Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/24489
DOI URL: http://dx.doi.org/10.21474/IJAR01/24489

ISSN (O): 2320-5407
ISSN (P): 3107-4928

RESEARCH ARTICLE

AI-DRIVEN PREDICTIVE COST AND SCHEDULE ANALYTICS FOR IMPROVING PERFORMANCE IN SAUDI GIGA-PROJECTS

Mahmoud Abdel Aziz

Manuscript Info

Manuscript History
Received: 18 August 2026
Final Accepted: 20 September 2026
Published: October 2026

Key words:-
Artificial intelligence; predictive analytics; giga-projects; construction cost; schedule forecasting; Saudi Arabia; project controls; digital twins

Abstract

Saudi giga-projects combine exceptional scope, interdependent delivery packages and evolving resource demands, making cost and schedule control a problem of continuous inference rather than periodic reporting. This critical review examines how artificial intelligence can support earlier recognition of budget exposure, completion risk and feasible recovery options. Thirty peer-reviewed studies published between 2021 and 2025 were selected through a focused search of scholarly publisher and institutional records, with article identities checked against their digital object identifiers. Evidence was organised around baseline estimation, dynamic forecasting, progress sensing, integrated modelling, and implementation governance. Ensemble learning, explainable prediction, continuously updated performance indicators and digital twin data streams show practical promise; however, accuracy reported within individual case studies cannot be treated as transferable performance for Saudi giga-projects. In particular, incompatible work breakdown structures, changing project scope, rare disruption events and weak external validation threaten forecast reliability. A governance-led architecture is therefore proposed that combines package-level data harmonisation, uncertainty-aware cost and time forecasts, causal review of emerging risks, and human-authorised interventions. The synthesis contributes a distinction between prediction quality and decision value: models improve delivery only when forecasts arrive early enough, contain calibrated uncertainty and trigger accountable corrective action. Future research should test cross project transfer, integrated cost and schedule dependencies and prospective operational benefits in Saudi portfolios.

"© 2026 by the Author(s). Published by IJAR under CC BY 4.0. Unrestricted use allowed with credit to the author."

Introduction:-

Delivering large construction programmes is difficult not simply because individual activities are complex, but because apparently local changes can propagate through thousands of contractual, spatial and logistical dependencies. Saudi Arabia's giga-project environment intensifies this coordination problem. Multiple design disciplines, international supply networks, simultaneous infrastructure and building packages, demanding handover interfaces and phased development decisions produce moving baselines. In this paper, giga-projects denote the very large, multi-asset development programmes launched under Saudi Vision 2030, in which destination-scale

Corresponding Author:- Mahmoud Abdel Aziz

infrastructure, real estate, tourism and utility assets are delivered as interdependent packages [31]. Cost reports and master schedules are indispensable; nevertheless, conventional month-end variance summaries reveal outcomes that may already be difficult to reverse. The central management question is consequently whether available information can anticipate emerging exposure before it crystallises into claims, lost productivity or delayed commissioning.
Recent studies position machine learning as a complement to established project control methods rather than a replacement for engineering judgement. A construction lifecycle review identifies planning and execution as prominent settings for artificial intelligence applications [1]. Research on cost estimation methods similarly finds considerable technical experimentation but persistent difficulties concerning input quality and practical adoption [2]. Construction-specific delay models have demonstrated the potential of ensemble methods [3], while work package scheduling research shows how operational indicators can inform forecasts during execution [4]. Any such model must, however, outperform established earned value forecasting methods, including earned schedule [32], whose accuracy has been evaluated empirically across real project portfolios [33]. Taken together, these studies justify considering predictive analytics across the whole delivery cycle, not merely at contract award.

Saudi relevance requires caution. Evidence on capital project material delays identifies design document errors, inaccurate estimates and price inflation as important sources of disruption [5]. Studies of Saudi pipeline projects likewise link risk exposure to cost and time outcomes, although their sectoral setting differs from mixed-use giga-project programmes [6]. Research on construction delays in Makkah indicates that procurement, coordination and site circumstances cannot be reduced to a universal technical variable [7]. These findings support the need for locally contextualised prediction, but they do not establish that any particular algorithm has already reduced overruns across Saudi giga-projects.

The literature offers three fragmented perspectives. Estimation research tests whether algorithms improve numerical accuracy at a defined project stage. Progress monitoring research investigates how actual work can be captured with less manual effort. Digital twin studies consider whether information from physical processes can support integrated decisions. Yet a forecast built from an isolated cost dataset may ignore schedule-critical interfaces; conversely, an accurate activity duration model may overlook financing constraints, variations and escalation. This review treats cost and time as coupled managerial outcomes and asks under what conditions predictive systems could improve decisions rather than simply produce more sophisticated dashboards.

Aim, objectives and review questions:-
The aim is to critically assess the applicability, limitations and implementation requirements of AI-driven predictive cost and schedule analytics for Saudi giga-project delivery, and to develop a conceptually defensible decision architecture grounded in recent construction research. The review pursues four distinct objectives: to compare evidence for baseline and continuously updated forecasts; to examine the information link between physical progress, cost exposure and completion risk; to evaluate interpretability, uncertainty and external validity; and to translate the findings into governance arrangements suited to multi-package programmes.

Three questions organise the analysis. RQ1: Which combinations of data and modelling methods offer decision relevant warning of cost or schedule deterioration? RQ2: How should heterogeneous forecasts be reconciled with contractual baselines and programme interfaces? RQ3: What validation and governance conditions must be met before a prediction can influence funding, resequencing or recovery decisions? These questions intentionally prioritise the quality of inference and its operational consequences over a ranking of algorithms by reported accuracy alone.

Review Methodology:-

A structured critical review was undertaken for literature published between January 2020 and December 2025, with the final analytical corpus comprising thirty peer-reviewed articles dated 2021 to 2025, as no eligible 2020 study was retained after screening. Searches used combinations of “construction cost prediction”, “schedule performance forecasting”, “machine learning delay”, “earned value forecasting”, “digital twin construction”, “progress monitoring”, “explainable AI”, and “Saudi construction risk”. Discovery drew on publicly accessible scholarly publisher pages, DOI-linked article records and university or indexing repositories. This process was deliberately focused rather than presented as an exhaustive database search. It did not involve unpublished project files, direct access to confidential Saudi portfolios or a formal bibliometric census.Screening applied four criteria: a documented scholarly publication within the date window; a relevant construction or project control application; an identifiable analytical mechanism or empirically examined delivery risk; and a traceable article title linked to a DOI. Articles centred solely on generic artificial intelligence, unrelated manufacturing applications or unsupported marketing

projections were excluded from the interpretive corpus. Several adjacent studies were retained where they illuminated a necessary enabling mechanism, including offsite production scheduling and contractual information processing. Publication details and DOI identities were checked against discoverable publisher or institutional records; this verifies bibliographic correspondence but is not an independent replication of reported results.

The unit of synthesis was the decision task rather than the named algorithm. Each source was classified under early cost estimation, in-flight financial forecasting, duration and delay prediction, progress measurement, digital integration, or Saudi risk context. Evidence was then compared across five dimensions: decision horizon, data provenance, target definition, evaluation design and operational transferability. Table 3 (Appendix A) summarises each study's decision task, method, setting, data and validation design. Special attention was paid to whether findings came from a bounded case study, historical records, survey perceptions, simulation or a general literature review. These distinctions matter because a precise estimate of repetitive wall-panel production does not prove comparable accuracy for a first-of-its-kind transport interchange.

A narrative synthesis was selected because the studies differ materially in outputs, units of analysis and evaluation metrics. Combining their error figures into a single pooled improvement percentage would create false comparability. Instead, convergent findings were identified, contradictions were examined and design implications were derived with explicit restrictions on what the evidence can establish. The analytical framework in Figure 1 and operational controls in Table 2 are propositions of this review; they are not results from a newly trained model, a participant survey or a deployed Saudi system.

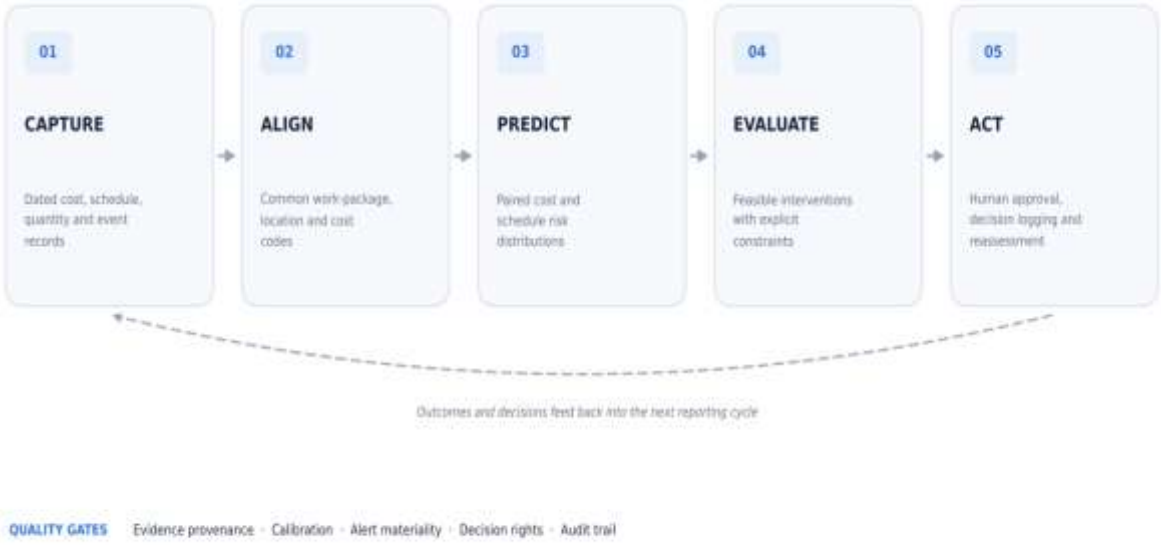


Figure 1. Proposed data-to-decision architecture. The diagram is a conceptual synthesis of the reviewed evidence, not an implemented system. Stage 04 corresponds to scenario evaluation in Section 9.

Predictive cost analytics:-

Construction cost prediction operates at several different levels, and confusing these targets can make a seemingly impressive model operationally misleading. Conceptual estimates assess affordability while scope remains uncertain; tender forecasts address pricing and allocation; forecast-at-completion models revise expected out-turn cost during delivery. A model trained on completed project attributes cannot automatically support weekly decisions about variations, labour productivity or unsettled claims. Wang and colleagues show that model design must be related to available building attributes and to gaps in comparative datasets [2]. Early infrastructure cost modelling using optimised gradient boosting adds a second requirement: explainability should clarify why an estimate changes, not merely provide a lower error statistic [8].

The evidence suggests that nonlinear ensembles can represent interactions among size, complexity, duration and technical characteristics more flexibly than a single linear relation. A road construction study linked predictive

estimates to explanations across several work types, demonstrating the importance of task specific calibration [9]. Combinatorial modelling for high-rise buildings also identified interactions involving duration and architectural characteristics [10]. These findings are informative for distinct project packages within a giga-project, but models fitted to homogeneous historical assets cannot be treated as ready-made portfolio estimators. Station structures, hospitality assets, tunnels and utilities have different cost drivers and measurement conventions.

During execution, the more consequential target may be the probability that an approved package budget is exceeded after recognised commitments, changes and productivity trends are taken into account. Continuous earned value research shows why relying on isolated reporting points conceals evolving performance trajectories [11]. Hybrid final-cost and completion time modelling illustrates an alternative to treating traditional performance indices as sufficient forecasts [12]. Nonetheless, neither statistical sophistication nor continuous updating can compensate for cost data that omit unapproved variations, claims exposure, currency escalation or incomplete purchase orders. A predictive system must therefore distinguish actual cost, committed cost, authorised change and modelled residual exposure.

Explainable predictions can make cost analysis auditable, although explanatory attribution is not proof of causation. Research using confidence intervals and feature attributions shows how uncertainty and interpretability can accompany construction estimates [13]. At programme level, however, a variable such as “planned duration” may function as both an input and a decision subject to change. A model could mislead managers if it treats a management revision as an exogenous driver. Forecasts should be reported with target date, assumptions, data cut-off, prediction range and comparison with a credible baseline; point estimates alone invite unjustified confidence.

Schedule prediction and recovery:-

Schedule analytics must distinguish duration estimation, probability of delay, remaining time prediction and the selection of a recovery strategy. The first three are predictive tasks; the fourth is prescriptive and requires evaluating consequences under resource, safety, access and contractual constraints. An ensemble-based delay study provides support for nonlinear classification of project delay exposure [3]. A work-area duration comparison indicates that model performance can differ according to the aggregation of activities and the construction setting [14]. Neither finding warrants an assumption that all schedule risks are visible in historical progress data.

A useful forecast should recognise leading indicators before completion milestones are breached. The integration of last planner measures with machine learning demonstrates the value of including short-cycle production commitments alongside conventional performance variables [4]. By contrast, historic master schedule updates may encode optimistic recovery plans, outdated logic or approvals that have not yet been secured. Training on such data risks reproducing human planning bias. Responsible model design therefore separates the original baseline, approved revisions, current working programme and independently evidenced physical progress.

The decision value of a schedule warning depends on its causal pathway. An alert that a mechanical package may finish late is incomplete unless the system identifies the handover interfaces exposed and the interventions still feasible. A self-recovering scheduling study used retrospective productivity information to update subsequent work plans and assess resource adjustments [15]. That result supports periodic learning, but its bounded masonry setting differs markedly from interdependent, multi-contractor giga-project activities. Re-sequencing a trade may lower one contractor's delay while creating congestion, rework or additional cost elsewhere.

Schedule predictions should accordingly be expressed as distributions of feasible completion dates and linked to critical or near-critical interfaces. Model evaluation should favour lead time, calibration and avoidable delay over the proportion of retrospectively correct classifications. When only rare failures are labelled, high classification accuracy can coexist with poor detection of consequential events. Validation should include precision at an actionable alert threshold, false alarm burden, recall of material slippages and the time remaining to intervene. Without these measures, a dashboard may seem technically successful while being operationally irrelevant.

Table 1. Evidence-to-decision mapping across predictive project control functions.

Decision task	Core inputs	Interpretation / value	Evidence	Caveat
Conceptual cost	Asset quantities; scope maturity; location; design characteristics	Baseline affordability and alternative appraisal	Explainable estimates [8], [9]	Limited early-stage comparators
Dynamic cost-to-complete	Actuals; commitments; change register; progress	Forward exposure, not merely historic variance	Continuous earned value [11]	Claims and unapproved changes require care
Milestone delivery	Actual productivity; look-ahead metrics; logic	Slippage probabilities and intervention lead time	Planning signals [4]; recovery cases [15]	Context-specific
Physical progress	Images; accepted quantities; BIM objects	Independent field evidence and revised remaining work	Computer vision [16], [17]	Recognition is not acceptance
Integrated interfaces	Package identifiers; dependencies; resources	Joint cost and schedule scenarios and cross-package exposure	Coupled modelling [24], [12]	Causal transfer unproven

Note. References indicate illustrative studies; reliability and transferability depend on project context.

Progress sensing, digital twins and data integration:-

Prediction becomes more useful when field observations are sufficiently timely and trustworthy to update forecasts. Visual construction monitoring research has established methods for connecting images, work recognition and progress assessment [16]. Indoor activity detection linked to building information models demonstrates how observed installed quantities can be compared with an intended design state [17]. A platform independent installation monitoring approach further illustrates the importance of collecting comparable progress evidence in variable site environments [18]. These developments address a persistent weakness of manual percentage complete entries: they can reflect subjective judgement rather than independently verifiable quantities.

Even accurate visual recognition is not equivalent to earned progress. Installed material may fail inspection, be outside the approved work scope, or remain inaccessible for commissioning. Physical progress records therefore require acceptance status, measurement rules and traceability to work packages. A digital representation of cast-in-place construction demonstrates the practical effort needed to link physical processes, data pipelines and decision displays [19]. Integration cannot be reduced to placing computer vision outputs beside cost charts. Each observation must map to a controlled information object with a consistent location, responsible party, activity code and cost account.

Digital twin literature offers useful integration patterns but also warns against treating a three-dimensional model as a complete representation of project behaviour. Reviews identify unresolved definitions, maturity differences and interoperability challenges [28–30]. Site-specific taxonomies distinguish monitoring, prediction and control applications, each of which requires different data and assurance [20]. For Saudi giga-projects, a federated information architecture is more credible than an assumption that every contractor will operate the same platform. A common interface can preserve supplier-specific systems while harmonising identifiers and evidence quality.

Production-focused applications show both opportunity and boundary conditions. An offsite construction twin integrated sensors, computer vision and simulation for scheduling and monitoring [21]. Semantic digital twin research addressed dynamic scheduling of precast elements affected by resource and process uncertainty [22]. Material traceability research connected on site fit out information to a service-oriented twin [23]. These studies support more responsive work package control; they do not demonstrate that production-level improvements automatically aggregate into better total programme outcomes. A portfolio may remain constrained by land access, permitting, design decisions or external supply disruptions after local productivity improves.

Coupled cost and schedule reasoning:-

Cost and schedule should not be modelled as independent scorecards because their drivers often overlap. Extended working hours may accelerate a late package but increase labour costs, lower productivity or heighten safety exposure. Material substitutions can protect a milestone while introducing redesign and quality risk. Deferred design decisions may reduce immediate spending yet increase later rework. Combined cost and duration forecasting has been explored using Bayesian optimisation and interpretable ensemble methods [24], and earned value based hybrid modelling offers a different route to joined out-turn assessment [12]. These contributions make integrated prediction plausible, although they do not resolve the causal identification problem.

A robust architecture should distinguish three relationships. First, common causes such as scope change, material scarcity or access restrictions can affect both outcomes. Second, schedule consequences can generate cost through prolongation, acceleration and claims. Third, cost controls can change time performance when procurement or mobilisation is postponed. These pathways imply that simply concatenating cost and schedule datasets may embed correlations that shift when managers intervene. Scenario analysis should explicitly model intervention assumptions and compare forecast changes against a clearly stated “no action” reference trajectory.

Cross-package dependencies create another source of risk. A late civil works handover can affect electrical installation and systems testing even when each downstream contractor reports acceptable local progress. Shared cranes, utility corridors, commissioning teams and inspection resources create congestion that is largely invisible to isolated activity predictors. A programme-level graph should therefore combine precedence constraints, spatial interference, contractual interfaces and resource conflicts. Its output need not be a single master prediction; a more credible representation is a set of linked exposure bands for affected milestones and budgets.

Interpretation demands restraint. Attribution methods can locate variables strongly associated with a forecast, while commercial risk assessment must determine whether the relationship is actionable. Construction cost studies from other national contexts demonstrate interpretable alternatives to black-box estimation [8–10]; they do not prove that an attributed factor is a lawful or effective contractual basis for changing a package scope. Forecasts should be reviewed jointly by planners, cost managers, engineering leads and commercial teams before recommending irreversible actions. The appropriate role for artificial intelligence is evidence synthesis and structured anticipation, with accountable people retaining authority over scope, payments, contractual notices and acceleration decisions.



Figure 2. Hypothesized feedback among physical production, remaining time, financial exposure and recovery decisions.

Transferability to Saudi giga-project conditions:-

External validity is the decisive challenge in applying the reviewed methods to Saudi giga-projects. Evidence from ordinary buildings, transport works and repetitive offsite manufacturing is not interchangeable with integrated, first-

of-their-kind destinations. Project scale changes the density of interfaces; novelty changes the availability of comparable training cases; and evolving investment decisions change the meaning of the baseline itself. Large portfolios may also contain packages that are simultaneously in concept design, procurement, construction and commissioning. A single project-level forecast cannot represent these distinct information states.

Locally observed constraints should guide feature design. Saudi capital project evidence links material delays to design document quality, weak estimation and inflation [5]; pipeline research identifies risk pathways to time and cost outcomes [6]. Evidence from Makkah shows that local construction delays remain sensitive to project management context [7]. These studies justify testing features relating to design maturity, approved change volume, procurement lead times, payment reliability, site access and workforce availability. They do not license precise claims about the frequency or financial impact of those factors across giga-projects, where programme-level data remain unevenly disclosed.

Model transfer should be treated as a testable hypothesis. This logic parallels reference class forecasting, which anchors predictions in the outcomes of comparable completed projects rather than in project-specific optimism [34]. A baseline could be trained on earlier packages, frozen before later observations and evaluated on genuinely subsequent projects or contractors. Accuracy from random splits of related work packages may be inflated when nearly identical design elements or shared procurement conditions appear in training and test sets. Results should also be separated by asset class, construction phase and commercial arrangement. A model that performs well for repetitive residential units may fail on signature architecture or complex underground works, and a truthful system must be allowed to abstain when no relevant comparator exists.

Portfolio oversight introduces a further governance issue: reporting incentives. Contractors may have legitimate differences in how they recognise progress, contingencies and potential compensation events. If a predictive score affects funding or management scrutiny, data providers may change coding behaviour. Independent evidence, field audit trails and transparent definitions are therefore necessary to prevent the analytics process from degrading the information on which it depends. A federated model can assign data ownership while specifying the minimum shared signals needed for coherent programme forecasts.

Proposed decision architecture and validation:-

Figure 1 translates the synthesis into five connected control functions: data capture, semantic alignment, predictive inference, scenario evaluation and authorised action. The architecture begins with immutable reporting snapshots and the preservation of original schedule and budget baselines. It then reconciles activity, cost account, location and package identifiers before producing updated probabilities of budget breach, milestone slippage and resource pressure. Every prediction carries a data quality flag, reference date and uncertainty range. An exception is passed for review when both its materiality and the remaining intervention window justify attention.

Figure 2 explains the coupled feedback logic. Physical progress and production rates inform remaining duration estimates; schedule changes alter forecast preliminaries, indirect costs and exposure to claims; funding and procurement choices, in turn, modify resource availability and future progress. External disruptions and changes of scope enter as distinct events rather than being silently absorbed into random error. A scenario engine compares feasible interventions while preserving explicit constraints, including safety, quality, access and contractual authorisation. Its purpose is to illuminate consequences, not automate commercial commitments.

Validation must precede operational reliance. First, out-of-time back testing should reproduce what was actually knowable at each historical reporting date, excluding data created later. Second, leave-project-out and leave-contractor-out tests should measure transfer across assets and organisations. Third, prediction intervals should be assessed for empirical coverage, because an apparently accurate point estimate with systematically narrow bands is unsafe for contingency decisions. Fourth, forecasts must be compared with credible baselines: planner estimates, earned value projections and simple persistence models. Complexity is worthwhile only when improvement is material, stable and useful.

Finally, a prospective pilot should record both alerts and actions. Metrics should include avoided milestone exposure, forecast bias, useful warning lead time, decision latency, false alarm workload and documented reasons for rejecting model advice. The comparison should account for concurrent controls and programme changes, rather than attributing every favourable outcome to software. Table 2 sets out corresponding controls and accountable

roles. This validation pathway differentiates technical feasibility, organisational acceptance and demonstrated delivery benefit.

Table 2. Model assurance and operational accountability for a giga-project pilot.

Control point	Required validation	Operational threshold / output	Accountability
Data acceptance	Reconcile package codes, snapshots and source timestamps	Quarantine unmatched or unapproved records	Information manager + controls lead
Model release	Out-of-time and leave-project-out tests; simple benchmarks	Document bias, calibration and exceptions	Analytics lead + independent assurance
Forecast alert	Materiality, interval width and available lead time	Rank as review / watch / escalate	Project controls manager
Recovery option	Test cost, completion, safety and contractual trade-offs	Explicit assumptions and feasible alternatives	Planning + commercial + engineering
Human decision	Approval limits, rationale, audit trail	Record accepted, changed or declined recommendation	Authorisedprogramme sponsor
Post-action learning	Track realised outcomes, drift and data changes	Retain, recalibrate, suspend or retire model	Portfolio assurance function

Discussion: performance, uncertainty and organisational learning:-

The reviewed evidence supports the proposition that predictive systems can improve information available to project managers, but not the stronger claim that algorithm adoption alone improves giga-project performance. Empirical studies establish useful performance for specific tasks: schedule estimation with production indicators [4], interpretable conceptual cost modelling [8], dynamic offsite scheduling [21] and continuous project tracking [11]. Their value lies in showing mechanisms that can be adapted and evaluated. Generalising their numerical gains across unrelated asset classes would neglect differences in baselines, reporting incentives, decision rights and physical constraints.

A distinction between epistemic and operational uncertainty is useful. Epistemic uncertainty arises from sparse or unrepresentative data, incorrect assumptions and shifting relationships; additional evidence or model redesign may reduce it. Operational uncertainty reflects genuinely variable weather, productivity, approvals and market conditions, some of which remains even with perfect measurement. Both should appear in decision reports. When a forecast confidence interval expands because a new contractor operates in unfamiliar conditions, the system should communicate reduced certainty instead of presenting a false impression of precision.

Generative AI introduces a further category of tool. Large language models are increasingly used to draft progress narratives, summarise variance reports and query project data in natural language. These capabilities can shorten reporting cycles, but they do not perform forecasting in the statistical sense reviewed here, and they may produce fluent explanations that are not supported by the underlying records. Where such tools are used, outputs should be restricted to traceable source data, flagged where uncertain and reviewed by the accountable controls lead before issue. Evaluating these tools against the lead-time, calibration and decision-value criteria proposed in this review is a priority for future research.

Organisational learning is equally important. A closed-loop system records what alert was raised, which intervention was approved and what occurred afterwards. This creates evidence about whether particular decisions actually changed cost or time outcomes. Without intervention records, a model may later learn from post-treatment data as if management actions were natural conditions. Moreover, improvements should be evaluated at programme level: an intervention that protects one milestone can displace scarce resources and create losses elsewhere. Effectiveness therefore depends on balancing local package interests with shared portfolio constraints.

Behavioural factors also condition predictive value. Daily crew reports can provide an early signal of productivity drift [25], but only if contractors submit them consistently and promptly; incentive mechanisms influence whether

parties share timely, fine-grained records [26], and adoption decisions depend on how each stakeholder perceives the balance of costs and benefits [27]. These conditions should therefore be audited alongside predictive accuracy, because a technically sound model fed by delayed or selectively coded data will degrade over time.

These considerations challenge a technology-first implementation strategy. Investing immediately in high-complexity models without harmonised controls can produce unreliable alerts at scale. A staged sequence is more defensible: consolidate trusted package data, establish interpretable baselines, pilot targeted forecasting, test predictive utility and only then connect advanced sensing or digital twin features. Research on digital twins and their architecture reinforces the need for integration standards and fit-for-purpose applications rather than indiscriminate platform expansion [28, 29, 30].

Answers to the review questions. In relation to RQ1, ensemble and explainable models trained on dated cost, commitment and production data, such as last planner metrics, offer the most decision-relevant warning when combined with verified field progress, although the evidence remains task-specific (Sections 4–6). In relation to RQ2, heterogeneous forecasts should be reconciled through harmonised package, location and cost-account identifiers, preserved original baselines and a programme-level dependency graph that reports linked exposure bands rather than a single master prediction (Sections 6 and 7; Figure 1). In relation to RQ3, out-of-time back-testing, leave-project-out and leave-contractor-out tests, calibrated prediction intervals, comparison with planner, earned value and persistence baselines, and documented human authorisation are prerequisites before predictions influence funding, resequencing or recovery decisions (Section 9; Table 2).

Research gaps and future investigations:-

Four gaps warrant priority. The first concerns joint targets. Studies frequently predict cost or time separately, while linked distributions of final cost and completion date remain less mature. New work should test multivariate and graph-based methods against simpler coupled baselines, measuring whether dependency modelling improves decisions. The second gap is real-world validation: transparent, multi-project Saudi benchmarks with dated data snapshots, target definitions and lawful access protocols are needed to assess generalisation. An anonymised portfolio spanning infrastructure, buildings and systems packages would support stronger evidence than isolated retrospective examples.

The third gap concerns intervention effects. Predicting that a milestone will slip is different from estimating whether overtime, resequencing or alternate procurement will improve the result. Future studies should evaluate counterfactual approaches with engineering constraints and carefully documented decision histories. The fourth gap involves model governance under change. As scope and procurement strategies evolve, the features that predict delay may change meaning. Drift detection, recalibration and fallback procedures should be investigated alongside fairness between contractors and traceability of human overrides.

Research designs should favour prospective comparisons where feasible, repeated across varied packages and implementation maturity levels. Reports should publish not only model errors but also uncertainty coverage, warning lead time, work required to maintain the pipeline and instances where models were withheld. This would allow researchers and clients to distinguish compelling demonstrations from deployable controls.

Practical and policy implications:-

For programme sponsors, the immediate priority is not selecting a particular algorithm but defining an enforceable data contract across delivery packages. Contractual information requirements should specify stable work breakdown codes, reporting timestamps, progress measurement rules, event identifiers and the governance of approved changes. A well-defined information standard reduces ambiguity without requiring all organisations to abandon their existing software. Cybersecurity, commercial confidentiality and proportionate access must be built into the data-sharing arrangement.

For project controls teams, predictive outputs should be integrated into existing assurance gates. Cost forecasts should disclose inclusions for commitments, changes, escalation and contingency; schedule forecasts should identify affected logic paths and interfaces. Every proposed recovery measure should be assessed for the combined effects on cost, completion, safety, quality and contractual obligations. Decision papers can compare forecast distributions before and after a proposed action, with assumptions recorded and challenged by the relevant specialists.For procurement and technology leaders, pilot selection should target clearly measurable problems with repeated observations and actionable lead times, rather than the most visually impressive digital demonstration. A package

with reliable daily quantities and meaningful resource flexibility may provide a better proof of value than a high-profile asset whose decisions are dominated by infrequent strategic changes. Procurement evaluation should require transparent model documentation, portability of project data, explainability for material alerts and the ability to discontinue models that underperform.

At policy and industry level, voluntary common taxonomies and evaluation protocols could make anonymised learning across programmes possible. Independent benchmarking would help prevent inflated claims based on incomparable test sets. The wider benefit is a more credible evidence base for investment, delivery assurance and workforce development. It should be recognised, however, that no analytics framework can substitute for timely approvals, capable contractors, adequate funding or coherent design management. Predictive intelligence is most valuable when embedded in institutions able to act on it.

Limitations:-

This review synthesises thirty DOI-traceable publications but does not claim an exhaustive enumeration of every relevant study. The corpus is uneven across project types and countries, and some of the strongest demonstrations concern bounded tasks rather than whole programmes. Its conclusions therefore describe plausible mechanisms and implementation conditions, not measured improvements in Saudi giga-project performance. No confidential cost records, live schedules, site images or contractor databases were analysed. The figures represent conceptual relationships, and the tables express a critical synthesis rather than new statistical results. Publication bias, inconsistent reporting of failed deployments and uncertainty about causal attribution further restrict generalisation. These limitations make prospective, independently evaluated pilots essential.

Conclusion:-

AI-driven analytics can extend cost and schedule control from retrospective variance reporting towards earlier, evidence based anticipation. Recent studies support the usefulness of interpretable cost estimates, machine learning duration forecasts, continuously updated project indicators and field connected digital information. Their combined lesson is conditional: technically accurate predictions become operationally valuable only when source data are trustworthy, uncertainty is visible, package interfaces are modelled and decisions remain accountable. Saudi giga-projects present a demanding setting because asset diversity, strategic change and contractual interdependence weaken naïve transfer from isolated case studies.

The proposed architecture combines harmonised package records, linked cost and schedule forecasts, scenario assessment and human-authorised recovery within a transparent governance cycle. Its merit must ultimately be established through out-of-time testing, cross project validation and prospective measurement of decisions improved. Rather than promising automatic control of complex developments, the defensible objective is to identify emerging exposure sufficiently early for project leaders to compare feasible responses and intervene with greater clarity.

Declarations:-

Author contributions: [Author name]: conceptualisation, methodology, analysis, writing – original draft, writing – review and editing.

Funding: This research received no external funding. [Confirm.]

Conflict of interest: The author is employed as a project controls professional by a project management consultancy engaged on Saudi giga-project programmes. The views expressed are the author's own and do not represent those of the employer or its clients. [Confirm wording and any employer/client approval requirements.]

Data availability: No new data were created or analysed in this study. [Update if an empirical component is added.]

Use of generative AI: [Describe any AI tools used, e.g., for language editing or structuring, in line with the target journal's policy.] The author reviewed and edited all content and takes full responsibility for the manuscript.

Appendix A. Characteristics of the reviewed studies

Table 3. Decision task, method and evidence characteristics of the thirty reviewed studies.

Ref.	Study	Decision task	Method / focus	Setting, data and validation design
[1]	Datta et al. (2024)	Lifecycle overview	Review of AI/ML across the construction project lifecycle	Literature review

[2]	Wang et al. (2025)	Early cost estimation	Review of ML techniques for building cost estimation	Literature review
[3]	Egwim et al. (2021)	Delay prediction	Ensemble machine learning	[to complete]
[4]	Lagos et al. (2024)	Schedule performance	Last Planner System metrics with ML	[to complete]
[5]	Shash and AbuAlnaja (2023)	Saudi risk context	Causes of material delays in capital projects	Saudi Arabia; [to complete]
[6]	Al-Gahtani et al. (2023)	Saudi risk context	Risk factors for time and cost overruns, pipeline projects	Saudi Arabia; [to complete]
[7]	Abdulghafour and Salman (2023)	Saudi risk context	Construction delays from the consultant's perspective	Makkah, Saudi Arabia; [to complete]
[8]	Zhang et al. (2025)	Early cost estimation	XGBoost with Bayesian optimisation, infrastructure projects	[to complete]
[9]	Larocque et al. (2025)	Early cost estimation	Explainable ML, road construction costs	[to complete]
[10]	Liu and Lin (2025)	Early cost estimation	Combinatorial ML with interpretability, high-rise buildings	[to complete]
[11]	Ngo et al. (2022)	In-flight forecasting	Continuous earned value management with singularity functions	[to complete]
[12]	Bakhshi et al. (2022)	In-flight forecasting	Hybrid ML with metaheuristic optimisers, final cost and time	[to complete]
[13]	Chen et al. (2025)	Early cost estimation	Advanced ML with explainable AI and confidence intervals	[to complete]
[14]	Zhang and Li (2024)	Duration prediction	Comparison of ML regression models	[to complete]
[15]	AlJassmi and Abduljalil (2023)	Schedule recovery	Periodic plan updating and recovery optimisation	[to complete]
[16]	Reja et al. (2022)	Progress measurement	Computer vision-based progress monitoring (review)	Literature review
[17]	Wei et al. (2022)	Progress measurement	Mask R-CNN with BIM, indoor construction	[to complete]
[18]	Zhao et al. (2023)	Progress measurement	Platform-independent visual installation monitoring	[to complete]
[19]	Chacón et al. (2024)	Digital integration	Digital twin of cast-in-place reinforced concrete works	Case study; [to complete]
[20]	Saif et al. (2024)	Digital integration	Construction digital twin taxonomy	Literature review
[21]	Alsakka et al. (2024)	Digital integration	Offsite digital twin for estimation, scheduling and monitoring	[to complete]
[22]	Kosse et al. (2024)	Digital integration	Semantic digital twin, dynamic scheduling of precast production	[to complete]
[23]	Yang et al. (2024)	Digital integration	Digital twin for material traceability in fit-out	[to complete]
[24]	Elmousalami et al. (2025)	Coupled cost and duration	Bayesian-optimised ensemble learning	[to complete]
[25]	Sadatnya et al. (2023)	Productivity prediction	ML using daily work reports	[to complete]
[26]	Xiao et al. (2024)	Data sharing and incentives	Incentive mechanism for digital twin data sharing	[to complete]

[27]	Yang and Ng (2024)	Adoption decisions	Game-theory analysis of digital twin adoption, modular construction	[to complete]
[28]	Moshood et al. (2024)	Digital integration	Infrastructure digital twin (review)	Literature review
[29]	Long et al. (2024)	Digital integration	Systematic review of digital twin applications	Literature review
[30]	AlBalkhy et al. (2024)	Digital integration	Digital twins in the built environment (review)	Literature review

References:-

1. Datta, S.D.; Islam, M.; Sobuz, M.H.R.; Ahmed, S.; Kar, M. (2024). Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. Heliyon, 10(5), e26888. https://doi.org/10.1016/j.heliyon.2024.e26888
2. Wang, R.; Salleh, H.; Lyu, J.; Abdul-Samad, Z.; Mohd Radzuan, N.F.; Wen, K.C. (2025). Application and prospect of machine learning techniques in cost estimation of building projects. Engineering, Construction and Architectural Management, 32(12), 8445–8471. https://doi.org/10.1108/ECAM-05-2024-0595
3. Egwim, C.N.; Alaka, H.; Toriola-Coker, L.O.; Balogun, H.; Sunmola, F. (2021). Applied artificial intelligence for predicting construction projects delay. Machine Learning with Applications, 6, 100166. https://doi.org/10.1016/j.mlwa.2021.100166
4. Lagos, C.I.; Herrera, R.F.; Mac Cawley, A.F.; Alarcón, L.F. (2024). Predicting construction schedule performance with last planner system and machine learning. Automation in Construction, 167, 105716. https://doi.org/10.1016/j.autcon.2024.105716
5. Shash, A.A.; AbuAlnaja, F.M. (2023). Causes of material delays in capital projects in Saudi Arabia. International Journal of Construction Management, 23(7), 1109–1117. https://doi.org/10.1080/15623599.2021.1956294
6. Al-Gahtani, K.; Shafaay, M.; Ahmed, O.; Alawshan, M.; Alsanabani, N.; Aljadhai, S. (2023). Risk factors for time and cost overruns of pipeline projects in Saudi Arabia. Advances in Civil Engineering, 2023, 9497451. https://doi.org/10.1155/2023/9497451
7. Abdulghafour, A.B.; Salman, K. (2023). Delays of construction projects in Makkah from the point of view of the consultant. Journal of Umm Al-Qura University for Engineering and Architecture, 14, 122–134. https://doi.org/10.1007/s43995-023-00024-2
8. Zhang, J.; Yuan, J.; Mahmoudi, A.; Ji, W.; Fang, Q. (2025). A data-driven framework for conceptual cost estimation of infrastructure projects using XGBoost and Bayesian optimization. Journal of Asian Architecture and Building Engineering, 24(2), 751–774. https://doi.org/10.1080/13467581.2023.2294871
9. Larocque, R.; Boulé, A.-M.; Cappart, Q. (2025). Estimating road construction costs with explainable machine learning. INFORMS Journal on Applied Analytics, 55(2), 137–153. https://doi.org/10.1287/inte.2023.0027
10. Liu, Z.; Lin, J. (2025). Combinatorial machine learning approaches for high-rise building cost prediction and their interpretability analysis. Journal of Asian Architecture and Building Engineering, 24(4), 2817–2828. https://doi.org/10.1080/13467581.2024.2378001
11. Ngo, K.A.; Lucko, G.; Ballesteros-Pérez, P. (2022). Continuous earned value management with singularity functions for comprehensive project performance tracking and forecasting. Automation in Construction, 143, 104583. https://doi.org/10.1016/j.autcon.2022.104583
12. Bakhshi, R.; Fard Moradinia, S.; Jani, R.; Vafaei Poor, R. (2022). Presenting a hybrid scheme of machine learning combined with metaheuristic optimizers for predicting final cost and time of project. KSCE Journal of Civil Engineering, 26(8), 3188–3203. https://doi.org/10.1007/s12205-022-1424-3
13. Chen, L.; Xu, C.; Lim, W.H.; et al. (2025). Transparent and reliable construction cost prediction using advanced machine learning and explainable AI. Engineering Science and Technology, an International Journal, 70, 102159. https://doi.org/10.1016/j.jestch.2025.102159
14. Zhang, S.; Li, X. (2024). A comparative study of machine learning regression models for predicting construction duration. Journal of Asian Architecture and Building Engineering, 23(6), 1980–1996. https://doi.org/10.1080/13467581.2023.2278887
15. AlJassmi, H.; Abduljalil, Y. (2023). Towards self-recovering construction schedules: A new method for periodically updating project plans and optimizing recovery actions. Journal of Asian Architecture and Building Engineering, 22(4), 2335–2347. https://doi.org/10.1080/13467581.2022.2153055

16. Reja, V.K.; Varghese, K.; Ha, Q.P. (2022). Computer vision-based construction progress monitoring. Automation in Construction, 138, 104245. https://doi.org/10.1016/j.autcon.2022.104245
17. Wei, W.; Lu, Y.; Zhong, T.; Li, P.; Liu, B. (2022). Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM. Automation in Construction, 140, 104327. https://doi.org/10.1016/j.autcon.2022.104327
18. Zhao, X.; Jin, Y.; Selvaraj, N.M.; Ilyas, M.; Cheah, C.C. (2023). Platform-independent visual installation progress monitoring for construction automation. Automation in Construction, 154, 104996. https://doi.org/10.1016/j.autcon.2023.104996
19. Chacón, R.; Posada, H.; Ramonell, C.; et al. (2024). Digital twinning of building construction processes. Case study: A reinforced concrete cast-in structure. Journal of Building Engineering, 84, 108522. https://doi.org/10.1016/j.jobe.2024.108522
20. Saif, W.; RazaviAlavi, S.; Kassem, M. (2024). Construction digital twin: A taxonomy and analysis of the application technology data triad. Automation in Construction, 167, 105715. https://doi.org/10.1016/j.autcon.2024.105715
21. Alsakka, F.; Yu, H.; El-Chami, I.; Hamzeh, F.; Al-Hussein, M. (2024). Digital twin for production estimation, scheduling and real-time monitoring in offsite construction. Computers & Industrial Engineering, 191, 110173. https://doi.org/10.1016/j.cie.2024.110173
22. Kosse, S.; Betker, V.; Hagedorn, P.; König, M.; Schmidt, T. (2024). A semantic digital twin for the dynamic scheduling of Industry 4.0-based production of precast concrete elements. Advanced Engineering Informatics, 62, 102677. https://doi.org/10.1016/j.aei.2024.102677
23. Yang, Y.; Li, M.; Yu, C.; Zhong, R.Y. (2024). Digital twin-enabled visibility and traceability for building materials in onsite fitout construction. Automation in Construction, 166, 105640. https://doi.org/10.1016/j.autcon.2024.105640
24. Elmousalami, H.; Elshaboury, N.; Ibrahim, A.H.; Elyamany, A.H. (2025). Bayesian optimized ensemble learning system for predicting conceptual cost and construction duration of irrigation improvement systems. KSCE Journal of Civil Engineering, 29(3), 100014. https://doi.org/10.1016/j.kscej.2024.100014
25. Sadatnya, A.; Sadeghi, N.; Sabzekar, S.; et al. (2023). Machine learning for construction crew productivity prediction using daily work reports. Automation in Construction, 152, 104891. https://doi.org/10.1016/j.autcon.2023.104891
26. Xiao, J.; Ma, S.; Wang, S.; Huang, G.Q. (2024). Fine-grained digital twin sharing framework for smart construction through an incentive mechanism. International Journal of Production Economics, 276, 109382. https://doi.org/10.1016/j.ijpe.2024.109382
27. Yang, J.; Ng, S.T. (2024). Prospects for digital twin technology in the building modular construction and operation phases: A game theory-based analysis. Journal of Cleaner Production, 470, 143344. https://doi.org/10.1016/j.jclepro.2024.143344
28. Moshood, T.D.; Rotimi, J.O.B.; Shahzad, W.; Bamgbade, J.A. (2024). Infrastructure digital twin technology: A new paradigm for future construction industry. Technology in Society, 77, 102519. https://doi.org/10.1016/j.techsoc.2024.102519
29. Long, W.; Bao, Z.; Chen, K.; Ng, S.T.; Wuni, I.Y. (2024). Developing an integrative framework for digital twin applications in the building construction industry: A systematic literature review. Advanced Engineering Informatics, 59, 102346. https://doi.org/10.1016/j.aei.2023.102346
30. AlBalkhy, W.; Karmaoui, D.; Ducoulombier, L.; Lafhaj, Z.; Linner, T. (2024). Digital twins in the built environment: Definition, applications, and challenges. Automation in Construction, 162, 105368. https://doi.org/10.1016/j.autcon.2024.105368
31. Kingdom of Saudi Arabia (2016). Saudi Vision 2030. Council of Economic and Development Affairs, Riyadh. Available online: https://www.vision2030.gov.sa
32. Lipke, W. (2003). Schedule is different. The Measurable News, March, 31–34.
33. Batselier, J.; Vanhoucke, M. (2015). Empirical evaluation of earned value management forecasting accuracy for time and cost. Journal of Construction Engineering and Management, 141(11), 05015010. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001008
34. Flyvbjerg, B. (2006). From Nobel Prize to project management: Getting risks right. Project Management Journal, 37(3), 5–15. https://doi.org/10.1177/875697280603700302