2023); Vasquez et al. (2024)
3. Limited explainability	Explanations rare outside material prediction	Explainable outputs required for decisions	Love et al. (2023); Ekanayake et al. (2022)
4. Role of human judgement	Rule translation and prioritisation need experts	Human control in the decision layer	Eber (2020); Ismail et al. (2023)
5. Unproven value	Benefits of digital twins seldom demonstrated	Value-based appraisal of digital investments	Vieira et al. (2024); Pregnolato et al. (2022)
6. Broken lifecycle links	No study follows quality data into operation	Structured handover baseline and feedback loop	Bassier et al. (2020); Iqbal et al. (2024)

Second, generalisation and validation. Pavement models that did well on their own data did badly on independent data (Zihan et al., 2023), and construction-phase authors call for testing on more sites (Vasquez et al., 2024). Our appraisal points the same way: only two of the 16 primary studies covered more than one site or case, and four were demonstrations without a quantitative test.

Third, explainability and trust. Explainable AI has had little attention in construction, and Love et al. (2023) suggested that attending to what stakeholders need, and to data fusion, could ease scepticism about adoption. The material studies show that methods such as SHAP can reveal whether predictions square with engineering knowledge (Ekanayake et al., 2022; Elshaarawy et al., 2024). Comparable explanations are rare in defect detection and SHM.

Fourth, human judgement. Eber (2020) argued that AI cannot supply the ability to set priorities or the creativity behind new approaches, so its role in construction management is strong support rather than replacement. Compliance checking bears this out, since rule translation still depends on experts (Ismail et al., 2023; Peng & Liu, 2023).

Fifth, unproven value. Digital twins and monitoring systems compete with physical works for funding, and their value is rarely demonstrated (Vieira et al., 2024). Sixth, broken lifecycle links. Construction data are seldom designed for reuse in operation, even though as-built geometry and early-age material data are just what a monitoring system needs as a baseline (Bassier et al., 2020; Iqbal et al., 2024).

UNDER PEER REVIEW IN IJAR

Figure 1b maps the corpus by phase and task. Construction-phase evidence is spread thinly over four tasks with two to four studies each. Operation-phase evidence leans on SHM and digital twin reviews, and the four digital twin studies are counted within the operation phase (9 + 4 = 13). No study in the corpus follows quality information from construction into operation.

7. A Quality Continuity Framework

Figure 2 shows the framework proposed in response to that gap. Its central idea is quality continuity: the information that establishes quality during construction becomes the baseline for monitoring in operation, and lessons from operation feed back into design and construction.

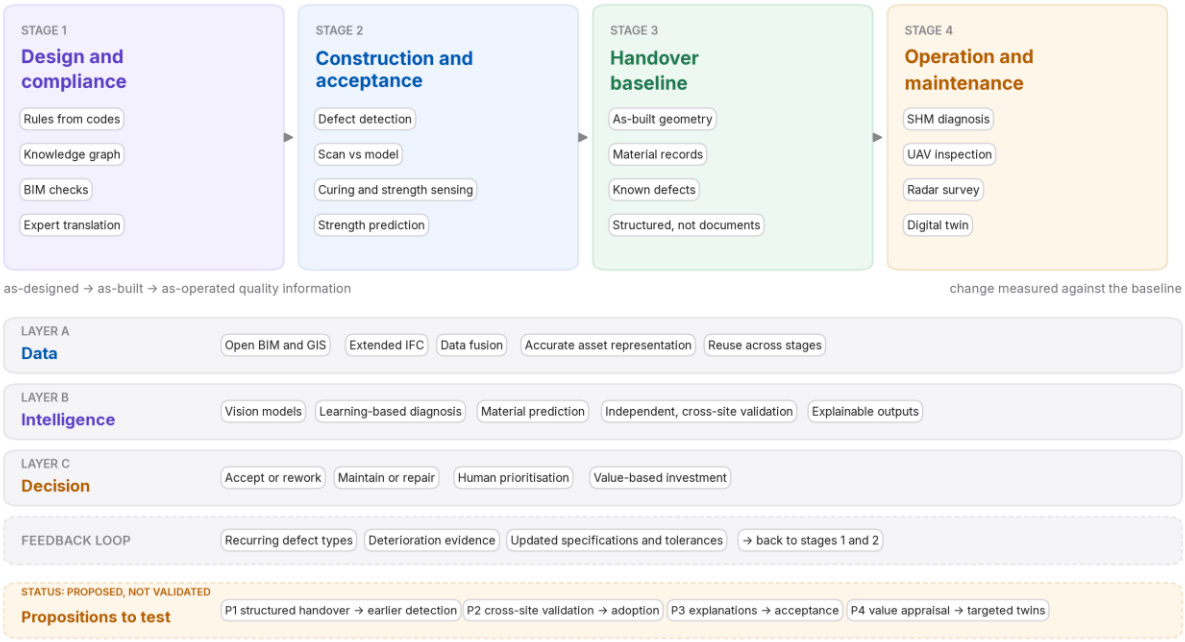


Figure 2. Proposed quality continuity framework. Propositions P1–P4 are untested.

7.1 Structure

The framework has four lifecycle stages and three layers. At design and compliance, codes are checked through knowledge graphs and rule translation, with experts involved (Ismail et al., 2023; Peng & Liu, 2023). During construction and acceptance, defect detection, scan-versus-model comparison and material sensing build an evidence record of as-built quality (Bassier et al., 2020; Lee et al., 2022; Lo et al., 2021; Tan et al., 2020). At handover, that record is delivered as a structured baseline of geometry, material data and known defects rather than as a stack of documents. In operation and maintenance, monitoring, UAV inspection and radar survey

measure change against the baseline (Ayele et al., 2020; Feroz & Abu Dabous, 2021; Yang et al., 2022).

The data layer carries shared, interoperable information between stages, drawing on open BIM and extended industry foundation classes (Hakimi et al., 2023) and on accurate asset representation (Aragón et al., 2025). Its one firm requirement is that data captured for acceptance are structured for reuse. The intelligence layer holds the models, under two rules: validate on data independent of training, across sites where possible (Zihan et al., 2023), and make outputs explainable to the engineers who act on them (Love et al., 2023). The decision layer turns outputs into acceptance, rework, maintenance or investment, keeps prioritisation in human hands (Eber, 2020), and uses value-based analysis to decide where digital spending is justified (Vieira et al., 2024). A feedback loop returns recurring defect types and similar findings to specifications.

Adoption can be staged. An organisation might start with tasks that already give objective measurements, such as point cloud checks and sensor-based curing control, and insist on structured outputs. Learning-based inspection can follow once local data allow independent validation, and full digital twins only where a value case supports them. The strength twin of Iqbal et al. (2024) already creates a record that later stages could reuse.

7.2 Evidence, Interpretation and Propositions

The framework mixes three kinds of statement, and they need to be kept apart. Some are findings supported by the reviewed evidence: geometric and material checks produce objective, structured measurements (Tan et al., 2020; Elshaarawy et al., 2024); vision models can lose accuracy on independent data (Zihan et al., 2023); and the value of digital twins is seldom demonstrated (Vieira et al., 2024). Others are our interpretation: that these findings add up to a broken link between construction and operation, and that a structured handover baseline is the natural place to repair it. No reviewed study tests that claim directly.

The rest are propositions that still need testing. P1: Assets handed over with structured as-built quality data will have deterioration detected earlier and at lower cost than assets handed over with documents only. P2: AI inspection models validated across independent sites will be adopted more widely, and trusted for longer, than models validated on a single dataset. P3:

UNDER PEER REVIEW IN IJAP

Explanations of model outputs will raise engineers' acceptance of AI-based quality decisions without making them over-rely on the models. P4: Value-based appraisal will lead to more targeted investment in digital twins than technology-led adoption. P1 and P4 rest mainly on interpretation; P2 and P3 extend patterns the evidence already shows.

7.3 How the Framework Could Be Tested

Each proposition implies a different test. P1 calls for a longitudinal comparison of matched assets, some handed over with structured baselines and some with documents, measuring time to detect deterioration and inspection cost. P2 could be tested by external validation of the same models on several sites, followed by tracking their use. P3 suits a controlled experiment in which engineers judge model outputs with and without explanations, measuring acceptance and whether they catch the model's errors. P4 could be examined through case studies comparing portfolios that did and did not use value-based appraisal. A first step could be a Delphi study with contractors, asset owners and inspectors to test whether the layers and stages are complete and workable. Until such evidence exists, the framework should be read as a research agenda.

8. Discussion

Table 3 sets this review against earlier ones. Reviews of AI in construction map techniques and applications but say little about quality and almost nothing about operation. SHM, inspection and digital twin reviews concentrate on assets in service, with little on the construction baseline. Luo et al. (2022) come closest, yet their focus is digital technology for quality management rather than AI across the lifecycle. The novelty here lies in placing the two strands in one analysis and in treating handover as the hinge between them.

Table 3. Comparison with prior reviews in the corpus and the gap this paper addresses

Prior research	Main focus	Gap this paper addresses
AI in construction: Chen et al. (2025); Xu et al. (2021); Jiang and Messner (2023)	AI and computer vision applications across construction, led by safety and process management	Quality is a minor theme and the operation phase is treated separately
Digital quality management: Luo et al. (2022)	Digital technologies for quality management in construction	Centred on the project; AI-based condition management in service is not the focus
Explainable AI: Love et al. (2023)	Taxonomy of explainable AI and research opportunities in construction	Explainability is not tied to specific quality and inspection tasks

UNDER PEER REVIEW IN IJAR

Prior research	Main focus	Gap this paper addresses
SHM and inspection: Gomez-Cabrera and Escamilla-Ambrosio (2022); Zinno et al. (2022); Feroz and Abu Dabous (2021); Zihan et al. (2023)	Machine learning for damage diagnosis; UAV and vision-based inspection	Centred on assets in service; the as-built quality baseline is not the focus
Digital twins: Hakimi et al. (2023); Aragón et al. (2025)	Data integration for predictive maintenance; definitions and processes	Construction-phase quality records as an input to the twin are not the focus
This paper	AI across construction quality and operation, linked through handover	Integrated quality continuity framework with testable propositions and a testing plan

The findings confirm earlier work and add to it. Like Chen et al. (2025), we find quality outside the main focus of AI research in construction. The gaps Luo et al. (2022) found, namely concealed work, prevention before construction and post-completion testing, match closely the missing links between construction and operation found here. The evidence is also thinner than the attention AI receives suggests: reviews make up nearly half the corpus.

For practitioners, the quickest gains are likely where AI and sensing give objective measurements, such as scan-versus-model checks and material monitoring. Inspection models should be tried on an organisation's own data before anyone relies on them. Asset owners should specify at procurement the as-built data they will need in operation, so that the quality evidence built up during construction is not lost at handover.

For researchers, the first need is longitudinal studies that follow assets from construction into operation, the only way to test P1. Shared, independent datasets and transparent reporting of performance would make results comparable (Zihan et al., 2023). Explainable methods should spread from material prediction to defect detection and SHM (Love et al., 2023), and the value of AI and digital twins should be measured in the terms owners use for investment (Vieira et al., 2024).

9. Limitations

The review has several limitations. Only one source, OpenAlex, was searched; Scopus, Web of Science, conference proceedings and grey literature were not. Conference papers are common in this field, so relevant work has certainly been missed. Only English-language articles were included, and publication bias towards positive results is likely.

UNDER PEER REVIEW IN IJAR

The documented search was run on 16 November 2026, after the corpus had been assembled through exploratory searching whose logs were not kept. It recovered 22 of the 30 studies; the other eight lie within its hit sets but below the export cut-off. Three of the strings were added after we saw that six included studies fell outside the core strings. The documented search therefore confirms and describes the corpus but did not generate it, and others rerunning it would not necessarily choose the same 30 articles. Lower-ranked records beyond the top 50 per string were never screened.

One reviewer did all the screening, coding and appraisal, so there was no check on agreement and some judgements may be idiosyncratic. Eligibility and appraisal relied on abstracts and metadata rather than full texts. Many eligible records were not retained because of the 30-study cap, so the corpus is illustrative of each task, not exhaustive. Reviews make up almost half of it, which widens coverage at the cost of second-hand evidence. The primary studies use different datasets and metrics, and most come from a few countries and asset types. Finally, the framework comes from the literature and has not been tested.

10. Conclusion

This review examined how AI supports construction quality management and intelligent infrastructure systems, drawing on 30 journal articles from 2020 to 2025 selected through a documented search. During construction, deep learning, scan-versus-model comparison, knowledge graphs and sensing with machine learning address defects, geometry, compliance and materials. In operation, machine learning, UAV inspection, ground penetrating radar and digital twins support diagnosis, inspection and asset decisions. The evidence in both phases is mostly prototype-based. Models often generalise poorly, explanation is rare, value is seldom shown, and quality information is not carried from construction into operation.

The paper's contribution is to join these two strands, which earlier reviews treat separately, and to propose a quality continuity framework that runs from as-designed through as-built to as-operated information. The framework separates what the evidence supports from our interpretation and from four propositions that remain untested, and it sets out how each could be tested. It is offered as a basis for empirical work, not as a validated model.

UNDER PEER REVIEW IN IJAR

Author Contributions. Syed Qalab Abbas Syed is the sole author and carried out the conception, search, screening, appraisal, synthesis, framework development and writing.

Funding. This research received no external funding.

Conflicts of Interest. The author declares no conflicts of interest.

Data Availability. No new data were created or analysed in this study. The review is based on published literature. The search strings, hit counts, screening log and appraisal coding are available from the author on request.

Ethics Statement. This study did not involve human participants, animals or personal data, and ethical approval was not required.

Acknowledgements. None.

UNDER PEER REVIEW IN IJAR

References

1. Aragón, A., Arquier, M., Tokdemir, O. B., Enfedaque, A., Alberti, M. G., Lieval, F., Loscos, E., Pavón, R. M., Novischi, D. M., Legazpi, P. V., & Yagüe, Á. (2025). Seeking a definition of digital twins for construction and infrastructure management. *Applied Sciences*, *15*(3), Article 1557. https://doi.org/10.3390/app15031557

2. Ayele, Y. Z., Aliyari, M., Griffiths, D., & Droguett, E. L. (2020). Automatic crack segmentation for UAV-assisted bridge inspection. *Energies*, *13*(23), Article 6250. https://doi.org/10.3390/en13236250

3. Bao, Y., Sun, H., Xu, Y., Guan, X., Pan, Q., & Liu, D. (2025). Recent advances in structural health diagnosis: A machine learning perspective. *Advances in Bridge Engineering*, *6*(1), Article 7. https://doi.org/10.1186/s43251-024-00155-z

4. Bassier, M., Vincke, S., De Winter, H., & Vergauwen, M. (2020). Drift invariant metric quality control of construction sites using BIM and point cloud data. *ISPRS International Journal of Geo-Information*, *9*(9), Article 545. https://doi.org/10.3390/ijgi9090545

5. Chen, K., Zhou, X., Bao, Z., Skibniewski, M. J., & Fang, W. (2025). Artificial intelligence in infrastructure construction: A critical review. *Frontiers of Engineering Management*, *12*(1), 24–38. https://doi.org/10.1007/s42524-024-3128-5

6. Eber, W. (2020). Potentials of artificial intelligence in construction management. *Organization, Technology and Management in Construction: An International Journal*, *12*(1), 2053–2063. https://doi.org/10.2478/otmcj-2020-0002

7. Ekanayake, I., Meddage, D., & Rathnayake, U. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials*, *16*, Article e01059. https://doi.org/10.1016/j.cscm.2022.e01059

8. Elshaarawy, M. K., Alsaadawi, M. M., & Hamed, A. K. (2024). Machine learning and interactive GUI for concrete compressive strength prediction. *Scientific Reports*, *14*(1), Article 16694. https://doi.org/10.1038/s41598-024-66957-3

UNDER PEER REVIEW IN IJAR

9. Feroz, S., & Abu Dabous, S. (2021). UAV-based remote sensing applications for bridge condition assessment. *Remote Sensing*, *13*(9), Article 1809. https://doi.org/10.3390/rs13091809

10. Gomez-Cabrera, A., & Escamilla-Ambrosio, P. J. (2022). Review of machine-learning techniques applied to structural health monitoring systems for building and bridge structures. *Applied Sciences*, *12*(21), Article 10754. https://doi.org/10.3390/app122110754

11. Hakimi, O., Liu, H., Abudayyeh, O., Houshyar, A., Almatared, M., & Alhawiti, A. (2023). Data fusion for smart civil infrastructure management: A conceptual digital twin framework. *Buildings*, *13*(11), Article 2725. https://doi.org/10.3390/buildings13112725

12. Iqbal, F., Ahmed, S., Tariq, M. A. B., Waqas, H. A., Al-Ammar, E. A., Wabaidur, S. M., & Fawad, M. (2024). BIM-IoT integration for remote real-time concrete compressive strength monitoring. *Ain Shams Engineering Journal*, *15*(7), Article 102863. https://doi.org/10.1016/j.asej.2024.102863

13. Ismail, A. S., Ali, K. N., Iahad, N. A., Kassem, M. A., & Al-Ashwal, N. T. (2023). BIM-based automated code compliance checking system in Malaysian fire safety regulations: A user-friendly approach. *Buildings*, *13*(6), Article 1404. https://doi.org/10.3390/buildings13061404

14. Jiang, Z., & Messner, J. I. (2023). Computer vision applications in construction and asset management phases: A literature review. *Journal of Information Technology in Construction*, *28*, 176–199. https://doi.org/10.36680/j.itcon.2023.009

15. Lee, S., Jeong, M., Cho, C.-S., Park, J., & Kwon, S. (2022). Deep learning-based PC member crack detection and quality inspection support technology for the precise construction of OSC projects. *Applied Sciences*, *12*(19), Article 9810. https://doi.org/10.3390/app12199810

16. Li, H., Zhang, C., Song, S., Demirkesen, S., & Chang, R. (2020). Improving tolerance control on modular construction project with 3D laser scanning and BIM: A case study of removable floodwall project. *Applied Sciences*, *10*(23), Article 8680. https://doi.org/10.3390/app10238680

UNDER PEER REVIEW IN IJAR

17. Lo, K.-C., Kwok, H.-W. T., Siu, M.-F. F., Shen, Q. G., & Lau, C.-K. (2021). Internet of Things-based concrete curing invention for construction quality control. *Advances in Civil Engineering*, *2021*(1), Article 9933615. https://doi.org/10.1155/2021/9933615

18. Love, P. E., Fang, W., Matthews, J., Porter, S., Luo, H., & Ding, L. (2023). Explainable artificial intelligence (XAI): Precepts, models, and opportunities for research in construction. *Advanced Engineering Informatics*, *57*, Article 102024. https://doi.org/10.1016/j.aei.2023.102024

19. Luo, H., Lin, L., Chen, K., Antwi-Afari, M. F., & Chen, L. (2022). Digital technology for quality management in construction: A review and future research directions. *Developments in the Built Environment*, *12*, Article 100087. https://doi.org/10.1016/j.dibe.2022.100087

20. Peng, J., & Liu, X. (2023). Automated code compliance checking research based on BIM and knowledge graph. *Scientific Reports*, *13*(1), Article 7065. https://doi.org/10.1038/s41598-023-34342-1

21. Pregnolato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J., & Taylor, C. (2022). Towards Civil Engineering 4.0: Concept, workflow and application of digital twins for existing infrastructure. *Automation in Construction*, *141*, Article 104421. https://doi.org/10.1016/j.autcon.2022.104421

22. Tan, Y., Li, S., & Wang, Q. (2020). Automated geometric quality inspection of prefabricated housing units using BIM and LiDAR. *Remote Sensing*, *12*(15), Article 2492. https://doi.org/10.3390/rs12152492

23. Vasquez, J., Furuhata, T., & Shimada, K. (2024). Image-enhanced U-Net: Optimizing defect detection in window frames for construction quality inspection. *Buildings*, *14*(1), Article 3. https://doi.org/10.3390/buildings14010003

24. Vieira, J., Almeida, N. M. D., Poças Martins, J., Patrício, H., & Morgado, J. G. (2024). Analysing the value of digital twinning opportunities in infrastructure asset management. *Infrastructures*, *9*(9), Article 158. https://doi.org/10.3390/infrastructures9090158

UNDER PEER REVIEW IN IJAR

25. Wang, Y.-W., Ni, Y.-Q., & Wang, S.-M. (2022). Structural health monitoring of railway bridges using innovative sensing technologies and machine learning algorithms: A concise review. *Intelligent Transportation Infrastructure*, *1*, Article liac009. https://doi.org/10.1093/iti/liac009

26. Xu, Y., Zhou, Y., Sekula, P., & Ding, L. (2021). Machine learning in construction: From shallow to deep learning. *Developments in the Built Environment*, *6*, Article 100045. https://doi.org/10.1016/j.dibe.2021.100045

27. Yang, J., Ruan, K., Gao, J., Yang, S., & Zhang, L. (2022). Pavement distress detection using three-dimension ground penetrating radar and deep learning. *Applied Sciences, 12*(11), Article 5738. https://doi.org/10.3390/app12115738

28. Yu, Y., Rashidi, M., Samali, B., Mohammadi, M., Nguyen, T. N., & Zhou, X. (2022). Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm. *Structural Health Monitoring*, *21*(5), 2244–2263. https://doi.org/10.1177/14759217211053546

29. Zihan, Z. U. A., Smadi, O., Tilberg, M., & Yamany, M. S. (2023). Synthesizing the performance of deep learning in vision-based pavement distress detection. *Innovative Infrastructure Solutions*, *8*(11), Article 299. https://doi.org/10.1007/s41062-023-01250-2

30. Zinno, R., Haghshenas, S. S., Guido, G., & Vitale, A. (2022). Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art. *IEEE Access*, *10*, 88058–88078. https://doi.org/10.1109/access.2022.3199443

UNDER PEER REVIEW IN IJAR