Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

An Integrated Machine Learning Framework for Predictive and Sustainable Road Construction Project Management

Sayed Baset Hashemi[1] and Sayed Mohammad Meraj Salehy[2]
[1,2]Peter the Great St. Petersburg Polytechnic University, RUSSIA

[1]Corresponding Author: sayedbasethashemi5@gmail.com

www.sjmars.com || Vol. 5 No. 3 (2026): June Issue

Date of Submission: 26-05-2026	**Date of Acceptance:** 04-06-2026	**Date of Publication:** 27-06-2026

ABSTRACT

Road construction projects are often affected by cost overruns, schedule delays, technical uncertainty, and coordination challenges, making proactive project control difficult. This study proposes an integrated machine learning framework for predictive and sustainable road construction project management. The framework uses structured project data to predict key outcomes, including actual cost, actual duration, risk level, and completion percentage. It combines target-specific models, overrun-based target engineering, logical output constraints, and an interpretable explanation layer.

The framework was developed using a dataset of 1000 road construction projects and evaluated on 100 holdout projects. For visual clarity, a representative 10-project sample is presented in the results section, while the reported performance metrics are based on the full holdout set. The results show strong predictive performance for the main targets. Cost prediction achieved a MAPE of 2.89% and an R^2 of 0.99, while duration prediction achieved a MAPE of 3.09% and an R^2 of 0.97. Risk classification reached an accuracy of 90%. Completion percentage showed comparatively weaker performance and is therefore treated as a supporting indicator.

Overall, the findings show that the proposed framework can provide accurate, interpretable, and practically useful support for road project forecasting. The study contributes to intelligent construction management by offering an integrated decision-support framework for more proactive and sustainable infrastructure delivery.

Keywords- Cost overrun, Decision support, Machine learning, Predictive analytics, Risk prediction, Road construction projects, Schedule prediction, Sustainable construction management.

I. INTRODUCTION

Road construction projects play a vital role in economic development, regional connectivity, and social welfare. However, their delivery is often affected by cost overruns, schedule delays, technical uncertainty, environmental conditions, and stakeholder coordination challenges. These issues reduce project efficiency and make effective project control more difficult. [5]–[7]

Traditional project assessment methods mainly rely on expert judgment, past experience, and static evaluation approaches. While these methods remain useful, they may not fully capture the complex and interacting effects of technical, environmental, and managerial factors. In recent years, machine learning has attracted growing attention in construction management because of its ability to learn patterns from project data and support more proactive prediction. [1]–[4]

Despite this progress, many existing studies focus on only one outcome, such as cost, time, or risk. Fewer studies provide an integrated framework that predicts several key project outcomes together while also offering interpretable explanations for decision-makers. This limits their practical value in road construction management, where project control requires a combined understanding of cost, schedule, and risk.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

To address this gap, this study proposes an integrated machine learning framework for predictive and sustainable road construction project management. The framework uses project-level data related to design, pavement characteristics, geotechnical conditions, environmental exposure, stakeholder constraints, and planning information to predict actual cost, actual duration, risk level, completion percentage, and selected supplementary indicators. It combines target-specific models, overrun-based target engineering, logical output constraints, and an explanation layer to improve both prediction quality and interpretability.

The framework was developed using a dataset of 1000 road construction projects and evaluated on 100 holdout projects. For visual clarity, a representative 10-project sample is presented in the results section, while the reported performance metrics are based on the full holdout set. By supporting earlier identification of likely project deviations and risk conditions, the proposed framework can help improve planning, resource allocation, and decision-making in road construction projects.

The main contributions of this study are as follows:

1. It proposes an integrated machine learning framework for simultaneous prediction of cost, duration, risk level, and selected performance indicators in road construction.
2. It introduces an overrun-based target engineering strategy with logically constrained outputs to improve realism and consistency.
3. It incorporates an explanation layer that links predicted outcomes to project-specific conditions, improving interpretability and decision support.
4. It demonstrates the effectiveness of the framework through holdout validation using standard regression and classification metrics.

II. LITERATURE REVIEW

Cost overruns and schedule delays remain major challenges in road and infrastructure projects. Prior studies show that these problems are often linked to project complexity, planning deficiencies, design changes, contractor performance, and coordination constraints. As a result, conventional reactive management approaches are often insufficient for early intervention in road construction projects. [5]–[7]

With the increasing availability of project data, machine learning has gained attention in construction management as a tool for improving project forecasting. Recent studies show that machine learning has been applied to cost estimation, schedule prediction, safety monitoring, and risk assessment. However, much of the existing work remains fragmented, with many studies focusing on a single project outcome rather than providing an integrated prediction framework. [1]–[4]

In the cost domain, previous research has shown that machine learning can improve forecasting accuracy and better capture nonlinear relationships among project variables. Similar progress has been reported in delay prediction and construction risk modeling, where data-driven methods have been used to support earlier identification of problematic project conditions. Despite these advances, cost, schedule, and risk are still often modeled separately, which limits their usefulness for integrated project management decisions. [9]–[13]

A further limitation of many existing models is limited interpretability. In practical construction management, decision-makers require not only accurate forecasts but also understandable reasons for those forecasts. This is especially important in road construction, where project outcomes are influenced by technical, environmental, and managerial factors at the same time. [10], [11]

Accordingly, this study addresses an important gap by proposing an integrated machine learning framework for road construction projects that simultaneously predicts cost, duration, risk level, completion percentage, and selected supplementary indicators. In addition, the framework incorporates logical output constraints and an explanation layer to improve both realism and interpretability. This positions the study as a practical extension of current predictive construction research toward a more integrated and decision-oriented road project management framework.

III. DATA DESCRIPTION AND PROBLEM FORMULATION

3.1 Data Description

The dataset used in this study consists of 1000 road construction projects prepared for predictive modeling. Each project record includes structured variables describing design characteristics, site and geotechnical conditions, planning-related information, and environmental factors. The data were organized in tabular form, making them suitable for supervised machine learning.

The input variables can be grouped into four categories: project geometry and design (such as road length, number of lanes, pavement type, pavement thickness, and bridges/culverts), site and geotechnical conditions (such as terrain type, drainage complexity, soil/subgrade condition, and load-bearing capacity), planning and corridor-related factors (such as right-of-way status, utility relocation, planned cost, and planned duration), and environmental indicators (such as

ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

temperature, humidity, weather condition, air quality, vibration, and crack width). Together, these variables capture the engineering, environmental, and managerial dimensions of road construction projects.

The framework predicts several project outcomes. The main outputs are actual cost, actual duration, risk level, and completion percentage. In addition, supplementary indicators such as cost overrun amount, schedule deviation, material usage, labor hours, equipment utilization, and safety-related measures are also generated. This output structure allows the framework to provide a broader view of project performance.

Table 1: Input and output variables used in the proposed framework

Group	Variables	Role
Geometry and design	Location, road length, number of lanes, pavement type, pavement thickness, bridges/culverts count	Input
Site and geotechnical	Terrain type, soil/subgrade CBR, drainage complexity, load-bearing capacity	Input
Planning and corridor	Planned cost, planned duration, planned dates, right-of-way status, utility relocation required	Input
Environmental and condition	Average temperature, average humidity, weather condition, air quality index, baseline vibration, baseline crack width	Input
Main outputs	Actual cost, actual duration, risk level, completion percentage	Output
Supplementary outputs	Cost overrun amount, schedule deviation, energy consumption, material usage, labor hours, equipment utilization, accident count, safety risk score, image analysis score, anomaly detection	Output

3.2 Problem Formulation

This study formulates road construction prediction as a multi-target supervised learning problem. The objective is to learn the relationship between project input variables and key project outcomes related to cost, time, and risk. Since the outputs differ in nature, the problem includes both regression tasks and a classification task. Regression is used for continuous outputs such as actual cost, actual duration, and completion percentage, while classification is used for risk level prediction.

A key feature of the proposed framework is that cost and duration are modeled through their deviation behavior relative to planned values, rather than being treated only as direct final outputs. The final estimates are then derived from these predicted deviations. This improves consistency between planned and actual performance and makes the outputs more meaningful for project management applications.

Overall, the problem addressed in this study is whether project-specific technical, environmental, and managerial variables can be used to accurately predict major road construction outcomes through an integrated and interpretable machine learning framework.

IV. METHODOLOGY

4.1 Framework Overview

This study proposes an integrated machine learning framework for predicting key road construction outcomes from structured project data. The framework takes project-level inputs related to design, site conditions, planning, and environmental factors, and produces multiple decision-support outputs, including actual cost, actual duration, risk level, completion percentage, and selected supplementary indicators. Unlike single-target approaches, the proposed framework uses separate prediction modules for different outputs while maintaining consistency across the final results.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

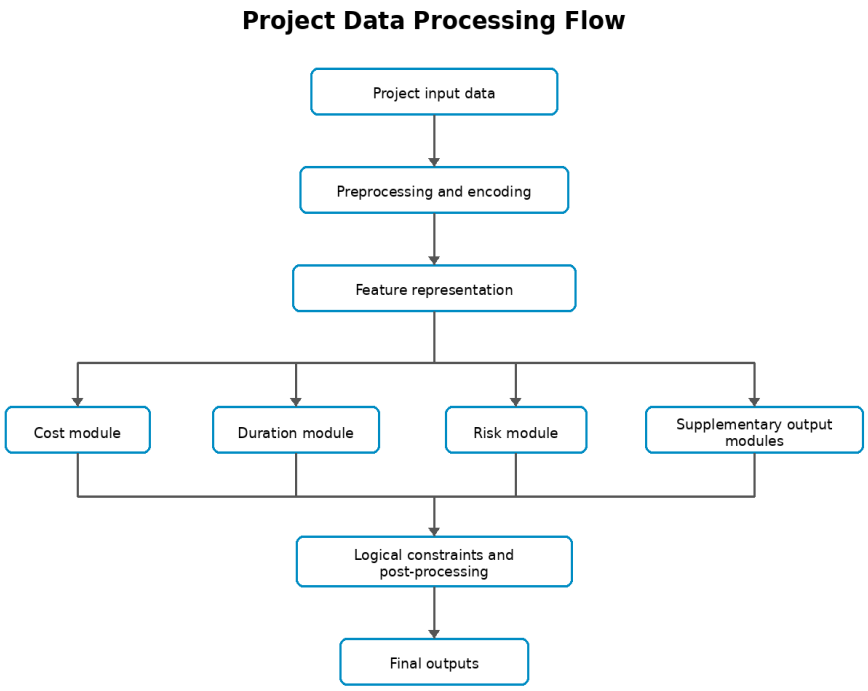

Figure 1: Overall architecture of the proposed framework

The workflow consists of five main stages: data preprocessing, feature representation, target-specific prediction, logical post-processing, and explanation generation. First, the project inputs are prepared and checked for modeling. Next, the processed features are passed to dedicated modules for cost, duration, risk, and supplementary outputs. The raw predictions are then refined through logical constraints to ensure realistic results. Finally, the framework generates interpretable explanations linked to the main project drivers.

4.2 Data Preparation and Feature Representation

The dataset contains both numerical and categorical variables, so preprocessing was performed to convert the inputs into a machine-readable form. Numerical variables such as road length, pavement thickness, planned cost, planned duration, load-bearing capacity, temperature, humidity, air quality index, vibration, and crack width were treated as continuous predictors. Categorical variables such as location, pavement type, terrain type, drainage complexity, right-of-way status, utility relocation, and weather condition were encoded before model training and prediction.

The feature space was designed to represent the project from multiple perspectives. It includes engineering attributes such as road size and pavement specification, geotechnical attributes such as terrain, drainage, and soil condition, planning-related attributes such as budget, schedule, right-of-way readiness, and utility coordination, and environmental attributes such as weather, temperature, humidity, and air quality. This representation reflects the idea that road project performance depends on interacting technical, managerial, and environmental factors.

4.3 Target Engineering

A key feature of the framework is its overrun-based target engineering strategy. Instead of predicting actual cost and actual duration only as direct final values, the framework first models their deviation behavior relative to planned cost and planned duration. The final estimates are then derived from these predicted deviations. This design improves consistency between planned and actual performance and makes the outputs more meaningful for project management applications.

In addition to cost and duration, the framework predicts project risk level as a three-class output: Low, Medium, and High. Completion percentage is treated as a bounded continuous output. Supplementary indicators, including energy consumption, material usage, labor hours, equipment utilization, accident count, safety risk score, image analysis score, and anomaly detection, are generated through separate target-specific models.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

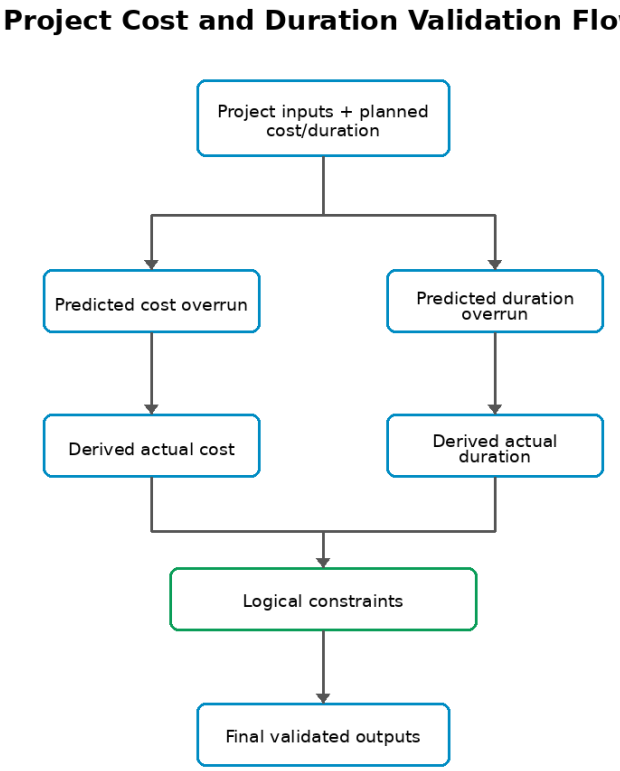

Figure 2: Overrun-based prediction logic for cost and duration estimation

4.4 Model Architecture

Because the framework addresses outputs with different statistical characteristics, separate predictive modules were developed for the main targets. Tree-based and ensemble learning methods were adopted because they are suitable for tabular construction data and can capture nonlinear relationships among project variables. [9], [11]–[13]

The cost module uses Extra Trees to model cost-overrun behavior and derive actual project cost. The duration module uses Random Forest to model duration-overrun behavior and derive actual project duration. For risk classification, the framework uses a soft voting ensemble that combines Random Forest, Extra Trees, and K-Nearest Neighbors to predict Low-, Medium-, or High-risk classes. Supplementary outputs are predicted using separate target-specific tree-based models.

Table 2: Target-specific prediction modules used in the framework

Output target	Modeling strategy	Algorithm
Actual cost	Cost-overrun-based regression	Extra Trees
Actual duration	Duration-overrun-based regression	Random Forest
Risk level	Three-class classification	Soft voting ensemble (Random Forest + Extra Trees + KNN)
Completion percentage	Bounded regression	Target-specific tree model
Supplementary indicators	Target-specific prediction	Tree-based models

4.5 Logical Constraints and Explanation Layer

After prediction, the framework applies logical constraints to improve the realism of the outputs. This step is used to keep bounded variables within meaningful limits, prevent invalid count-based outputs, and preserve consistency between planned values, predicted overruns, and final estimated outcomes. As a result, the system produces outputs that are not only predictive, but also practically interpretable.

To improve interpretability, the framework also includes an explanation layer. This layer generates input-grounded reason statements for the prediction, typically expressed as primary, secondary, and tertiary contributors. The explanations are derived from project-specific characteristics such as terrain, drainage complexity, right-of-way status, utility relocation, and environmental exposure. In this way, the framework supports not only prediction, but also decision-making.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

4.6 Implementation Summary

The proposed methodology combines structured project data, preprocessing, overrun-based target engineering, target-specific machine learning modules, logical post-processing, and an explanation layer within one integrated framework. This design allows the system to predict cost, duration, risk, and related performance indicators in a consistent and interpretable way, making it suitable for road construction decision support.

V. EXPERIMENTAL SETUP AND EVALUATION

5.1 Experimental Setup

The proposed framework was implemented in a Python-based environment using supervised machine learning methods for structured tabular data. The evaluation focused on the main project outcomes: actual cost, actual duration, risk level, and completion percentage. Although the framework also generated supplementary indicators, the primary analysis in this study emphasizes cost, duration, and risk prediction.

To assess generalization performance, the framework was evaluated using a holdout validation strategy based on unseen road construction projects. The main validation results are reported using a holdout set of 100 projects. For visual clarity, a representative 10-project sample is presented in the results section, while the reported performance metrics are based on the full holdout set.

For each holdout project, the framework received the complete set of input variables and generated predictions for the target outputs. These predictions were then compared with the corresponding actual project values to assess predictive performance.

5.2 Evaluation Metrics

For continuous outputs such as actual cost, actual duration, and completion percentage, the framework was evaluated using four standard regression metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). MAE and RMSE measure the magnitude of prediction error, MAPE expresses error in percentage terms, and R^2 indicates the agreement between predicted and actual values.

For risk level prediction, classification accuracy was used as the main evaluation metric. In addition, a confusion matrix was used to examine the classification performance across the three risk classes: Low, Medium, and High.

To improve the clarity of the results, the evaluation is presented using both numerical and visual summaries. Metric tables are used to report overall performance, while actual-versus-predicted plots are used for cost and duration, and a confusion matrix is used for risk classification.

This evaluation design provides a clear basis for assessing the predictive effectiveness of the proposed framework and its practical usefulness for road construction project management.

VI. RESULTS AND DISCUSSION

The proposed framework was evaluated using holdout validation on unseen road construction projects. In the paper, the overall performance can be summarized using the full 100-project holdout set, while a representative 10-project sample may be presented in figures for visual clarity. Across the evaluated cases, the framework performed strongest in cost prediction, duration prediction, and risk classification, while completion percentage showed comparatively weaker performance and is therefore better treated as a supporting indicator.

Table 3: Overall validation performance of the proposed framework

Output	MAE	RMSE	MAPE	R^2	Accuracy
Actual cost	342,990	458,743	2.89%	0.99	—
Actual duration	15 days	21 days	3.09%	0.97	—
Risk level	—	—	—	—	90%
Completion percentage	5 points	6 points	6.90%	0.43	—

6.1 Cost and Duration Prediction Performance

The cost prediction module achieved strong results, with an MAE of $342,990, an RMSE of $458,743, a MAPE of 2.89%, and an R^2 of 0.99. These results indicate excellent agreement between predicted and actual project cost values. From a practical perspective, an average percentage error below 3% suggests that the framework can provide reliable cost forecasts for road construction decision-making.

The duration prediction module also performed well, with an MAE of 15 days, an RMSE of 21 days, a MAPE of 3.09%, and an R^2 of 0.97. This indicates that the framework can estimate project duration with a high degree of accuracy.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

Given that road construction schedules commonly extend over several months, an average error of about two weeks remains practically acceptable for planning and control purposes.

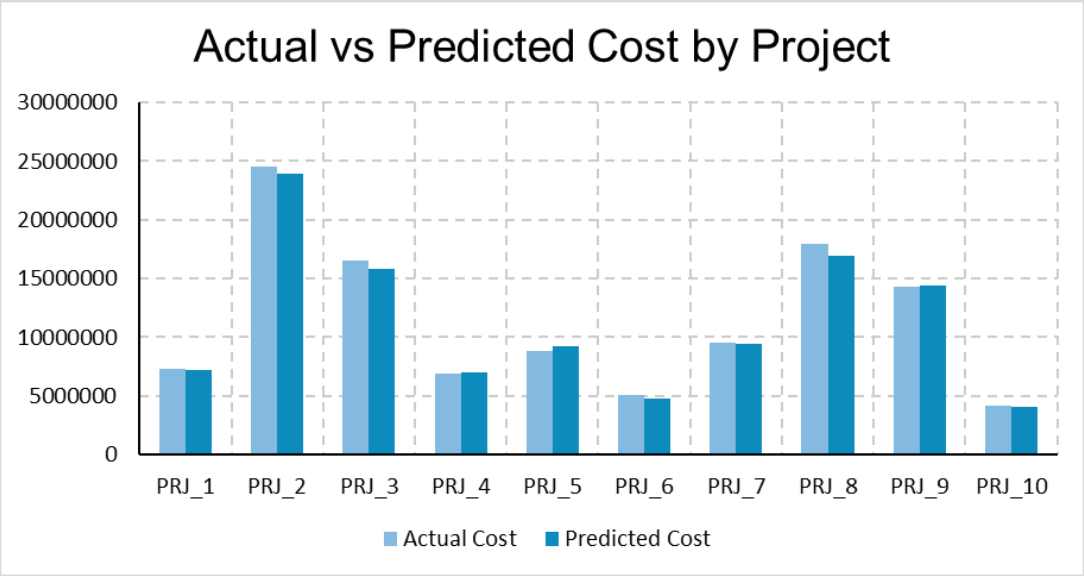

Figure 3: Actual versus predicted project cost

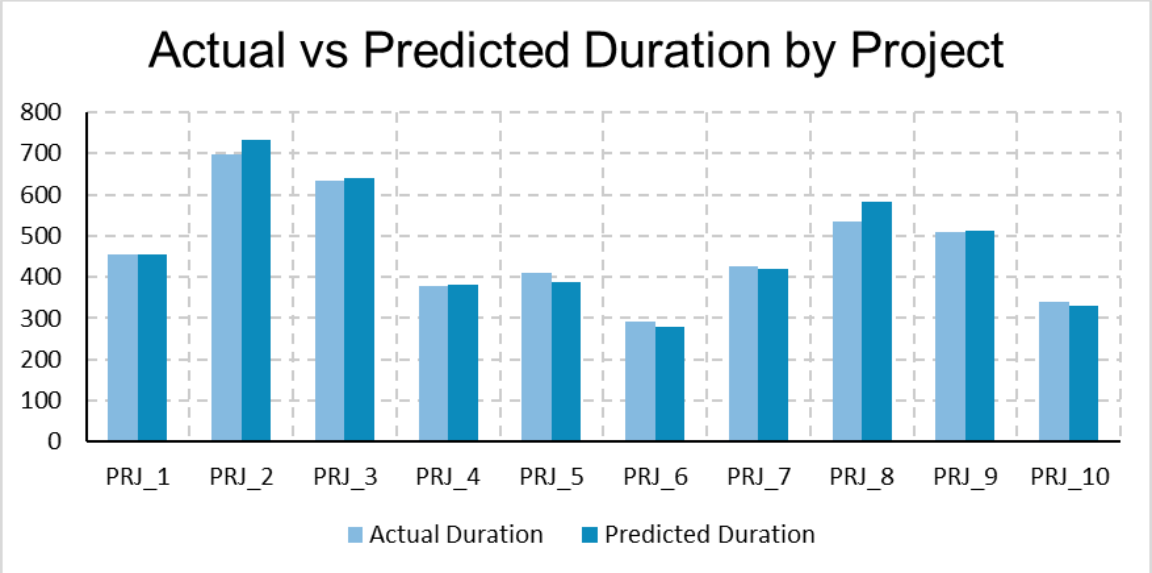

Figure 4: Actual versus predicted project duration

The strong performance of both modules is consistent with the overrun-based target engineering strategy adopted in the framework. By modeling deviation behavior relative to planned cost and planned duration, the system preserves a more realistic relationship between planned and actual project outcomes.

6.2 Risk Classification and Completion Percentage

The risk classification module achieved an overall accuracy of 90%, correctly classifying 8 of 10 projects in the representative sample, while achieving 90% accuracy on the full holdout set. The confusion matrix showed particularly strong performance in identifying High-risk projects, with all actual High-risk cases correctly classified. This is an important result from a project management perspective, as early identification of high-risk projects supports more timely intervention and mitigation planning.

The completion percentage output was less accurate than cost and duration. It achieved an MAE of 5 percentage points, an RMSE of 6, a MAPE of 6.90%, and an R^2 of 0.43. These results suggest that completion percentage is more difficult to predict from the current feature set and should therefore be interpreted as a supplementary performance indicator rather than a primary strength of the framework.

Stallion Journal for Multidisciplinary Associated Research Studies
ISSN (Online): 2583-3340
Volume-5 Issue-3 || June 2026 || PP. 30-38 https://doi.org/10.55544/sjmars.5.3.5

Table 4: Representative sample of holdout projects used for visual comparison (cost in US$, duration in days)

Project ID	Actual Cost ($)	Pred. Cost ($)	Cost Error	Actual Dur.	Pred. Dur.	Dur. Error	Actual Risk	Pred. Risk
PRJ_1	7,280,938.4	7,246,165.9	-34772.4	453.9	455.5	1.6	Med	Med
PRJ_2	24,520,632	23,970,983.2	-549649	698.6	734.1	35.	High	High
PRJ_3	16,574,323.5	15,838,093.5	-736230	634.9	641.2	6.3	High	High
PRJ_4	6,941,181.8	7,037,389.8	96207.98	379.1	382.4	3.3	Low	Med
PRJ_5	8,819,766.8	9,214,859.6	395092.8	409.6	388.7	-20.9	Med	Med
PRJ_6	5,093,296.7	4,792,035.2	-301261	292.1	279.5	-12.6	Med	Med
PRJ_7	9,549,019.6	9,471,210.6	-77809	426.7	419.7	-7.1	Med	Low
PRJ_8	17,926,740.4	16,945,440.9	-981299	535.4	584.1	48.6	Low	Low
PRJ_9	14,291,317.4	14,440,108.9	148791.6	509.1	511.9	2.8	Med	Med
PRJ_10	4,125,827.6	4,017,040.7	-108787	338.4	328.5	-9.9	High	High

6.3 Discussion of the Main Findings

The results show that the main strength of the proposed framework lies in its ability to predict actual cost, actual duration, and risk level with good practical accuracy. These are the outputs most directly related to project planning and control, and they are also the outcomes for which the framework produced the most reliable results. The modular architecture of the framework likely contributed to this performance, since each target was modeled using a method suited to its statistical characteristics rather than relying on a single model for all outputs.

The results also highlight the value of combining prediction with interpretability. In addition to producing forecasts, the framework generates explanation statements linked to project-specific drivers. This makes the system more useful for practitioners, as it provides not only predicted values but also insight into the likely causes of risk and performance variation.

From a practical standpoint, accurate cost and duration prediction can support better budgeting, scheduling, and resource allocation, while reliable risk classification can serve as an early warning mechanism for projects requiring closer managerial attention. In this sense, the framework contributes to more proactive and sustainable road construction management.

6.4 Summary

Overall, the results confirm that the proposed framework performs strongly for the most important road project prediction tasks. Cost prediction achieved very low percentage error and excellent fit, duration prediction also showed strong agreement with actual outcomes, and risk classification demonstrated good practical value, particularly for High-risk detection. Completion percentage was less robust and is therefore presented as a supporting indicator. Taken together, these findings show that the framework offers a useful and interpretable decision-support tool for predictive road construction project management.

VII. CONCLUSION AND FUTURE WORK

This study proposed an integrated machine learning framework for predictive and sustainable road construction project management. The framework combines structured project data, target-specific prediction modules, logical output constraints, and an explanation layer to estimate major project outcomes, including actual cost, actual duration, risk level, and completion percentage.

The results show that the framework performs strongly in cost prediction, duration prediction, and risk classification, which are the most important outputs for project planning and control. In particular, the low cost and time errors and the strong risk classification results indicate that the framework can provide practical support for forecasting and early decision-making in road construction projects. Completion percentage showed weaker performance and is therefore better treated as a supporting indicator.

Overall, the study demonstrates that road project prediction can be improved through an integrated and interpretable machine learning framework. By combining prediction accuracy with practical explanation, the proposed approach offers useful support for more proactive and sustainable road construction management.

Future work should focus on extending the validation to larger unseen project sets, improving the prediction of supplementary outputs, and incorporating additional project progress variables. The framework may also be extended to other infrastructure project types and enhanced with more advanced explainability and real-time updating capabilities.

REFERENCES

[1] Abioye SO, Oyedele LO, Akanbi LA, Ajayi AO, Davila Delgado JM, Bilal M, Akinade OO, Ahmed AA. Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. Journal of Building Engineering. 2021;44:103299. doi:10.1016/j.jobe.2021.103299.

[2] Darko A, Chan APC, Adabre MA, Edwards DJ, Hosseini MR, Ameyaw EE. Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. Automation in Construction. 2020;112:103081. doi:10.1016/j.autcon.2020.103081.

[3] Datta SD, Islam M, Sobuz MHR, Ahmed S, Kar M. Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. Heliyon. 2024;10(5):e26888. doi:10.1016/j.heliyon.2024.e26888.

[4] Gao Y, Antwi-Afari MF, Huang Y, Chen Z-S, Manzoor B. Artificial Intelligence in Construction Project Management: A Systematic Literature Review of Cost, Time, and Safety Management. Buildings. 2026;16(5):1061. doi:10.3390/buildings16051061.

[5] Flyvbjerg B, Holm MKS, Buhl SL. How common and how large are cost overruns in transport infrastructure projects? Transport Reviews. 2003;23(1):71–88. doi:10.1080/01441640309904.

[6] Kumar M, Kumari S. Causes of delays in road construction projects: A systematic review. Journal of Financial Management of Property and Construction. 2025;30(2):256–294. doi:10.1108/JFMPC-01-2023-0004.

[7] Orya F, Calahorra-Jimenez M. Delays in Infrastructure Projects: Main Reasons in the Design, Procurement, and Construction Phases. Public Works Management & Policy. 2025;30(2):261–280. doi:10.1177/1087724X241308310.

[8] Hanafy NO, Hanafy NO. An Extensive Examination of Uses of Machine Learning and Artificial Intelligence in The Construction Industry's Project Life Cycle. Energy and Buildings. 2025;345:116094. doi:10.1016/j.enbuild.2025.116094.

[9] Ashtari MA, Ansari R, Hassannayebi E, Jeong J. Cost Overrun Risk Assessment and Prediction in Construction Projects: A Bayesian Network Classifier Approach. Buildings. 2022;12(10):1660. doi:10.3390/buildings12101660.

[10] Larocque R, Boulé A-M, Cappart Q. Estimating Road Construction Costs with Explainable Machine Learning. Interfaces. 2025;55(2):137–153. doi:10.1287/inte.2023.0027.

[11] Gondia A, Moussa A, Ezzeldin M, El-Dakhakhni W. Machine learning-based construction site dynamic risk models. Technological Forecasting and Social Change. 2023;189:122347. doi:10.1016/j.techfore.2023.122347.

[12] Zhang Y, Minchin RE, Flood I, Ries RJ. Preliminary Cost Estimation of Highway Projects Using Statistical Learning Methods. Journal of Construction Engineering and Management. 2023;149(5):04023026. doi:10.1061/JCEMD4.COENG-12773.

[13] Shoar S, Chileshe N, Edwards JD. Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. Journal of Building Engineering. 2022;50:104102. doi:10.1016/j.jobe.2022.104102.