From Algorithmic Accuracy to Organisational Value: A Systematic Overview of Artificial Intelligence-Enabled Defect Detection in Facilities Management

Ahmad Shahsyafie Muhamad[1*], Mat Naim Abdullah@Asmoni[1], Ahmad Sha'rainon Md Shaarani[1], Norshaliza Kamaruddin[2]

[1]Department of Real Estate, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia

[2]Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100700430

Received: 22 July 2026; Accepted: 28 July 2026; Published: 04 August 2026

ABSTRACT

Artificial intelligence (AI) has become an important enabler of defect detection, fault diagnosis and predictive maintenance in Facilities Management (FM). Despite rapid technological advancements, existing studies predominantly emphasise algorithmic accuracy while giving comparatively less attention to the organisational, operational and managerial conditions required to translate AI predictions into sustainable FM value. This paper aims to provide a systematic overview of AI-enabled defect detection in FM and to develop an integrative Technology–Process–Management (TPM) perspective for understanding how AI capabilities are transformed into organisational value. The study adopts a systematic overview methodology based on a previously completed systematic literature review covering publications from January 2019 to March 2025. The source review identified 94 eligible studies through a transparent PRISMA-based review process. Evidence was synthesised using a structured narrative approach by categorising findings into three analytical dimensions: Technology, Process and Management. The review identifies five major AI technology families, namely conventional machine learning, deep learning, computer vision, semantic and language-enabled methods, and digital twins. The findings demonstrate that high predictive accuracy alone is insufficient to ensure successful implementation in FM. Instead, AI effectiveness depends on data quality, interoperability, workflow integration, governance, organisational readiness, operator participation and lifecycle economic considerations. Based on these findings, the paper proposes a Technology–Process–Management (TPM) framework that conceptualises the relationships between implementation readiness, operational mechanisms and multi-level organisational outcomes. The review also outlines future research priorities, including prospective multi-site validation, human–AI collaboration, interoperability, lifecycle economic evaluation and implementation-oriented reporting. The findings suggest that AI-enabled defect detection should be viewed as a socio-technical transformation rather than solely a technological innovation. For FM practitioners, successful AI adoption requires aligning technological capability with organisational processes, governance structures and maintenance decision-making to achieve sustainable operational value.

Keywords: Artificial intelligence; Facilities management; Defect detection; Predictive maintenance; Technology–Process–Management framework

INTRODUCTION

Facilities management (FM) is responsible for sustaining the functionality, safety, comfort and efficiency of the built environment throughout the longest and often most resource-intensive stage of an asset's lifecycle. In this stage, defects range from visible deterioration of envelopes and finishes to latent faults in heating, ventilation and air-conditioning (HVAC), electrical, plumbing and control systems. Conventional inspection and maintenance practices are frequently reactive, periodic and dependent on individual experience. They may therefore identify faults only after performance has degraded, energy has been wasted or service has been disrupted. Digital FM research has long argued that building information modelling (BIM), sensor networks, reality capture and interoperable information systems can improve the timeliness and quality of operational decisions [24, 12, 16, 21].

Artificial intelligence (AI) extends this digitalisation trajectory by converting operational and visual data into estimates of equipment condition, fault type, defect location, failure probability or remaining useful life. In this paper, AI-enabled defect detection is used as the umbrella term for fault detection and diagnosis, anomaly and condition monitoring, physical-defect recognition, and predictive-maintenance applications in FM. These functions encompass data-driven classifiers, anomaly-detection algorithms, deep neural networks, computer vision, semantic models, and hybrid systems that combine physical or expert knowledge with learning algorithms. Reviews of building fault detection and diagnosis consistently show rapid methodological development and strong performance under experimental conditions [25, 22, 4]. Predictive-maintenance demonstrations have also established technically feasible workflows for data collection, processing, model development, notification, and model improvement [2, 1].

Technical progress, however, has not translated automatically into routine FM practice. Buildings are heterogeneous, sparsely labelled and dynamically operated. Sensors drift, metadata are inconsistent, control sequences change, and the frequency of genuine failures is low relative to normal operation. Models trained on one plant or building may not transfer to another without adaptation. Even a statistically accurate alert creates little value if it is not understandable to an operator, linked to the correct asset, prioritised against other work, and converted into a work order with an accountable response. Studies of commercial FDD and field implementation consequently emphasise benchmarking, integration effort, operator interpretation and measurement of realised savings rather than accuracy alone [6, 14, 8].

Existing reviews illuminate important parts of this problem but remain fragmented. Some classify algorithms and FDD processes [25, 22, 4]; others examine BIM, digital twins or disruptive technologies in FM [20, 9, 7]. The organisational conditions through which an AI output becomes an FM outcome are less often treated as the central unit of analysis. This fragmentation risks encouraging a technology-push view in which adoption is equated with model installation and success is equated with classification performance.

This paper addresses that limitation by providing a systematic overview of AI-enabled defect detection through a Technology-Process-Management (TPM) lens. Its purpose is not to identify a universally superior algorithm. Instead, it asks how technological capability, implementation processes and organisational arrangements interact to produce or prevent operational value. Four questions guide the paper:

Table 1. Review questions and analytical contribution

Review question	Analytical purpose
RQ1. Which AI technology families are used for defect detection and diagnosis in FM?	Distinguish their data needs, strengths, limitations and appropriate operational roles.
RQ2. Which process conditions mediate the translation of model outputs into maintenance action?	Examine data pipelines, system integration, alert handling, work-order execution and learning loops.
RQ3. Which management conditions shape adoption, sustained use and value realisation?	Analyse readiness, capability, culture, stakeholder engagement, governance, risk and economics.

Review question	Analytical purpose
RQ4. How should technical, operational, human and strategic outcomes be connected in future evaluation?	Develop an integrative framework and a research agenda for implementation-oriented evidence.

The paper makes four contributions. First, it consolidates technically and managerially oriented evidence into a single implementation-centred account. Second, it introduces an outcome ladder that separates model performance from operational, human and strategic value. Third, it develops a framework in which data and organisational readiness are antecedents, integration and human-AI work design are mechanisms, and FM performance is a multidimensional outcome. Fourth, it demonstrates how a systematic overview can transparently re-analyse implementation-oriented evidence from an existing systematic review while preserving the boundaries of the underlying record. Together, these contributions support research design and organisational decision making without presenting the overview as a substitute for a de novo systematic search.

Review Approach

Design and Scope

The paper was developed as a systematic overview and critical synthesis derived from a previously completed systematic literature review of AI-enabled defect detection in FM. The source review covered January 2019 to March 2025. It documented 2,561 records before deduplication, 1,675 unique records at title-and-abstract screening, 112 full texts and 94 included publications. A systematic review undertakes and reports a complete, reproducible search, selection, appraisal and extraction process for a defined question. By contrast, a systematic overview re-analyses an established review record to integrate evidence at a broader conceptual or implementation level. This approach was selected because the present study asks how technological capability, operational processes and management conditions interact, rather than which individual algorithm performs best. Its value lies in connecting otherwise fragmented technical and organisational findings through the Technology-Process-Management lens while preserving the evidentiary boundaries of the source review. Accordingly, the paper does not claim that a new database search was conducted for this manuscript. The substantive scope remains building operations and maintenance, including HVAC and other building services, building envelopes and multi-system FM applications. Construction-phase quality control and non-building industrial maintenance are used only as adjacent evidence when they offer a directly transferable methodological insight.

The reporting logic follows PRISMA 2020, particularly its emphasis on transparent identification, selection, appraisal and synthesis [18]. The source review searched SciSpace Deep Search, SciSpace Full Text Search, Google Scholar and PubMed in December 2024, with an update in March 2025. Its core Boolean logic was ("artificial intelligence" OR "machine learning" OR "deep learning" OR "computer vision" OR "neural network" OR "predictive analytics" OR "digital twin") AND ("facilities management" OR "building maintenance" OR "asset management") AND ("defect detection" OR "fault diagnosis" OR "anomaly detection" OR "condition monitoring" OR "predictive maintenance"). Application terms, including HVAC, building envelope and building systems, were used to refine retrieval where supported by the search interface. Google Scholar used the narrower documented combination ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("facilities management" OR "building maintenance") AND ("defect detection" OR "predictive maintenance") with a 2019-2025 date limit. Backward and forward citation chaining supplemented database retrieval. These details improve transparency but do not make the source search equivalent to a de novo search across Scopus, Web of Science and IEEE Xplore.

Figure 1 reproduces the documented PRISMA 2020-aligned study selection process from the source systematic review that underpins the present systematic overview. The figure is included to improve methodological transparency regarding the evidence base synthesised in this paper and should not be interpreted as representing a newly conducted database search.

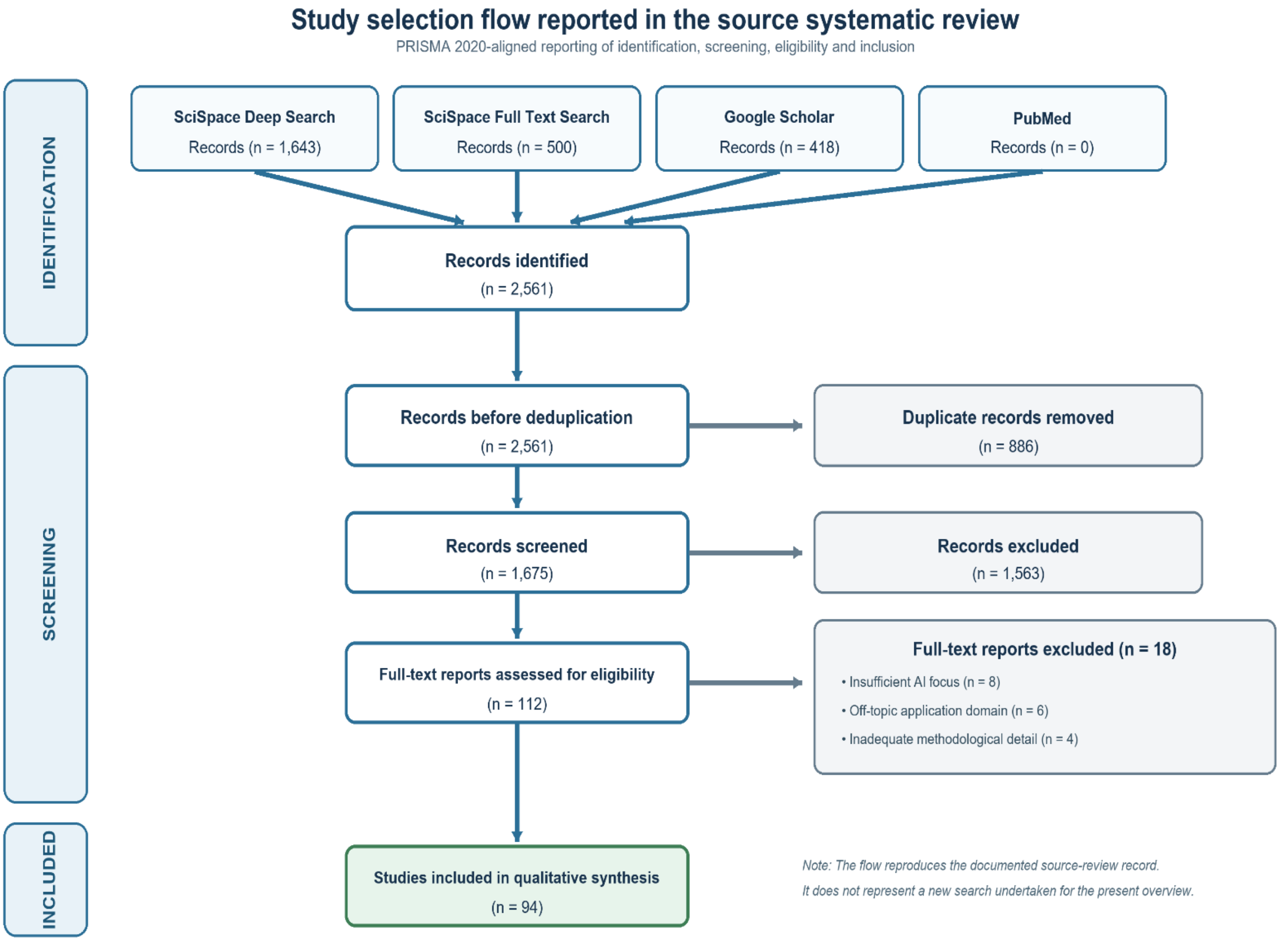

Figure 1. PRISMA 2020-aligned study selection flow reproduced from the source systematic review underpinning the present systematic overview.

Note. The study selection flow reproduces the documented source-review record reported in the underlying systematic literature review. It is presented solely to illustrate the origin of the evidence base synthesised in this systematic overview and should not be interpreted as representing a new literature search conducted for the present study.

Source: Reproduced from the underlying systematic literature review conducted by the authors.

As illustrated in Figure 1, the present paper synthesises evidence from the 94 studies included in the documented source systematic review rather than conducting an independent literature search. The figure therefore provides transparency regarding the origin and scope of the evidence synthesised throughout this overview.

Eligibility Logic

Eligible evidence addressed at least one of four functions: detection of abnormal building-system behaviour; diagnosis or classification of the underlying fault; recognition or localisation of a physical defect; or prediction of failure or maintenance need. The operational context had to relate to occupied buildings, building services, FM information systems, condition assessment or maintenance decision making. Peer-reviewed research articles and reviews were prioritised. Conceptual studies were retained when they specified an implementation architecture, adoption mechanism or testable management relationship.

Evidence was excluded from the core synthesis when it concerned only design optimisation, construction scheduling, energy forecasting without a fault or maintenance component, or industrial assets with no credible transfer to FM. This boundary is important because defect detection during construction and fault management during occupation have different users, decision cycles, liability structures and information systems. Bridge, tunnel and transport studies appearing in the source report were therefore not used to calculate FM-specific prevalence or performance.

The documented selection process comprised deduplication by DOI and title similarity, title-and-abstract screening, and full-text assessment. The source report states that two reviewers screened titles and abstracts using a structured eligibility rubric, with disagreements resolved through discussion and, when required, consultation with a third reviewer; it reports Cohen's kappa of 0.82. Full-text exclusions were attributed to insufficient AI focus (n = 8), an off-topic application domain (n = 6) or inadequate methodological detail (n = 4). Eligible records required an AI method, an FM or building-operations context, a defect-detection or maintenance function, publication within the stated period and sufficient scholarly detail. The review excluded design optimisation, construction-only quality control, energy forecasting without a maintenance component and non-building applications without credible FM transfer. These figures reproduce the source-review record and should be interpreted as reported rather than independently reconstructed in the present overview.

Synthesis strategy and evidence interpretation

The synthesis combined deductive and inductive coding. Deductive categories were Technology, Process and Management. Technology covered algorithms, sensing, computing, data and technical architecture. Process covered data engineering, integration, validation, alert triage, work-order execution and feedback. Management covered strategy, readiness, competence, stakeholder engagement, governance, culture, risk and economics. Inductive comparison was then used to identify cross-cutting mechanisms such as transferability, interpretability and operator actionability.

The source review used a standardised extraction form, piloted on five studies. It recorded bibliographic details, study design and context, AI technique and input data, building system and defect type, validation approach, reported outcomes, TPM theme, limitations and research gaps. The report describes duplicate extraction for 20% of included studies. Discrepancies were discussed, and periodic checks were applied to the remaining records. Quality appraisal adapted dimensions from the Mixed Methods Appraisal Tool and Critical Appraisal Skills Programme to the heterogeneous corpus: methodological rigour, empirical validation, conceptual contribution and reporting quality. The source review assigned holistic high, moderate or low ratings rather than calculating a pooled numerical score across unlike study designs. No study was excluded solely because of a low rating. Instead, higher-quality studies received greater interpretive weight in the narrative synthesis, whereas lower-quality studies were retained for contextual comparison but were not relied upon for major conclusions. Because the complete 94-study extraction workbook and item-level ratings were not supplied with the source report, this overview does not present them as independently auditable supplementary data.

Given the marked heterogeneity of assets, faults, datasets, building types, validation designs and performance measures, no meta-analysis or direct ranking of algorithm families was attempted. Accuracy values were not pooled because a high score on a balanced experimental dataset is not commensurate with performance under low fault prevalence in an operational building, where false-alarm burden, diagnostic time and avoided cost are material. The synthesis therefore uses structured narrative comparison: evidence is compared within technology families and interpreted against four contextual dimensions - data provenance, validation setting, operational workflow and outcome level. Quantitative findings are reported only when their context and denominator are identifiable and are treated as illustrative rather than comparative effect estimates. Publication-year counts in the source review were recalculated: 68 of 94 studies were assigned to 2023-2025, equivalent to 72.3%, rather than the 68% stated in the source narrative. Other frequency estimates remain descriptive until the complete coding workbook is audited.

Quality and credibility considerations

Credibility was considered at four levels. Method credibility concerns the appropriateness and transparency of data collection, modelling and validation. Deployment credibility concerns whether a method was tested prospectively in an operational building, across time and, where relevant, across sites under realistic fault prevalence. Decision credibility concerns whether intended users assessed the interpretability and actionability of outputs. Value credibility concerns whether technical performance was connected to verified operational or economic outcomes using an explicit baseline and attribution method. Claims and numerical values central to the synthesis were checked against identifiable publisher or institutional records where available; unsupported or irreconcilable claims were omitted, softened or explicitly identified as source-review estimates. This layered approach avoids treating peer review or model accuracy as a sufficient quality signal.

Technology landscape for AI-enabled defect detection

Conventional machine learning

Conventional supervised machine-learning methods including decision trees, random forests, support vector machines and gradient-boosting models remain attractive for structured building-automation data. They can perform fault classification with moderate computational requirements and, in the case of trees and rule-enhanced models, provide decision paths that operators can inspect. Almobarek et al. [1], for example, implemented a predictive-maintenance framework for a university chilled-water system using decision-tree models. Publisher-reported results exceeded 98% accuracy for the individual chilled-water components and improved fault prediction relative to the existing building-management control logic. These results are stronger than the 92% figure reproduced in the source report and illustrate why performance claims should be checked against primary publications.

Conventional models are not intrinsically simple in operational use. Their apparent interpretability can be undermined by unstable features, unrecorded control changes, imbalanced fault classes and building-specific calibration. A model may learn operating regimes or seasonal proxies rather than causal fault signatures. Furthermore, supervised learning depends on labelled examples, while facilities typically contain abundant normal data and comparatively few verified faults. The most appropriate role of conventional machine learning is therefore often a bounded classification or prioritisation task supported by explicit feature governance and operator validation.

Deep learning and representation learning

Deep learning can learn complex nonlinear and temporal relationships without relying entirely on hand-crafted features. Recurrent architectures, convolutional networks, autoencoders and related representation-learning methods have been applied to sensor streams, chiller data and multi-variable HVAC behaviour. Bouabdallaoui et al. [2] demonstrated an autoencoder-based predictive-maintenance workflow in a sports facility and, importantly, documented obstacles involving data availability and feedback collection. This illustrates that model development is only one stage of an operational pipeline.

Deep models intensify three challenges. First, they typically require substantial and representative training data. Second, performance can degrade across equipment, buildings, climates and operating policies. Third, opaque decision logic may impede trust and fault investigation. Transfer learning offers a partial response to the second problem. In experiments on two water-cooled chillers, Liu et al. [15] reported maximum accuracy improvements of 12.63 and 8.18 percentage points in two transfer tasks when target-domain data were limited. These results show potential, but transferability must be evaluated prospectively and should not be assumed merely because source and target assets share a label.

Explainable AI can help connect predictions to engineering knowledge. Li et al. [13] compared feature-level explanation methods for a convolutional chiller diagnostic model and reported 98.5% classification accuracy, with SHAP and layer-wise relevance propagation performing better than alternative explanation methods on their proposed criteria. Explanations, however, are not equivalent to causal proof. Their operational value depends on whether they reveal physically credible features, support troubleshooting and remain stable under changing conditions.

Computer vision and remote inspection

Computer vision enables defect recognition from photographs, video, thermal imagery, point clouds and unmanned-aerial-vehicle surveys. It is particularly relevant to envelopes, roofs, finishes and accessible structural components, where manual inspection is slow, hazardous or inconsistent. Computer-vision research in the built environment covers classification, object detection, segmentation and tracking, with quality control and defect recognition forming a major application family [19].

The apparent precision of image models must be interpreted in relation to defect scale, material diversity, lighting, occlusion and dataset provenance. Crack-only datasets with controlled backgrounds cannot establish performance across spalling, debonding, moisture staining and glazing defects on heterogeneous façades. Moreover, a pixel mask is not yet a maintenance decision. Operational deployment requires reliable asset localisation, severity estimation, linkage to inspection history, risk prioritisation and integration with the work-management process. Computer vision should consequently be evaluated as an inspection system, not solely as a segmentation model.

Semantic, knowledge-based and language-enabled methods

Building data are semantically fragmented. The same equipment may be labelled differently across BIM, building-automation, energy-management and computerised maintenance-management systems. Semantic models and ontologies can map points, assets, relationships and fault logic into a machine-readable representation. This is critical for portable analytics because a technically capable model cannot scale if every site requires manual point mapping and bespoke asset matching. Reviews of BIM and digital FM consistently identify information exchange and interoperability as persistent barriers [24, 20, 16].

Natural-language processing can extract recurring failure modes, urgency and asset information from work orders, inspection narratives and manuals. These sources contain valuable maintenance knowledge but are often inconsistent, abbreviated and affected by local documentation practices. Language-enabled tools are therefore most credible when combined with controlled vocabularies, asset identifiers and human review. The immediate opportunity is not autonomous maintenance instruction, but improved coding, retrieval, triage and organisational learning.

Digital twins as an integration architecture

A digital twin is not a single AI technique. In FM it is better understood as an integration architecture that connects a representation of the asset with operational data, analytical models and decision processes. Digital twins can provide spatial context for faults, support scenario evaluation and connect analytics to BIM or asset information. Systematic reviews nevertheless show that definitions and maturity levels vary widely, and that many reported “twins” are closer to static digital models or one-way data shadows [9, 7]. At the urban scale, AI-enabled city-information models extend this logic to federated assets, although the evidence remains weighted toward proposed architectures and demonstrators rather than autonomous maintenance at scale [11].

Digital-twin value depends on the quality of identifiers, data synchronisation, semantic interoperability and process integration. BIM-based twins and extended-reality interfaces may improve access to instructions and context for field operators, but evidence of combined, sustained use remains limited [5]. A digital twin should therefore be justified by a specific maintenance decision and information flow, not adopted as an end in itself.

This use-case discipline is also a response to recurring adoption barriers. Reviews identify high initial investment, skills scarcity, interoperability difficulties and organisational resistance as connected rather than independent constraints [10]. A technically ambitious twin can therefore create less value than a narrower architecture that solves a well-defined maintenance problem with manageable information requirements. Collectively, the reviewed technological evidence shows that analytical capability is necessary but insufficient: deployment value depends on the processes that translate model outputs into maintenance action.

Table 2. Technology families, operational roles and principal evidence risks

Technology family	Typical FM role	Primary strength	Principal evidence or deployment risk
Conventional machine learning	Fault classification, prioritisation and anomaly scoring	Strong performance on structured data; some models offer inspectable logic	Dependence on labelled, site-specific data and unstable operating regimes
Deep learning and representation learning	Time-series diagnosis, anomaly detection and prognosis	Learns nonlinear and temporal patterns with less manual feature engineering	Data intensity, opacity, class imbalance and weak cross-building transfer
Computer vision and thermal analytics	Envelope and surface-defect inspection; localisation and segmentation	Scalable and safer coverage of visual defects	Dataset bias, environmental sensitivity and weak linkage from masks to work orders
Semantic and language-enabled methods	Point mapping, work-order coding, knowledge retrieval and explanation	Connects heterogeneous information and captures unstructured knowledge	Ontology maintenance, inconsistent terminology and hallucination or extraction error
Digital twins	Spatial context, integration, simulation and decision support	Unifies asset representation, live data and analytics	High integration burden, ambiguous maturity and unclear incremental value

Process factors: translating predictions into maintenance action

Data engineering is part of the intervention

AI implementation begins before model selection. Asset identification, sensor coverage, timestamp alignment, missing-data handling, calibration, fault labelling and contextual metadata determine what can be learned. Chen et al. [4] describe data-driven FDD as a multi-stage process including collection, cleansing, preprocessing, baseline establishment, detection, diagnosis and potential prognosis. This process view is important because many apparent modelling failures originate upstream.

FM data present a structural imbalance: the organisation needs examples of failures to learn failure signatures, but competent maintenance aims to prevent failures. Labels may also reflect technician judgement rather than confirmed root cause. Organisations should therefore create explicit label-governance procedures, preserve uncertainty, and distinguish alarms, observed symptoms, diagnosed causes and completed remedies. Synthetic data and simulation can supplement evidence but should not be treated as substitutes for field validation.

Interoperability and workflow integration

An AI output must be connected to the asset register, BIM or spatial record, building-management system and maintenance workflow. A minimum integration chain includes a stable asset identifier, time-stamped evidence, diagnostic confidence, priority or consequence, recommended verification, work-order creation, assigned responsibility, closure code and feedback to the model. Without this chain, alerts accumulate in dashboards and are detached from accountability.

BIM can support visualisation, knowledge storage and asset context, but systematic reviews show that information exchange between design, construction and FM systems remains difficult [20, 3]. Standards and open interfaces reduce integration effort, yet semantic alignment and data ownership still require organisational decisions. Integration should be tested end to end using realistic failure scenarios rather than demonstrated only through data transfer.

Alert design, triage and operator actionability

Maintenance teams operate under constrained labour, access windows, spare-parts availability and competing risks. An accurate model can still increase workload if alerts are repetitive, poorly prioritised or not accompanied by evidence. Effective alert design should communicate the affected asset, likely fault, confidence, observed deviation, consequence of delay and recommended verification. It should also make uncertainty visible.

Operator involvement is essential because technicians contribute contextual knowledge that is rarely present in training data. Their feedback can reveal maintenance work already under way, temporary overrides, sensor failures or operational requirements that explain an anomaly. Human review is therefore not merely a concession to immature AI; it is an information source and a control mechanism. The design goal is calibrated reliance: operators should neither dismiss all model outputs nor treat them as unquestionable.

Validation, monitoring and continuous learning

Validation must match the intended use. Randomly splitting observations from the same building can overestimate generalisation because near-identical operating conditions appear in both training and test data. Stronger designs include temporal hold-outs, equipment-level hold-outs, external-building validation and prospective deployment. Evaluation should also reflect real fault prevalence and the cost of false positives and false negatives.

Frank et al. [6] proposed a systematic performance-evaluation framework for building fault-detection and diagnosis algorithms, while Lin et al. [14] combined field evidence on costs and savings with a repeatable evaluation methodology and a public test dataset. These contributions show that benchmarking requires shared definitions, ground truth and test conditions. After deployment, data drift, sensor replacement, control changes and model updates should be monitored through a documented model lifecycle. Collectively, the process evidence indicates that technical predictions create FM value only when they are integrated, prioritised, acted upon and monitored within an accountable maintenance workflow. This process dependence leads directly to the management conditions that sustain capability, user participation and investment.

Management Factors: Readiness, Governance and Value

Organisational readiness and capability

AI readiness combines infrastructure, data, people and decision authority. An organisation may own a modern building-management system yet lack stable metadata, data-engineering capability or personnel who can translate model outputs into maintenance decisions. Digital-transformation research in FM similarly identifies technology, policy and infrastructure as mutually reinforcing readiness dimensions [17]. Qualitative evidence from FM professionals identifies manpower, infrastructure, knowledge and budget as practical capability conditions for digital-twin adoption [23]. Readiness assessment should therefore precede procurement and should be specific to the intended use case.

Capability is interdisciplinary. Data scientists need knowledge of operating modes and fault mechanisms; FM engineers need sufficient analytical literacy to challenge outputs; information-technology teams need to support integration and security; and managers need to define value, risk and accountability. Vendor support can accelerate implementation but creates dependency if models, data transformations and diagnostic rules remain opaque. Contracts should address data portability, model documentation, change control and knowledge transfer.

Culture, stakeholder engagement and change

AI-enabled maintenance alters professional judgement, work allocation and sometimes the perceived status of expertise. Resistance may signal fear of surveillance or job displacement, but it may also be a rational response to poor alerts and previous technology projects. Engagement should begin with problem definition rather than end-user training after procurement. Technicians should participate in selecting use cases, defining actionable outputs, testing interfaces and reviewing false alarms.

The lack of user evidence is a recurring weakness. Gunay et al. [8] found that 90% of reviewed field-implementation papers did not include feedback from building operators or occupants after implementation. This omission matters because operational value is co-produced by technology and users. A technically successful intervention that technicians bypass, misunderstand or cannot act on is not a successful FM implementation.

Governance, accountability and responsible use

Governance should specify who owns the data, approves a model, responds to an alert, overrides a recommendation and reviews adverse outcomes. Cybersecurity and privacy are particularly relevant where building systems expose operational data, access patterns or occupancy information. Model governance should include versioning, performance thresholds, audit logs and escalation rules.

Interpretability is both a technical and governance issue. Explainable models can assist diagnosis, but organisations must determine what level of explanation is needed for the decision's consequence. A low-risk maintenance prompt may require less evidence than an automated shutdown or safety-critical intervention. Governance should be proportional to risk and should preserve meaningful human authority.

Business case and outcome measurement

Accuracy is not a business case. Costs include sensing, connectivity, data engineering, software, systems integration, cybersecurity, model maintenance, training and operator time. Benefits may include avoided failures, reduced energy waste, improved labour productivity, longer equipment life, improved comfort and better regulatory evidence. These outcomes occur on different timescales and may be difficult to attribute.

Field evidence demonstrates that value is possible. Lin et al. [14] reported median whole-building savings of 8% among 26 organisations using fault-detection and diagnosis tools across 550 buildings, while also quantifying software and labour costs per monitoring point. The finding is more decision-relevant than an isolated accuracy statistic because it connects technology with actual operations. However, the authors caution that savings are not solely attributable to the diagnostic tool. Gunay et al. [8] similarly found positive post-implementation energy outcomes in many reported cases but noted inconsistent measurement and verification. Future studies should report total cost of ownership, counterfactual logic and uncertainty, not only gross savings. Collectively, the management evidence suggests that governance, capability, stakeholder involvement and credible economic evaluation determine whether implementation persists beyond a technical pilot.

Table 3. Implementation barriers, mechanisms and decision-relevant metrics

Barrier	Value-realisation mechanism	Illustrative metric
Sparse, noisy or poorly labelled data	Data governance, fault taxonomy, calibration and uncertainty-preserving labels	Missingness; label confidence; fault coverage; data latency
Weak cross-building generalisation	Temporal and external validation; transfer learning; local recalibration	External F1/recall; calibration error; performance degradation by site
Opaque model output	Domain-grounded explanations and evidence-rich alerts	Explanation stability; operator-rated usefulness; diagnostic time
Disconnected FM systems	Stable asset identifiers, open interfaces and end-to-end workflow tests	Successful asset matching; work-order conversion rate; integration downtime

Barrier	Value-realisation mechanism	Illustrative metric
Alert burden and resistance	Co-design, consequence-based triage and feedback capture	Alerts per asset; false alarms; acknowledgement and completion rates
Uncertain economics	Prospective baseline, total-cost accounting and measurement and verification	Avoided cost; energy or downtime reduction; payback; net present value
Weak accountability	Model lifecycle governance, escalation and audit trail	Override rate; unresolved critical alerts; time to review; incident recurrence

An integrative framework for AI value realisation in FM

Figure 2 presents the three-stage value-realisation framework supported by the synthesis. Antecedent readiness determines whether an organisation can initiate a credible AI use case. Implementation mechanisms determine whether technical outputs enter the maintenance system and influence action. Outcomes should be evaluated at multiple levels, with feedback from completed work updating data, models and organisational knowledge.

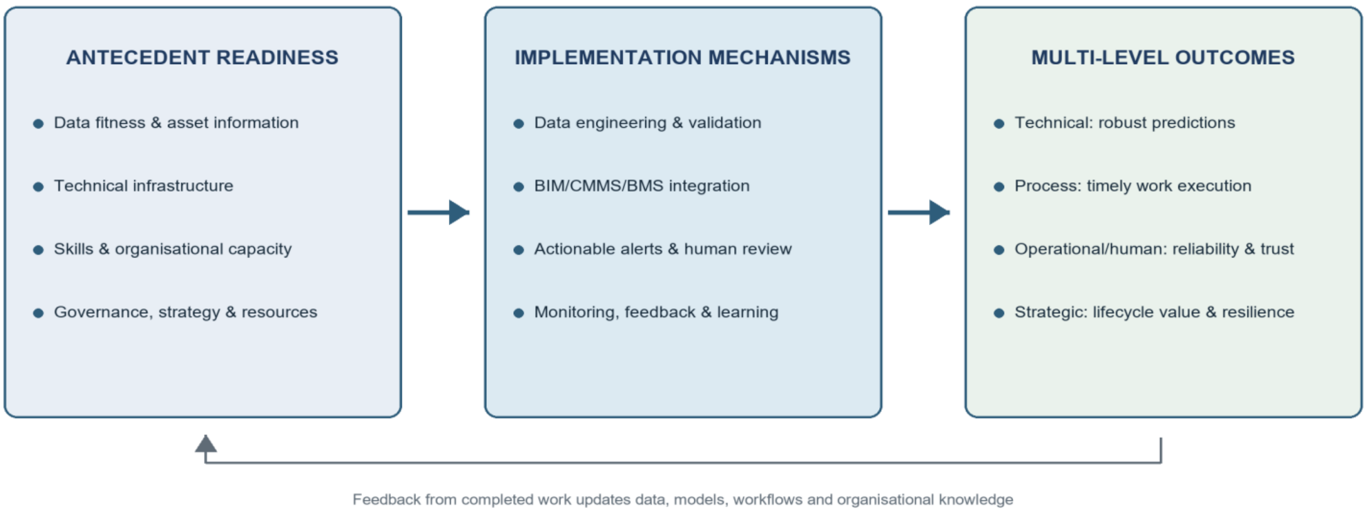

Figure 2. Technology-Process-Management framework for AI value realisation in facilities management

The proposed TPM framework is a conceptual synthesis derived from the reviewed literature and should not be interpreted as an empirically validated causal model.

The framework makes four distinctions. First, AI readiness is broader than technology availability: it includes data fitness, asset information, skills, governance and resources. Second, integration is a mediating mechanism rather than a background technical detail. Third, organisational culture and stakeholder engagement moderate whether recommendations are used and whether learning persists after a pilot. Fourth, FM outcomes form a ladder:

1. Technical outcomes: discrimination, calibration, robustness, lead time, interpretability and computational performance.
2. Process outcomes: alert quality, work-order conversion, response time, diagnostic effort and closure quality.
3. Operational and human outcomes: avoided failures, downtime, maintenance cost, energy use, safety, comfort, workload and user trust.

4. Strategic outcomes: lifecycle value, resilience, sustainability, organisational learning and scalable digital capability.

Success at a lower level enables but does not guarantee success at a higher level. A model may classify faults accurately but generate no savings because alerts arrive too late, cannot be mapped to an asset or are ignored. Conversely, a simpler and less accurate model may create greater value if its outputs are trusted, well prioritised and integrated into an effective workflow. Model selection should therefore optimise expected decision value rather than accuracy in isolation.

The framework clarifies propositions for subsequent empirical testing rather than presenting validated causal relationships. Data readiness is expected to improve technical performance and reduce implementation rework; organisational readiness is expected to improve integration quality and sustained use; and integration quality is expected to mediate the relationship between AI capability and operational performance. Operator involvement and governance are expected to moderate the relationship between alert quality and maintenance action, while verified operational benefits and transparent costs should influence continued investment and scaling. These propositions require validation through longitudinal case studies, multi-site surveys and mixed-method designs. Comparative case studies can examine process mechanisms and boundary conditions, while surveys and structural models can test relationships across organisations; neither form of evidence is supplied by the present literature-based overview.

Research Agenda

Prospective, multi-site and lifecycle validation

The field needs studies that begin before deployment, define baseline performance and follow implementations through stabilisation and routine use. Multi-site validation should separate changes in climate, equipment, control policy and organisational practice. Reporting only the best-performing model on a retrospective dataset cannot establish deployment readiness. Studies should document model updates, sensor changes and performance decay over time.

Common benchmarks and implementation reporting

Shared datasets and evaluation frameworks are necessary but should represent operational complexity. Benchmarks should include realistic fault prevalence, simultaneous faults, changing regimes, missing data and asset metadata. Alongside algorithm metrics, authors should report integration architecture, labour input, alert policy and intended user. Frank et al. [6] and Lin et al. [14] provide foundations for repeatable FDD evaluation; future work should extend these foundations beyond HVAC and connect tests to operational consequence.

Human–AI decision making

Research should examine how technicians interpret confidence, explanations and competing alerts; when they override recommendations; and how trust changes after correct and incorrect predictions. Appropriate methods include observation, interviews, controlled decision experiments and log analysis. Outcome measures should include diagnostic time, decision quality, workload and calibrated reliance. User involvement should be treated as an empirical variable, not simply a recommendation.

Interoperability, semantics and portable analytics

Portable AI requires machine-readable descriptions of assets, points, relationships, units and control intent. Research should evaluate whether semantic models reduce configuration time and improve cross-site performance. BIM and digital-twin studies should report the incremental value of spatial and semantic integration compared with simpler architectures. Open interfaces and vendor-neutral asset identifiers are especially important for lifecycle continuity.

Economic and sustainability evaluation

Cost-benefit studies should adopt a lifecycle perspective and distinguish gross technical potential from realised and attributable benefit. Prospective evaluations should record implementation cost, operator time, avoided maintenance, service continuity, equipment life and energy effects. Sustainability claims should consider rebound, additional sensing and computing, asset longevity and avoided material replacement. Economic evidence should be stratified by building type, portfolio size and baseline digital maturity. Collectively, these research priorities are expected to strengthen the practical maturity, transferability and evidentiary credibility of AI-enabled defect detection in FM.

Table 4. Priority research designs for an implementation-oriented evidence base

Priority question	Recommended design	Minimum reporting requirement
Does the model generalise beyond its development site?	External validation across buildings, equipment and seasons	Site characteristics, fault prevalence, recalibration and confidence intervals
Does the alert improve maintenance decisions?	Prospective human-in-the-loop study	Alert content, user role, action taken, override and diagnostic time
Does integration create operational value?	Before–after or stepped implementation with process tracing	System interfaces, work-order workflow, baseline and implementation effort
Is the intervention economically viable?	Lifecycle cost-benefit study with sensitivity analysis	Total cost, attribution method, avoided cost and uncertainty
Can the solution remain effective over time?	Longitudinal post-deployment study	Model versions, drift indicators, retraining triggers and sustained outcomes

IMPLICATIONS FOR PRACTICE AND POLICY

For FM organisations, the central implication is to start with a decision problem rather than a technology category. A strong use case has a material consequence, observable signals, an identifiable asset, a defined response and a measurable outcome. Data readiness and workflow mapping should then establish whether AI is needed and whether the organisation can act on its output. A phased implementation can begin with decision support for one system, incorporate technician feedback and scale only after operational benefit is verified.

Procurement should evaluate more than demonstration accuracy. Organisations should request evidence of external validation, uncertainty handling, interoperability, model monitoring, data ownership and total cost. Acceptance testing should include false-alarm burden, asset mapping, work-order integration and operator usability. Contracts should preserve access to data and diagnostic history and should require sufficient documentation for continuity if a vendor changes.

For technology providers, product design should expose evidence, uncertainty and recommended verification rather than only fault labels. Integration effort should be treated as a core performance characteristic. For professional bodies and policymakers, priorities include common terminology, minimum reporting standards, workforce development and neutral test datasets. Standards should support innovation while making claims comparable and risks governable. In practice, the paper contributes a role-specific decision logic. Facility managers can use it to select actionable use cases and redesign workflows, while building owners can evaluate lifecycle value and investment risk. Software developers can prioritise interoperability, explainability and model monitoring. Policymakers and professional bodies can develop reporting standards, workforce capabilities and neutral benchmarks.

LIMITATIONS

This paper inherits important limitations from the source review. The original search relied heavily on SciSpace and Google Scholar and did not document a complementary search of Scopus, Web of Science and IEEE Xplore. Database-specific syntax, export files, the complete screening log, item-level quality ratings and the full 94-study extraction workbook were not supplied with the manuscript materials. The present revision reports the available search logic, screening stages, extraction fields and appraisal dimensions, but these additions cannot reconstruct missing records or establish full reproducibility. Some source references were incomplete, application boundaries were inconsistent and several numerical claims required correction. Central claims were therefore checked against traceable publications where possible, unsupported prevalence estimates were avoided, and source-review counts are labelled as descriptive.

A future de novo update should search Scopus, Web of Science and IEEE Xplore alongside relevant FM and building-science sources. It should archive the protocol, database-specific strings, dated exports, deduplication decisions, full-text exclusions and item-level appraisal, and provide a complete included-study extraction matrix. The date boundary also requires care: a search last updated in March 2025 cannot claim complete coverage through December 2025 without an additional update. Finally, heterogeneity in defect definitions, building types, datasets and outcomes prevents causal comparison or pooled ranking of algorithm families. The TPM framework is a synthesis-derived conceptual contribution and remains to be validated empirically through prospective field studies, multi-site testing, human-AI decision research, lifecycle cost-benefit analysis and longitudinal monitoring of model performance.

CONCLUSION

AI-enabled defect detection in FM has progressed from rule-based and conventional machine-learning tools to deep learning, computer vision, semantic models and digital-twin architectures. The evidence demonstrates substantial technical capability, but it also shows that high accuracy under controlled conditions is an incomplete and sometimes misleading indicator of readiness. The decisive question is whether a prediction can be translated into a timely, trusted and economically justified maintenance action.

This systematic overview reframes that translation as a Technology-Process-Management problem. Data and organisational readiness enable credible implementation; integration, alert design, human judgement and continuous learning mediate value; and outcomes extend from technical performance to process efficiency, operational reliability, occupant experience and strategic lifecycle value. The field should now prioritise prospective field validation, common benchmarks, human-AI decision research, interoperable information architectures and rigorous lifecycle economics. For practitioners, the most defensible path is a phased, use-case-led programme in which data, workflow and accountability are designed before models are scaled. Rather than viewing AI-enabled defect detection as a purely technological innovation, this paper demonstrates that sustainable organisational value can only be achieved through the alignment of technology, operational processes and management capabilities.

DECLARATIONS

Ethical approval: Not applicable. This systematic overview used published literature and did not involve human participants or animals.

Conflicts of interest: The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability: The present overview is based on the supplied source-review report and the traceable publications cited in this manuscript. The complete bibliographic master file, screening log, item-level appraisal and 94-study extraction matrix were not available for deposit and are therefore not represented as supplementary data. Any future de novo update should archive these records in a public repository or provide them as journal supplementary material.

REFERENCES

1. Almobarek, M., Mendibil, K., & Alrashdan, A. (2023). Predictive maintenance 4.0 for chilled water system at commercial buildings: A methodological framework. *Buildings*, *13*(2), 497. https://doi.org/10.3390/buildings13020497
2. Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., & Bennadji, B. (2021). Predictive maintenance in building facilities: A machine learning-based approach. *Sensors*, *21*(4), 1044. https://doi.org/10.3390/s21041044
3. Cao, Y., Kamaruzzaman, S. N., & Aziz, N. M. (2022). Building information modeling (BIM) capabilities in the operation and maintenance phase of green buildings: A systematic review. *Buildings*, *12*(6), 830. https://doi.org/10.3390/buildings12060830
4. Chen, Z., O'Neill, Z., Wen, J., Pradhan, O., Yang, T., Lu, X., Lin, G., Miyata, S., Lee, S., Shen, C., Chiosa, R., Piscitelli, M. S., Capozzoli, A., Hengel, F., Kührer, A., Pritoni, M., Liu, W., Clauß, J., Chen, Y., & Herr, T. (2023). A review of data-driven fault detection and diagnostics for building HVAC systems. *Applied Energy*, *339*, 121030. https://doi.org/10.1016/j.apenergy.2023.121030
5. Coupry, C., Noblecourt, S., Richard, P., Baudry, D., & Bigaud, D. (2021). BIM-based digital twin and XR devices to improve maintenance procedures in smart buildings: A literature review. *Applied Sciences*, *11*(15), 6810. https://doi.org/10.3390/app11156810
6. Frank, S., Lin, G., Jin, X., Singla, R., Farthing, A., & Granderson, J. (2019). A performance evaluation framework for building fault detection and diagnosis algorithms. *Energy and Buildings*, *192*, 84–92. https://doi.org/10.1016/j.enbuild.2019.03.024
7. Ghansah, F. (2024). Digital twins for smart building at the facility management stage: A systematic review of enablers, applications, and challenges. *Smart and Sustainable Built Environment*, *14*(4), 1194–1229. https://doi.org/10.1108/SASBE-10-2023-0298
8. Gunay, B., Markus, A., Parsaei, A., Darwazeh, D., Bursill, J., & Pellerin, L. (2025). A critical review of field implementation of data-driven operation and maintenance technologies to reduce building energy use and greenhouse-gas emissions. *Energy and Buildings*, *346*, 116087. https://doi.org/10.1016/j.enbuild.2025.116087
9. Hodavand, F., Ramaji, I. J., & Sadeghi, N. (2023). Digital twin for fault detection and diagnosis of building operations: A systematic review. *Buildings*, *13*(6), 1426. https://doi.org/10.3390/buildings13061426
10. Jahangir, M. F., Schultz, C. P. L., & Kamari, A. (2024). A review of drivers and barriers of digital twin adoption in building project development processes. *Journal of Information Technology in Construction*, *29*, 141–178. https://doi.org/10.36680/j.itcon.2024.008
11. Lawal, O. O., Nawari, N. O., & Lawal, O. (2025). AI-enabled cognitive predictive maintenance of urban assets using city information modeling—Systematic review. *Buildings*, *15*(5), 690. https://doi.org/10.3390/buildings15050690
12. Lee, J. Y., Irisboev, I. O., & Ryu, Y.-S. (2021). Literature review on digitalization in facilities management and facilities management performance measurement: Contribution of Industry 4.0 in the global era. *Sustainability*, *13*(23), 13432. https://doi.org/10.3390/su132313432
13. Li, G., Chen, L., Fan, C., Li, T., Xu, C., & Fang, X. (2023). Interpretation and explanation of convolutional neural network-based fault diagnosis model at the feature-level for building energy systems. *Energy and Buildings*, *295*, 113326. https://doi.org/10.1016/j.enbuild.2023.113326
14. Lin, G., Kramer, H., & Granderson, J. (2020). Building fault detection and diagnostics: Achieved savings, and methods to evaluate algorithm performance. *Building and Environment*, *168*, 106505. https://doi.org/10.1016/j.buildenv.2019.106505
15. Liu, J., Zhang, Q., Li, X., Li, G., Liu, Z., Xie, Y., Li, K., & Liu, B. (2021). Transfer learning-based strategies for fault diagnosis in building energy systems. *Energy and Buildings*, *250*, 111256. https://doi.org/10.1016/j.enbuild.2021.111256
16. Marocco, M., & Garofolo, I. (2021). Integrating disruptive technologies with facilities management: A literature review and future research directions. *Automation in Construction*, *131*, 103917. https://doi.org/10.1016/j.autcon.2021.103917

17. Naji, K. K., Gunduz, M., & Al-Qahtani, A. (2024). Unveiling digital transformation: Analyzing building facility management's preparedness for transformation using structural equation modeling. *Buildings*, *14*(9), 2794. https://doi.org/10.3390/buildings14092794
18. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71
19. Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities and challenges. *Automation in Construction*, *132*, 103940. https://doi.org/10.1016/j.autcon.2021.103940
20. Pascual, L.-L. (2022). A systematic review of the extent to which BIM is integrated into operation and maintenance. *Sustainability*, *14*(14), 8692. https://doi.org/10.3390/su14148692
21. Sampaio, R. P., Costa, A. A., & Flores-Colen, I. (2022). A systematic review of artificial intelligence applied to facility management in the building information modeling context and future research directions. *Buildings*, *12*(11), 1939. https://doi.org/10.3390/buildings12111939
22. Shi, Z., & O'Brien, W. (2019). Development and implementation of automated fault detection and diagnostics for building systems: A review. *Automation in Construction*, *104*, 215–229. https://doi.org/10.1016/j.autcon.2019.04.002
23. Shuhaimi, A. M., Mohamed Yusof, L., & Rahman, R. A. (2024). Drivers, capabilities, and challenges for adopting digital twin in facility management: A profound qualitative investigation. *Planning Malaysia*, *22*(32), 589–606. https://doi.org/10.21837/pm.v22i32.1528
24. Wong, J. K. W., Ge, J., & He, S. X. (2018). Digitisation in facilities management: A literature review and future research directions. *Automation in Construction*, *92*, 312–326. https://doi.org/10.1016/j.autcon.2018.04.006
25. Zhao, Y., Li, T., Zhang, X., & Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews*, *109*, 85–101. https://doi.org/10.1016/j.rser.2019.04.021