African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 5, Issue 4, 2026
Page No: 241-263
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2025: 1.126 SJIFactor 2024: 6.752 معامل التأثير العربي 2025: 1.65

A Predictive Framework for Infrastructure Project Failure in Libya: Integrating Instability Indicators, Temporal Machine Learning, and Explainable Early Warning (Simulation-Informed Proof-of-Concept Study)

Imsaad Muftah Assalahi [*1], Hamad Heebah Hasan [2]
[1] Higher Institute of Engineering Technologies, Bani Walid, Libya
[2] Higher Institute of Science and Technology, Harawa, Libya

إطار تنبؤي لإخفاق مشاريع البنية التحتية في ليبيا: دمج مؤشرات عدم الاستقرار والتعلّم الآلي الزمني والإنذار المبكر القابل للتفسير (دراسة إثبات مفهوم قائمة على المحاكاة)

امساعد مفتاح الصلاحي [*1]، حمد هيبة حسن هيبة [2]
[1] المعهد العالي للتقنيات الهندسية، بني وليد، ليبيا
[2] المعهد العالي للعلوم والتقنية، هراوة، ليبيا

[*]Corresponding author: aalslahe@yahoo.com

Received: August 04, 2026	Accepted: October 16, 2026	Published: October 29, 2026

Abstract:
Projects in fragile settings face political, financial, institutional, procurement, and control disruptions. This article proposes a simulation-informed framework to estimate six-month infrastructure-project failure risk in Libya using instability, financing, approval delays, organizational capacity, and project-performance signals. No real Libyan project data were available; therefore, this is a proof of concept, not a field-validated study. A synthetic longitudinal dataset (seed 20260902) covered 420 projects and 8,631 project-month records; 324 projects were used for training and 96 for temporal testing. Four classifiers were compared. In this simulation only, logistic regression achieved the highest ROC-AUC (0.700), with PR-AUC 0.381, recall 0.567, specificity 0.709, and Brier score 0.204. SHAP highlighted instability, progress gap, procurement and payment delays, financing stability, and management capacity; these are model attributions, not causal effects or evidence of Libyan field performance. The proposed Project Failure Risk Index uses normalized variables and training-derived weights. The framework offers a transparent architecture for future validation with Libyan data. Its real-world accuracy and decision value remain unknown; field validation, calibration, and governance review are prerequisites for deployment.

Keywords: Infrastructure project failure, Libya, Temporal prediction, Machine learning, Explainable AI, SHAP, Early warning.

الملخص:
تواجه مشاريع البنية التحتية في البيئات الهشّة مخاطر سياسية ومالية ومؤسسية وتعاقدية. وتقترح الدراسة إطارًا تنبؤيًا قائمًا على المحاكاة لتقدير احتمال إخفاق المشروع في ليبيا خلال ستة أشهر، بدمج مؤشرات عدم الاستقرار والتمويل وتأخر الموافقات والقدرات التنظيمية وأداء المشروع. وبسبب غياب بيانات فعلية عن المشاريع الليبية لا تمثل الدراسة بحثًا ميدانيًا، بل تستخدم بيانات طولية اصطناعية مولدة بالبذرة 20260902 شملت 420 مشروعًا و8,631 سجلًا شهريًا. خُصص 324 مشروعًا للتدريب و96 للاختبار الزمني. وقورنت أربعة نماذج؛ وفي المحاكاة وحدها، حقق الانحدار اللوجستي أداءً: ROC-AUC قدره 0.700، وPR-AUC قدره 0.381، واستدعاء 0.567، ونوعية 0.709، ودرجة Brier 0.204. وأبرز تحليل SHAP عدم الاستقرار وفجوة التقدم وتأخر المشتريات والدفع واستقرار التمويل والقدرة الإدارية بوصفها أهم متغيرات تنبؤ النموذج؛ ولا تثبت هذه الإسنادات السببية أو تمثل الواقع الليبي. ويقترح البحث مؤشرًا

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

لمخاطر الإخفاق بأوزان التدريب، وهيكلية وبروتوكولًا للتحقق مستقبلًا. ولا يُعتمد الإطار تشغيليًا قبل التحقق الميداني والمعايرة ومراجعة الحوكمة.

الكلمات المفتاحية: إخفاق مشاريع البنية التحتية، ليبيا، التنبؤ الزمني، التعلّم الآلي، الذكاء الاصطناعي القابل للتفسير، شاب (SHAP)، الإنذار المبكر.

1 Introduction

Infrastructure delivery in fragile and conflict-affected environments is shaped by interdependent risks that extend beyond the boundaries of the project organization. Funding interruptions, institutional fragmentation, delayed approvals, procurement disruption, exchange-rate instability, security constraints, and weak project controls may combine to produce schedule slippage, cost escalation, suspension, abandonment, or termination. Libya provides an important setting for studying these interactions because infrastructure projects operate within a political and institutional environment where instability can affect public finance, procurement continuity, supply chains, administrative decision making, and contractor performance. The problem is not merely geographical. Libya represents a theoretically relevant case of infrastructure delivery under fragile institutional conditions, where macro-level disturbances can propagate into project-level failure mechanisms.

Existing Libyan construction research has made valuable contributions by identifying and prioritizing risk factors. Studies have examined risk-management practices and project performance in the Libyan construction industry [1], risk management under political unrest [2], cost-risk appraisal [3], and political and financial risk prioritization using fuzzy analytic hierarchy process methods [4]. These studies are important because they document the salience of political, financial, and managerial risks. Yet they mainly address risk identification, risk assessment, or risk ranking. They do not establish a temporally validated predictive system that uses information available at a given month to estimate whether a project will fail within a defined future horizon.

This distinction is central. **Risk identification** asks what hazards or uncertainties exist. **Risk ranking** orders factors by perceived or analytically derived importance. **Risk assessment** evaluates probability and consequence, often for planning or prioritization. **Prediction** estimates a future outcome from information available at a specified time, typically with out-of-sample validation. **Early warning** translates prediction into timely decisions, thresholds, escalation rules, and intervention pathways. A ranked list of political and financial risks may guide managerial attention, but it does not automatically provide a calibrated probability that a specific project will fail within the next six months, nor does it specify the decision threshold at which a ministry, owner, consultant, or contractor should intervene.

Internationally, construction-management research has increasingly adopted predictive analytics for delay, cost overrun, and project performance. Machine-learning models have been used for construction delay prediction [5, 6], cost-overrun risk prediction [7, 8], capital project failure prediction [9], and broader construction risk assessment [10, 11, 12]. Early-warning systems have also been developed in construction-related domains, including metro safety risk warning [13] and earthquake-enabled smart infrastructure responses [14]. However, fewer studies integrate macro-political and economic instability indicators with project-month administrative records in a temporally structured failure-prediction system for fragile infrastructure contexts. This creates a methodological and practical gap: infrastructure owners need a way to combine contextual instability and project-control signals into an explainable early-warning process.

This article responds to that gap by developing a predictive framework for infrastructure-project failure in Libya. The article is deliberately framed as a **simulation-informed proof-of-concept** because no real Libyan project-level dataset was supplied. Therefore, every numerical result in this article is a simulation output. Synthetic observations are not presented as field evidence. The simulation demonstrates how the proposed architecture, modelling workflow, validation logic, explainability tools, and Project Failure Risk Index (PFRI) could operate once real project records are available. Field validation is pending and is treated as an essential requirement before any operational deployment.

The article makes five contributions. First, it provides a parsimonious theoretical framework linking macro instability, institutional and financing disruption, procurement and decision delays, project-performance deterioration, and future failure. Second, it reconstructs the research design as a temporal prediction problem rather than a cross-sectional risk-ranking exercise. Third, it proposes a longitudinal data architecture for Libyan infrastructure projects, including administrative records, monthly project-control data, project characteristics, contractor and owner attributes, and contextual indicators. Fourth, it demonstrates, using explicitly synthetic simulation outputs, how logistic regression and machine-learning models can be compared using discrimination, precision–recall performance, calibration, and decision-oriented measures. Fifth, it rejects arbitrary PFRI weights and proposes a transparent, empirically derived, calibration-based index for early warning.

The remainder of the article is organized as follows. Section 2 reviews related literature. Section 3 develops the theoretical and conceptual framework. Section 4 states research questions and field-validation hypotheses. Section 5 presents the methodology, including data architecture, variables, failure definition, modelling, validation, explainability, PFRI development, and robustness procedures. Section 6 reports simulation-only

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

results. Section 7 discusses interpretation and implications. Sections 8 and 9 set out theoretical, practical, and policy implications. Section 10 identifies limitations and future research requirements. Section 11 concludes.

2 Literature Review

Research on construction-project risk has evolved from qualitative identification and ranking toward data-driven prediction and decision support. In the Libyan context, available studies have emphasized the importance of political instability, financial constraints, weak risk-management practices, and cost uncertainty. Hossen [3] addressed cost-risk appraisal in Libyan construction projects using a project-risk-management lens. Nasreddine Ali Algremazy et al. [1] examined the effect of risk-management practices on construction project performance. Dwa [2] considered risk management during political unrest, while Shetwan et al. [4] used fuzzy analytic hierarchy process methods to prioritize political and financial drivers in a fragile environment. These studies help establish the relevance of political and financial risk factors in Libya, but they do not provide a temporally validated predictive model using project-month records and future failure outcomes.

A broader international literature has developed predictive approaches for construction delay, cost, and performance. Gondia et al. [5] applied machine-learning algorithms to construction-project delay-risk prediction. Plebankiewicz and Wieczorek [7, 8] developed models for cost-overrun risk prediction and adapted such models to facility type. Chen [9] examined performance measurement and prediction of capital project failure, showing the value of measurable project-performance indicators for anticipating adverse outcomes. Heravi et al. [41] further linked project-risk-management maturity and organizational learning with power-plant project success, while Hosny et al. [42] demonstrated the relevance of claims prediction and decision awareness. Andrade et al. [43] used real schedule data to compare forecasting capabilities, reinforcing the importance of objective project-control data. Delay-analysis research has also refined methods for diagnosing schedule disruption [15, 16].

Machine learning has attracted increasing attention in construction engineering and management. Random forests [17], support vector machines [18], gradient boosting approaches such as XGBoost [19], and automated machine-learning pipelines [20] have been used or proposed for project cost, duration, risk, and performance prediction. Recent studies discuss ensemble learning for construction risk assessment [21], neural-network-based construction-risk prediction [22], and comparative artificial-intelligence applications in construction project management [23], infrastructure risk management with machine learning and optimization [11], and AI integration pathways for construction risk management [12]. These developments are relevant, but predictive modelling must be evaluated under appropriate validation designs. Leakage-free modelling requires strict separation of information used for prediction from information that becomes available only after the prediction date. The importance of calibration and decision-curve analysis has been emphasized in prediction research [24, 25, 26, 27]. Machine-learning approaches have also been applied to improve cost and duration forecasts in green building projects [28].

Prediction should also be distinguished from explanation. Shmueli [44] argues that explanatory modelling and predictive modelling pursue different objectives. In the present article, explanatory theory is used to specify plausible pathways and variables, whereas predictive modelling is used to estimate future failure risk from available information. A model may predict well without identifying causal effects; conversely, a theoretically meaningful factor may not improve out-of-sample prediction. This distinction is especially important in fragile contexts where political and institutional variables may be correlated with many unobserved conditions.

Explainable artificial intelligence is increasingly relevant in construction-management applications because project stakeholders require actionable and auditable explanations. SHAP methods provide local and global model-attribution tools for understanding feature contributions in fitted models [29]. Recent reviews identify explainable AI as an emerging area in construction engineering and management [30]. However, model explanations should not be interpreted as causal effects unless supported by a causal identification design. In an early-warning setting, SHAP values can indicate which variables contributed most to a model's risk estimate for a project-month, helping project-control teams decide what to investigate.

Organizational resilience and dynamic capabilities provide a complementary theoretical basis. Construction projects require the ability to absorb disruption, reconfigure resources, coordinate stakeholders, and adapt to changing constraints. Recent studies examine resilience in prefabricated construction projects [31], employee and organizational resilience in construction projects [32], organizational learning and resilience [33], social capital and contractor resilience [34], and network embeddedness and safety resilience [35]. In fragile contexts, resilience may determine whether macro instability becomes a manageable disturbance or escalates into project failure.

Fragile and conflict-affected settings also require special attention to institutional and macroeconomic conditions. The IMF's strategy for fragile and conflict-affected states emphasizes the importance of institutional capacity and macro-fiscal constraints in such environments [36]. Infrastructure and energy planning in fragile settings must account for conflict, uncertainty, and institutional limitations [37]. For Libyan infrastructure projects, this means that project-level prediction should not rely solely on internal schedule and cost metrics. It

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

should integrate contextual indicators that may disrupt finance, approvals, security, procurement, and supply chains.

Table 1 summarizes the positioning of the present framework relative to selected prior literature.

Table 1 ***Literature comparison and positioning***

Stream	Representative studies	Main contribution	Limitation addressed by this article
Libyan construction risk	Hossen [3]; Nasreddine Ali Algremazy et al. [1]; Dwa [2]; Shetwan et al. [4]	Identifies, appraises, or ranks political, financial, managerial, and cost risks	Limited temporal prediction of future failure using project-month administrative data
Construction delay and cost prediction	Gondia et al. [5]; Plebankiewicz and Wieczorek [7, 8]; Kryukov [6]; Ottaviani et al. [20]	Applies predictive analytics to delay, cost, and forecasting	Often not focused on fragile macro-instability pathways or failure early warning
Capital project failure and performance measurement	Chen [9]; Andrade et al. [43]	Links performance metrics to failure or forecasting	Needs integration with contextual instability and operational early-warning thresholds
Early-warning systems	Ding and Zhou [13]; Lin et al. [14]	Demonstrates warning logic in construction-related systems	Requires adaptation to infrastructure-project failure governance
Explainable AI in construction	Lundberg et al. [29]; Naghdi et al. [30]; Zawaideh [45]	Provides model-attribution tools	Needs careful distinction between attribution and causality
Decision-analytic prediction	Vickers and Elkin [26]; Vickers et al. [27]; Rousson and Zumbrunn [25]; Saito and Rehmsmeier [39]	Emphasizes calibration, PR-AUC, and decision consequences	Rarely integrated into Libyan infrastructure early-warning protocols

3 Theoretical and Conceptual Framework

The proposed framework combines three perspectives: systems theory, organizational resilience and dynamic capabilities, and early-warning logic. The combination is intentionally parsimonious. Systems theory is used because infrastructure projects are embedded in interacting political, institutional, financial, supply-chain, and organizational subsystems. Organizational resilience and dynamic capabilities are used because project actors differ in their ability to absorb and adapt to disruption. Early-warning logic is used because the practical objective is not merely to describe risk but to trigger timely intervention before irreversible failure.

The conceptual pathway is as follows:

Macro instability → funding, institutional, and procurement disruption → decision and execution problems → performance deterioration → future project failure.

Macro instability may include political uncertainty, conflict or security disruption, inflation, exchange-rate volatility, fiscal stress, and institutional fragmentation. These conditions may affect project financing, payment regularity, administrative approvals, procurement reliability, imported-material availability, and contractor liquidity. Disruptions then appear at the project level as delayed payments, slow decisions, procurement delays, change-order accumulation, claims, reduced site productivity, and management overload. Over time, these execution problems produce progress gaps, schedule variance, cost variance, prolonged suspension, or abandonment.

The framework does not claim that these pathways are proven causal effects in Libya. In the absence of field data and a causal identification design, the relationships remain theoretically grounded explanatory propositions. Prediction is a different task: the model estimates whether a project will fail within a future horizon using information available at the prediction date. Predictive importance does not establish causal influence, and causal relevance does not guarantee predictive value.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Figure 1. Conceptual framework.

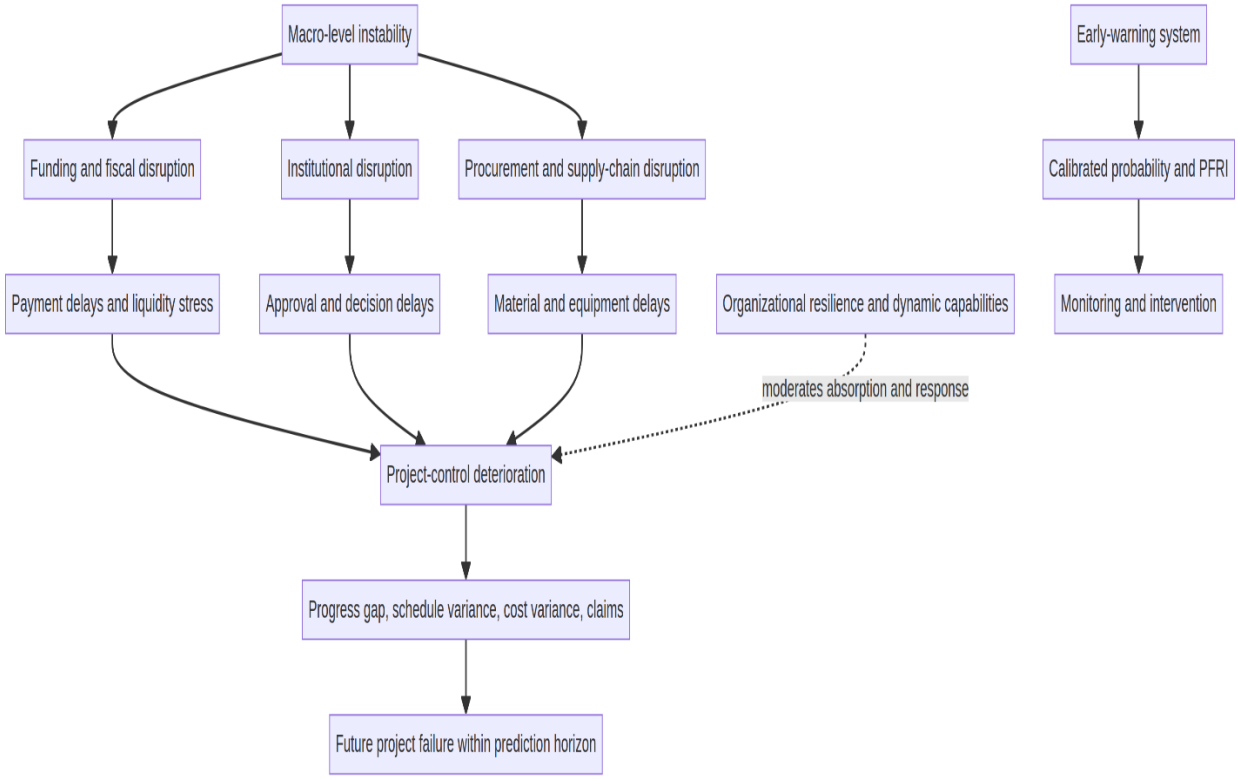


Figure 1 presents the conceptual framework.

Note. The framework is theoretical and does not establish causal effects.

Macro instability indicators affect financing stability, institutional stability, procurement continuity, and decision speed. These intermediate disruptions influence project execution, including progress gap, payment delay, approval delay, procurement delay, cost variance, schedule variance, management capacity, supply stability, change orders, and claims. Deteriorating project performance increases the predicted risk of future failure. Organizational resilience and dynamic capabilities moderate the extent to which disruption is absorbed or escalates. The early-warning system converts predicted probabilities into calibrated decision thresholds and escalation actions.

The framework also supports the design of an explainable early-warning system. If the model identifies a project-month as high risk, decision makers need to know whether the warning is driven mainly by macro instability, payment delay, procurement delay, progress gap, or management-capacity indicators. SHAP-based attribution can support such review, provided it is treated as model explanation rather than causal diagnosis.

4 Research Questions and Hypotheses

Because no real Libyan project-level dataset was supplied, the hypotheses below are formulated for future field validation. They are not tested empirically in this article. The simulation reported later demonstrates the analytical pipeline only.

RQ1. How can macro-instability, institutional, financing, procurement, and project-performance indicators be organized into a temporally structured framework for predicting infrastructure-project failure in Libya?

RQ2. When real project-month data become available, how accurately can interpretable statistical models and machine-learning models predict failure within a six-month future horizon?

RQ3. Which variables contribute most to model predictions in an explainable early-warning system, and how can these attributions support managerial review without being misinterpreted as causal effects?

RQ4. How can a Project Failure Risk Index be constructed without arbitrary weights and linked to calibrated intervention thresholds?

RQ5. What validation, robustness, and governance procedures are required before deploying such a system in Libyan infrastructure agencies or project-control units?

For future field validation, the following hypotheses may be tested as predictive and associational propositions:

H1. Higher macro instability observed before the prediction date is associated with a higher probability of infrastructure-project failure within the future prediction horizon.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

H2. Lower financing stability and longer payment delays observed before the prediction date are associated with a higher probability of future project failure.
H3. Greater procurement delay, approval delay, and lower decision speed observed before the prediction date are associated with a higher probability of future project failure.
H4. Larger progress gaps, schedule variance, and cost variance observed before the prediction date are associated with a higher probability of future project failure.
H5. Higher management capacity, institutional stability, supply stability, and organizational resilience indicators are associated with a lower probability of future project failure.
H6. A temporally validated predictive model integrating macro and project-level indicators will outperform models using only static project characteristics, subject to field validation.
These hypotheses are not causal claims. Testing them would require observed project outcomes, properly timed predictors, and sensitivity analyses for confounding, measurement error, and alternative failure definitions.

5 Methodology

5.1 Research Design

The study is designed as a simulation-informed proof-of-concept and field-validation protocol. It specifies a longitudinal prediction problem in which project-month observations are used to predict whether an infrastructure project will fail within a future six-month horizon. The design requires predictors measured at or before time (t), and an outcome measured over the interval (t+1) to (t+6). No post-(t) information may be used as a predictor.
The article contains two distinct components:

1. **Proposed field study design:** a data architecture, sampling strategy, failure definition, modelling workflow, validation plan, explainability protocol, PFRI construction method, and robustness plan for real Libyan infrastructure data.
2. **Synthetic proof-of-concept:** a simulation using seed 20260902, 420 synthetic projects, 8,631 project-month rows, 324 training projects, and 96 temporally held-out test projects.

All numerical results in Section 6 are synthetic simulation outputs. They are not empirical findings about Libya.

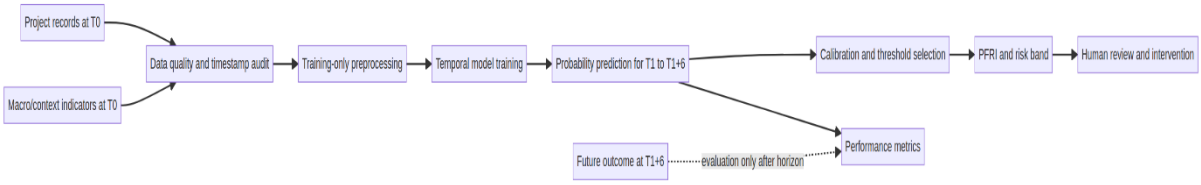


Figure 2. Data architecture and temporal prediction workflow.
Note. T0 denotes information available at prediction time; the future outcome is evaluated only after the outcome window closes.

Project-level master records are linked to monthly progress, payment, procurement, approval, change-order, claim, and performance records. Contractor, consultant, owner, project type, location, funding source, and contextual instability indicators are merged by project and month. Prediction uses information available up to month (t) to estimate failure within months (t+1) to (t+6). Training uses earlier projects or periods; testing uses later projects or periods.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

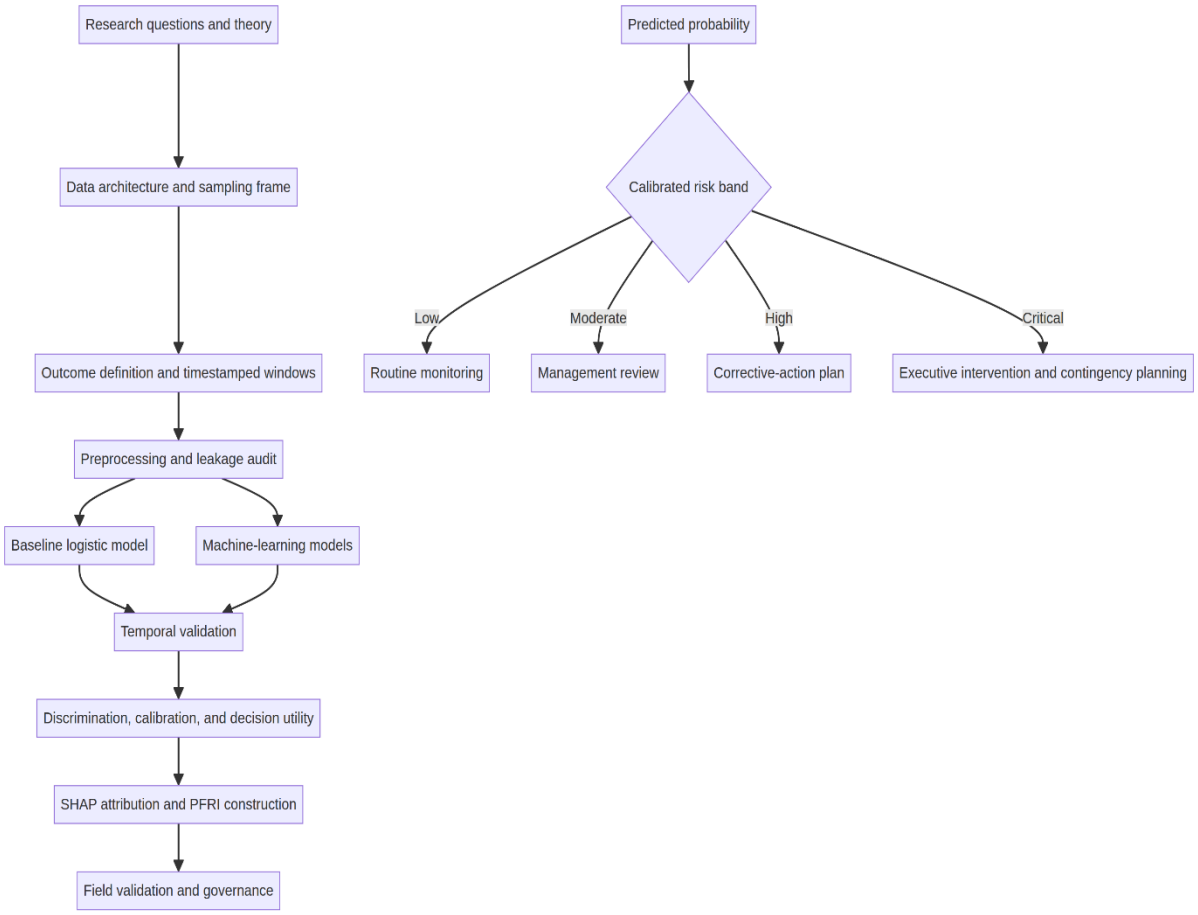


Figure 3. Methodology.

Note. The final deployment stage requires field validation and governance safeguards.

The workflow includes data acquisition, eligibility screening, temporal indexing, variable construction, missing-data diagnosis, leakage audit, training/test split, preprocessing, model training, threshold selection, calibration assessment, SHAP explanation, PFRI construction, robustness checks, and early-warning deployment protocol.

5.2 Study Context

The intended field context is Libyan infrastructure delivery, including public and private projects in sectors such as transport, roads, bridges, water, wastewater, energy, public buildings, and municipal infrastructure. Libya is relevant because infrastructure delivery is exposed to political, fiscal, institutional, and security conditions that may disrupt project continuity. Contextual fragility matters for prediction because project distress can originate outside the construction site, including delayed budget releases, exchange-rate volatility, interrupted imports, institutional changes, approval bottlenecks, or security incidents.

The study does not claim to have collected Libyan field data. Instead, it identifies the data needed for a valid field study.

5.3 Data Sources

A field implementation should combine the following sources:

- **Project-level administrative records:** contract value, contract type, planned duration, award date, sector, location, client type, funding source, procurement route, contractor classification, consultant role, and ownership structure.
- **Monthly project-control records:** planned progress, actual progress, earned value indicators where available, schedule variance, cost variance, site productivity, payment certificates, payment delays, approved variations, pending change orders, claims, disputes, suspensions, and completion status.
- **Procurement and approval records:** tendering delays, material procurement lead times, import delays, approval cycle time, design-review delay, permit delay, and decision turnaround.
- **Organizational characteristics:** contractor experience, financial capacity indicators, management capacity, consultant capacity, owner decision capacity, prior performance, and staffing continuity.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

- **Contextual indicators:** political instability, security events, inflation, exchange-rate volatility, fiscal conditions, public-expenditure constraints, fuel or material shortages, and institutional stability.
- **Outcome records:** cancellation, termination, abandonment, prolonged suspension, severe cost or schedule overrun, or failure to achieve planned completion within specified thresholds.

Missing variables should be recorded as data-collection requirements rather than silently omitted. If management capacity or institutional stability cannot be obtained from administrative sources, structured expert assessment may be used as a supplementary instrument, but it must undergo expert review, pilot testing, reliability assessment, and construct-validation procedures.

5.4 Sampling

No achieved sample is claimed. A field study should use a stratified sampling design or census approach where records permit. The target population is infrastructure projects initiated, active, completed, suspended, abandoned, terminated, or cancelled within a defined observation period in Libya.

Table 2 ***Target sampling plan, not fabricated sample characteristics***

Dimension	Proposed field sampling requirement
Target population	Libyan infrastructure projects with traceable administrative and monthly project-control records
Sectors	Roads, bridges, transport, water, wastewater, energy, public buildings, municipal infrastructure
Ownership	Public, public–private, and private infrastructure projects where records are accessible
Geography	Stratification across western, eastern, southern, and central regions, subject to data access and safety
Project size	Small, medium, large, and megaprojects based on contract value and duration thresholds defined before analysis
Project status	Completed, active, delayed, suspended, abandoned, cancelled, and terminated projects
Inclusion criteria	Known start date, contract baseline, monthly or periodic progress records, identifiable outcome status
Exclusion criteria	Projects lacking baseline scope or outcome status; duplicate records; records with unverifiable project identity
Failure representation	Oversampling or targeted retrieval of failed projects may be necessary for model training, but prevalence must be reweighted or evaluated transparently
Validation sample	Later projects or later time periods reserved for temporal testing; external validation by agency or region when feasible

A defensible target sample size should be based on expected failure prevalence, number of candidate predictors, temporal clustering, model complexity, and the need for an adequate number of failure events. The target should be specified before data extraction; it should not be replaced by a convenience sample without disclosure.

5.5 Variables

Variables are organized into five groups: contextual instability, institutional and financing conditions, procurement and decision processes, project-performance indicators, and organizational capacity.

Table 3 ***Variable definitions and operationalization***

Variable	Level	Risk orientation	Proposed operational definition

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Variable	Level	Risk orientation	Proposed operational definition
macro_instability	Context-month	Higher = higher risk	Composite indicator of political, security, fiscal, inflation, and exchange-rate instability before month (t)
institutional_stability	Context/owner-month	Lower = higher risk	Continuity and reliability of administrative approvals, institutional authority, and governance processes
financing_stability	Project-month	Lower = higher risk	Regularity and predictability of project funding and budget releases
payment_delay	Project-month	Higher = higher risk	Days or months between certified payment and actual payment
procurement_delay	Project-month	Higher = higher risk	Delay in procuring materials, equipment, subcontractors, or imported inputs relative to plan
approval_delay	Project-month	Higher = higher risk	Delay in owner, consultant, regulatory, or design approvals
decision_speed	Project-month	Lower = higher risk	Average time required for formal project decisions, requests for information, and approvals
progress_gap	Project-month	Higher = higher risk	Planned cumulative progress minus actual cumulative progress at month (t)
schedule_variance	Project-month	Higher adverse variance = higher risk	Schedule deviation relative to baseline or approved update
cost_variance	Project-month	Higher adverse variance = higher risk	Cost deviation relative to budget or earned-value baseline
supply_stability	Project-month	Lower = higher risk	Reliability of material, equipment, labor, and logistics supply
management_capacity	Project/organization	Lower = higher risk	Contractor/consultant/owner capability, staffing, planning, and coordination capacity
change_order_intensity	Project-month	Higher = higher risk	Number or value of change orders relative to contract value
claim_intensity	Project-month	Higher = higher risk	Number or value of claims and disputes relative to project size
project_size	Project	Context-dependent	Contract value, duration, and technical complexity
project_type	Project	Control	Sector and facility type
location	Project	Control	Region, municipality, and exposure to contextual disruption
funding_source	Project	Control	Public budget, donor, private, public–private, or mixed source
failure_6m	Outcome	1 = failure	Failure event occurring within months (t+1) to (t+6) according to the operational definition

All time-varying predictors must be lagged or measured at the prediction date. Outcome-window information must not enter the feature set.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

5.6 Operational Failure Definition

A rigorous field study should define infrastructure-project failure using measurable administrative outcomes. The primary binary outcome is:

[failure_6m = 1]

if, within the six months after prediction month (t), the project experiences at least one of the following:

- formal cancellation or termination;
- abandonment or documented non-continuation;
- prolonged suspension exceeding a pre-specified duration threshold;
- failure to achieve planned completion within a severe delay threshold;
- severe cost overrun beyond a pre-specified and contractually meaningful threshold;
- failure to resume after funding, security, or institutional stoppage.

Thresholds should be defined before analysis in consultation with owners, ministries, contractors, consultants, and auditors. Sensitivity analyses should test alternative definitions, such as stricter failure based only on cancellation, termination, abandonment, or prolonged suspension, and broader failure including severe schedule or cost overrun.

5.7 Preprocessing

The preprocessing protocol should include:

- project identity resolution and duplicate removal;
- temporal indexing by project-month;
- verification that predictor values are available at or before the prediction month;
- coding of categorical variables using training-data procedures only;
- imputation of missing values using methods fitted within training folds;
- missingness indicators where missingness itself may carry operational meaning;
- normalization or standardization for models requiring scaling;
- multicollinearity diagnostics for interpretable models;
- outlier review to distinguish data errors from genuine extreme project distress;
- class-imbalance handling using class weights or fold-contained resampling only.

SMOTE or related methods, if used, must be applied only within training folds and never before temporal splitting, to avoid leakage [38].

5.8 Predictive Modelling

The prediction task is binary classification: failure within six months. Four model families are considered:

- **Logistic regression:** interpretable baseline suitable for calibrated probability estimation and coefficient inspection, subject to assumptions and multicollinearity diagnostics.
- **Random forest:** nonlinear ensemble method capable of modelling interactions and nonlinearities [17].
- **Histogram gradient boosting:** efficient gradient-boosted tree method suitable for tabular data; conceptually related to boosting approaches widely used in predictive modelling.
- **Support vector machine:** margin-based classifier with probability calibration when probabilistic outputs are required [18].

Neural networks are not proposed as a default because field datasets in this domain may be moderate in size and heterogeneous. They should be considered only if sample size, feature richness, and validation design justify their use.

5.9 Validation

Validation must be temporal. Earlier projects or earlier time periods should be used for model development, while later projects or later time periods should be reserved for testing. Cross-validation for tuning should occur within the training data only. Test data must remain untouched until final evaluation.

Evaluation metrics should include:

- ROC-AUC for discrimination across thresholds;
- PR-AUC because failure may be imbalanced and precision–recall curves are informative under class imbalance [39];
- precision;
- recall/sensitivity;
- F1 score;
- specificity;
- Brier score;
- calibration curve;
- confusion matrix at selected thresholds;
- decision-curve analysis where intervention consequences can be specified [26, 27].

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Thresholds should be selected using decision consequences. Missing a genuinely failing project may be more costly than investigating a false alarm, but false positives also consume scarce managerial and financial resources. Therefore, early-warning thresholds should be determined with stakeholder-defined costs and capacity constraints.

Table 4*Validation strategy*

Validation issue	Required procedure
Temporal leakage	Predictors must be observed at or before month (t); outcome measured only in (t+1) to (t+6)
Train/test split	Earlier projects or periods for training; later projects or periods for testing
Hyperparameter tuning	Nested or internal cross-validation within training data only
Calibration	Reliability curves, Brier score, recalibration if needed
Class imbalance	PR-AUC, recall, precision, class weighting, training-fold resampling only
Clustered observations	Project-level splitting to avoid the same project appearing in both train and test sets
External validity	Validation by region, sector, owner type, and later time period when data permit
Decision utility	Decision-curve analysis and comparison with treat-all, treat-none, and existing monitoring

5.10 Explainable AI

Explainability is required for adoption. The preferred approach is to apply SHAP to the best-performing fitted model, or to the selected operational model if interpretability and calibration justify a different choice. SHAP can provide:

- global feature importance;
- direction of feature effects;
- project-month-level explanations;
- dependence plots for nonlinear relationships;
- audit trails for warnings.

SHAP results must be interpreted as model-attribution patterns. They do not prove that changing a feature will causally reduce failure risk. For example, if macro_instability has high SHAP importance, the model relied heavily on that variable in the fitted prediction function; it does not prove that a specific political event caused a specific project to fail.

5.11 PFRI Development

The original use of arbitrary PFRI weights is rejected. A publication-quality PFRI should use transparent and reproducible weighting. The proposed index is:

[PFRI_{it} = \sum_{j=1}^{p} w_j z_{ijt}^{risk}]

where (PFRI_{it}) is the Project Failure Risk Index for project (i) at month (t), (z_{ijt}^{risk}) is a normalized risk-oriented predictor, and (w_j) is a non-negative weight derived from the training procedure such that (\sum_j w_j = 1).

Risk orientation means that higher normalized values always indicate higher risk. For protective variables such as financing_stability, management_capacity, institutional_stability, supply_stability, and decision_speed, the normalized value is reversed before aggregation.

Weights can be derived and compared using:

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

- **Equal weighting:** transparent but assumes all variables contribute equally.
- **Regression weighting:** based on absolute standardized logistic coefficients, subject to multicollinearity and model assumptions.
- **Permutation-derived weighting:** based on loss in predictive performance when each variable is permuted within the training procedure.
- **Model-importance weighting:** based on feature importance from tree-based models, with caution regarding bias.
- **SHAP-based weighting:** based on mean absolute SHAP values from the fitted model.
- **Expert-informed weighting:** useful when data are sparse, but must be documented, elicited systematically, and tested for sensitivity.

For the proposed empirical PFRI, permutation-derived weights are preferred because they are linked to predictive performance in the training procedure. SHAP-based weights are useful for interpretability and comparison but should not be treated as causal weights.

Risk bands should not use arbitrary universal cutoffs. In a field deployment, Low, Moderate, High, and Critical bands should be derived from calibration, stakeholder-defined intervention consequences, operational capacity, and decision-curve analysis. If provisional bands are used during piloting, they must be labelled provisional until real calibration is completed.

Table 5 ***PFRI construction***

Step	Procedure
1. Select variables	Use only predictors available before the prediction date
2. Orient risk	Reverse protective variables so higher values indicate higher risk
3. Normalize	Apply training-data normalization parameters
4. Estimate weights	Derive permutation-based weights within the training procedure
5. Aggregate	Compute weighted sum of normalized risk variables
6. Calibrate	Map PFRI scores to observed failure probabilities using field data
7. Define bands	Set Low/Moderate/High/Critical thresholds based on calibration and decision consequences
8. Validate	Test discrimination, calibration, decision utility, and subgroup stability
9. Monitor	Recalibrate periodically as conditions and data quality change

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

5.12 Robustness

A field study should include the robustness plan in **Table 6**.

Table 6 ***Robustness and sensitivity plan***

Domain	Robustness procedure
Failure definition	Test strict and broad failure definitions
Prediction horizon	Compare 3-, 6-, and 12-month horizons
Missing data	Diagnose missingness mechanisms; compare complete-case and imputed analyses
Multicollinearity	Variance inflation factors and correlation diagnostics for interpretable models
Model stability	Bootstrap intervals and temporal resampling where valid
Calibration	Recalibration by period, sector, and agency if needed
Subgroups	Sector, region, project size, funding source, public/private ownership
Thresholds	Sensitivity to false-negative and false-positive costs
Feature sets	Macro-only, project-only, and integrated macro-project models
Leakage	Formal audit of timestamped variables and outcome windows
Measurement bias	Compare administrative, consultant, contractor, and owner records where possible
Reporting bias	Examine whether failed projects have systematically poorer documentation
External validation	Test on later periods, different agencies, or withheld regions

6 Results

This section reports only **synthetic proof-of-concept simulation outputs**. The simulation does not constitute empirical evidence about Libyan infrastructure projects. It demonstrates the proposed analytical workflow under controlled synthetic assumptions. Field validation is pending.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The simulation used seed 20260902. The synthetic dataset contained 420 synthetic projects and 8,631 project-month rows. The temporal split assigned 324 projects to training and 96 projects to temporal testing. The outcome was simulated failure within a six-month future horizon. The simulated overall failure prevalence was 0.2128.

Table 7 ***Synthetic proof-of-concept descriptive summary***

Item	Simulation output
Random seed	20260902
Synthetic projects	420
Synthetic project-month rows	8,631
Training projects	324
Temporal test projects	96
Prediction horizon	Six-month future failure
Simulated overall failure prevalence	0.2128

Four models were compared: logistic regression, random forest, histogram gradient boosting, and support vector machine. The temporal test metrics are reported in **Table 8**. These are simulation metrics and must not be generalized to Libya.

Table 8 ***Synthetic proof-of-concept model comparison on temporal test data***

Model	ROC-AUC	PR-AUC	Precision	Recall	F1	Specificity	Sensitivity	Brier score
Logistic regression	0.700	0.381	0.308	0.567	0.399	0.709	0.567	0.204
Random forest	0.693	0.357	0.345	0.424	0.380	0.816	0.424	0.179
Histogram gradient boosting	0.692	0.352	0.457	0.137	0.211	0.963	0.137	0.141
Support vector machine	0.670	0.314	0.359	0.045	0.079	0.982	0.045	0.143

In this synthetic scenario, logistic regression had the highest ROC-AUC, at 0.700. This does not prove that logistic regression is superior for Libyan infrastructure data. It only indicates that, under the synthetic data-generating scenario used here, the logistic model ranked failure and non-failure observations slightly better than the other models by ROC-AUC. Random forest had a similar ROC-AUC of 0.693 and a lower Brier score of 0.179. Histogram gradient boosting and support vector machine had high specificity but very low recall/sensitivity in this simulation, indicating that they missed many simulated future failures at the selected threshold. In an early-warning application, such missed failures may be unacceptable if the cost of inaction is high.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

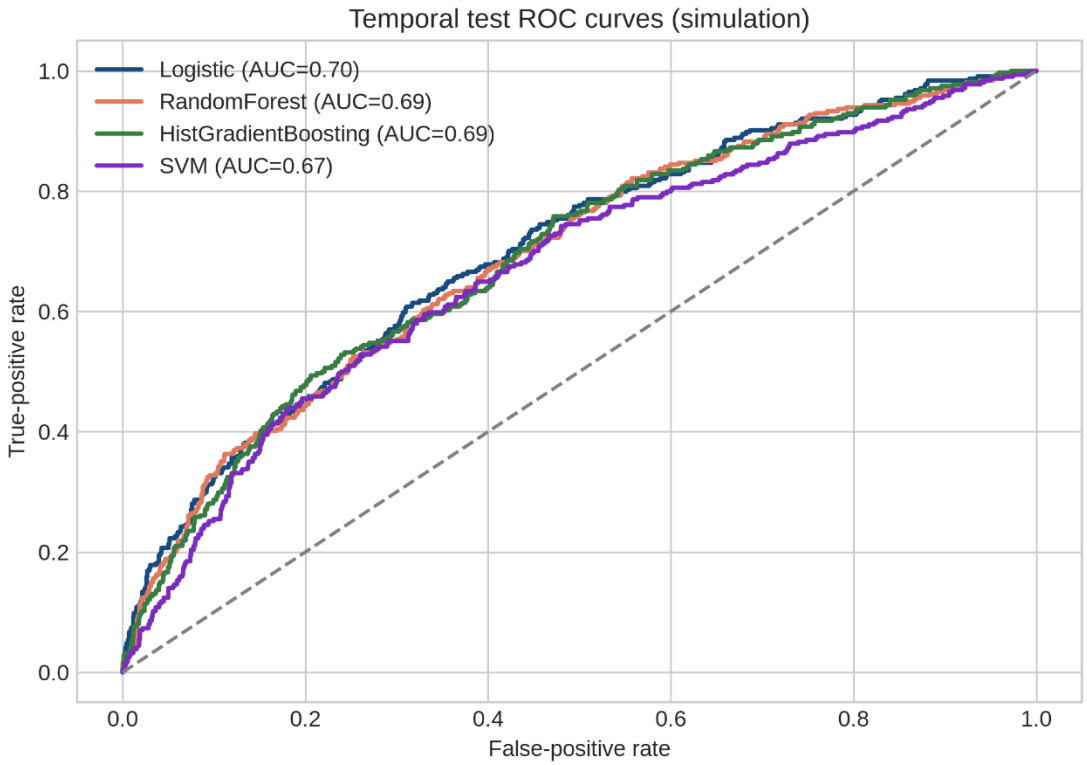


Figure 4. ROC curves.
The temporal test ROC curves are shown below. In this simulation only, logistic regression had the highest ROC-AUC.

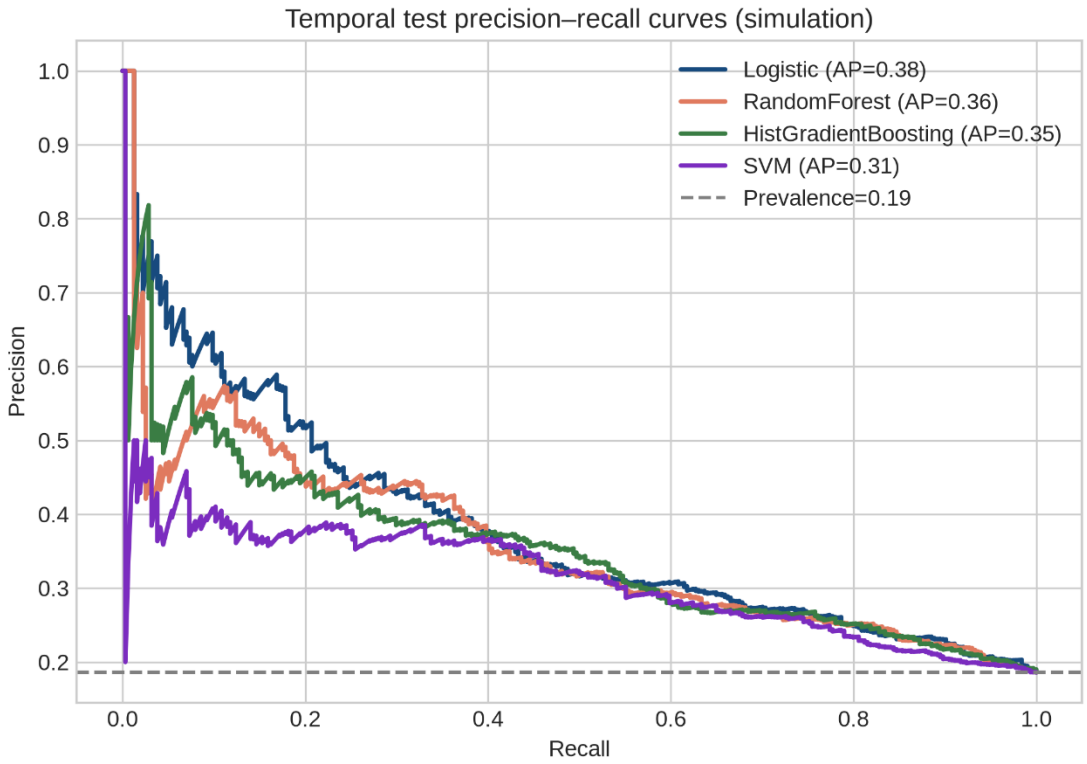


Figure 5. Precision–recall curves.
The precision–recall curves are emphasized because the simulated failure prevalence was 0.2128. Logistic regression had the highest PR-AUC in this synthetic scenario.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

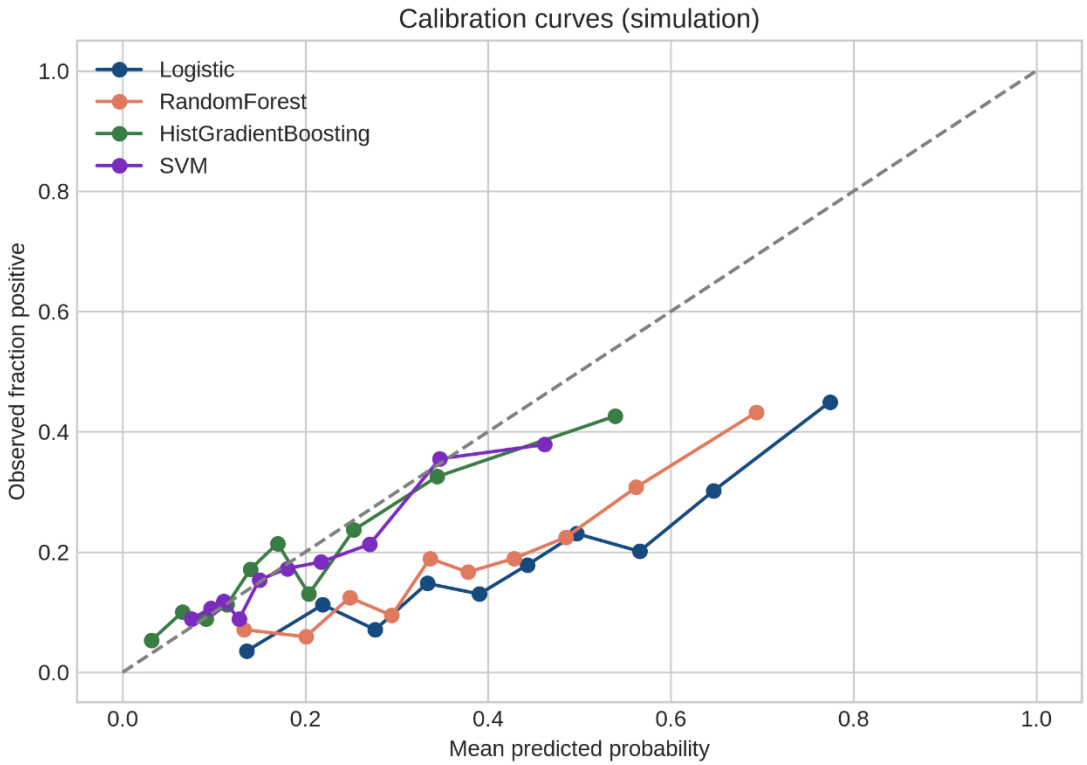


Figure 6. Calibration.

Calibration curves compare predicted probabilities with observed simulated failure frequencies. The Brier scores in Table 8 suggest differences in probability accuracy, but calibration plots and intercept/slope diagnostics are required before selecting an operational model. Because the underlying results are synthetic, calibration findings cannot be transferred to Libyan practice.

SHAP analysis was conducted for the simulation as a model-attribution exercise. The supplied mean absolute SHAP importance values are reported in **Table 9**.

Table 9 ***Synthetic proof-of-concept SHAP model-attribution patterns***

Rank	Feature	Mean absolute SHAP importance
1	macro_instability	0.4092
2	progress_gap	0.1745
3	procurement_delay	0.1650
4	financing_stability	0.1343
5	payment_delay	0.1266
6	management_capacity	0.1243
7	approval_delay	0.0923
8	cost_variance	0.0910
9	institutional_stability	0.0718
10	schedule_variance	0.0489
11	supply_stability	0.0410
12	decision_speed	0.0398

The leading simulated attribution was macro_instability, followed by progress_gap, procurement_delay, financing_stability, payment_delay, and management_capacity. These patterns are consistent with the conceptual logic that contextual instability and project-control deterioration may jointly signal future failure. However, these SHAP values are not causal effects. They show how the fitted model used synthetic variables to generate predictions.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

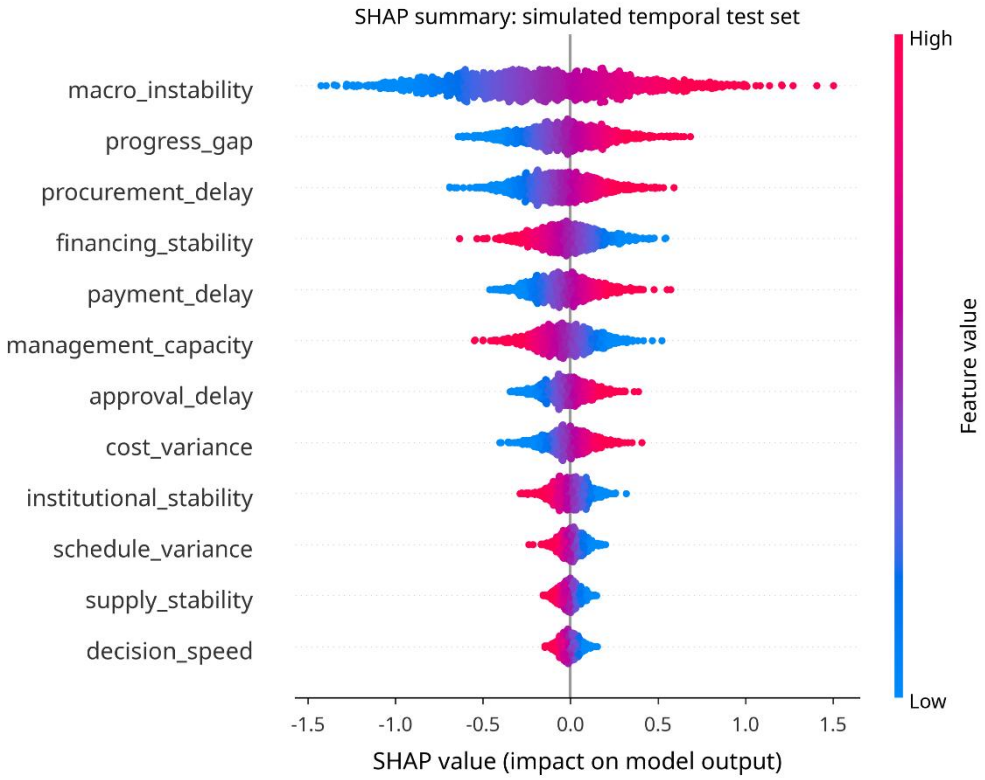


Figure 7. SHAP summary.

The SHAP summary displays feature importance and the direction of model contributions for the synthetic proof-of-concept model.

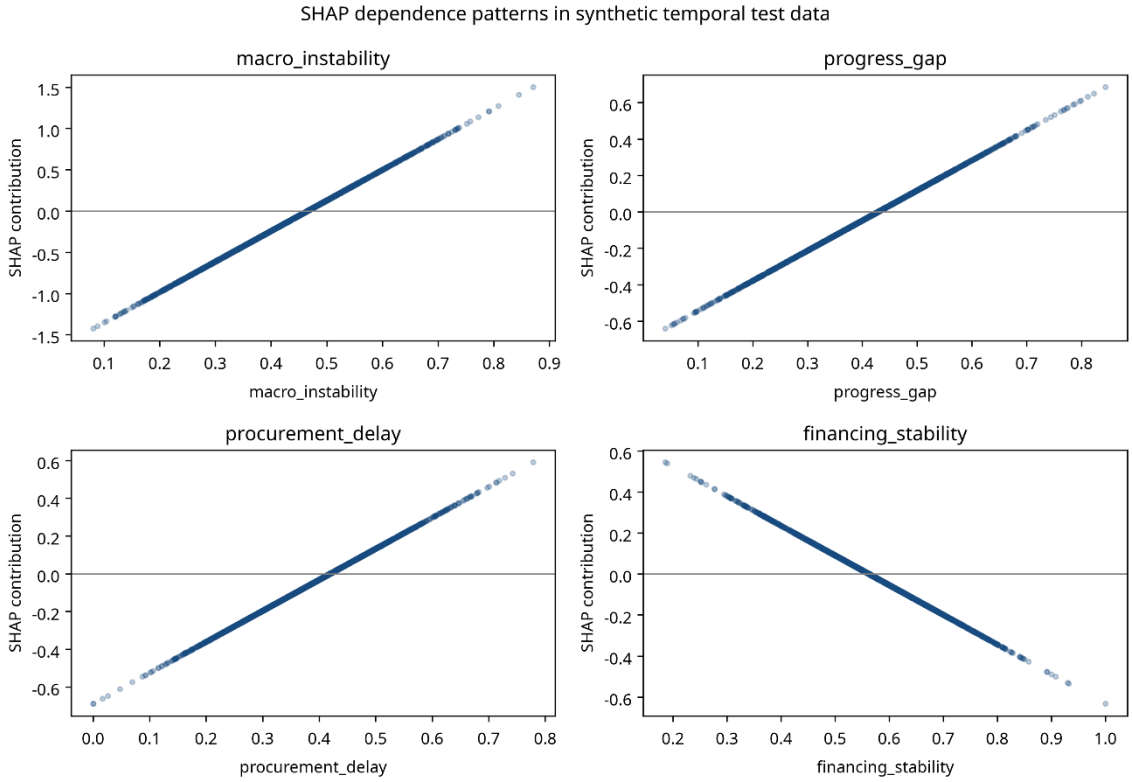


Figure 8. SHAP dependence and feature effects.

The dependence plots show how simulated feature values relate to SHAP contributions. They are synthetic model-attribution graphics, not empirical Libyan evidence.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The PFRI was reconstructed conceptually by using normalized risk-oriented variables and weights derived from the training procedure rather than arbitrary weights. The simulation demonstrates the feasibility of computing such an index, but no field-calibrated risk bands are available. **Figure 9** should therefore be read as a synthetic illustration only.

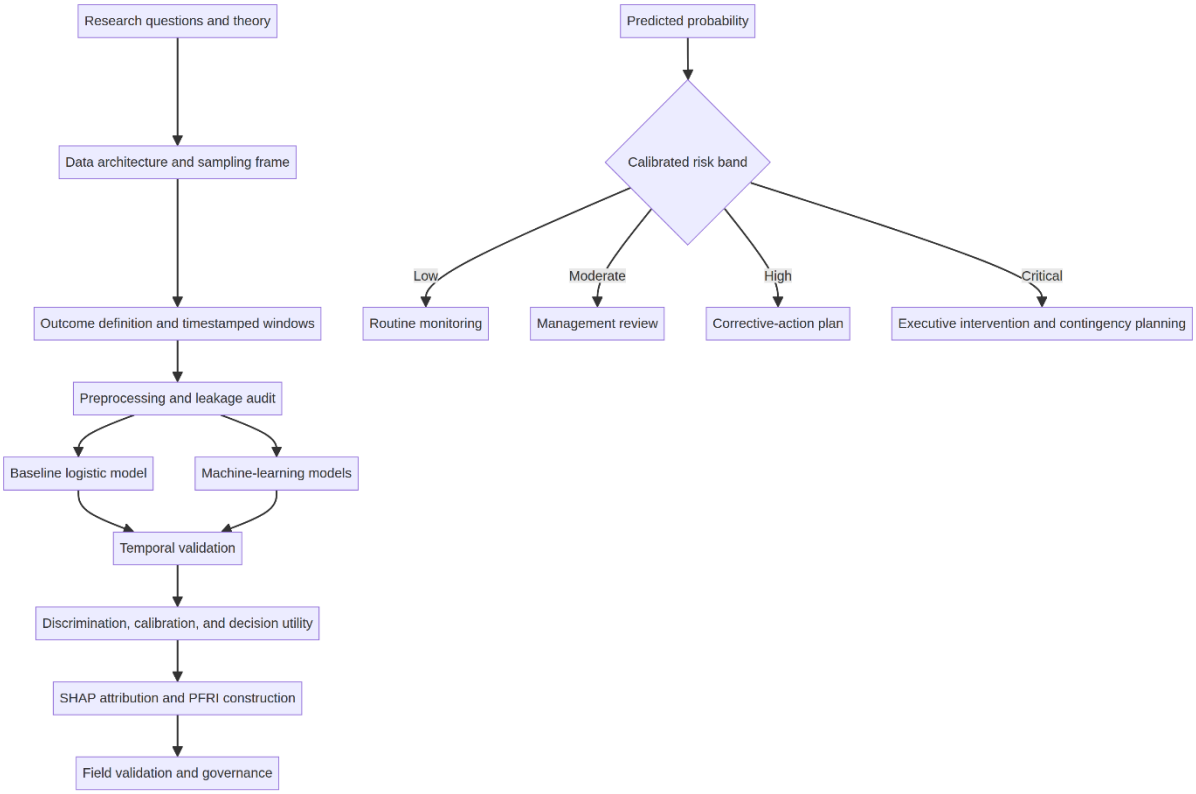


Figure 9. PFRI distribution and classification, synthetic proof-of-concept. The simulated PFRI distribution is shown below. No universal risk bands are claimed; any operational categories must be recalibrated with real project outcomes. **Figure 10. Early-warning decision workflow.** The operational workflow translates predicted probabilities and PFRI bands into monitoring, management review, corrective action, and executive intervention. Thresholds must be calibrated using field data and decision consequences.

7 Discussion

The study discovered, within a synthetic proof-of-concept only, that a temporally structured early-warning workflow can integrate macro-instability indicators, financing and institutional variables, procurement and approval delays, management capacity, and project-performance indicators to estimate future infrastructure-project failure risk. The simulation showed that the workflow can compare interpretable and machine-learning models, evaluate them using discrimination, precision–recall performance, recall, specificity, calibration-related metrics, and explain model predictions using SHAP. The most important finding is not that any reported model is empirically accurate for Libya. Rather, the finding is that a transparent and reproducible architecture can be specified for future field validation.

This matters because Libyan construction research has already shown that political, financial, and managerial risks are important [1, 2, 3, 4]. Yet a ministry or owner responsible for an active portfolio needs more than a ranked list of risks. It needs to know which projects are likely to fail within a decision horizon, which warning signals are driving the forecast, and what intervention threshold is justified. A predictive early-warning system can support portfolio triage, payment prioritization, procurement escalation, approval acceleration, contractor-support decisions, and executive review.

The simulation patterns are broadly compatible with prior studies. Macro instability and financing-related variables were prominent in the synthetic SHAP outputs, which aligns with Libyan studies emphasizing political and financial drivers. Progress gaps, procurement delays, payment delays, and approval delays are also consistent with construction delay and performance literatures that identify schedule and administrative disruptions as important project-control signals [5, 15, 16]. Cost variance and schedule variance are consistent with performance-measurement approaches to capital-project failure prediction [9, 40]. However, because the present numerical results are simulated, the comparison is conceptual rather than empirical.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The proposed mechanism is systemic. Macro instability does not affect project failure only through a single route. It may reduce fiscal reliability, delay public payments, interrupt imports, slow institutional decisions, and increase uncertainty for contractors and suppliers. These conditions may create liquidity stress, resource shortages, unresolved variations, claims, and weak site progress. Projects with stronger organizational resilience and dynamic capabilities may absorb disruptions through resource reallocation, stakeholder coordination, alternative procurement, and faster decision making. Projects with weak capacity may convert the same external disturbance into prolonged delay or failure. This mechanism is plausible but requires field validation.

The theoretical novelty is restrained. The article does not claim to introduce a universal theory of infrastructure failure. Its contribution is to connect systems theory, resilience and dynamic capabilities, and early-warning logic in a specific predictive architecture for fragile infrastructure contexts. Systems theory explains interdependence across macro and project subsystems. Resilience and dynamic capabilities explain why some project organizations absorb shocks better than others. Early-warning logic translates predictive estimates into operational action.

The methodological contribution is the reconstruction of the problem as a temporal prediction task. Many risk studies identify or rank risks using cross-sectional judgments. Such work is useful but cannot by itself establish predictive performance. A temporal model requires a prediction date, a future outcome window, leakage controls, project-level train/test separation, calibration assessment, and threshold selection based on consequences. The framework therefore moves from risk awareness toward auditable decision support.

The empirical contribution is deliberately limited. The simulation demonstrates feasibility and produces synthetic metrics, but it does not estimate real Libyan failure prevalence, real coefficients, real p-values, real model performance, or real intervention utility. The reported ROC-AUC, PR-AUC, Brier scores, SHAP values, and prevalence are simulation outputs. They should be used only to understand the proposed workflow and to plan field validation.

The practical usefulness lies in the proposed operating model. Owners and ministries could use the system to identify projects requiring closer monitoring. Consultants could use model explanations to focus review on procurement, approvals, progress gaps, or payment bottlenecks. Contractors could use the outputs to document emerging distress and request timely decisions. Project-control units could combine probability estimates and PFRI bands with existing dashboards. However, deployment should not occur before validation, calibration, and governance safeguards are in place.

The Libya-specific relevance is that the framework explicitly incorporates political, fiscal, institutional, procurement, and project-control conditions rather than treating projects as isolated technical systems. Its transferability is likely to be highest in settings with similar fragility, data constraints, and public-infrastructure governance challenges. Nevertheless, variables, thresholds, failure definitions, and calibration must be localized. A model trained in one country, agency, or period should not be transferred without external validation.

8 Theoretical Implications

The article offers three theoretical implications.

First, it clarifies the distinction between explanatory association and prediction in construction risk research. The proposed causal pathway is theoretically plausible, but the predictive model is evaluated by out-of-sample performance. This distinction responds to Shmueli's [44] argument that explanation and prediction are different scientific tasks. In construction management, this distinction is often blurred when ranked risks are interpreted as predictive evidence.

Second, the framework extends project-failure thinking beyond internal project controls. It conceptualizes infrastructure projects as open systems embedded in political, fiscal, institutional, and supply-chain environments. Macro instability may propagate through funding, approvals, procurement, and decision channels into performance deterioration. This systems view is particularly relevant in fragile contexts, where external shocks are not exceptional disturbances but recurring features of the delivery environment.

Third, the framework integrates organizational resilience and dynamic capabilities into early-warning logic. Resilience is not treated as an abstract virtue; it becomes measurable through capacity, decision speed, supply stability, learning, staffing continuity, and coordination indicators. Future field studies can test whether resilience-related variables improve prediction and whether they moderate the relationship between macro disruption and project deterioration.

9 Practical and Policy Implications

For infrastructure owners and ministries, the proposed framework supports portfolio-level triage. A calibrated early-warning system could identify projects with rising probability of six-month failure and allocate scarce managerial attention accordingly. High-risk projects could be reviewed for payment bottlenecks, procurement constraints, approval delays, security exposure, or contractor-capacity problems.

For consultants and project-control units, the framework provides an auditable structure for monthly risk review. Rather than relying only on narrative progress reports, teams could track normalized risk-oriented indicators and

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

investigate model explanations. SHAP outputs may help identify whether a warning is driven by progress gaps, procurement delay, payment delay, or contextual instability.

For contractors, a transparent warning system could support earlier communication of distress. If payment delay, approval delay, or procurement disruption is driving predicted failure risk, contractors can document the issue and request timely corrective action. This may reduce adversarial escalation if the system is governed fairly.

For policymakers, the framework highlights that project failure is not only a contractor-performance issue. Fiscal instability, delayed budget releases, institutional fragmentation, procurement restrictions, and slow approvals may increase portfolio-level failure risk. A national or agency-level dashboard could aggregate risk signals across sectors and regions, helping policymakers identify systemic bottlenecks.

Risk bands should be linked to action. A possible provisional field protocol is:

- **Low:** routine monitoring;
- **Moderate:** project-control review and documentation of emerging issues;
- **High:** formal management review, corrective-action plan, and follow-up deadline;
- **Critical:** executive intervention, funding/procurement/approval escalation, and contingency planning.

These bands are provisional. They must be calibrated against observed field outcomes and intervention consequences before operational use.

Ethical and governance safeguards are essential. False alarms may unfairly stigmatize contractors or projects. Missed warnings may lead to preventable failure. Data access, confidentiality, model auditability, appeal mechanisms, and human oversight should be specified before deployment.

10 Limitations and Future Research

The most important limitation is the absence of real Libyan project-level data. The article is a simulation-informed proof-of-concept. Synthetic observations are not field evidence. All numerical results are simulation outputs, and field validation is pending.

Data availability is a major challenge. Administrative records may be incomplete, inconsistent, fragmented across agencies, or unavailable for failed projects. Failed, abandoned, or suspended projects may have poorer documentation, creating selection and reporting bias. Payment, procurement, approval, and progress records may be recorded differently across owners or consultants. Contextual indicators may be difficult to align temporally with project-month observations.

Measurement bias is also likely. Management capacity, institutional stability, and decision speed may require subjective assessment unless administrative timestamps are available. If surveys are used, common-method bias must be addressed through multiple respondents, objective records, temporal separation, validated constructs, expert review, pilot testing, confirmatory factor analysis where appropriate, composite reliability, convergent and discriminant validity, and careful questionnaire design.

Temporal stability is uncertain. A model trained during one political or fiscal period may not perform well during another. Libya's institutional and macroeconomic conditions may change rapidly, requiring periodic recalibration and external validation. Context specificity also limits generalization. A model developed for Libyan public infrastructure may not transfer to private building projects or to other countries without validation.

External validity remains untested. The simulation used 420 synthetic projects and 8,631 synthetic project-month rows, but these numbers do not represent an achieved Libyan sample. Future research should collect longitudinal project records across sectors, regions, project sizes, funding sources, and project statuses, including completed and failed projects. Validation should include temporal test sets, agency-level or regional external validation, and subgroup analysis.

Future research should also evaluate decision utility. ROC-AUC alone is insufficient for early warning. Researchers should conduct decision-curve analysis, compare the model with existing monitoring, and estimate the consequences of false positives and false negatives. Thresholds should reflect intervention capacity and the relative cost of missed failures.

A reproducibility protocol should accompany field implementation. Supplementary files should include a codebook, data dictionary, outcome-definition document, preprocessing scripts, temporal split specification, modelling scripts, model parameters, calibration plots, SHAP procedures, PFRI construction code, and robustness analyses. The proof-of-concept simulation used Python with scikit-learn-compatible modelling, SHAP-compatible explainability, seed 20260902, project-level temporal splitting, preprocessing fitted on training data, and a six-month future outcome. Exact package versions, scripts, and synthetic data-generation code should be archived in supplementary materials for review and replication.

11 Conclusions

This article developed a simulation-informed predictive framework for infrastructure-project failure in Libya. It addressed a gap between risk identification or ranking and operational early warning. The proposed framework

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

integrates macro instability, institutional and financing conditions, procurement and approval delays, organizational capacity, and project-performance indicators into a temporal prediction workflow.

The study is not a validated empirical assessment of Libyan infrastructure projects. No real Libyan project-level dataset was supplied. All numerical results are synthetic proof-of-concept outputs. In the simulation using seed 20260902, logistic regression achieved the highest temporal test ROC-AUC, but this does not establish real-world superiority. SHAP attribution patterns highlighted macro_instability, progress_gap, procurement_delay, financing_stability, payment_delay, and management_capacity, but these are model attributions, not causal effects.

The main contribution is a proposed, transparent simulation-based architecture that remains to be validated with real Libyan project data. The framework specifies data sources, sampling requirements, variables, operational failure definitions, preprocessing, leakage prevention, temporal validation, explainability, PFRI construction, calibration, robustness, and early-warning governance. Before practical deployment, the system must be validated using real Libyan project records, calibrated against observed outcomes, and governed through responsible decision protocols.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

1. Nasreddine Ali Algremazy, Ideris, Z., Alferjany, M. A., & Akram, A. (2023). The effect of risk management practices on project performance: A case study of the Libyan construction industry. International Journal of Professional Business Review, 8(6), e01420. https://doi.org/10.26668/businessreview/2023.v8i6.1420
2. Dwa, M. (2025). Risk management in Libyan construction projects during times of political unrest. Libya Journal of Applied Sciences and Technology, 13(1), 135–150. https://doi.org/10.67652/ljast.v13i1.60
3. Hossen, F. A. (2013). Cost risk appraisal: An application of project risk management process in Libyan construction projects. Journal of Civil Engineering and Architecture, 7(5). https://doi.org/10.17265/1934-7359/2013.05.009
4. Shetwan, A. G., Rekik, S., Richards, D., Nassar, Y., Ajina, A. B. A., & Mensah, J. H. R. (2026). Risk prioritization in Libyan construction projects using FAHP: Political and financial drivers in a fragile environment. Discover Civil Engineering, 3(1), 31. https://doi.org/10.1007/s44290-026-00419-9
5. 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
6. Kryukov, K. (2025). Machine learning for construction delay prediction: A model built on statistically validated key factors. International Journal of Construction Management, 26(11), 2512–2530. https://doi.org/10.1080/15623599.2025.2599363
7. Plebankiewicz, E., & Wieczorek, D. (2020). Adaptation of a cost overrun risk prediction model to the type of construction facility. Symmetry, 12(10), 1739. https://doi.org/10.3390/sym12101739
8. Plebankiewicz, E., & Wieczorek, D. (2020). Prediction of cost overrun risk in construction projects. Sustainability, 12(22), 9341. https://doi.org/10.3390/su12229341
9. Chen, H. L. (2015). Performance measurement and the prediction of capital project failure. International Journal of Project Management, 33(6), 1393–1404. https://doi.org/10.1016/j.ijproman.2015.02.004
10. Banerjee Chattapadhyay, D., Putta, J., & Rao P, R. M. (2021). Risk identification, assessments, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical-machine learning model. Buildings, 11(4), 172. https://doi.org/10.3390/buildings11040172
11. Moussa, A., Ezzeldin, M., & El-Dakhakhni, W. (2025). Machine learning and optimization strategies for infrastructure projects risk management. Construction Management and Economics, 43(8), 557–582. https://doi.org/10.1080/01446193.2025.2479764
12. Tian, K., Zhu, Z., Mbachu, J., Moorhead, M., & Ghanbaripour, A. (2025). Artificial intelligence in construction risk management: A decade of developments, challenges, and integration pathways. Journal of Risk Research, 28(9–10), 1078–1110. https://doi.org/10.1080/13669877.2025.2512080
13. Ding, L. Y., & Zhou, C. (2013). Development of web-based system for safety risk early warning in urban metro construction. Automation in Construction, 34, 45–55. https://doi.org/10.1016/j.autcon.2012.11.001
14. Lin, Y. S., Chan, R. W. K., & Tagawa, H. (2020). Earthquake early warning-enabled smart base isolation system. Automation in Construction, 115, 103203. https://doi.org/10.1016/j.autcon.2020.103203

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

15. Çevikbaş, M., Okudan, O., & Işık, Z. (2022). New delay-analysis method using modified schedule and modified updated schedule for construction projects. Journal of Construction Engineering and Management, 148(11), 04022119. https://doi.org/10.1061/(asce)co.1943-7862.0002394
16. Yang, J. B., Yang, C. C., & Kao, C. K. (2010). Evaluating schedule delay causes for private participating public construction works under the Build-Operate-Transfer model. International Journal of Project Management, 28(6), 569–579. https://doi.org/10.1016/j.ijproman.2009.10.005
17. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324
18. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/bf00994018
19. Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
20. Ottaviani, F. M., Ballesteros-Pérez, P., & Narbaev, T. (2025). Automated machine learning pipeline for robust project cost and duration forecasting. Automation in Construction, 178, 106426. https://doi.org/10.1016/j.autcon.2025.106426
21. Al Mnaseer, R. (2025). Ensemble machine learning for risk assessment in construction project management: Applications in infrastructure development. Asian Journal of Civil Engineering, 27(2), 485–497. https://doi.org/10.1007/s42107-025-01515-4
22. Zhong, D., Chen, J., Chen, K., & Li, H. (2021). Construction project risk prediction model based on EW-FAHP and one dimensional convolution neural network. PLOS ONE, 16(2), e0246539. https://doi.org/10.1371/journal.pone.0246539
23. Mali, A. S., Kolhe, A., Gorde, P., Kolekar, A., Umbrajkar, A., Solepatil, S., & Zare, K. (2025). Application of artificial intelligence and machine learning in construction project management: A comparative study of predictive models. Asian Journal of Civil Engineering, 26(6), 2671–2686. https://doi.org/10.1007/s42107-025-01335-6
24. Gedik, N. (2026). Leakage-free evaluation of heart disease prediction models under nested cross-validation: Discrimination, calibration, and decision-curve analysis. Scientific Reports. https://doi.org/10.1038/s41598-026-56565-8
25. Rousson, V., & Zumbrunn, T. (2011). Decision curve analysis revisited: Overall net benefit, relationships to ROC curve analysis, and application to case-control studies. BMC Medical Informatics and Decision Making, 11(1), 45. https://doi.org/10.1186/1472-6947-11-45
26. Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. Medical Decision Making, 26(6), 565–574. https://doi.org/10.1177/0272989x06295361
27. Vickers, A. J., Cronin, A. M., Elkin, E. B., & Gonen, M. (2008). Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Medical Informatics and Decision Making, 8(1), 53. https://doi.org/10.1186/1472-6947-8-53
28. Darko, A., Glushakova, I., Boateng, E. B., & Chan, A. P. C. (2023). Using machine learning to improve cost and duration prediction accuracy in green building projects. Journal of Construction Engineering and Management, 149(8), 04023061. https://doi.org/10.1061/jcemd4.coeng-13101
29. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9
30. Naghdi, M., Mahmoudi, N., Erfani, A., & Ilbeigi, M. (2026). Explainable artificial intelligence in construction engineering and management: Review of applications, trends, and opportunities. Automation in Construction, 190, 107112. https://doi.org/10.1016/j.autcon.2026.107112
31. Wang, B., Geng, L., Dang, P., & Zhang, L. (2022). Developing a framework for dynamic organizational resilience analysis in prefabricated construction projects: A project life cycle perspective. Journal of Construction Engineering and Management, 148(10), 04022110. https://doi.org/10.1061/(asce)co.1943-7862.0002381
32. Wang, D., & Wang, P. (2023). Effect of employee resilience on organizational resilience in construction projects: Considering the role of project tasks. Engineering, Construction and Architectural Management, 31(11), 4236–4254. https://doi.org/10.1108/ecam-08-2022-0797
33. Jiao, P., & Bu, W. (2024). The impact of organizational learning on organizational resilience in construction projects. Buildings, 14(4), 975. https://doi.org/10.3390/buildings14040975
34. Wang, D., & Sun, Q. (2025). How social capital influences organizational resilience in construction projects from the perspective of contractors. Engineering, Construction and Architectural Management, 1–23. https://doi.org/10.1108/ecam-03-2025-0463
35. Kou, Y., & Liu, K. (2025). Effects of dual organizational network embeddedness on safety resilience in construction projects. Engineering, Construction and Architectural Management, 33(6), 4305–4329. https://doi.org/10.1108/ecam-11-2024-1567

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

36. International Monetary Fund. (2022). The IMF strategy for fragile and conflict-affected states. Policy Papers, 2022(004), 1. https://doi.org/10.5089/9798400201820.007
37. Spyrou, E., Hobbs, B. F., Bazilian, M. D., & Chattopadhyay, D. (2019). Planning power systems in fragile and conflict-affected states. Nature Energy, 4(4), 300–310. https://doi.org/10.1038/s41560-019-0346-x
38. 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
39. 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
40. Hwang, B.-G., Zhao, X., & Ng, S. Y. (2013). Capital project performance measurement and benchmarking in Singapore. Engineering, Construction and Architectural Management, 20(2), 143–159. https://doi.org/10.1108/09699981311303017
41. Heravi, A., Coffey, V., & Trigunarsyah, B. (2018). The influence of project risk management maturity and organizational learning on the success of power plant construction projects. Project Management Journal, 49(5), 22–37. https://doi.org/10.1177/8756972818786661
42. Hosny, H. E., Ibrahim, A. H., & Elrazek, M. A. (2015). Construction claims prediction and decision awareness framework using artificial neural networks and backward optimization. Journal of Construction Engineering and Project Management, 5(1), 11–19. https://doi.org/10.6106/jcepm.2015.5.1.011
43. Andrade, P. A. d., Martens, A., & Vanhoucke, M. (2019). Using real project schedule data to compare earned schedule and earned duration management project time forecasting capabilities. Automation in Construction, 99, 68–78. https://doi.org/10.1016/j.autcon.2018.11.030
44. Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3). https://doi.org/10.1214/10-sts330
45. Zawaideh, F. (2026). BIM-integrated explainable machine learning for cost, schedule, and carbon risk prediction in high-rise building construction. Asian Journal of Civil Engineering. https://doi.org/10.1007/s42107-026-01761-0

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **AJAPAS** and/or the editor(s). **AJAPAS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

 Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).