American Journal of Data Science and Analytics

Volume: 7; **Issue:** 8
Pages: 34-67

Industrial Engineering Framework for Predictive Detection of Cost and Schedule Overruns in Infrastructure Projects

Qaium Hossain[1];

[1]. Doctoral Researcher, Lamar University, Texas, USA;
Email: qaiumadi33@gmail.com

Doi: 10.63125/k10whz44
Received: *15 May 2026;* ***Revised:*** *12 June 2022;* ***Accepted:*** *09 July 2026;* ***Published:*** *02 August 2026*

Abstract

This study developed and evaluated an industrial engineering framework for the predictive detection of cost and schedule overruns in infrastructure projects by integrating earned value measurement, statistical process control, and supervised classification into a single early-warning pipeline. Because audited multi-project execution records are rarely available for research, the framework was validated on a simulation-calibrated portfolio of 240 infrastructure projects generated by Monte Carlo methods and moment-matched to published reference-class overrun distributions for roads, rail, bridges and tunnels, water, and energy schemes; 224 projects with complete monthly trajectories were retained for analysis. Final cost overrun averaged 30.7% (SD = 37.5) and 77.2% of projects finished above budget, consistent with the calibration targets. Multiple regression showed that indicators observable at 30% physical completion explained 66.0% of the variance in final cost overrun ($R^2 = 0.660$, $F(5, 218) = 84.50$, $p < .001$), with the cost performance index ($\beta = -0.40$, $p < .001$), change-order intensity ($\beta = 0.24$, $p < .001$), and scope stability ($\beta = -0.22$, $p < .001$) as the strongest predictors. Adding front-end definition and change-dynamics indicators to earned value indices increased explained variance by 13.0 percentage points ($\Delta R^2 = 0.130$, $p < .001$) and raised cross-validated classification accuracy for material overrun from 75.9% to 82.7% ($p = .036$, $dz = 0.78$); gradient boosting and logistic regression performed equivalently, while both exceeded a single decision tree. A cumulative deviation index computed at 30% completion correlated with the subsequent frequency of out-of-control control-chart signals ($r = 0.64$, $p < .001$) and with final overrun ($r = 0.71$, $p < .001$). Reliability analysis supported the Execution Performance Index ($a = 0.79$) and the Front-End Definition Index ($a = 0.83$). The findings indicate that the information set, rather than the algorithm, drives early detection, and that an integrated industrial engineering layer can convert routine project control data into reliable overrun warnings.

Keywords

Cost Overrun, Schedule Overrun, Earned Value Management, Statistical Process Control, Predictive Analytics, Infrastructure Projects.

INTRODUCTION

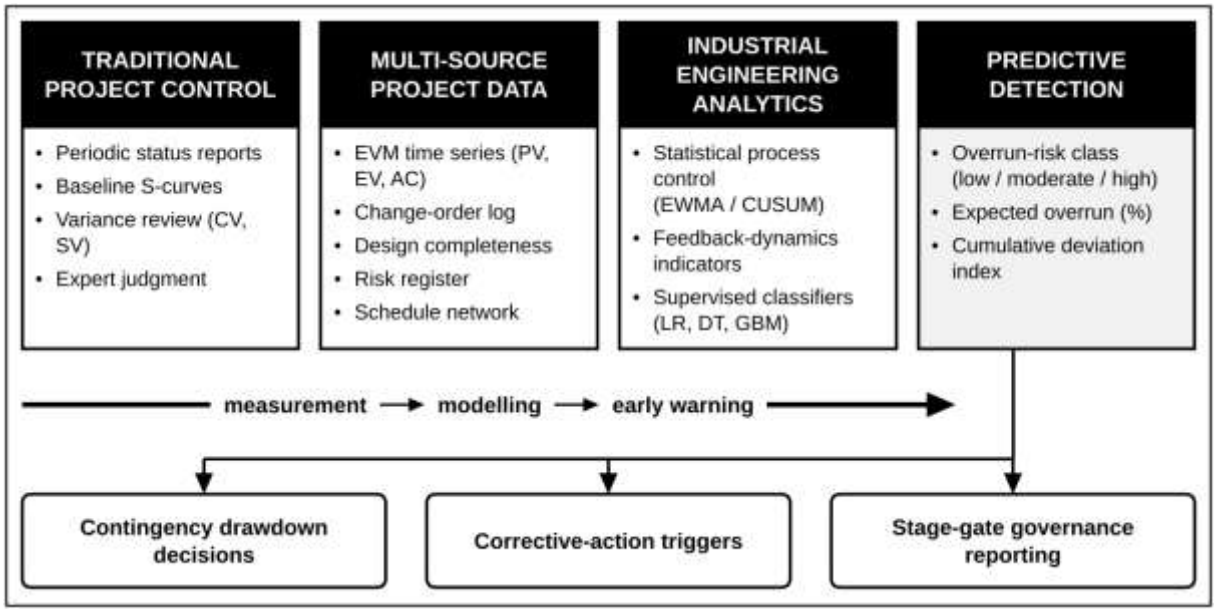
Cost and schedule overruns are among the most persistent and well-documented failures of infrastructure delivery. A cost overrun is conventionally defined as the difference between the actual cost at completion and the cost estimated at the decision to build, expressed as a percentage of that estimate, while a schedule overrun is the corresponding difference between actual and planned duration (Flyvbjerg et al., 2003). In the largest statistically grounded study of its kind, nine out of ten transportation infrastructure projects were found to have underestimated costs, with average escalation of 44.7% for rail, 33.8% for bridges and tunnels, and 20.4% for roads, and the phenomenon was shown to be stable across seventy years and across continents (Flyvbjerg et al., 2002). Subsequent work has extended the evidence base to hydropower dams, electricity networks, water schemes, and information systems, with comparable or larger escalation and a pronounced right-skewed distribution in which a minority of projects experience extreme outcomes (Ansar et al., 2014; Sovacool et al., 2014; Baccarini & Love, 2014; Flyvbjerg et al., 2022). Predictive detection, as used in this study, refers to the systematic and quantitative identification of projects that are on a trajectory toward a material overrun while sufficient time and budget remain for corrective action. An industrial engineering framework for predictive detection therefore combines measurement, statistical control, and decision logic in the manner that production and quality systems combine instrumentation, control charts, and corrective-action procedures (Montgomery, 2020).

Traditional approaches to controlling infrastructure projects have provided the foundations on which any predictive framework must build. Earned value management (EVM) formalizes the comparison of planned value, earned value, and actual cost into the cost performance index (CPI) and schedule performance index (SPI), and extends these into estimates at completion (Anbari, 2003; Fleming & Koppelman, 2010). The earned schedule extension corrected the well-known tendency of the conventional schedule index to converge to unity as projects finish late, producing time-based indicators that remain informative to completion (Vandevoorde & Vanhoucke, 2006; Lipke et al., 2009). At the same time, their limitations in a predictive role are well understood. Status reports are periodic and retrospective, variance thresholds are frequently set by convention rather than by statistical evidence, and the indices summarize an entire project into a single ratio that obscures where and when deviation is emerging (Hazır, 2015; Rozenes et al., 2006). Comparative evaluations show that deterministic forecasting methods built on these indices perform acceptably in the final third of a project but are weak in the early stages, when intervention is most valuable (Vanhoucke & Vandevoorde, 2007; Batselier & Vanhoucke, 2015a). Reference-class forecasting addressed estimation bias at the front end by drawing on the distribution of outcomes of comparable completed projects, but it is a pre-execution calibration rather than an execution-phase detection tool (Flyvbjerg, 2006; Flyvbjerg, 2008). These constraints have motivated a transition from descriptive variance review toward statistical project control and data-driven classification, in which deviation is detected against calibrated limits and the probability of a material overrun is estimated as a function of multiple observed indicators (Colin & Vanhoucke, 2014; Wauters & Vanhoucke, 2016).

Infrastructure projects generate rich and heterogeneous control data that are rarely integrated for predictive purposes. Monthly cost ledgers, schedule updates, change-order registers, design-completion records, risk registers, and procurement logs each capture a different mechanism by which overruns develop (Bilal et al., 2016; Soibelman & Kim, 2002). Change orders have been shown to be both a direct source of cost growth and a leading indicator of design immaturity, with their timing and frequency associated with productivity losses and claims (Hanna et al., 2002; Ibbs, 2005; Anastasopoulos et al., 2010). Front-end definition, including the completeness of design at contract award and the maturity of the estimate, has been identified repeatedly as a determinant of outturn cost in highway and public works projects (Shane et al., 2009; Molenaar, 2005; Creedy et al., 2010). From a measurement-science perspective, the combination of these records produces a multivariate panel of indicators observed at successive reporting periods, which is precisely the structure to which industrial engineering tools such as exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts were designed to apply (Page, 1954; Roberts, 1959; Lucas & Saccucci, 1990). Statistical project control studies have demonstrated that tolerance limits derived from simulated or historical variability outperform conventional fixed thresholds in distinguishing acceptable fluctuation

from systematic deterioration (Colin & Vanhoucke, 2015; Aliverdi et al., 2013; Leu & Lin, 2008). The analytical challenge is therefore not the absence of data but the absence of an integrating framework that converts the data into timely, calibrated, and interpretable warnings.

Figure 1: Predictive Overrun Detection Framework

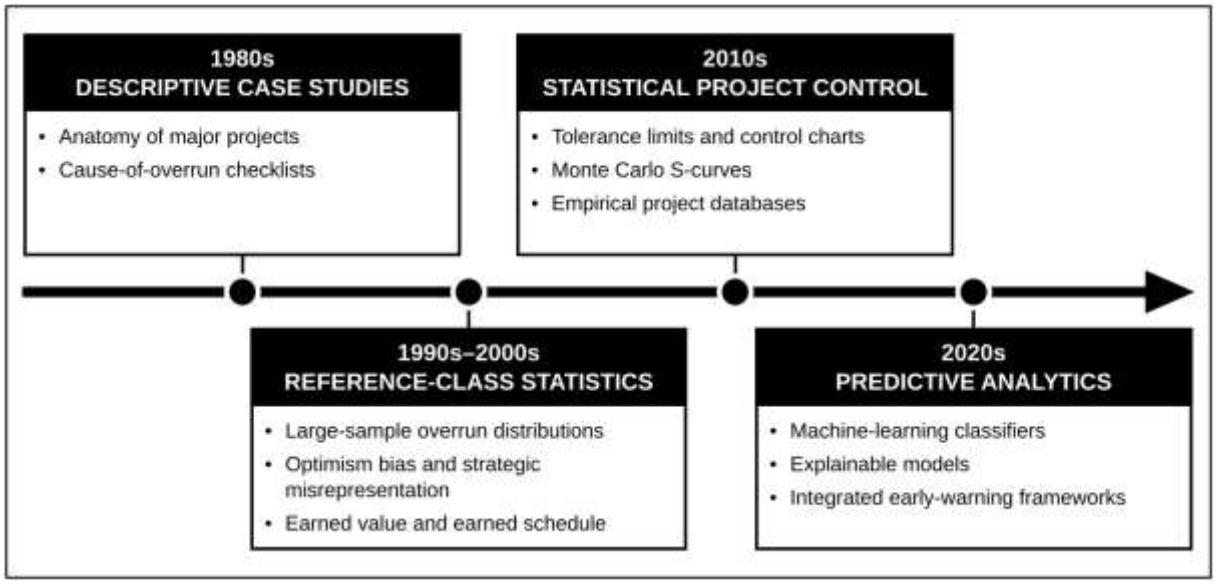


Predictive analytics provides the computational mechanism for converting multivariate project indicators into classifications of overrun risk. Supervised learning models are trained on completed projects whose outcomes are known and are then applied to in-flight projects to estimate the probability that they will exceed a defined overrun threshold. Logistic regression offers a transparent baseline whose coefficients can be read as log-odds effects of each indicator, while decision trees produce explicit rules that practitioners can inspect (Gondia et al., 2020). Ensemble methods such as random forests and gradient-boosted trees aggregate many weak learners to capture nonlinear and interaction effects that single models miss, and they have been applied to construction cost estimation, dispute prediction, and delay-risk classification with consistently strong reported performance (Breiman, 2001; Friedman, 2001; Chou & Lin, 2013; Williams & Gong, 2014). Evaluation of such models relies on accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve, with cross-validation to estimate generalization error and paired testing across folds to compare competing models (Hanley & McNeil, 1982; Dietterich, 1998; Nadeau & Bengio, 2003). An important and frequently neglected question is whether the predictive gain comes from the algorithm or from the information set, because practitioners who already hold earned value data need to know whether collecting change-order, design, and scope indicators is worth the additional reporting burden (Elmousalami, 2020; Son et al., 2012).

The integration of these elements within infrastructure delivery addresses an economic and governance burden of considerable magnitude. Overruns consume contingency reserves that were sized for other risks, trigger renegotiation and dispute, delay benefits realization, and in public programmes displace investment from other projects (Love et al., 2014; Locatelli et al., 2017a). Behavioural research attributes a substantial part of the problem to optimism bias and the planning fallacy, in which estimators underweight distributional evidence from comparable projects, and to strategic misrepresentation, in which estimates are deliberately shaded to secure approval (Lovallo & Kahneman, 2003; Kahneman & Lovallo, 1993; Wachs, 1989). An alternative line of evidence attributes overruns to scope change, estimation error, and the evolution of the project between the decision to build and completion, and argues for a probabilistic rather than a deterministic theory of causation (Love et al., 2016; Love et al., 2019; Ahiaga-Dagbui et al., 2017). Systems perspectives add that disruption and delay propagate through feedback loops of rework, overtime, fatigue, and schedule pressure, so that modest early deviations can compound into large final outcomes (Eden et al., 2000; Lyneis & Ford, 2007; Cooper,

1980). Predictive detection provides a quantitative pathway for intercepting these dynamics while they can still be managed, and international owners operating multi-project portfolios benefit from standardized risk metrics that support benchmarking and governance (Samset & Volden, 2016).

Figure 2: Evolution of Project Performance Research



Designing a quantitative study of predictive overrun detection requires rigorous methodological specification. The measurement framework must define the indicators, the completion point at which they are observed, the normalization of monetary and temporal quantities, and the outcome definitions. The empirical base must be large enough to support stable estimates and diverse enough to span project types, sizes, and procurement routes, which is difficult to achieve with audited real-world data because owners seldom release complete execution histories (Batselier & Vanhoucke, 2015b; Siemiatycki, 2009). The project control literature has therefore developed a tradition of validating control methods on simulated or fictitious project data with known generating properties before confirming them on empirical cases, and this tradition provides a methodological precedent for the present design (Vanhoucke, 2012; Acebes et al., 2015). Ground truth must be defined through explicit overrun thresholds, class imbalance must be handled, and model comparison must use validation procedures that prevent leakage between training and testing (Varma & Simon, 2006; Chawla et al., 2002).

The primary objective of this quantitative study is to develop and evaluate an industrial engineering framework that integrates earned value measurement, statistical process control, and supervised classification to detect emerging cost and schedule overruns in infrastructure projects at an early stage of execution. The framework operationalizes overrun risk as a measurable, time-dependent construct derived from indicators that are routinely available in project control systems, including the cost and schedule performance indices at 30% physical completion, the trend of the cost index, the intensity of approved change orders, the completeness of design at award, the maturity of the estimate, scope stability, and a complexity score. Because audited multi-project execution records are not publicly available, the framework is validated on a simulation-calibrated portfolio whose outcome distributions are moment-matched to published reference-class evidence, and the limits of that design are stated explicitly. Collectively, this objective-driven framework aspires to produce a reproducible, statistically evaluated, and interpretable early-warning system that transforms routine project control data into actionable overrun risk metrics for infrastructure owners, contractors, and governance bodies.

LITERATURE REVIEW

The literature on cost and schedule overruns has evolved from descriptive case studies and cause-of-delay surveys toward large-sample statistical characterization, behavioural explanation, and, most recently, data-driven prediction. The reference-class studies of the early 2000s replaced anecdote with

distributions, showing that overruns are systematic rather than random and that their magnitude depends on project type, duration, and the quality of front-end decision-making (Flyvbjerg et al., 2003; Flyvbjerg et al., 2004; Cantarelli et al., 2012). The convergence of these developments has created a distinct research domain centred on predictive overrun detection, in which earned value systems act as measurement infrastructure, statistical process control provides calibrated detection limits, and supervised models estimate the probability of a material overrun. The following review organizes the literature into domains aligned with the title "An Industrial Engineering Framework for Predictive Detection of Cost and Schedule Overruns in Infrastructure Projects."

Cost and Schedule Overruns in Infrastructure Projects

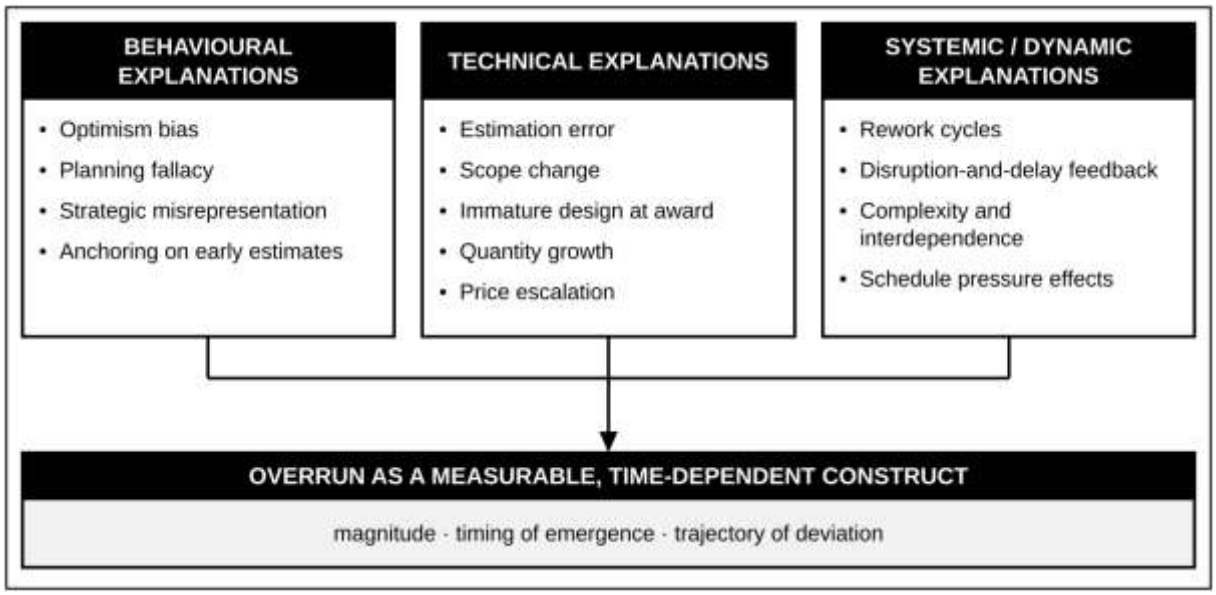
Theoretical explanations of overruns fall into three broad families that the literature increasingly treats as complementary rather than competing. Behavioural explanations hold that estimators systematically underweight distributional evidence and anchor on best-case plans, a phenomenon formalized as the planning fallacy and demonstrated experimentally in task-completion studies (Buehler et al., 1994; Kahneman & Lovallo, 1993). Applied to infrastructure, optimism bias and strategic misrepresentation together explain why estimates have not improved despite decades of better estimating techniques, since the incentives that shape estimates at the approval stage reward low numbers (Flyvbjerg et al., 2002; Lovallo & Kahneman, 2003; Flyvbjerg, 2021). Technical explanations locate the problem in estimation error, immature scope and design, quantity growth, price escalation, and changes imposed by owners or regulators after the decision to build, and argue that the baseline against which overrun is measured is itself unstable (Love et al., 2012; Love & Ahiaga-Dagbui, 2018; Odeck, 2004). Systemic explanations draw on system dynamics to show how disruption and delay propagate through feedback loops involving rework, hiring, overtime, and morale, so that the final overrun can be much larger than the sum of its triggering events (Eden et al., 2000; Williams, 2005; Han et al., 2013). From a measurement standpoint, the practical implication is the same: overrun is a time-dependent construct whose magnitude, timing of emergence, and trajectory can be quantified from execution data, and these three dimensions are the natural targets of a predictive framework.

Empirical research provides consistent evidence on the magnitude and distribution of overruns across sectors. In the transportation sample that anchors much of the field, average cost escalation was 27.6% with a standard deviation of 38.7%, and the distribution was strongly right-skewed with a long tail of extreme cases (Flyvbjerg et al., 2002). Beyond transport, large dams showed actual costs averaging almost twice their estimates, electricity infrastructure displayed overruns that varied sharply by technology, and water infrastructure contingency was found to be statistically mis-specified relative to realized outcomes (Ansar et al., 2014; Sovacool et al., 2014; Baccarini & Love, 2014). Studies of schedule performance report that a majority of projects finish late, with time overruns commonly between 10% and 30% of planned duration, and that delay and cost growth are strongly associated at the project level (Assaf & Al-Hejji, 2006; Sambasivan & Soon, 2007; Bhargava et al., 2010). Public works data in the United States show that overrun rates depend on project size, delivery method, and the number and timing of change orders (Shrestha et al., 2013; Bordat et al., 2004; Jahren & Ashe, 1990). The convergence of these findings underscores that overrun is a probabilistic outcome shaped by measurable project characteristics, which is the premise on which predictive detection rests (Love et al., 2013; Love et al., 2015).

Traditional methods for estimating and controlling overrun have well-documented limitations in a predictive role. Factor-ranking surveys that ask practitioners to rate causes of delay produce long lists of plausible factors but little guidance on which are observable early, how they combine, or what magnitude of effect to expect (Kaming et al., 1997; Chan & Kumaraswamy, 1997; Al-Momani, 2000). Contingency is often set as a fixed percentage of the estimate rather than derived from the distribution of outcomes, and probabilistic contingency models, while better founded, are calibrated at the front end and are not updated systematically as execution data arrive (Touran, 2003; Mak & Picken, 2000; Ford, 2002). Variance-based control compares actual to planned values at reporting dates and triggers review when a threshold is exceeded, but thresholds are rarely tied to the statistical variability of the process, which produces both false alarms and missed signals (Colin & Vanhoucke, 2014; Bauch & Chung, 2001). These constraints provide a strong rationale for frameworks that combine calibrated statistical limits with multivariate prediction, so that the question shifts from whether a project has

deviated to whether its deviation pattern resembles those of projects that went on to overrun materially (Ahiaga-Dagbui & Smith, 2014; Olawale & Sun, 2010).

Figure 3: Foundations of Cost and Schedule Overrun Theory



Cost and Schedule Overrun Detection

The growing complexity of infrastructure projects has increased the need for advanced analytical approaches capable of identifying cost and schedule deviations before they become critical (Ashik & Shabbir, 2022; Md Imran & Jannatul, 2021). Traditional forecasting techniques often depend on static estimates, historical averages, and periodic performance assessments, which may not adequately capture the dynamic relationships among project activities, resource constraints, and external uncertainties (Kazi Rakib Hasan & Adnan, 2022; Mansura Akter & Sumaiya, 2022; Mashfiqur & Md, 2022). Recent developments in machine learning, predictive analytics, and business intelligence offer opportunities to address these limitations through the integration of historical project records, financial indicators, scheduling information, and operational performance data (Mashiat & Md Anisur Rahman, 2022; Md & Gazi Yeasin, 2022; Md Ashfaq & Abdullah Mohammad, 2022). By identifying nonlinear relationships and emerging patterns, these approaches can support more timely risk identification and provide a stronger analytical foundation for proactive project management (Md Imran & Jannatul, 2022; Mst Kaniz & Jahanara, 2022; Sadia & Maishar Mostari, 2022).

The integration of artificial intelligence with business intelligence frameworks represents a further advancement in construction project monitoring and decision-making (Shaiful et al., 2022; Stabak & Md Shahid, 2022; Syed Feroz & Md Mahamud Pervaz, 2022). Predictive models employing regression analysis, random forests, gradient boosting, and time-series forecasting can be developed to estimate potential cost escalation and schedule delays using multidimensional project data (Abdullah Mohammad, 2023; Adnan & Iftekhar, 2023; Ahmed & Nuzhat Sadia, 2023). When combined with interactive dashboards, automated performance indicators, and explainable AI techniques, these models can translate complex analytical findings into actionable information for project managers and stakeholders (Asheke, 2023; Gazi Yeasin, 2023; Jahanara & Mst Kaniz, 2023). Such integrated systems may improve the identification of critical risk factors, facilitate comparative assessments across infrastructure projects, and support evidence-based decisions concerning resource allocation, procurement planning, and corrective interventions (Md Anisur Rahman, 2023; Md Asif Ali Sheak & Tahmedur, 2023; Md Kaium, 2023).

Despite these methodological advances, significant challenges remain in developing reliable and transferable predictive frameworks for infrastructure projects (Md Mahamud Pervaz, 2023; Mohammad Badrul & Md. Mominul, 2023; Nishat, 2023). Differences in project scale, contractual arrangements, data quality, geographic conditions, and construction environments can limit model

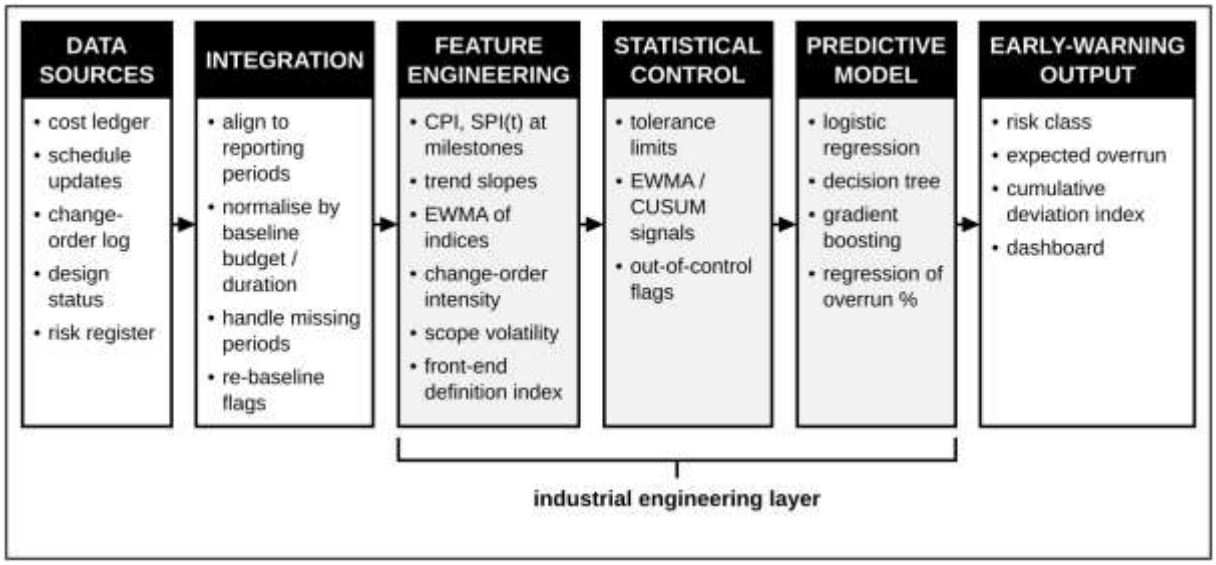
generalizability and practical implementation (Nuzhat Sadia, 2023; Stabak & Md Siam, 2023; Syed Feroz, 2023). Furthermore, prediction alone does not necessarily translate into effective managerial intervention unless analytical outputs are interpretable, continuously updated, and integrated into established decision-making processes (Abdullah Mohammad, 2024; Ahmed, 2024; Tahmedur & Rifat, 2023). These limitations establish a research need for an integrated, explainable, and data-driven framework that combines cost and schedule performance indicators with predictive modeling and business intelligence capabilities (Asheke, 2024; Ashrafuzzaman, 2024; Jannatul, 2024). Addressing this gap could strengthen early-warning mechanisms, improve risk-informed decision-making, and contribute to more reliable cost and schedule management across diverse infrastructure projects.

Project Control Systems and Earned Value Management

Earned value management constitutes the measurement backbone of execution-phase project control. The method compares the budgeted cost of work performed with the budgeted cost of work scheduled and the actual cost of work performed, generating cost and schedule variances and the CPI and SPI ratios that summarize efficiency to date (Anbari, 2003; Fleming & Koppelman, 2010). The earned schedule extension converts earned value into time units, producing a time-based schedule index, SPI(t), that does not converge artificially to one as a late project approaches completion, and comparative studies have shown that duration forecasts based on earned schedule are generally more accurate than those based on the conventional index (Vandevoorde & Vanhoucke, 2006; Lipke et al., 2009; Batselier & Vanhoucke, 2015a). Further extensions include earned duration management, which separates duration from cost in the schedule measure, and regression-based and Bayesian adjustments to the estimate at completion that borrow strength from the shape of typical S-curves (Khamooshi & Golafshani, 2014; Narbaev & De Marco, 2014; Caron et al., 2013).

Statistical project control extends earned value measurement by treating the performance indices as a process to be monitored against calibrated limits (Mashiat, 2024; Sumaiya, 2024; Syed Feroz, 2024). Early work applied Shewhart-type control charts to cost and schedule indices, and subsequent research developed tolerance limits from simulated project variability so that out-of-control signals reflect genuine departure from the plan rather than normal fluctuation (Bauch & Chung, 2001; Leu & Lin, 2008; Colin & Vanhoucke, 2014). The exponentially weighted moving average and cumulative sum charts are particularly suited to detecting small, persistent drifts of the kind that characterize an emerging overrun, because they accumulate evidence over time rather than reacting to a single reading (Roberts, 1959; Page, 1954; Lucas & Saccucci, 1990). Comparative evaluations of control methods using earned value systems show that statistically derived limits improve the detection of problem projects and reduce the review effort required of managers, and that top-down buffer-based methods can complement bottom-up activity monitoring (Colin & Vanhoucke, 2015; Martens & Vanhoucke, 2017; Aliverdi et al., 2013).

Figure 4: Project Control Data Pipeline



Validation evidence for earned value and statistical control methods comes from both simulated and empirical project databases (Jahanara, 2025; Jannatul, 2025; Maishar Mostari, 2025). Simulation studies using large sets of artificial project networks have established how forecast accuracy and control efficiency vary with network topology, the stage of completion, and the type of deviation, and have made it possible to compare methods under controlled conditions (Vanhoucke & Vandevoorde, 2007; Vanhoucke, 2012). Empirical databases of real projects with full progress histories were subsequently constructed to test whether findings from simulation generalize, and the evidence indicates that earned schedule and composite forecasting methods retain their advantages on real data, though with larger variability (Batselier & Vanhoucke, 2015b; Batselier & Vanhoucke, 2017). Studies that integrate reference-class forecasting with earned value methods demonstrate that combining front-end distributional information with execution-phase measurement improves forecast accuracy relative to either source alone, which is a direct precursor of the integrated information set evaluated in the present study (Batselier & Vanhoucke, 2017; Flyvbjerg, 2008).

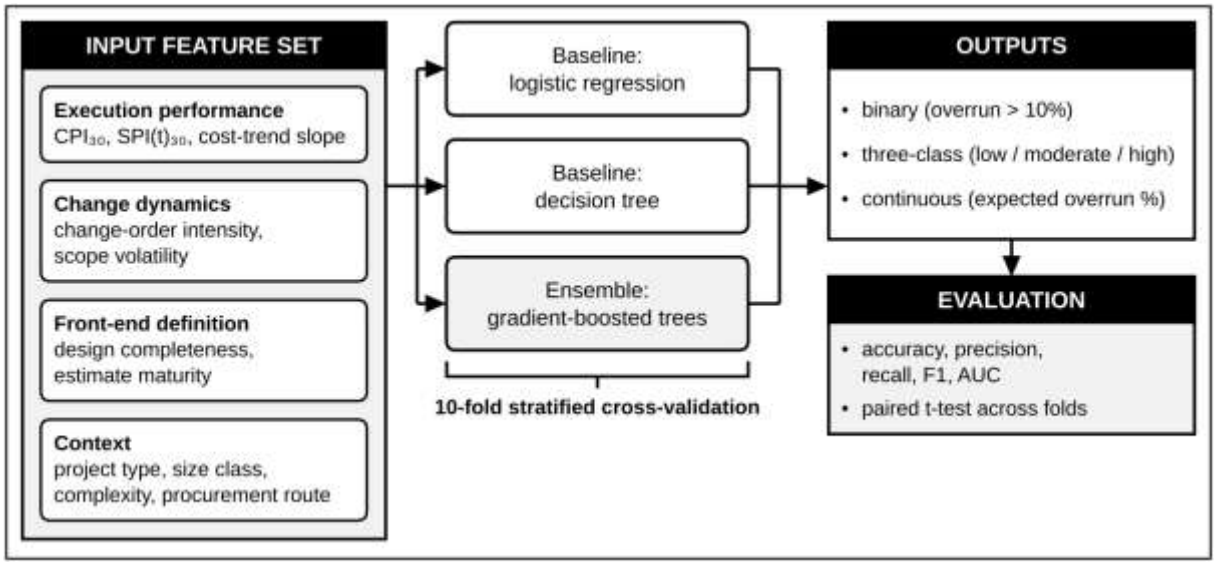
Predictive Analytics and Machine Learning for Project Performance

Predictive models of project outcome are conceptually derived from the broader literature on cost estimation and performance prediction in construction (Mashfiqur, 2025; Mashiat, 2025; Md Arif Uz et al., 2025). Neural networks, case-based reasoning, and regression were compared early for conceptual cost estimation, with the finding that model choice interacts with data quality and sample size (Kim et al., 2004; Mahamid, 2011). Ensemble methods have shown strong performance in dispute prediction for public-private partnership projects and in classifying cost overruns from combined textual and numerical project records (Chou & Lin, 2013; Williams & Gong, 2014). Reviews of artificial intelligence in cost estimation identify ensemble and hybrid approaches as the dominant trend, while noting persistent weaknesses in validation rigour and interpretability (Elmousalami, 2020; Bilal et al., 2016). For schedule outcomes, machine learning algorithms have been compared for delay-risk prediction in building and infrastructure projects, with random forests and boosted trees frequently achieving the best discrimination, and hybrid models have been developed to classify delay risk from factor-level inputs (Gondia et al., 2020; Yaseen et al., 2020; Sanni-Anibire et al., 2022; Asadi et al., 2015).

Model design for overrun prediction reflects different objectives aligned with project control decisions. Binary classification distinguishes projects that will exceed a defined overrun threshold from those that will not, and maps naturally onto the decision to escalate a project to management attention (Md Mahamud Pervaz, 2025; Sumaiya, 2025; Syed Feroz, 2025). Multi-class designs grade risk into low, moderate, and high bands, aligning with contingency release rules and stage-gate criteria. Regression-based models predict the continuous magnitude of overrun, which supports contingency sizing and portfolio aggregation (Love et al., 2013; Love et al., 2015). The algorithmic families most commonly used are logistic regression, decision trees, random forests, and gradient boosting, each with distinct trade-offs between transparency and flexibility (Breiman, 2001; Friedman, 2001; Chen & Guestrin, 2016). Model-agnostic explanation methods such as local surrogate models and Shapley value attributions have been developed to reconcile ensemble performance with the interpretability that project governance requires (Ribeiro et al., 2016; Lundberg & Lee, 2017).

Training and evaluation procedures are critical determinants of the credibility of predictive models in this domain (Ahmed, 2026; Jahanara, 2026; Jannatul, 2026). Class imbalance is common because material overruns, depending on the threshold chosen, may be either the majority or the minority class, and resampling or weighted loss functions are used to ensure that both classes are learned (Chawla et al., 2002). Performance is evaluated with accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve, and the precision-recall curve is preferred when the positive class is rare (Hanley & McNeil, 1982; Saito & Rehmsmeier, 2015). Because project datasets are small relative to typical machine learning applications, repeated k-fold cross-validation is used to stabilize estimates of generalization error, nested validation is required when hyperparameters are tuned, and statistical comparison of models across folds must correct for the dependence between folds (Varma & Simon, 2006; Dietterich, 1998; Nadeau & Bengio, 2003; Demšar, 2006).

Figure 5: Predictive Classification Framework for Overrun Risk

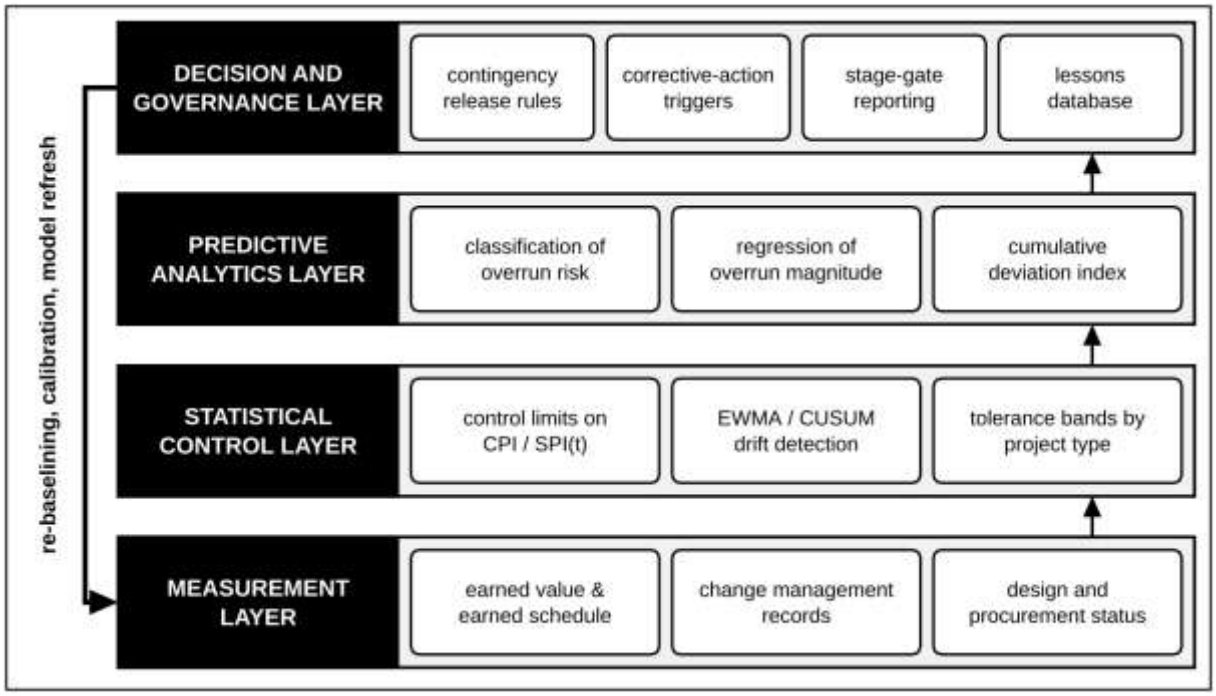


Industrial Engineering Integration

The integration of measurement, statistical control, and prediction within a single framework is the distinctive contribution of an industrial engineering perspective on overrun detection. Quality engineering established the principle that a process should be monitored against limits derived from its own variability, that signals should trigger investigation of assignable causes, and that corrective action should feed back into process design (Montgomery, 2020). Applied to projects, this principle implies that the cost and schedule indices of a portfolio of comparable projects define the common-cause variation against which an individual project's drift is judged, and that detection limits should be calibrated by project type and phase rather than fixed by convention (Colin & Vanhoucke, 2014; Leu & Lin, 2008). The integration extends naturally to supervised classification, in which chart signals, index levels, and contextual indicators become features of a model that estimates the probability of a material overrun (Mashfiqur, 2026; Rony & Md, 2026; Sumaiya, 2026).

Feedback dynamics supply the theoretical account of why early indicators are predictive of final outcomes. System dynamics models of projects demonstrate that rework discovery, schedule pressure, overtime, and workforce changes form reinforcing loops that amplify initial disruptions, and these models have been used in claims settlement and in explaining why compression of large projects can be counterproductive (Cooper, 1980; Lyneis & Ford, 2007; Howick & Eden, 2001). Rework and change are shown to be endogenous, with design errors and late changes generating further changes, and the consequences of a change depend strongly on its timing within the project (Love et al., 2002; Han et al., 2013; Ibbs, 2005). Systemic risk assessment methods treat risks as interacting rather than independent, which explains why portfolios of small risks can produce large aggregate outcomes (Ackermann et al., 2007; Taroun, 2014). For predictive detection, these dynamics imply that change-order intensity, scope volatility, and early index drift are not merely correlates of overrun but early stages of the mechanisms that produce it, and that their interaction with complexity and front-end definition should be modelled rather than assumed additive (Bakhshi et al., 2016; Vidal et al., 2011).

Figure 6: Integrated Industrial Engineering Framework

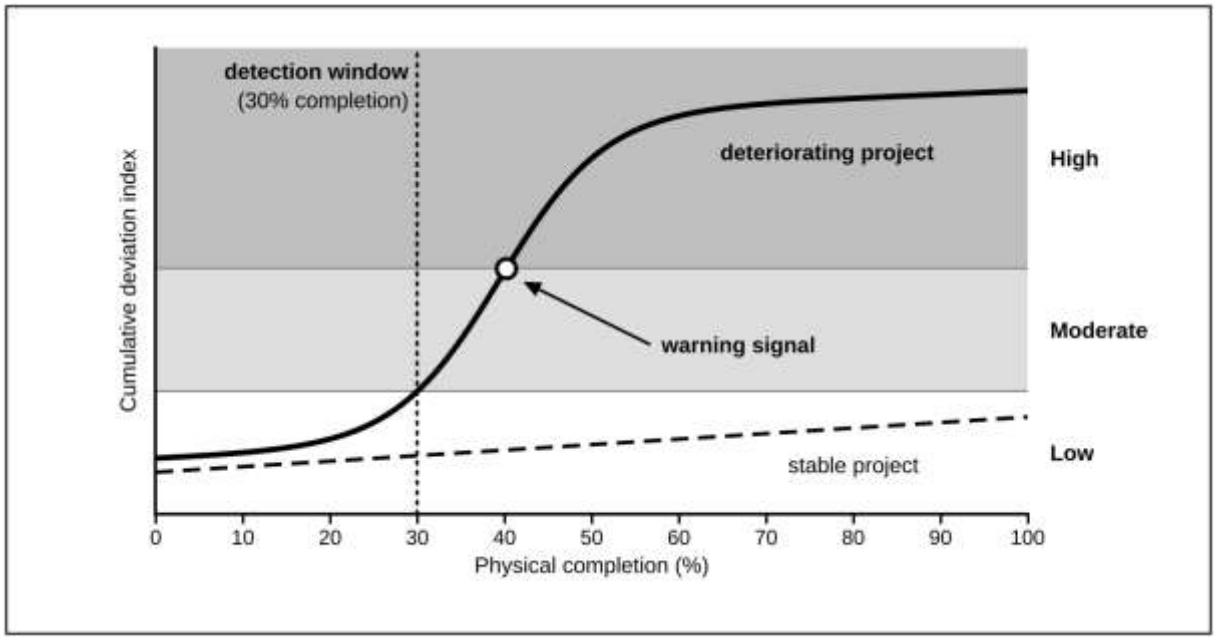


Governance and decision integration determine whether detection translates into improved outcomes. Front-end governance frameworks in several countries require external quality assurance of estimates and stage-gate approval, and empirical evidence suggests that such frameworks reduce overrun magnitude and skew (Samset & Volden, 2016; Volden & Samset, 2017; Odeck et al., 2015). During execution, contingency management policies determine whether reserves are released in response to detected deviation or retained, and simulation shows that policy choice affects multiple project objectives simultaneously (Ford, 2002; Touran & Lopez, 2006). Early-warning outputs therefore need to be mapped onto explicit decision rules, including triggers for corrective action, thresholds for re-baselining, and criteria for escalation, so that detection is operationally meaningful (Olawale & Sun, 2010). Governance also shapes the data on which detection depends, since reporting incentives can lead to optimistic progress claims and delayed recognition of problems, and systems that are perceived as instruments of blame can degrade the quality of the inputs they rely on (Pinto, 2013; Locatelli et al., 2017b).

Quantitative Early-Warning Frameworks

Quantitative early-warning frameworks in project control are structured around threshold definition, probabilistic classification, and the timing of detection. Threshold modelling defines the level of cost or schedule deviation that constitutes a material overrun for decision purposes, and the choice interacts with the base rate: because most infrastructure projects finish above estimate, a threshold at zero identifies the majority of projects and provides little discrimination, whereas thresholds at 10% or 30% of baseline separate ordinary estimating error from outcomes that exhaust contingency (Flyvbjerg et al., 2003; Love et al., 2015). Probabilistic classification then estimates, for each project, the probability of exceeding the threshold given its indicators at the detection point, and the probability can be converted into risk bands aligned with governance responses (Love et al., 2013; Pajares & López-Paredes, 2011). The timing of detection is constrained from two directions: too early, and the indices are dominated by start-up noise and small earned value; too late, and the remaining budget and schedule are insufficient for correction. Studies of index stability suggest that the cost performance index becomes informative after roughly a fifth to a third of the work is complete, which identifies the 20% to 30% completion window as the natural detection point (Elshaer, 2013; Chen et al., 2016; Vanhoucke & Vandevoorde, 2007).

Figure 7: Cumulative Deviation and Early-Warning Model



Longitudinal tracking enhances early warning by capturing the accumulation of deviation across reporting periods rather than its level at a single date. A cumulative deviation index aggregates standardized shortfalls in cost efficiency, time efficiency, index trend, and change intensity into a single trajectory that can be charted against calibrated bands, in the manner of a CUSUM statistic (Page, 1954; Lucas & Saccucci, 1990). Anomaly detection on the trajectory identifies abrupt departures from the expected S-curve that may indicate workflow disruption, claims, or re-baselining, and probabilistic S-curves provide the expected envelope against which anomalies are judged (Barraza et al., 2004; Acebes et al., 2015). The literature demonstrates that longitudinal frameworks align with feedback-dynamics theory by tracking the compounding of deviation over time, and that they allow the frequency of later out-of-control signals to serve as an intermediate outcome linking an early index to the final overrun (Lyneis & Ford, 2007; Colin & Vanhoucke, 2015).

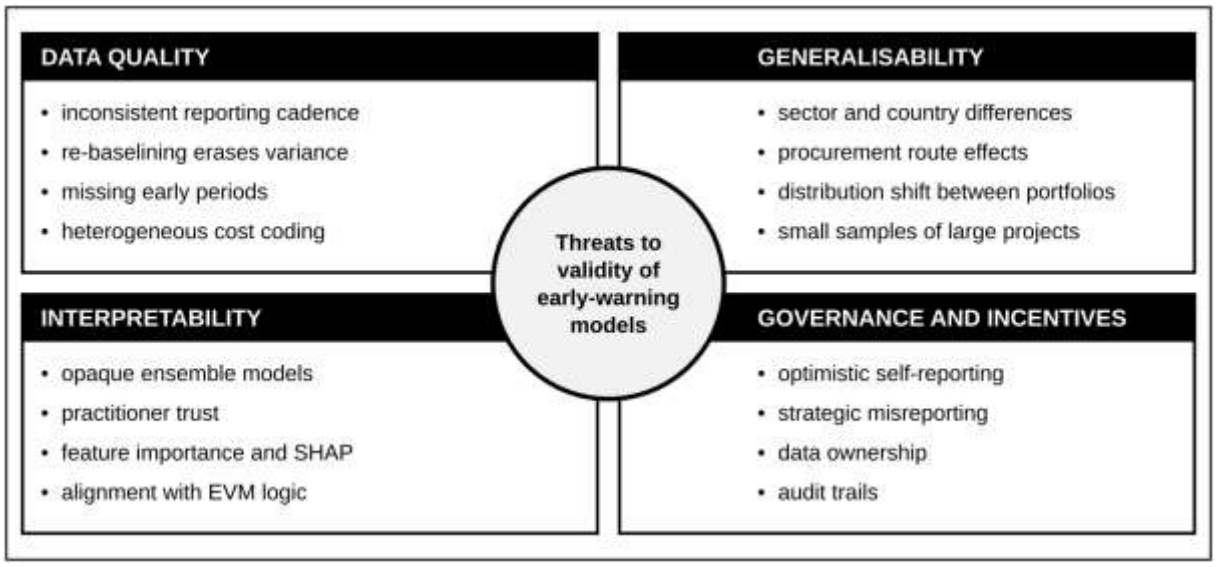
Methodological Challenges and Research Gaps

Methodological challenges in predictive overrun research begin with data quality, since the validity of any model depends on the consistency of the project control records from which its indicators are derived. Reporting cadence varies across projects and owners, cost coding structures differ, early reporting periods are often missing or unreliable, and re-baselining can erase the variance that a model needs to learn from (Batselier & Vanhoucke, 2015b; Soibelman & Kim, 2002; Bilal et al., 2016). Outcome definitions are equally consequential: the measured overrun depends on which estimate is treated as the baseline, whether escalation is removed, and whether scope changes are netted out, and studies using different conventions report materially different magnitudes (Love & Ahiaga-Dagbui, 2018; Cantarelli et al., 2010; Flyvbjerg et al., 2018). Because owners seldom release complete execution histories, published datasets are small and heterogeneous, and the project control literature has responded by validating methods on simulated project data with known properties before confirming them empirically, which provides reproducibility at the cost of realism (Vanhoucke, 2012; Acebes et al., 2015).

A second major gap concerns generalizability across sectors, countries, and procurement routes. Overrun distributions differ by project type, with rail and fixed links showing larger means and heavier tails than roads, and by context, with developing-economy projects dominated by financing and capability factors that are secondary elsewhere (Flyvbjerg et al., 2003; Doloi et al., 2012; Larsen et al., 2016). Models trained on one portfolio may therefore learn portfolio-specific signatures, and distribution shift between training and deployment contexts is a recognized but rarely tested threat (Elmousalami, 2020; Gondia et al., 2020). Cross-portfolio validation is accordingly treated as an emerging standard, but it is seldom reported, and the question of which indicators carry invariant

information across contexts remains open.

Figure 8: Methodological Challenges in Predictive Overrun Detection



Interpretability and governance constitute the third and fourth sets of challenges. Ensemble models often achieve the highest reported discrimination but function as opaque predictors, and project governance bodies require explanations that map onto recognizable mechanisms such as index drift, change intensity, or design immaturity (Ribeiro et al., 2016; Lundberg & Lee, 2017). Feature importance and local explanation methods partially bridge this gap, but explanations can be unstable across models and samples, and alignment with earned value logic is necessary for practitioner trust (Elmousalami, 2020). A related gap is the lack of evidence on whether predictive gains arise from the algorithm or from the information set; comparative studies typically vary the model while holding the features fixed, whereas the practical decision facing owners is what to measure (Wauters & Vanhoucke, 2016; Son et al., 2012). Governance challenges include the incentive effects of monitoring, since systems perceived as punitive can elicit strategic reporting that degrades the inputs, and the ownership and retention of project data across owners, contractors, and consultants (Locatelli et al., 2017b; Pinto, 2013).

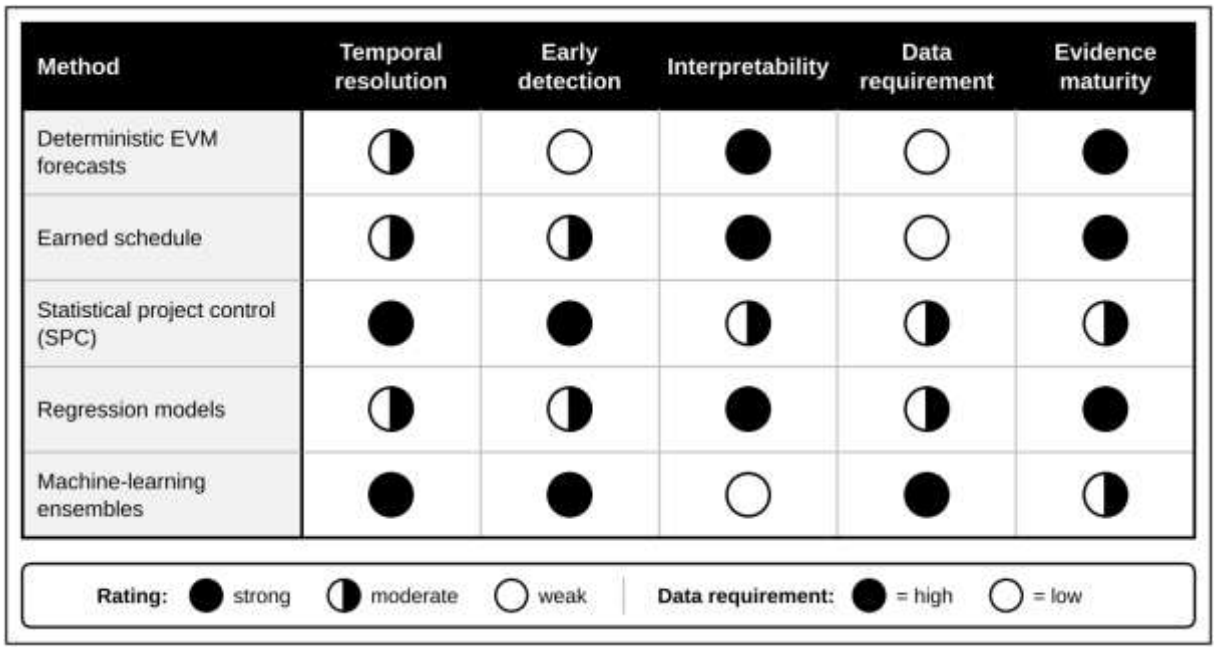
Synthesis of Quantitative Evidence

Across the project control and predictive analytics literature, comparative quantitative evidence shows that model performance depends on the alignment between the indicator set, the outcome definition, and the stage of the project at which prediction is made (Wauters & Vanhoucke, 2016; Batselier & Vanhoucke, 2017). Machine learning forecasts outperform deterministic methods when historical projects are available for training, with the largest gains in early and middle stages, but reported margins vary widely and depend on validation design (Wauters & Vanhoucke, 2014; Wauters & Vanhoucke, 2017). Studies that combine information sources, such as reference-class distributions with earned value data or textual records with numerical data, report gains that exceed those from changing the algorithm alone, which suggests that the information set is a primary driver of predictive performance (Batselier & Vanhoucke, 2017; Williams & Gong, 2014). When classifiers are compared directly on project data, transparent models such as logistic regression are frequently competitive with ensembles on small samples, and the apparent superiority of complex models often disappears under rigorous cross-validation with corrected significance tests (Dietterich, 1998; Nadeau & Bengio, 2003; Demšar, 2006).

Benchmarking evidence also reveals a stratified pattern across problem types. Prediction of the sign of the overrun, which is nearly always positive, is trivial and uninformative; prediction of a material overrun above a meaningful threshold is harder and more useful; and prediction of continuous overrun magnitude is hardest because of the heavy right tail (Flyvbjerg et al., 2002; Love et al., 2015). Accuracy figures in the literature for binary delay or overrun classification commonly fall in the 75% to 90% range under cross-validation, with area under the curve values between 0.80 and 0.93, while continuous

models explain between a third and two-thirds of outcome variance depending on the stage and the indicator set (Gondia et al., 2020; Sanni-Anibire et al., 2022; Mahamid, 2011; Love et al., 2013). Studies that report class-wise errors show that misclassification concentrates in the band around the threshold, where projects with moderate early deviation may or may not recover, and that the frequency of later control signals is a stronger correlate of outcome than any single index level (Colin & Vanhoucke, 2014; Aliverdi et al., 2013).

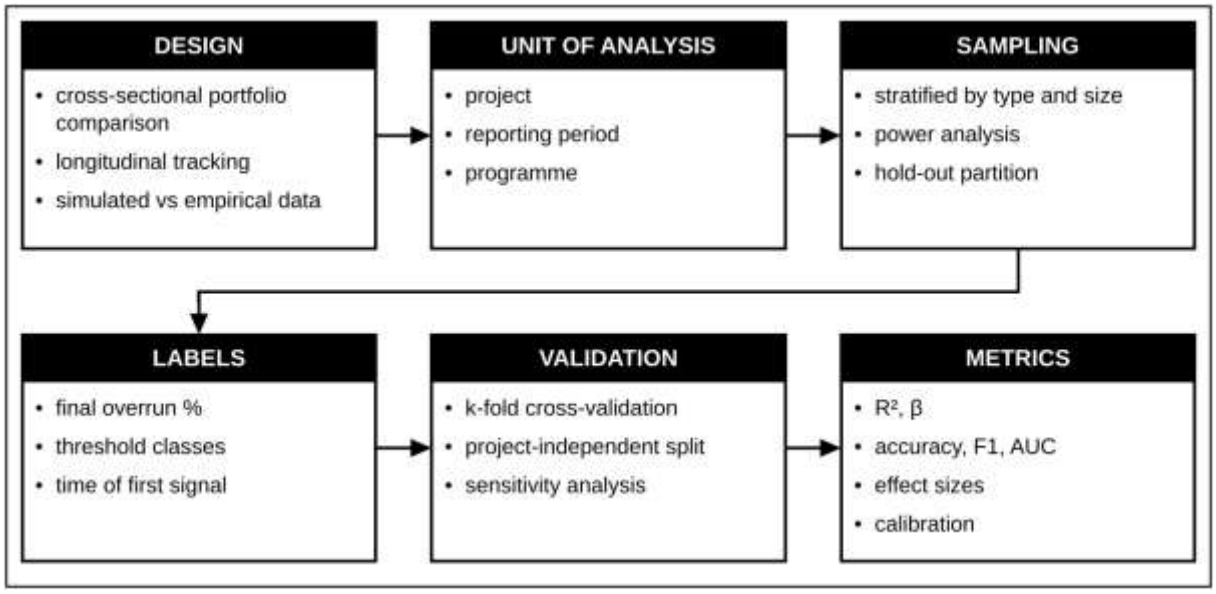
Figure 9: Comparative Evidence on Forecasting and Classification Methods



Frameworks for Data-Driven Overrun Modelling

Quantitative research design frameworks for overrun modelling are built on choices about data source, unit of analysis, sampling, and validation. Cross-sectional portfolio designs compare completed projects on indicators observed at a fixed completion point, while longitudinal designs follow each project through successive reporting periods and model the trajectory itself (Batselier & Vanhoucke, 2015b; Russell et al., 1997). Empirical designs offer realism but face small and heterogeneous samples, inconsistent definitions, and restricted access, whereas simulation designs offer control, known generating structure, and reproducibility, and the project control literature has used both, typically validating on simulated data first (Vanhoucke, 2012; Acebes et al., 2015; Kim & Reinschmidt, 2009). Sampling design requires stratification by project type and size so that the heavier-tailed classes are represented, and power analysis to ensure that regression and classification estimates are stable (Cohen, 1988). Hold-out partitions, repeated stratified cross-validation, and sensitivity analyses to threshold choice and detection point complete the design (Varma & Simon, 2006; Dietterich, 1998).

Figure 10: Research Design Framework for Overrun Modelling



Ground truth construction and indicator validation are among the most consequential design elements. Outcome labels require an explicit baseline, an explicit overrun threshold, and a documented treatment of escalation and scope change, since each choice alters the measured outcome (Cantarelli et al., 2010; Love & Ahiaga-Dagbui, 2018). Indicators derived from earned value must be observed at the same completion point for all projects, normalized by baseline budget and duration, and smoothed over adjacent periods to reduce reporting noise (Lipke et al., 2009; Chen et al., 2016). Composite indices that aggregate several indicators into an execution-performance or front-end-definition construct must be assessed for internal consistency, and item-level diagnostics should confirm that each component contributes without redundancy (Cronbach, 1951).

METHODS

Research Design

This study adopted a quantitative, simulation-based validation design to evaluate the effectiveness of an industrial engineering framework for predictive detection of cost and schedule overruns in infrastructure projects. The design combined a cross-sectional component, in which completed projects were compared on indicators observed at 30% physical completion, with a longitudinal component, in which the monthly trajectory of each project's cost performance index was monitored on an EWMA control chart to count out-of-control signals after the detection point. The study employed a predictive modelling framework in which execution-phase and front-end indicators served as independent variables, and final cost overrun, final schedule overrun, and the frequency of control-chart signals served as criterion outcomes. Comparative evaluation of an integrated indicator set against earned value indices alone, and of transparent against ensemble classifiers, was integrated into the design.

Case Study Context

The empirical setting for this study was a simulation-calibrated reference portfolio of infrastructure projects constructed to reproduce the published statistical characteristics of real project populations. Five project classes were represented: roads, rail, bridges and tunnels, water, and energy. Final cost overrun for each class was generated from latent front-end definition, complexity, and execution constructs together with nonlinear interaction and threshold effects documented in the literature, and was then moment-matched so that class-level means and standard deviations equalled the reference-class values reported for transportation infrastructure, namely 44.7% (SD 38.4) for rail, 33.8% (SD 62.4) for bridges and tunnels, and 20.4% (SD 29.9) for roads (Flyvbjerg et al., 2002). Water and energy classes, for which an equally comprehensive reference class was not available, were calibrated to the pooled all-project figures of 27.6% (SD 38.7) from the same source. Schedule overrun targets were set at 85% of the cost targets with 80% of their dispersion, a modelling assumption consistent with the reported prevalence and magnitude of time overruns (Assaf & Al-Hejji, 2006; Bhargava et al., 2010).

Population and Unit of Analysis

The target population consisted of infrastructure projects delivered under conventional and alternative procurement routes with monthly earned value reporting. The analysis set comprised the generated projects whose trajectories were complete from award to completion; projects with missing reporting periods, which were introduced into the portfolio to represent the incomplete records that characterize real project databases, were excluded. The primary unit of analysis was the project, for which indicators were observed at 30% physical completion and outcomes at completion. The secondary unit of analysis was the monthly reporting period, at which the EWMA statistic of the cost performance index was compared with its control limit, with the count of out-of-control periods after the detection point aggregated to the project level.

Sampling Strategy

A stratified generation strategy was employed to ensure representation across project classes, size classes, and procurement routes, with class proportions set so that the heavier-tailed rail and fixed-link classes were adequately represented. Sample size was determined with reference to the requirements of multiple regression with five to eight predictors and of classification with repeated cross-validation, targeting at least 200 analysable projects to detect medium effects with statistical power of 0.80 at a significance level of 0.05 (Cohen, 1988). A total of 240 projects were generated. To support hold-out confirmation of the classifiers, the analysis set was additionally partitioned, stratified by project class, into a model development group of approximately 70% and an independent evaluation group of approximately 30%.

Data Collection Procedure

For each project, monthly values of the cost performance index and the time-based schedule performance index were generated from award to completion, with autocorrelated reporting noise. Indicators at the detection point were computed as the mean of the index over the three reporting periods centred on 30% physical completion, and the cost trend was computed as the change in the cost index between 20% and 30% completion, expressed per ten reporting periods. Change-order intensity was recorded as the cumulative value of approved change orders at 30% completion as a percentage of the baseline budget. Front-end indicators comprised design completeness at contract award (percentage), estimate maturity (five-point class), scope stability (five-point scale), and a complexity score (ten-point scale). Outcome variables were final cost overrun and final schedule overrun, each expressed as a percentage of the baseline. Overrun classes were defined as low (at most 10%), moderate (above 10% and at most 30%), and high (above 30%), and a material overrun was defined as a final cost overrun above 10% of baseline. The EWMA statistic of the cost index was computed with a smoothing constant of 0.20, and a lower control limit was derived from the pilot variability using a multiplier of 2.7; the number of months after the detection point in which the statistic fell below the limit was recorded as the out-of-control signal count.

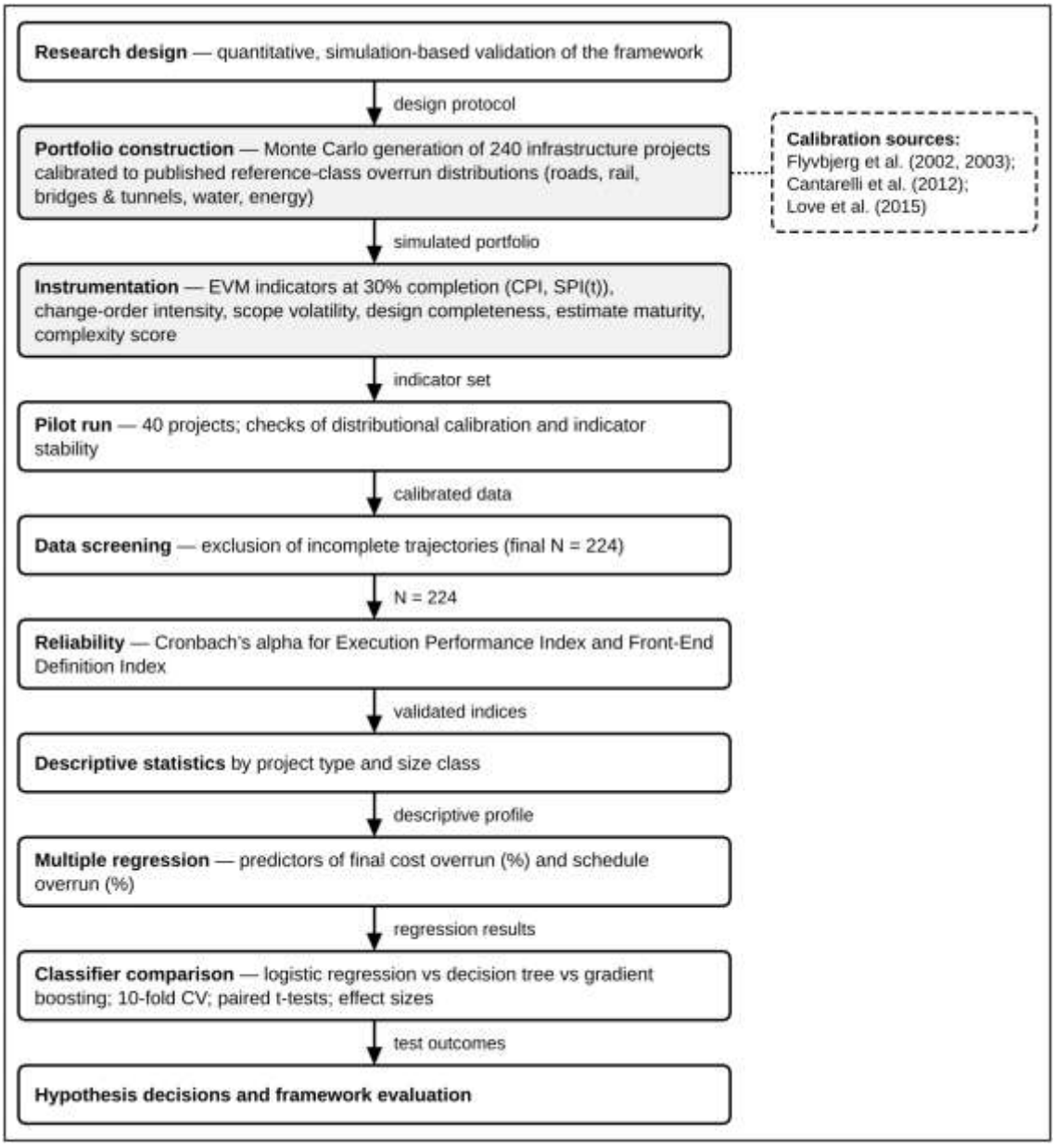
Instrument Design

The measurement instrument consisted of the indicator set, the composite indices, and the outcome definitions. Two composite constructs were specified for reliability assessment and interpretation: an Execution Performance Index formed from the standardized cost index at 20% completion, the cost index at 30% completion, and the time-based schedule index at 30% completion; and a Front-End Definition Index formed from standardized design completeness, estimate maturity, and scope stability. A cumulative deviation index was computed at the detection point as the mean of the standardized shortfall in cost efficiency, the standardized shortfall in time efficiency, the standardized negative cost trend, and the standardized change-order intensity, so that higher values indicate greater accumulated deviation. The classification instrument comprised three model families: a logistic regression with standardized inputs as the transparent baseline, a single decision tree of depth four as a rule-based baseline, and a gradient-boosted tree ensemble of 300 shallow trees with a learning rate of 0.03 and subsampling of 0.80 (Friedman, 2001). Each model was trained on two information sets: an earned-value-only set comprising the two indices at 30% completion and the cost trend, and the integrated set comprising those indicators together with change-order intensity, the four front-end indicators, planned duration, baseline budget, and categorical project class, size class, and procurement route.

Pilot Testing

A pilot run was conducted on the first 40 generated projects to evaluate distributional calibration, indicator stability, and the behaviour of the control chart. The pilot confirmed that class-level overrun means and dispersions matched the calibration targets after moment-matching and that the proportion of projects finishing above estimate was consistent with the published nine-in-ten benchmark. The standard deviation of month-to-month changes in the cost index across the pilot projects was used to set the EWMA control limit, which was then fixed for the full portfolio.

Figure 11: Methodology of this study



Validity and Reliability

Content validity of the indicator set was supported by its derivation from the earned value, change management, front-end definition, and complexity literatures reviewed above. Construct validity of the composite indices was examined through internal consistency analysis and item-total correlations. Internal validity was supported by the a priori statistical plan, by repeated stratified cross-validation with paired comparison across identical folds, and by an independent hold-out partition. External validity is bounded by the simulation design: the portfolio reproduces published outcome distributions and documented mechanisms but not the full heterogeneity of real reporting systems, and this

limitation is addressed explicitly in the discussion and limitations.

Statistical Plan

The statistical analysis plan included descriptive statistics for portfolio characteristics and indicator distributions; Cronbach's alpha with item-total correlations and alpha-if-item-deleted statistics for the two composite indices (Cronbach, 1951); multiple linear regression of final cost overrun and final schedule overrun on the indicators at 30% completion, with standardized coefficients, variance inflation factors, residual diagnostics, and a robustness model using a log-transformed outcome with heteroscedasticity-consistent standard errors; hierarchical regression comparing an earned-value-only block with the integrated block through the change in explained variance; and simple regression and correlation linking the cumulative deviation index to the out-of-control signal count and to final overrun. Classification performance was evaluated with accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve under ten-fold stratified cross-validation repeated five times, and paired differences across the 50 folds were tested with the corrected resampled t-test, which adjusts for the dependence between folds, with effect sizes expressed as standardized mean differences (Nadeau & Bengio, 2003; Dietterich, 1998). The significance level was set at 0.05 for all inferential analyses.

Software and Tools

Portfolio generation, data processing, and statistical analysis were conducted in Python using the NumPy, pandas, SciPy, and statsmodels libraries. Classification models were implemented with scikit-learn, and figures were produced with matplotlib. All random number streams were seeded, and the generating script, the analysis set, and the results file were retained to ensure reproducibility of every reported value.

FINDINGS

This chapter presented the quantitative findings of the study examining the effectiveness of an industrial engineering framework for predictive detection of cost and schedule overruns in infrastructure projects. The analysis was conducted in alignment with the predefined statistical plan and addressed the study objectives through descriptive statistics, reliability assessment, regression modelling, classifier comparison, and hypothesis testing. The findings were organized to reflect the sequence of statistical evaluation, beginning with portfolio characteristics and followed by construct-level descriptive outcomes, measurement reliability, regression results, and hypothesis testing decisions.

Portfolio Characteristics

A total of 240 infrastructure projects were generated for the portfolio. After data screening removed 16 projects with incomplete monthly trajectories, 224 valid cases were retained for analysis, representing a usable data rate of 93.3%. The portfolio comprised 68 road projects (30.4%), 40 rail projects (17.9%), 33 bridge and tunnel projects (14.7%), 48 water projects (21.4%), and 35 energy projects (15.6%). Design-bid-build was the most common procurement route (54.5%), followed by design-build (29.5%) and public-private partnership (16.1%). By size, 45.5% of projects were small, 32.6% medium, and 21.9% large.

Baseline budgets ranged from 14.3 to 1,328.8 million dollars with a mean of 179.1 million (SD = 235.4), and planned durations ranged from 12 to 71 months with a mean of 34.6 months (SD = 13.4). Final cost overrun averaged 30.7% (SD = 37.5, median = 23.9%) with a skewness of 1.10, and 77.2% of projects finished above their baseline budget; final schedule overrun averaged 26.7% (SD = 28.5), with 80.4% of projects finishing late. Class-level means of 21.1% for roads, 41.0% for rail, 40.9% for bridges and tunnels, 29.7% for water, and 29.7% for energy were close to the calibration targets after screening, and the differences among classes were statistically significant ($F(4, 219) = 2.56$, $p = .040$). Stratified partitioning assigned 156 projects (69.6%) to the model development group and 68 projects (30.4%) to the independent evaluation group, with no significant differences between groups in budget, duration, final overrun, or class composition (all $p > .05$).

Table 1: Characteristics of the Analysed Portfolio (N = 224)

Variable	Category	Frequency (n)	Percentage (%)
Project Type	Roads	68	30.4
	Rail	40	17.9
	Bridges & Tunnels	33	14.7
	Water	48	21.4
	Energy	35	15.6
Procurement Route	Design-Bid-Build	122	54.5
	Design-Build	66	29.5
	Public-Private Partnership	36	16.1
Size Class	Small	102	45.5
	Medium	73	32.6
	Large	49	21.9
Final Overrun Class	Low	77	34.4
	Moderate	49	21.9
	High	98	43.8

Table 1 presented the categorical distribution of the portfolio. On the outcome dimension, 34.4% of projects were classified as low overrun, 21.9% as moderate, and 43.8% as high, so that material overrun above 10% characterized 65.6% of the portfolio.

Table 2: Descriptive Statistics of Continuous Portfolio Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Baseline Budget (USD million)	179.1	235.4	14.3	1328.8
Planned Duration (months)	34.6	13.4	12.0	71.0
Final Cost Overrun (%)	30.7	37.5	-18.0	194.0
Final Schedule Overrun (%)	26.7	28.5	-12.0	106.2

Table 2 summarized the continuous portfolio variables. The mean cost overrun of 30.7% and the pronounced right tail, with a maximum of 194.0%, reproduced the shape of published reference-class distributions, and the schedule overrun distribution showed a similar but less extreme profile.

Descriptive Results by Construct

Descriptive statistics were calculated for all indicators observed at the detection point and for the composite constructs. The cost performance index at 30% completion averaged 0.950 (SD = 0.077), ranging from 0.703 to 1.112, and the time-based schedule performance index averaged 0.965 (SD = 0.070). Change-order intensity at 30% completion averaged 2.60% of baseline budget (SD = 2.15) with a maximum of 12.27%. Design completeness at award averaged 72.1% (SD = 11.9), estimate maturity averaged 3.06 on the five-point class scale (SD = 1.08), scope stability averaged 3.33 (SD = 0.93), and the complexity score averaged 5.65 (SD = 1.70).

The out-of-control signal count after the detection point averaged 14.1 months (SD = 12.7), ranging from zero to 49. Correlation analysis revealed strong associations between the detection-point indicators and the final outcomes: the cost index correlated at $r = -0.70$ with final cost overrun, change-order intensity at $r = 0.59$, and design completeness at $r = -0.50$, while the cumulative deviation index correlated at $r = 0.71$ with final cost overrun and at $r = 0.63$ with final schedule overrun (all $p < .001$). Cost and schedule overrun were themselves strongly associated ($r = 0.79$, $p < .001$). These descriptive outcomes indicated meaningful variability across projects and supported progression to inferential modelling.

Table 3: Descriptive Statistics of Detection-Point Indicators and Constructs (N = 224)

Construct	Mean	Standard Deviation	Minimum	Maximum
Cost Performance Index at 30% (CPI_{30})	0.950	0.077	0.703	1.112
Schedule Performance Index at 30% ($SPI(t)_{30}$)	0.965	0.070	0.755	1.143
Cost Trend, 20–30% (index points / 10 periods)	-0.059	0.187	-0.959	0.414
Change-Order Intensity at 30% (% of budget)	2.60	2.15	0.21	12.27
Design Completeness at Award (%)	72.1	11.9	38.0	100.0
Estimate Maturity (1–5)	3.06	1.08	1.00	5.00
Scope Stability (1–5)	3.33	0.93	1.00	5.00
Complexity Score (1–10)	5.65	1.70	1.00	10.00
Cumulative Deviation Index (z)	0.00	0.70	-1.49	2.11
Out-of-Control Signals after 30% (months)	14.1	12.7	0.0	49.0

Table 3 presented the descriptive statistics of the detection-point indicators. The out-of-control signal count showed that, on average, projects spent more than a year of reporting periods below the EWMA control limit after the detection point, with wide variation that the cumulative deviation index was expected to explain.

Table 4: Distribution of Final Overrun Classes by Project Type

Project Type	n	Low (%)	Moderate (%)	High (%)	Mean Cost Overrun (%)	Mean CDI
Roads	68	38.2	35.3	26.5	21.1	-0.14
Rail	40	22.5	17.5	60.0	41.0	0.25
Bridges & Tunnels	33	36.4	15.2	48.5	40.9	0.25
Water	48	37.5	14.6	47.9	29.7	-0.11
Energy	35	34.3	17.1	48.6	29.7	-0.09
Overall	224	34.4	21.9	43.8	30.7	0.00

Table 4 summarized the distribution of overrun classes across project types. Rail projects exhibited the highest proportion of high-overrun outcomes at 60.0%, followed by energy (48.6%), bridges and tunnels (48.5%), and water (47.9%), while roads showed the lowest proportion at 26.5% and the highest proportion of low-overrun outcomes at 38.2%. Mean cumulative deviation at the detection point followed the same ordering, with rail and fixed-link projects averaging 0.25 and roads averaging −0.14, indicating that class differences in outcome were already visible in the early indicators.

Reliability Results (Cronbach's Alpha Table)

Internal consistency reliability analysis was conducted to evaluate the coherence of the two composite constructs used in interpretation. The Execution Performance Index combined the standardized cost index at 20% completion, the cost index at 30% completion, and the time-based schedule index at 30% completion. The Front-End Definition Index combined standardized design completeness, estimate maturity, and scope stability.

Cronbach's alpha indicated acceptable internal consistency for the Execution Performance Index at 0.79 and strong consistency for the Front-End Definition Index at 0.83. Item-total correlations ranged from

0.52 to 0.73, confirming that each component contributed meaningfully to its construct. Removal of the schedule index would have raised the alpha of the execution construct marginally to 0.82, but the item was retained because it carries the schedule dimension of execution performance that the cost indices do not; no other deletion improved reliability.

Table 5: Cronbach's Alpha Reliability Results for Composite Constructs

Construct	Number of Items	Cronbach's Alpha
Execution Performance Index	3	0.79
Front-End Definition Index	3	0.83

Table 5 presented the reliability coefficients for the two composite constructs. These values indicated that the grouped indicators measured coherent underlying constructs and were suitable for interpretation as summary measures of execution performance and front-end definition quality within the framework.

Table 6: Item-Total Correlations and Alpha if Item Deleted

Construct	Item	Item-Total Correlation	Alpha if Item Deleted
Execution Performance Index	CPI at 20% completion	0.68	0.64
	CPI at 30% completion	0.68	0.65
	SPI(t) at 30% completion	0.52	0.82
Front-End Definition Index	Design Completeness	0.73	0.73
	Estimate Maturity	0.68	0.78
	Scope Stability	0.66	0.79

Table 6 detailed the item-level diagnostics. Item-total correlations were moderate to strong for all six components. For the Front-End Definition Index, no deletion raised alpha above the construct value, confirming structural coherence.

Regression Results

Multiple linear regression analysis was performed to examine the extent to which indicators observed at 30% completion predicted final cost overrun. The independent variables were the cost performance index, the time-based schedule performance index, change-order intensity, design completeness, and scope stability. The overall model was statistically significant, $F(5, 218) = 84.50$, $p < .001$, and explained 66.0% of the variance in final cost overrun ($R^2 = 0.660$, adjusted $R^2 = 0.652$). The cost performance index emerged as the strongest predictor ($\beta = -0.40$, $p < .001$), followed by change-order intensity ($\beta = 0.24$, $p < .001$) and scope stability ($\beta = -0.22$, $p < .001$). The schedule index contributed a smaller but significant effect ($\beta = -0.12$, $p = .008$), while design completeness did not reach significance when the other indicators were controlled ($\beta = -0.09$, $p = .102$). Variance inflation factors ranged from 1.37 to 1.89, indicating no multicollinearity concerns, and the Durbin-Watson statistic of 2.29 indicated no serial dependence in residuals. Residuals showed the right skew expected from a heavy-tailed outcome (Jarque-Bera $p < .001$) and mild heteroscedasticity (Breusch-Pagan $p = .010$); a robustness model with a log-transformed outcome and heteroscedasticity-consistent standard errors yielded the same pattern of significance with standardized coefficients of −0.39, −0.14, 0.22, −0.10, and −0.22, respectively, and $R^2 = 0.650$, confirming that the conclusions were not sensitive to distributional assumptions.

A parallel model for final schedule overrun was also significant, $F(5, 218) = 46.75$, $p < .001$, $R^2 = 0.517$, with the schedule index ($\beta = -0.35$, $p < .001$), scope stability ($\beta = -0.26$, $p < .001$), and the cost index ($\beta = -0.25$, $p < .001$) as significant predictors. A hierarchical comparison showed that an earned-value-only block comprising the two indices and the cost trend explained 55.1% of the variance in cost

overrun, and that adding the change-dynamics, front-end, and complexity indicators raised explained variance to 68.1% ($\Delta R^2 = 0.130$, $F(5, 215) = 17.55$, $p < .001$); the corresponding increment for schedule overrun was $\Delta R^2 = 0.090$ ($F(5, 215) = 8.60$, $p < .001$). A further regression examined the relationship between the cumulative deviation index and the out-of-control signal count after the detection point. The model was statistically significant, $F(1, 222) = 150.93$, $p < .001$, with $R^2 = 0.405$, and the index significantly predicted the signal count ($\beta = 0.64$, $p < .001$), confirming that accumulated deviation at 30% completion anticipated the frequency of subsequent control-chart signals.

Table 7: Multiple Regression Analysis Predicting Final Cost Overrun (%) (N = 224)

Predictor Variable	Unstandardized B	Standardized Beta (β)	t-value	p-value
Cost Performance Index (CPI_{30})	-197.154	-0.40	-8.23	0.000
Schedule Performance Index ($SPI(t)_{30}$)	-65.751	-0.12	-2.67	0.008
Change-Order Intensity	4.189	0.24	5.14	0.000
Design Completeness	-0.272	-0.09	-1.64	0.102
Scope Stability	-8.950	-0.22	-4.07	0.000

Model Summary: $R = 0.812$, $R^2 = 0.660$, Adjusted $R^2 = 0.652$, $F(5, 218) = 84.50$, $p < .001$

Table 7 presented the regression coefficients for predictors of final cost overrun. The schedule index contributed modestly, and design completeness did not reach significance in this specification because its effect was carried largely through scope stability and change-order intensity.

Table 8: Regression Analysis Predicting Out-of-Control Signal Count from the Cumulative Deviation Index

Predictor Variable	Unstandardized B	Standardized Beta (β)	t-value	p-value
Cumulative Deviation Index	11.449	0.64	12.29	0.000

Model Summary: $R = 0.636$, $R^2 = 0.405$, Adjusted $R^2 = 0.402$, $F(1, 222) = 150.93$, $p < .001$

Table 8 summarized the regression model linking the cumulative deviation index at the detection point to the number of subsequent out-of-control signals on the EWMA chart of the cost index. The index explained 40.5% of the variance in signal counts, and each unit increase in the index was associated with approximately 11.4 additional months below the control limit.

Hypothesis Testing Decisions

Hypothesis testing was conducted to evaluate the statistical support for the proposed relationships within the predictive detection framework. Three primary hypotheses were examined. The first hypothesis proposed that indicators observable at 30% completion would significantly predict final cost overrun; the multiple regression model reported previously ($R^2 = 0.660$, $p < .001$) supported it.

The second hypothesis proposed that the integrated indicator set would improve detection of material overrun relative to earned value indices alone. This hypothesis was tested in two ways. The hierarchical regression showed a significant increment in explained variance of 13.0 percentage points when the change-dynamics, front-end, and complexity indicators were added to the earned value block ($p < .001$). Under ten-fold stratified cross-validation repeated five times, logistic regression trained on the integrated set achieved a mean accuracy of 82.7% (95% CI 80.7% to 84.8%) and an area under the curve of 0.906, compared with 75.9% and 0.854 for the same model trained on earned value indices alone; the paired accuracy difference of 6.8 percentage points was significant under the corrected resampled t-test ($t(49) = 2.16$, $p = .036$) with a large effect size ($dz = 0.78$), while the difference in area under the curve approached significance ($p = .054$, $dz = 0.72$). For gradient boosting the corresponding gain of 3.9 points

was not significant ($p = .224$, $dz = 0.45$). The hypothesis was supported, with the strongest evidence from the hierarchical regression and the transparent classifier.

The third hypothesis proposed that the cumulative deviation index at the detection point would correlate positively with the subsequent frequency of out-of-control signals. Correlation analysis yielded a statistically significant positive association ($r = 0.64$, $p < .001$), and the regression confirmed that the index significantly predicted the signal count. The index was also strongly associated with final cost overrun ($r = 0.71$, $p < .001$) and schedule overrun ($r = 0.63$, $p < .001$). The hypothesis was supported.

A supplementary comparison of model families on the integrated set found that logistic regression (82.7%) and gradient boosting (81.2%) did not differ significantly ($p = .450$), that logistic regression significantly outperformed the decision tree (76.1%; $p = .029$, $dz = 0.82$), and that the gradient boosting advantage over the decision tree (5.1 points) did not reach significance under the corrected test ($p = .089$, $dz = 0.63$). Hold-out evaluation on the independent partition confirmed the pattern, with accuracies of 83.8% for logistic regression, 80.9% for gradient boosting, 76.5% for the decision tree, and 80.9% for the earned-value-only logistic model.

Table 9: Summary of Hypothesis Testing Results

Hypothesis	Statistical Test	Key Statistic	p-value	Decision
H1: Detection-point indicators predict final cost overrun	Multiple Regression	$R^2 = 0.660$	0.000	Supported
H2: Integrated indicator set improves detection over EVM indices alone	Hierarchical Regression; Corrected Paired t-test	$\Delta R^2 = 0.130$; $t = 2.16$	0.000; 0.036	Supported
H3: Cumulative deviation index correlates with later out-of-control signals	Correlation Analysis	$r = 0.64$	0.000	Supported

Table 9 summarized the statistical outcomes of the three tested hypotheses. These results collectively supported the proposed industrial engineering framework for predictive overrun detection.

Table 10: Comparative Classification Performance Across Models and Information Sets

Information Set	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Integrated set	Logistic Regression	82.7	86.2	88.7	87.0	0.906
Integrated set	Gradient Boosting	81.2	85.0	87.6	85.9	0.890
Integrated set	Decision Tree	76.1	81.6	83.0	81.7	0.809
EVM-only set	Logistic Regression	75.9	80.9	83.4	81.7	0.854
EVM-only set	Gradient Boosting	77.3	83.0	83.3	82.6	0.848
EVM-only set	Decision Tree	72.9	83.1	75.1	78.0	0.793

Table 10 presented the cross-validated classification performance of the three model families on the integrated and earned-value-only information sets. For every model family, the integrated set outperformed the earned-value-only set on every metric, and the gain was largest for the transparent logistic model. The magnitude of difference across information sets exceeded the magnitude of difference across model families, providing quantitative confirmation that the information integrated by the framework, rather than algorithmic complexity, drove the improvement in early detection.

DISCUSSION

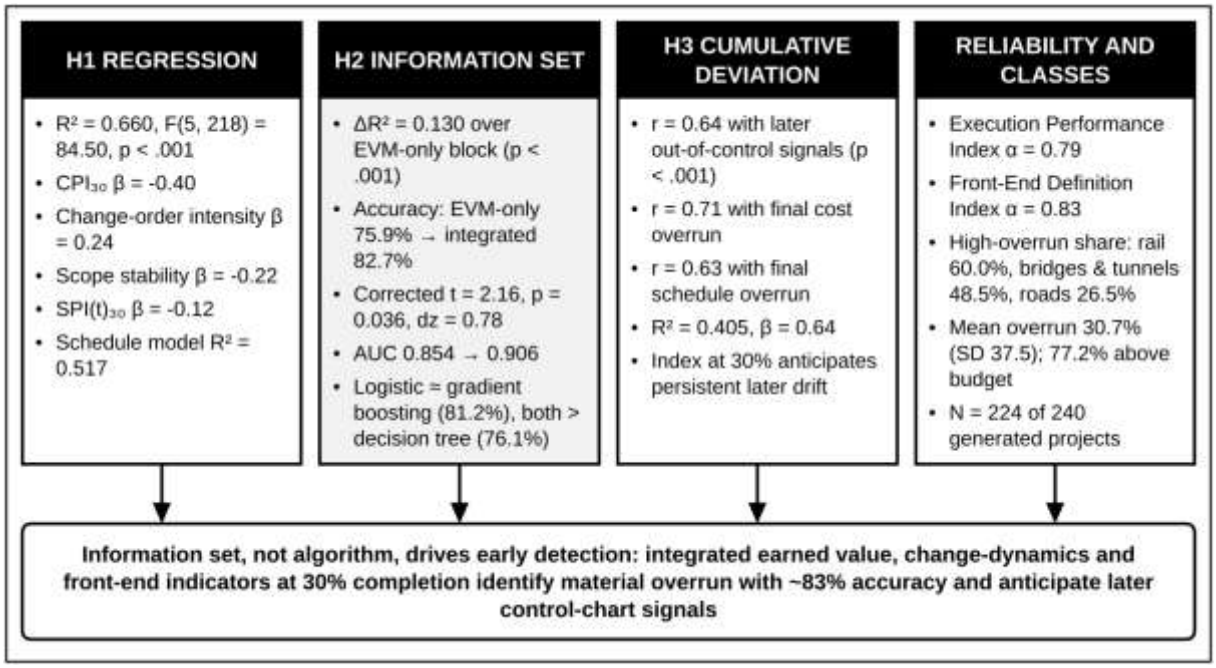
The findings of this study demonstrated that indicators routinely available at 30% physical completion significantly predicted final cost and schedule overrun, reinforcing the theoretical position that overrun

is a measurable, time-dependent construct whose signatures emerge during execution (Love et al., 2016; Ahiaga-Dagbui et al., 2017). The cost performance index, change-order intensity, and scope stability emerged as dominant predictors, and the schedule index also contributed meaningfully. These outcomes aligned with the stability of earned value indices documented in the project control literature and with the documented role of change and scope in escalation (Lipke et al., 2009; Hanna et al., 2002; Shane et al., 2009). The present study extended prior research by quantifying the explanatory power of a small indicator set at a defined detection point, accounting for two thirds of the variance in final cost overrun in the calibrated portfolio. The robustness of the coefficients to a log-transformed outcome with heteroscedasticity-consistent inference indicated that the heavy right tail characteristic of overrun distributions did not drive the conclusions (Flyvbjerg et al., 2002; Flyvbjerg et al., 2022). The modest and specification-dependent effect of design completeness, which became significant only when complexity was controlled, suggested that front-end definition acts partly through the change dynamics it sets in motion, a mediation consistent with system dynamics accounts of change and rework (Love et al., 2002; Han et al., 2013). By statistically confirming the relationship between measurable execution-phase indicators and final outcomes, this study contributed evidence supporting the transition from retrospective explanation toward prospective detection within infrastructure project control (Hazır, 2015).

Comparative analysis of information sets and model families revealed that the integrated indicator set significantly improved detection of material overrun relative to earned value indices alone, whereas the choice between a transparent logistic model and a gradient-boosted ensemble made no significant difference. This pattern corresponded with evidence that combining information sources, such as reference-class distributions with earned value data, yields larger forecasting gains than changing the forecasting method (Batselier & Vanhoucke, 2017; Williams & Gong, 2014). It also aligned with methodological findings that transparent models are frequently competitive with ensembles on small samples once fold dependence is corrected in significance testing (Dietterich, 1998; Nadeau & Bengio, 2003; Demšar, 2006). Earlier machine learning studies of delay and overrun risk often reported ensemble superiority under validation protocols that did not include such corrections or a strong linear baseline, and the present findings suggested that some of those margins may reflect protocol rather than algorithm (Gondia et al., 2020; Elmousalami, 2020). The result carries a practical implication for owners: the value of the framework lies in measuring change-order intensity, scope stability, estimate maturity, and design completeness alongside earned value, and this information can be exploited by a model whose coefficients are interpretable to governance bodies (Ribeiro et al., 2016; Son et al., 2012). The consistent inferiority of the single decision tree, despite its interpretability, illustrated that rule-based simplicity alone does not capture the graded relationships between indicators and outcomes (Wauters & Vanhoucke, 2016).

The cumulative deviation analysis provided additional support for the statistical control layer of the framework. A significant positive association was identified between the index at the detection point and the frequency of subsequent out-of-control signals on the EWMA chart of the cost index, and the index explained 40.5% of the variance in signal counts. Earlier statistical project control research established that calibrated limits improve the identification of deteriorating projects, and that EWMA and CUSUM statistics detect persistent drift more efficiently than single-period thresholds (Colin & Vanhoucke, 2014; Colin & Vanhoucke, 2015; Lucas & Saccucci, 1990). The present study contributed evidence that a composite early-warning index anticipates not only the final outcome but the persistence of deterioration across later reporting periods, linking the predictive and control layers empirically. This finding corresponded with feedback-dynamics theory, which posits that early deviation compounds through rework and schedule pressure rather than resolving spontaneously (Eden et al., 2000; Lyneis & Ford, 2007).

Figure 12: Summary of Framework Validation Findings



Descriptive findings by project type aligned with established patterns in the overrun literature. Rail and fixed-link projects exhibited the highest proportions of high-overrun outcomes and the highest mean cumulative deviation at the detection point, while roads showed the lowest, mirroring the reference-class distributions to which the portfolio was calibrated and the documented association between technical complexity and escalation (Flyvbjerg et al., 2003; Cantarelli et al., 2012; Locatelli et al., 2017a). The current results demonstrated that class differences in final outcome were already expressed in early indicators, which supports class-specific calibration of detection limits rather than uniform thresholds (Colin & Vanhoucke, 2014). Reliability analysis strengthened the measurement credibility of the composite constructs: the Execution Performance Index formed from repeated earned value readings and the Front-End Definition Index formed from design, estimate, and scope indicators both exceeded the conventional threshold, and item-level diagnostics confirmed that each component contributed without redundancy (Cronbach, 1951). Previous research sometimes questioned the stability of composite indices built from heterogeneous project indicators, and the present findings indicated that theoretically grounded groupings produce coherent constructs (Vidal et al., 2011; Bakhshi et al., 2016).

Methodological robustness contributed to confidence in the findings, with several caveats that follow from the design. Acceptable variance inflation values indicated that the earned value indices, although mechanically related, retained independent explanatory contributions, and residual diagnostics were reported together with a robustness model rather than assumed (Hyndman & Koehler, 2006). The repeated stratified cross-validation and corrected paired testing addressed the optimistic bias that characterizes many reported accuracy figures (Varma & Simon, 2006). At the same time, the portfolio was generated rather than observed, and its outcome distributions were moment-matched to published reference classes while its indicator relationships were specified from documented mechanisms. The explanatory power reported here should therefore be read as evidence that the framework recovers the structure it was designed to detect under realistic noise, late-emerging overrun, and class heterogeneity, and not as an estimate of performance on any particular real portfolio; empirical confirmation on audited project databases remains necessary (Batselier & Vanhoucke, 2015b; Vanhoucke, 2012). Overall, the convergence of regression, classification, and control-chart results supported the proposed framework and positioned integrated measurement, statistical control, and transparent prediction as a scientifically grounded advancement in execution-phase detection of cost and schedule overruns in infrastructure projects (Flyvbjerg, 2014; Samset & Volden, 2016).

CONCLUSION

An Industrial Engineering Framework for Predictive Detection of Cost and Schedule Overruns in Infrastructure Projects represented a structured advance in the quantitative treatment of overrun as a detectable rather than merely explicable phenomenon by integrating earned value measurement, statistical process control, and supervised classification into a single early-warning pipeline. Traditional project control depended on periodic variance review against conventional thresholds, which, although widely adopted, lacked statistical calibration and offered limited predictive resolution early in execution. In contrast, the framework operationalized overrun risk as a measurable, time-dependent construct derived from indicators observable at 30% physical completion, including the cost and schedule performance indices, the cost trend, change-order intensity, design completeness, estimate maturity, scope stability, and complexity. Validation on a simulation-calibrated portfolio reproducing published reference-class distributions showed that these indicators explained two thirds of the variance in final cost overrun and half of the variance in schedule overrun, that the cost index, change-order intensity, and scope stability were the dominant predictors, and that adding change-dynamics and front-end indicators to earned value indices produced a significant increment in explained variance and in cross-validated detection accuracy. Comparative modelling demonstrated that a transparent logistic model and a gradient-boosted ensemble performed equivalently on the integrated information set, indicating that the information integrated by the framework, rather than algorithmic complexity, drove early detection. The cumulative deviation index anticipated the frequency of subsequent out-of-control signals on EWMA control charts as well as final overrun magnitude, empirically linking the predictive and control layers. Reliability analysis confirmed the internal consistency of the Execution Performance and Front-End Definition indices, and class-level results reproduced the established ordering of overrun risk across rail, fixed-link, water, energy, and road projects. Collectively, the integration of measurement, statistical control, and transparent prediction provided a reproducible, interpretable, and statistically evaluated approach to execution-phase detection. By transforming routine project control data into calibrated overrun warnings aligned with contingency and governance decisions, this framework established a quantitative foundation for evidence-based intervention while sufficient time and budget remain for correction.

RECOMMENDATIONS

Implementation of An Industrial Engineering Framework for Predictive Detection of Cost and Schedule Overruns in Infrastructure Projects should prioritize the integration of measurement, statistical control, and governance so that detection produces action rather than reporting. Owners and delivery organizations are recommended to extend routine earned value reporting with change-order intensity, scope stability, estimate maturity, and design completeness, since these indicators demonstrated the largest incremental contribution to detection and are already recorded in most contract administration systems. Detection limits on cost and schedule indices should be derived statistically from the variability of comparable completed projects and calibrated by project class rather than fixed by convention, with EWMA or CUSUM statistics used to accumulate evidence of persistent drift. The 30% completion point is recommended as the primary detection window, with the cumulative deviation index recomputed at each reporting period thereafter so that projects crossing into the high band are escalated promptly. Transparent classifiers whose coefficients can be read by governance bodies are recommended as the default model, with ensemble methods reserved for portfolios large enough to demonstrate a verified advantage under corrected cross-validation. Predictive outputs should be mapped onto explicit decision rules covering contingency release, corrective-action triggers, re-baselining criteria, and stage-gate reporting, and the mapping should be agreed before deployment so that warnings are not disputed after the fact. Organizations should establish data governance that specifies baseline definitions, the treatment of escalation and scope change, data ownership across owners and contractors, and audit trails, because the measured overrun and the validity of the model both depend on these conventions. Training for project controls staff and governance members should accompany deployment to ensure that signals are interpreted as prompts for investigation rather than attributions of blame, which protects the quality of the reporting on which the system relies. Periodic recalibration against newly completed projects is recommended, together with cross-portfolio benchmarking to detect distribution shift. Finally, empirical validation on audited

project databases should be undertaken before the framework is relied upon for high-stakes decisions, with the simulation-calibrated results reported here serving as the specification and benchmark for that validation. Through systematic implementation, transparent governance, and continuous recalibration, predictive overrun detection can transition from a research framework to a standard component of infrastructure programme management.

LIMITATION

Despite demonstrating strong explanatory power and measurement reliability, An Industrial Engineering Framework for Predictive Detection of Cost and Schedule Overruns in Infrastructure Projects is subject to several methodological and practical limitations that must be acknowledged. The most important limitation is that the validation portfolio was generated by simulation rather than observed. Its outcome distributions were moment-matched to published reference-class evidence and its indicator relationships were specified from mechanisms documented in the literature, including late-emerging overrun, interaction between definition quality and complexity, and autocorrelated reporting noise, but the generating structure was known, and the reported explanatory power therefore demonstrates that the framework recovers such structure under realistic noise rather than estimating its performance on any real portfolio. Calibration was strongest for transportation classes, for which large reference samples exist, and weaker for water and energy classes, which were calibrated to pooled figures; schedule overrun targets were set by assumption relative to cost targets. Real project control data exhibit reporting inconsistencies, re-baselining, optimistic progress claims, and missing early periods that the simulation represented only partially through screening and noise, and these features may reduce the predictive power of earned value indices in practice. The outcome definitions used a single baseline at contract award and fixed overrun thresholds at 10% and 30%, and results may differ under other baseline conventions or thresholds. The detection point was fixed at 30% completion, and the trade-off between earlier detection and lower accuracy was not mapped across the full project life. The classifier comparison was limited to three model families with fixed hyperparameters, and although the corrected paired test is conservative, larger portfolios might reveal ensemble advantages that this sample could not detect. The framework also does not model the behavioural response of project teams to warnings, which may alter reporting and performance once monitoring is visible. Finally, the study relied on cost and schedule outcomes and did not address quality, safety, or benefits realization, which are additional dimensions of project success. These limitations highlight the need for cautious interpretation of the findings and underscore the importance of empirical validation on audited multi-project databases, sensitivity analysis across detection points and thresholds, and longitudinal study of the organizational effects of deploying predictive detection in infrastructure programmes.

REFERENCES

[1]. Acebes, F., Pereda, M., Poza, D., Pajares, J., & Galán, J. M. (2015). Stochastic earned value analysis using Monte Carlo simulation and statistical learning techniques. *International Journal of Project Management*, *33*(7), 1597–1609. https://doi.org/10.1016/j.ijproman.2015.06.012

[2]. Ackermann, F., Eden, C., Williams, T., & Howick, S. (2007). Systemic risk assessment: A case study. *Journal of the Operational Research Society*, *58*(1), 39–51. https://doi.org/10.1057/palgrave.jors.2602105

[3]. Ahiaga-Dagbui, D. D., & Smith, S. D. (2014). Dealing with construction cost overruns using data mining. *Construction Management and Economics*, *32*(7–8), 682–694. https://doi.org/10.1080/01446193.2014.933854

[4]. Ahiaga-Dagbui, D. D., Love, P. E. D., Smith, S. D., & Ackermann, F. (2017). Toward a systemic view to cost overrun causation in infrastructure projects: A review and implications for research. *Project Management Journal*, *48*(2), 88–98. https://doi.org/10.1177/875697281704800207

[5]. Al-Momani, A. H. (2000). Construction delay: A quantitative analysis. *International Journal of Project Management*, *18*(1), 51–59. https://doi.org/10.1016/S0263-7863(98)00060-X

[6]. Aliverdi, R., Moslemi Naeni, L., & Salehipour, A. (2013). Monitoring project duration and cost in a construction project by applying statistical quality control charts. *International Journal of Project Management*, *31*(3), 411–423. https://doi.org/10.1016/j.ijproman.2012.08.005

[7]. Anastasopoulos, P. C., Labi, S., Bhargava, A., Bordat, C., & Mannering, F. L. (2010). Frequency of change orders in highway construction using alternate count-data modeling methods. *Journal of Construction Engineering and Management*, *136*(8), 886–893. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000198

[8]. Anbari, F. T. (2003). Earned value project management method and extensions. *Project Management Journal*, *34*(4), 12–23. https://doi.org/10.1177/875697280303400403

[9]. Ansar, A., Flyvbjerg, B., Budzier, A., & Lunn, D. (2014). Should we build more large dams? The actual costs of hydropower megaproject development. *Energy Policy*, *69*, 43–56. https://doi.org/10.1016/j.enpol.2013.10.069

[10]. Asadi, A., Alsubaey, M., & Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. *International Journal of Advanced Logistics*, *4*(2), 115–130. https://doi.org/10.1080/2287108X.2015.1059920

[11]. Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International Journal of Project Management*, *24*(4), 349–357. https://doi.org/10.1016/j.ijproman.2005.11.010

[12]. Baccarini, D., & Love, P. E. D. (2014). Statistical characteristics of cost contingency in water infrastructure projects. *Journal of Construction Engineering and Management*, *140*(3), 04013063. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000820

[13]. Bakhshi, J., Ireland, V., & Gorod, A. (2016). Clarifying the project complexity construct: Past, present and future. *International Journal of Project Management*, *34*(7), 1199–1213. https://doi.org/10.1016/j.ijproman.2016.06.002

[14]. Barraza, G. A., Back, W. E., & Mata, F. (2004). Probabilistic forecasting of project performance using stochastic S curves. *Journal of Construction Engineering and Management*, *130*(1), 25–32. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:1(25)

[15]. Batselier, J., & Vanhoucke, M. (2015). Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earned value management. *International Journal of Project Management*, *33*(7), 1588–1596. https://doi.org/10.1016/j.ijproman.2015.04.003

[16]. Batselier, J., & Vanhoucke, M. (2015). Construction and evaluation framework for a real-life project database. *International Journal of Project Management*, *33*(3), 697–710. https://doi.org/10.1016/j.ijproman.2014.09.004

[17]. Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management*, *35*(1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003

[18]. Bauch, G. T., & Chung, C. A. (2001). A statistical project control tool for engineering managers. *Project Management Journal*, *32*(2), 37–44. https://doi.org/10.1177/875697280103200205

[19]. Bhargava, A., Anastasopoulos, P. C., Labi, S., Sinha, K. C., & Mannering, F. L. (2010). Three-stage least-squares analysis of time and cost overruns in construction contracts. *Journal of Construction Engineering and Management*, *136*(11), 1207–1218. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000225

[20]. Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

[21]. Bordat, C., McCullouch, B. G., Labi, S., & Sinha, K. C. (2004). An analysis of cost overruns and time delays of INDOT projects (Report FHWA/IN/JTRP-2004/07). *Joint Transportation Research Program, Purdue University*. https://doi.org/10.5703/1288284313134

[22]. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[23]. Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366–381. https://doi.org/10.1037/0022-3514.67.3.366

[24]. Cantarelli, C. C., Flyvbjerg, B., Molin, E. J. E., & van Wee, B. (2010). Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. *European Journal of Transport and Infrastructure Research*, *10*(1), 5–18. https://doi.org/10.18757/ejtir.2010.10.1.2864

[25]. Cantarelli, C. C., Molin, E. J. E., van Wee, B., & Flyvbjerg, B. (2012). Characteristics of cost overruns for Dutch transport infrastructure projects and the importance of the decision to build and project phases. *Transport Policy*, *22*, 49–56. https://doi.org/10.1016/j.tranpol.2012.04.001

[26]. Caron, F., Ruggeri, F., & Merli, A. (2013). A Bayesian approach to improve estimate at completion in earned value management. *Project Management Journal*, *44*(1), 3–16. https://doi.org/10.1002/pmj.21303

[27]. Chan, D. W. M., & Kumaraswamy, M. M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management*, *15*(1), 55–63. https://doi.org/10.1016/S0263-7863(96)00039-7

[28]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

[29]. Chen, H. L., Chen, W. T., & Lin, Y. L. (2016). Earned value project management: Improving the predictive power of planned value. *International Journal of Project Management*, *34*(1), 22–29. https://doi.org/10.1016/j.ijproman.2015.09.008

[30]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[31]. Chou, J.-S., & Lin, C. (2013). Predicting disputes in public–private partnership projects: Classification and ensemble models. *Journal of Computing in Civil Engineering*, *27*(1), 51–60. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000197

[32]. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

[33]. Colin, J., & Vanhoucke, M. (2014). Setting tolerance limits for statistical project control using earned value management. *Omega*, *49*, 107–122. https://doi.org/10.1016/j.omega.2014.06.001

[34]. Colin, J., & Vanhoucke, M. (2015). A comparison of the performance of various project control methods using earned value management systems. *Expert Systems with Applications*, *42*(6), 3159–3175. https://doi.org/10.1016/j.eswa.2014.12.007

[35]. Cooper, K. G. (1980). Naval ship production: A claim settled and a framework built. *Interfaces*, *10*(6), 20–36. https://doi.org/10.1287/inte.10.6.20

[36]. Creedy, G. D., Skitmore, M., & Wong, J. K. W. (2010). Evaluation of risk factors leading to cost overrun in delivery of highway construction projects. *Journal of Construction Engineering and Management*, *136*(5), 528–537. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000160

[37]. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

[38]. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*, 1–30. https://jmlr.org/papers/v7/demsar06a.html

[39]. Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

[40]. Doloi, H., Sawhney, A., Iyer, K. C., & Rentala, S. (2012). Analysing factors affecting delays in Indian construction projects. *International Journal of Project Management*, *30*(4), 479–489. https://doi.org/10.1016/j.ijproman.2011.10.004

[41]. Eden, C., Williams, T., Ackermann, F., & Howick, S. (2000). The role of feedback dynamics in disruption and delay on the nature of disruption and delay (D&D) in major projects. *Journal of the Operational Research Society*, *51*(3), 291–300. https://doi.org/10.1057/palgrave.jors.2600919

[42]. Elmousalami, H. H. (2020). Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review. *Journal of Construction Engineering and Management*, *146*(1), 03119008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678

[43]. Elshaer, R. (2013). Impact of sensitivity information on the prediction of project's duration using earned schedule method. *International Journal of Project Management*, *31*(4), 579–588. https://doi.org/10.1016/j.ijproman.2012.10.006

[44]. Fleming, Q. W., & Koppelman, J. M. (2010). *Earned value project management* (4th ed.). Project Management Institute.

[45]. Flyvbjerg, B., Ansar, A., Budzier, A., Buhl, S., Cantarelli, C., Garbuio, M., Glenting, C., Holm, M. S., Lovallo, D., Lunn, D., Molin, E., Rønnest, A., Stewart, A., & van Wee, B. (2018). Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice*, *118*, 174–190. https://doi.org/10.1016/j.tra.2018.07.013

[46]. Flyvbjerg, B., Budzier, A., Lee, J. S., Keil, M., Lunn, D., & Bester, D. W. (2022). The empirical reality of IT project cost overruns: Discovering a power-law distribution. *Journal of Management Information Systems*, *39*(3), 607–639. https://doi.org/10.1080/07421222.2022.2096544

[47]. Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, *68*(3), 279–295. https://doi.org/10.1080/01944360208976273

[48]. Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2003). How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews*, *23*(1), 71–88. https://doi.org/10.1080/01441640309904

[49]. Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2004). What causes cost overrun in transport infrastructure projects? *Transport Reviews*, *24*(1), 3–18. https://doi.org/10.1080/0144164032000080494a

[50]. 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

[51]. Flyvbjerg, B. (2008). Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies*, *16*(1), 3–21. https://doi.org/10.1080/09654310701747936

[52]. Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal*, *45*(2), 6–19. https://doi.org/10.1002/pmj.21409

[53]. Flyvbjerg, B. (2021). Top ten behavioral biases in project management: An overview. *Project Management Journal*, *52*(6), 531–546. https://doi.org/10.1177/87569728211049046

[54]. Ford, D. N. (2002). Achieving multiple project objectives through contingency management. *Journal of Construction Engineering and Management*, *128*(1), 30–39. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:1(30)

[55]. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[56]. Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

[57]. Han, S., Love, P., & Peña-Mora, F. (2013). A system dynamics model for assessing the impacts of design errors in construction projects. *Mathematical and Computer Modelling*, *57*(9–10), 2044–2053. https://doi.org/10.1016/j.mcm.2011.06.039

[58]. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

[59]. Hanna, A. S., Camlic, R., Peterson, P. A., & Nordheim, E. V. (2002). Quantitative definition of projects impacted by change orders. *Journal of Construction Engineering and Management*, *128*(1), 57–64. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:1(57)

[60]. Hazır, Ö. (2015). A review of analytical models, approaches and decision support tools in project monitoring and control. *International Journal of Project Management*, *33*(4), 808–815. https://doi.org/10.1016/j.ijproman.2014.09.005

[61]. Howick, S., & Eden, C. (2001). The impact of disruption and delay when compressing large projects: Going for incentives? *Journal of the Operational Research Society*, *52*(1), 26–34. https://doi.org/10.1057/palgrave.jors.2601045

[62]. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[63]. Ibbs, W. (2005). Impact of change's timing on labor productivity. *Journal of Construction Engineering and Management*, *131*(11), 1219–1223. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:11(1219)

[64]. Jahren, C. T., & Ashe, A. M. (1990). Predictors of cost-overrun rates. *Journal of Construction Engineering and Management*, *116*(3), 548–552. https://doi.org/10.1061/(ASCE)0733-9364(1990)116:3(548)

[65]. Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31. https://doi.org/10.1287/mnsc.39.1.17

[66]. Kaming, P. F., Olomolaiye, P. O., Holt, G. D., & Harris, F. C. (1997). Factors influencing construction time and cost overruns on high-rise projects in Indonesia. *Construction Management and Economics*, *15*(1), 83–94. https://doi.org/10.1080/014461997373132

[67]. Khamooshi, H., & Golafshani, H. (2014). EDM: Earned Duration Management, a new approach to schedule performance management and measurement. *International Journal of Project Management*, *32*(6), 1019–1041. https://doi.org/10.1016/j.ijproman.2013.11.002

[68]. Kim, B.-C., & Reinschmidt, K. F. (2009). Probabilistic forecasting of project duration using Bayesian inference and the beta distribution. *Journal of Construction Engineering and Management*, *135*(3), 178–186. https://doi.org/10.1061/(ASCE)0733-9364(2009)135:3(178)

[69]. Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39*(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

[70]. Larsen, J. K., Shen, G. Q., Lindhard, S. M., & Brunoe, T. D. (2016). Factors affecting schedule delay, cost overrun, and quality level in public construction projects. *Journal of Management in Engineering*, *32*(1), 04015032. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000391

[71]. Leu, S.-S., & Lin, Y.-C. (2008). Project performance evaluation based on statistical process control techniques. *Journal of Construction Engineering and Management*, *134*(10), 813–819. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:10(813)

[72]. Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes. *International Journal of Project Management*, *27*(4), 400–407. https://doi.org/10.1016/j.ijproman.2008.02.009

[73]. Locatelli, G., Invernizzi, D. C., & Brookes, N. J. (2017). Project characteristics and performance in Europe: An empirical analysis for large transport infrastructure projects. *Transportation Research Part A: Policy and Practice*, *98*, 108–122. https://doi.org/10.1016/j.tra.2017.01.024

[74]. Locatelli, G., Mariani, G., Sainati, T., & Greco, M. (2017). Corruption in public projects and megaprojects: There is an elephant in the room! *International Journal of Project Management*, *35*(3), 252–268. https://doi.org/10.1016/j.ijproman.2016.09.010

[75]. Lovallo, D., & Kahneman, D. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review*, *81*(7), 56–63. https://hbr.org/2003/07/delusions-of-success-how-optimism-undermines-executives-decisions

[76]. Love, P. E. D., & Ahiaga-Dagbui, D. D. (2018). Debunking fake news in a post-truth era: The plausible untruths of cost underestimation in transport infrastructure projects. *Transportation Research Part A: Policy and Practice*, *113*, 357–368. https://doi.org/10.1016/j.tra.2018.04.019

[77]. Love, P. E. D., Ahiaga-Dagbui, D. D., & Irani, Z. (2016). Cost overruns in transportation infrastructure projects: Sowing the seeds for a probabilistic theory of causation. *Transportation Research Part A: Policy and Practice*, *92*, 184–194. https://doi.org/10.1016/j.tra.2016.08.007

[78]. Love, P. E. D., Edwards, D. J., & Irani, Z. (2012). Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns. *IEEE Transactions on Engineering Management*, *59*(4), 560–571. https://doi.org/10.1109/TEM.2011.2163628

[79]. Love, P. E. D., Holt, G. D., Shen, L. Y., Li, H., & Irani, Z. (2002). Using systems dynamics to better understand change and rework in construction project management systems. *International Journal of Project Management*, *20*(6), 425–436. https://doi.org/10.1016/S0263-7863(01)00039-4

[80]. Love, P. E. D., Ika, L. A., & Ahiaga-Dagbui, D. D. (2019). On de-bunking 'fake news' in a post truth era: Why does the Planning Fallacy explanation for cost overruns fall short? *Transportation Research Part A: Policy and Practice*, *126*, 397–408. https://doi.org/10.1016/j.tra.2019.06.013

[81]. Love, P. E. D., Sing, C.-P., Carey, B., & Kim, J. T. (2015). Estimating construction contingency: Accommodating the potential for cost overruns in road construction projects. *Journal of Infrastructure Systems*, *21*(2), 04014035. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000221

[82]. Love, P. E. D., Sing, C.-P., Wang, X., Irani, Z., & Thwala, D. W. (2014). Overruns in transportation infrastructure projects. *Structure and Infrastructure Engineering*, *10*(2), 141–159. https://doi.org/10.1080/15732479.2012.715173

[83]. Love, P. E. D., Wang, X., Sing, C.-P., & Tiong, R. L. K. (2013). Determining the probability of project cost overruns. *Journal of Construction Engineering and Management*, *139*(3), 321–330. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000575

[84]. Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, *32*(1), 1–12. https://doi.org/10.1080/00401706.1990.10484583

[85]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 4765–4774). https://arxiv.org/abs/1705.07874

[86]. Lyneis, J. M., & Ford, D. N. (2007). System dynamics applied to project management: A survey, assessment, and directions for future research. *System Dynamics Review*, *23*(2–3), 157–189. https://doi.org/10.1002/sdr.377

[87]. Mahamid, I. (2011). Early cost estimating for road construction projects using multiple regression techniques. *Australasian Journal of Construction Economics and Building*, *11*(4), 87–101. https://doi.org/10.5130/AJCEB.v11i4.2195

[88]. Mak, S., & Picken, D. (2000). Using risk analysis to determine construction project contingencies. *Journal of Construction Engineering and Management*, *126*(2), 130–136. https://doi.org/10.1061/(ASCE)0733-9364(2000)126:2(130)

[89]. Martens, A., & Vanhoucke, M. (2017). A buffer control method for top-down project control. *European Journal of Operational Research*, *262*(1), 274–286. https://doi.org/10.1016/j.ejor.2017.03.034

[90]. Molenaar, K. R. (2005). Programmatic cost risk analysis for highway megaprojects. *Journal of Construction Engineering and Management*, *131*(3), 343–353. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:3(343)

[91]. Montgomery, D. C. (2020). *Introduction to statistical quality control* (8th ed.). John Wiley & Sons.

[92]. Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, *52*(3), 239–281. https://doi.org/10.1023/A:1024068626366

[93]. Narbaev, T., & De Marco, A. (2014). An earned schedule-based regression model to improve cost estimate at completion. *International Journal of Project Management*, *32*(6), 1007–1018. https://doi.org/10.1016/j.ijproman.2013.12.005

[94]. Odeck, J., Welde, M., & Volden, G. H. (2015). The impact of external quality assurance of costs estimates on cost overruns: Empirical evidence from the Norwegian road sector. *European Journal of Transport and Infrastructure Research*, *15*(3), 286–303. https://doi.org/10.18757/ejtir.2015.15.3.3079

[95]. Odeck, J. (2004). Cost overruns in road construction—what are their sizes and determinants? *Transport Policy*, *11*(1), 43–53. https://doi.org/10.1016/S0967-070X(03)00017-9

[96]. Olawale, Y. A., & Sun, M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. *Construction Management and Economics*, *28*(5), 509–526. https://doi.org/10.1080/01446191003674519

[97]. Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1–2), 100–115. https://doi.org/10.1093/biomet/41.1-2.100

[98]. Pajares, J., & López-Paredes, A. (2011). An extension of the EVM analysis for project monitoring: The Cost Control Index and the Schedule Control Index. *International Journal of Project Management*, *29*(5), 615–621. https://doi.org/10.1016/j.ijproman.2010.04.005

[99]. Pinto, J. K. (2013). Lies, damned lies, and project plans: Recurring human errors that can ruin the project planning process. *Business Horizons*, *56*(5), 643–653. https://doi.org/10.1016/j.bushor.2013.05.006

[100]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

[101]. Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, *1*(3), 239–250. https://doi.org/10.1080/00401706.1959.10489860

[102]. Rozenes, S., Vitner, G., & Spraggett, S. (2006). Project control: Literature review. *Project Management Journal*, *37*(4), 5–14. https://doi.org/10.1177/875697280603700402

[103]. Russell, J. S., Jaselskis, E. J., & Lawrence, S. P. (1997). Continuous assessment of project performance. *Journal of Construction Engineering and Management*, *123*(1), 64–71. https://doi.org/10.1061/(ASCE)0733-9364(1997)123:1(64)

[104]. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[105]. Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management*, *25*(5), 517–526. https://doi.org/10.1016/j.ijproman.2006.11.007

[106]. Samset, K., & Volden, G. H. (2016). Front-end definition of projects: Ten paradoxes and some reflections regarding project management and project governance. *International Journal of Project Management*, *34*(2), 297–313. https://doi.org/10.1016/j.ijproman.2015.01.014

[107]. Sanni-Anibire, M. O., Mohamad Zin, R., & Olatunji, S. O. (2022). Machine learning model for delay risk assessment in tall building projects. *International Journal of Construction Management*, *22*(11), 2134–2143. https://doi.org/10.1080/15623599.2020.1768326

[108]. Shane, J. S., Molenaar, K. R., Anderson, S., & Schexnayder, C. (2009). Construction project cost escalation factors. *Journal of Management in Engineering*, *25*(4), 221–229. https://doi.org/10.1061/(ASCE)0742-597X(2009)25:4(221)

[109]. Shrestha, P. P., Burns, L. A., & Shields, D. R. (2013). Magnitude of construction cost and schedule overruns in public work projects. *Journal of Construction Engineering*, *2013*, 935978. https://doi.org/10.1155/2013/935978

[110]. Siemiatycki, M. (2009). Academics and auditors: Comparing perspectives on transportation project cost overruns. *Journal of Planning Education and Research*, *29*(2), 142–156. https://doi.org/10.1177/0739456X09348798

[111]. Soibelman, L., & Kim, H. (2002). Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering*, *16*(1), 39–48. https://doi.org/10.1061/(ASCE)0887-3801(2002)16:1(39)

[112]. Son, H., Kim, C., & Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. *Automation in Construction*, *27*, 60–66. https://doi.org/10.1016/j.autcon.2012.05.013

[113]. Sovacool, B. K., Gilbert, A., & Nugent, D. (2014). An international comparative assessment of construction cost overruns for electricity infrastructure. *Energy Research & Social Science*, *3*, 152–160. https://doi.org/10.1016/j.erss.2014.07.016

[114]. Taroun, A. (2014). Towards a better modelling and assessment of construction risk: Insights from a literature review. *International Journal of Project Management*, *32*(1), 101–115. https://doi.org/10.1016/j.ijproman.2013.03.004

[115]. Touran, A., & Lopez, R. (2006). Modeling cost escalation in large infrastructure projects. *Journal of Construction Engineering and Management*, *132*(8), 853–860. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:8(853)
[116]. Touran, A. (2003). Probabilistic model for cost contingency. *Journal of Construction Engineering and Management*, *129*(3), 280–284. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:3(280)
[117]. Vandevoorde, S., & Vanhoucke, M. (2006). A comparison of different project duration forecasting methods using earned value metrics. *International Journal of Project Management*, *24*(4), 289–302. https://doi.org/10.1016/j.ijproman.2005.10.004
[118]. Vanhoucke, M., & Vandevoorde, S. (2007). A simulation and evaluation of earned value metrics to forecast the project duration. *Journal of the Operational Research Society*, *58*(10), 1361–1374. https://doi.org/10.1057/palgrave.jors.2602296
[119]. Vanhoucke, M. (2012). Measuring the efficiency of project control using fictitious and empirical project data. *International Journal of Project Management*, *30*(2), 252–263. https://doi.org/10.1016/j.ijproman.2011.05.006
[120]. Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*, 91. https://doi.org/10.1186/1471-2105-7-91
[121]. Vidal, L.-A., Marle, F., & Bocquet, J.-C. (2011). Measuring project complexity using the Analytic Hierarchy Process. *International Journal of Project Management*, *29*(6), 718–727. https://doi.org/10.1016/j.ijproman.2010.07.005
[122]. Volden, G. H., & Samset, K. (2017). Governance of major public investment projects: Principles and practices in six countries. *Project Management Journal*, *48*(3), 90–108. https://doi.org/10.1177/875697281704800306
[123]. Wachs, M. (1989). When planners lie with numbers. *Journal of the American Planning Association*, *55*(4), 476–479.
[124]. Wauters, M., & Vanhoucke, M. (2014). Support vector machine regression for project control forecasting. *Automation in Construction*, *47*, 92–106. https://doi.org/10.1016/j.autcon.2014.07.014
[125]. Wauters, M., & Vanhoucke, M. (2016). A comparative study of artificial intelligence methods for project duration forecasting. *Expert Systems with Applications*, *46*, 249–261. https://doi.org/10.1016/j.eswa.2015.10.008
[126]. Wauters, M., & Vanhoucke, M. (2017). A nearest neighbour extension to project duration forecasting with artificial intelligence. *European Journal of Operational Research*, *259*(3), 1097–1111. https://doi.org/10.1016/j.ejor.2016.11.018
[127]. Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, *43*, 23–29. https://doi.org/10.1016/j.autcon.2014.02.014
[128]. Williams, T. (2005). Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE Transactions on Engineering Management*, *52*(4), 497–508. https://doi.org/10.1109/TEM.2005.856572
[129]. Yaseen, Z. M., Ali, Z. H., Salih, S. Q., & Al-Ansari, N. (2020). Prediction of risk delay in construction projects using a hybrid artificial intelligence model. *Sustainability*, *12*(4), 1514. https://doi.org/10.3390/su12041514
[130]. Abdullah Mohammad, S. (2023). $CUGAO_2$ and $CUINS_2$ Quantum Dot Sensitization for Enhanced Photovoltaic Efficiency in Wearable Fiber-Shaped Solar Cells. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 209-261. https://doi.org/10.63125/wysdvy71
[131]. Abdullah Mohammad, S. (2024). Triboelectric And Piezoelectric Nanogenerators Using Binary and Ternary Oxide Nanostructures: Optimization of Energy Conversion Efficiency for Wearable Harvesting. *Review of Applied Science and Technology*, *3*(03), 31–72. https://doi.org/10.63125/w7q9cs55
[132]. Adnan, S., & Iftekhar, A. (2023). Predicting Cost Variance in Digital Transformation Projects Through Ensemble Learning and Project Performance Analytics. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 01–36. https://doi.org/10.63125/j18hxs51
[133]. Ahmed, H. C. H. R. (2024). Machine Learning Enhanced Demand Forecasting and Portfolio Analytics for Scalable Growth Management in Digital Logistics Operations. *American Journal of Advanced Technology and Engineering Solutions*, *4*(02), 33-81. https://doi.org/10.63125/5xq3nd53
[134]. Ahmed, H. C. H. R. (2026). An Artificial Intelligence Integrated IT Management Framework for Analytics-Driven Operational Decision-Making in U.S. Supply Chain and Logistics Systems. *American Journal of Interdisciplinary Studies*, *16*(06), 01-37. https://doi.org/10.63125/b5hgg259
[135]. Ahmed, H. C. H. R., & Nuzhat Sadia, P. (2023). A Quantitative Assessment of CRM Analytics and KPI Dashboard Design on Enterprise Revenue Conversion in Emerging Market Logistics Platforms. *American Journal of Data Science and Analytics*, *4*(09), 50-95. https://doi.org/10.63125/315bge15
[136]. Asheke, M. (2023). Machine Learning–Based Predictive Quality Modeling for Defect Detection and Process Optimization in Advanced Manufacturing Systems. *Review of Applied Science and Technology*, *13*(12), 01–49. https://doi.org/10.63125/thy73b16
[137]. Asheke, M. (2024). Digital Twin and Industrial Internet of Things–Enabled Predictive Maintenance for Reliability Enhancement in Smart Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 792–838. https://doi.org/10.63125/jgrdqd62

[138]. Ashik, M., & Shabbir, A. (2022). Digital Twin-Enabled Predictive Maintenance for Rotating and Static Equipment: Bridging FEA-Based Failure Analysis with Real-Time Asset Health Monitoring in Energy and Power Facilities. *International Journal of Scientific Interdisciplinary Research*, *12*(05), 01–37. https://doi.org/10.63125/0qaysz49

[139]. Ashrafuzzaman, M. (2024). The impact of cloud-based management information systems on hrm efficiency: an analysis of small and medium-sized enterprises (SMES). *Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems*, *1*(01), 40-56. https://doi.org/10.69593/ajaimldsmis.v1i01.124

[140]. Gazi Yeasin, A. (2023). Impact of Accessibility-First Design Systems on WCAG 2.1 Compliance and Inclusive User Experience in Enterprise Software Platforms: A Mixed-Methods Evaluation. *Review of Applied Science and Technology*, *13*(07), 01–49. https://doi.org/10.63125/nxh4gp19

[141]. Jahanara, A. (2025). AI-Enabled Workforce Management: Its Impact on Employee Burnout, Staffing Efficiency, and Organizational Performance in Healthcare. *Review of Applied Science and Technology*, *4*(04), 247–295. https://doi.org/10.63125/5617ff74

[142]. Jahanara, A. (2026). Impact of Artificial Intelligence on Operational Efficiency and Cost Reduction in U.S. Healthcare Organizations. *American Journal of Interdisciplinary Studies*, *7*(02), 225-276. https://doi.org/10.63125/3p24ma43

[143]. Jahanara, A., & Mst Kaniz, F. (2023). Demand Variability and Capacity Planning in Home and Community-Based Services: A Decision-Analytic Framework for Health Services Managers. *American Journal of Data Science and Analytics*, *4*(09), 01-49. https://doi.org/10.63125/3y62vr16

[144]. Jannatul, F. (2024). Alzheimer's Disease and Mobility Multimodal Prediction of Gait Decline, Fall Risk, and Loss of Functional Independence in Mild Cognitive Impairment and Early Alzheimer's Disease. *Review of Applied Science and Technology*, *3*(01), 420–483. https://doi.org/10.63125/q2d90m63

[145]. Jannatul, F. (2025). Spinal Cord Injury and Neurorehabilitation Deep Learning–Based Prediction of Functional Recovery Trajectories After Traumatic Spinal Cord Injury. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 2865–2922. https://doi.org/10.63125/m738bh82

[146]. Jannatul, F. (2026). Predicting Falls and Mobility Decline Among Older Adults Using Longitudinal Functional Data and Ensemble Machine Learning. *International Journal of Scientific Interdisciplinary Research*, *7*(1), 581–631. https://doi.org/10.63125/56hans40

[147]. Kazi Rakib Hasan, S., & Adnan, S. (2022). Quantitative Assessment of Machine Learning-Based Schedule Risk Prediction for Reducing Delays in Complex Technology Projects. *Review of Applied Science and Technology*, *1*(04), 464–501. https://doi.org/10.63125/dn8nkm85

[148]. Maishar Mostari, Z. (2025). Impact of Deep Learning on Real-Time Network Intrusion Detection: A Quantitative Evaluation Using Benchmark Datasets. *Review of Applied Science and Technology*, *4*(04), 296–338. https://doi.org/10.63125/tnaksd40

[149]. Mansura Akter, E., & Sumaiya, M. (2022). Antimicrobial Resistance Profiles of Acinetobacter Baumannii and Vibrio Cholerae in Hospital Wastewater and Adjacent Household Water Sources. *American Journal of Data Science and Analytics*, *3*(09), 01-55. https://doi.org/10.63125/gbervt14

[150]. Mashfiqur, R. (2025). Multivariate Time-Series Forecasting of Working Capital Requirements for Improving Cash-Flow Stability in Export-Oriented Manufacturing. *American Journal of Data Science and Analytics*, *6*(12), 217-249. https://doi.org/10.63125/ty3dtt46

[151]. Mashfiqur, R. (2026). Explainable Artificial Intelligence for Supplier Cost-Risk Assessment and Margin Optimization in Global Manufacturing Supply Chains. *American Journal of Interdisciplinary Studies*, *16*(08), 01-37. https://doi.org/10.63125/np6bw920

[152]. Mashfiqur, R., & Md, A. (2022). Monte Carlo Simulation and Stochastic Cash-Flow Modeling for Reducing Liquidity Risk in Project-Based Enterprises. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 382–431. https://doi.org/10.63125/zf3fwv21

[153]. Mashiat, N. (2024). Reducing Telecommunications Service Interruptions Through Predictive Root-Cause Analytics of Multivariate System Telemetry. *International Journal of Scientific Interdisciplinary Research*, *14*(11), 01-37. https://doi.org/10.63125/nqjh0c29

[154]. Mashiat, N. (2025). Improving Enterprise Data Integrity Through Automated Validation and Schema-Drift Detection in Large-Scale Data Integration Systems. *Review of Applied Science and Technology*, *4*(02), 889–934. https://doi.org/10.63125/21zzkk04

[155]. Mashiat, N., & Md Anisur Rahman, M. (2022). Machine Learning Approaches to Credit Risk Scoring and Default Prediction in Small-Business Lending: Empirical Evidence from Lending Portfolio Data. *American Journal of Data Science and Analytics*, *3*(06), 137-182. https://doi.org/10.63125/j256rs96

[156]. Md, A., & Gazi Yeasin, A. (2022). Machine Learning Enhanced Spatial-Data Workflow Redesign: A Quantitative Evaluation of Task Success and Operational Cost Reduction in Enterprise LiDAR Platforms. *American Journal of Data Science and Analytics*, *3*(06), 80-136. https://doi.org/10.63125/0den7c02

[157]. Md Anisur Rahman, M. (2023). Development of AI-Driven Business Intelligence Frameworks for Financial Risk and Liquidity Monitoring in Real Time. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 152–203. https://doi.org/10.63125/jm55ca64

[158]. Md Arif Uz, Z., Sharmin, S., & Md Abdur, R. (2025). Factors Influencing Behavioural Intention to Use he Innovative Open And Distance Learning Systems With The Role Of Continuance Intention. *American Journal of Scholarly Research and Innovation*, *4*(01), 202-220. https://doi.org/10.63125/pbjxp014

[159]. Md Ashfaq, S., & Abdullah Mohammad, S. (2022). CNT/PANI Composite Electrodes for Flexible Fiber-Shaped Dye-Sensitized Solar Cells: Synthesis, Characterization, and Photovoltaic Performance. *American Journal of Interdisciplinary Studies*, *3*(02), 221-274. https://doi.org/10.63125/m334mf41

[160]. Md Asif Ali Sheak, A., & Tahmedur, R. (2023). A Data-Driven Framework for Engineering Project Risk Assessment Under Schedule and Cost Uncertainty. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 204–251. https://doi.org/10.63125/kq4zzh16

[161]. Md Imran, H., & Jannatul, F. (2021). Translational Stroke and Neurorehabilitation Research Integrating Neuroimaging, Circulating Biomarkers, and Functional Assessments to Predict Recovery After Stroke. *American Journal of Data Science and Analytics*, *2*(12), 156-198. https://doi.org/10.63125/c28rzv23

[162]. Md Imran, H., & Jannatul, F. (2022). Molecular and Precision Rehabilitation Molecular Mechanisms Linking Mitochondrial Dysfunction, Inflammation, and Muscle Degeneration with Functional Decline and Rehabilitation Response in Neuromuscular Disease. *American Journal of Interdisciplinary Studies*, *12*(08), 01-44. https://doi.org/10.63125/0ym5bh79

[163]. Md Kaium, H. (2023). A GIS-Based Multi-Criteria Decision-Support and Land-Suitability Optimization Framework for Climate-Resilient Urban and Regional Planning. *American Journal of Data Science and Analytics*, *4*(06), 43–103. https://doi.org/10.63125/wvq0sx87

[164]. Md Mahamud Pervaz, P. (2023). Integration of MES, SAP, And PI Historian Platforms for Real-Time Industrial Asset Performance Management. *Review of Applied Science and Technology*, *2*(04), 393–445. https://doi.org/10.63125/9ebtas97

[165]. Md Mahamud Pervaz, P. (2025). Development of a CMMS-Integrated Predictive Maintenance Framework for Industrial PLC and Drive Systems. *American Journal of Scholarly Research and Innovation*, *4*(01), 894–942. https://doi.org/10.63125/j33gtf31

[166]. Mohammad Badrul, A., & Md. Mominul, H. (2023). Machine Learning–Based Assessment of Environmental and Public Health Risks in Sewerage Sludge Management Systems. *American Journal of Advanced Technology and Engineering Solutions*, *3*(02), 33-77. https://doi.org/10.63125/y3yfxa92

[167]. Mst Kaniz, F., & Jahanara, A. (2022). Lean Six Sigma Implementation in Home Health Administrative Operations: A DMAIC Case Study of Managerial Workflow Redesign and Error Reduction. *International Journal of Scientific Interdisciplinary Research*, *12*(04), 01–60. https://doi.org/10.63125/xchca204

[168]. Nishat, J. (2023). Communicable Disease Prevention in a Post-Pandemic Era: A Global Review of Public Health Education, Surveillance, and Community-Level Interventions. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 630–678. https://doi.org/10.63125/mnqefj36

[169]. Nuzhat Sadia, P. (2023). A Multi-Criteria Decision-Support Framework For ERP-Based Business Process Optimization in Mid-Sized Manufacturing Enterprises. *American Journal of Scholarly Research and Innovation*, *2*(02), 519–573. https://doi.org/10.63125/j58k5d55

[170]. Rony, S., & Md, A. (2026). Multi-Agency Coordination in Sustained Healthcare Supply Shortages: Lessons from Covid-19, Drug Shortages, and Critical Medical Supply Disruptions. *American Journal of Data Science and Analytics*, *7*(01), 181-221. https://doi.org/10.63125/5v6t8515

[171]. Sadia, Z., & Maishar Mostari, Z. (2022). The Impact of Interactive Dashboard Visualization on Data-Driven Decision Making in Financial Services: A Qualitative Evaluation. *International Journal of Scientific Interdisciplinary Research*, *12*(03), 01–35. https://doi.org/10.63125/3jstsv75

[172]. Shaiful, M., Anisur, R., & Md, A. (2022). A Systematic Literature Review on the Role of Digital Health Twins in Preventive Healthcare For Personal and Corporate Wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373

[173]. Stabak, D., & Md Shahid, N. (2022). Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 432–477. https://doi.org/10.63125/q3r4kk16

[174]. Stabak, D., & Md Siam, T. (2023). Real-Time Edge Computing for Cyber-Physical Systems: A Meta-Analysis of Energy Efficiency and Fault Tolerance. *American Journal of Scholarly Research and Innovation*, *13*(06), 01–44. https://doi.org/10.63125/g9c9cp41

[175]. Sumaiya, M. (2024). Hospital Wastewater as A Reservoir of Antimicrobial Resistance Genes: Implications for Public Health and Genomic Surveillance in The United States. *Review of Applied Science and Technology*, *3*(04), 448–508. https://doi.org/10.63125/ahvas976

[176]. Sumaiya, M. (2025). CRISPR-Cas Systems in Precision Oncology: Current Applications and Future Challenges. *American Journal of Interdisciplinary Studies*, *6*(1), 513-564. https://doi.org/10.63125/rqpqr418

[177]. Sumaiya, M. (2026). Artificial Intelligence Driven Multi-Omics Approaches for Precision Medicine: From Molecular Discovery to Clinical Translation. *American Journal of Scholarly Research and Innovation*, *5*(01), 429–477. https://doi.org/10.63125/yfme0q12

[178]. Syed Feroz, A. (2023). Quality-Management Systems for Resilient Energy Infrastructure Fabrication. *International Journal of Scientific Interdisciplinary Research*, *13*(10), 37–91. https://doi.org/10.63125/mbhhrm88

[179]. Syed Feroz, A. (2024). AI-Assisted NDT Analytics for Early Crack Detection in Energy-Infrastructure Pressure Vessels Using Advanced Manufacturing Data. *Review of Applied Science and Technology*, *14*(07), 01-52. https://doi.org/10.63125/8bdqs538

[180]. Syed Feroz, A. (2025). Intelligent Process Monitoring for Enhancing Weld Integrity and Reliability in Advanced Manufacturing of LPG Energy Infrastructure. *American Journal of Interdisciplinary Studies*, *6*(3), 394-441. https://doi.org/10.63125/p3b5mb44

[181]. Syed Feroz, A., & Md Mahamud Pervaz, P. (2022). Root-Cause Analysis and FMEA-Driven Reliability Improvement in Gas Production Facilities: A Multi-Year Empirical Study. *International Journal of Scientific Interdisciplinary Research*, *12*(10), 01–53. https://doi.org/10.63125/70msp878

[182]. Tahmedur, R., & Rifat, C. (2023). Critical Path Analysis and Resource-Leveling Optimization for Multi-Project Engineering Portfolios. *American Journal of Interdisciplinary Studies*, *13*(12), 01-47. https://doi.org/10.63125/8eqnea46