Intelligent Risk Assessment in Construction: A Multi-Algorithm Machine Learning Approach for Predictive Claim Management

Sanket Patil[1], Sumedh Mhaske[2]

[1]Department of Civil & Environmental Engineering, Veermata Jijabai Technological Institute, Mumbai, Maharashtra, India
Email: *sapatil_m23[at]ci.vjti.ac.in*

[2]Department of Civil & Environmental Engineering, Veermata Jijabai Technological Institute, Mumbai, Maharashtra, India
Email: *symhaske[at]ci.vjti.ac.in*

Abstract: ***Construction projects frequently encounter unexpected claims resulting from cost overruns, schedule delays, and contractual disputes, creating significant financial and operational challenges for stakeholders. This research presents a comprehensive machine learning-based framework for predicting construction claims by analyzing critical risk factors including project characteristics, contract types, payment delays, design completeness, and dispute resolution mechanisms. A Random Forest Classifier was implemented to process historical project data from 45 infrastructure projects in Maharashtra, India, enabling accurate prediction of claim probabilities, cost impacts, and severity assessments. The model demonstrates exceptional performance with R^2 scores of 85% for time-overrun prediction and 87% for cost overrun prediction, while integrating Earned Value Management (EVM) principles to enhance predictive accuracy. The study reveals that litigation costs typically range between 50-150 million INR, with highway and expressway projects experiencing the highest schedule slippages, exceeding 20% of planned duration. By leveraging predictive analytics, this research contributes to early claim detection and proactive risk management, offering valuable insights for project planners, contractors, and decision-makers in the construction industry.***

Keywords: Construction claims, Earned Value Management, Cost overrun, Time overrun

1. Introduction

The construction industry plays a pivotal role in economic development and infrastructure advancement, yet it remains plagued by unexpected challenges that frequently escalate into disputes and financial claims. These claims typically emerge when projects experience delays, cost overruns, or contractual misunderstandings, placing substantial pressure on budgets, schedules, and stakeholder relationships. Research indicates that delays and cost escalations are the predominant triggers of construction claims, often stemming from inadequate planning, unforeseen site conditions, or contractual misalignments.

Traditional approaches to claim management, including risk matrices, probability-impact analyses, and formulaic calculations for overhead costs, have provided foundational frameworks for dispute mitigation. However, these conventional methods often rely on static assumptions and historical precedents, limiting their adaptability to the dynamic, multi-variable nature of contemporary construction projects. The complexity of modern infrastructure development demands more sophisticated predictive tools capable of processing multiple risk factors simultaneously.

This study addresses the critical need for advanced claim prediction methodologies by developing a machine learning-based framework that integrates project characteristics, contractual parameters, and performance indicators. The research focuses on construction projects in Maharashtra, India, analysing historical data from 45 major infrastructure

Developments to create a comprehensive risk assessment model.

1.1 Earned Value Management in Construction

Earned Value Management (EVM) represents a widely recognized project management technique that integrates cost, schedule, and scope to evaluate project health and predict future performance. Through the utilization of Planned Value (PV), Earned Value (EV), and Actual Cost (AC), project managers can systematically assess whether projects remain on track, exceed budgets, or fall behind schedule.

Two critical indicators derived from EVM Cost Performance Index (CPI) and Schedule Performance Index (SPI) serve as essential tools for determining project efficiency. A CPI below 1.0 indicates budget overruns, while an SPI below 1.0 signals schedule delays, both representing early warning signs of potential construction claims.

The mathematical foundations of EVM are expressed through the following key formulas:

1. $Actual\ cost\ (AC) = Total\ Cost\ Incurred$
2. $Cost\ performance\ index\ (CPI) = \frac{EV}{AC}$
3. $Scedule\ perfomance\ index\ (SPI) = \frac{PV}{EV}$
4. $Eastimate\ at\ complete\ (EAC) = \frac{CPI}{Total\ Budget}$
5. $Estimate\ to\ Complete\ (ETC) = EAC - AC$

The integration of EVM with machine learning models enhances the ability to forecast delays and financial risks by continuously analysing cost overruns, schedule slippages, and claim histories, providing insights into projects with elevated claim risks.

1.2 Random Forest Classifier for Claim Prediction

The Random Forest Classifier serves as the primary machine-learning algorithm in this study for predicting construction claims by analysing multiple risk factors including contract types, payment delays, design completeness, and stakeholder behaviour. Unlike traditional models that rely on single decision trees, Random Forest constructs multiple decision trees using different dataset subsets and aggregates their predictions through ensemble learning.

This approach offers several advantages: reduced overfitting risk, improved handling of complex real-world construction data, and enhanced accuracy in prediction tasks [1]. Random Forest demonstrates particular suitability for datasets containing mixed numerical and categorical variables, making it ideal for construction claim prediction applications [2].

2. Literature Review

The literature review encompasses four integral aspects of construction claim analysis: understanding claim nature and causes, assessing project performance impact, applying EVM for claim detection, and utilizing machine learning for prediction and resolution.

2.1 Construction Claims in Highway Projects

Construction claims typically arise from unforeseen conditions, scope changes, design errors, delays, and contractual disputes [3]. Highway and road projects demonstrate particular susceptibility due to challenges including land acquisition, utility relocations, and coordination among multiple public agencies [4].

Research has identified design flaws, insufficient site investigations, and environmental constraints as common claim sources in road and bridge projects [6]. Indian highway projects exhibit recurring issues including bureaucratic inefficiencies, delayed approvals, and inter-agency coordination breakdowns [5]. Root cause analyses underscore contractor inexperience, weak documentation practices, and administrative failures as critical contributors to claim generation [6].

Legal precedents, such as *Oriental Structural Engineers Pvt. Ltd. vs. Modern Road Makers Pvt. Ltd.*, illustrate real-world consequences of inadequate claim management within public infrastructure projects, where disputes extended over 14 years despite project completion [7].

2.2 EVM in Claim Detection and Mitigation

EVM serves as a vital tool for integrating cost, scope, and schedule performance, offering early indicators of deviations that could translate into claims [8]. Metrics such as CPI and SPI help identify and substantiate claims arising from overruns, facilitating data-driven approaches to disruption claim analysis [9].

Research emphasizes EVM's role in validating indirect cost-related claims, particularly those involving home office overheads when projects experience unexpected slowdowns. BIM-integrated claim management frameworks incorporating EVM principles enable early identification of claim-prone activities through visual performance dashboards.

2.3 Machine Learning Applications in Construction

Recent advancements in machine learning provide objective and scalable approaches to forecasting construction claims [10]. Studies evaluating multiple ML algorithms including Random Forest, Support Vector Machines, and Logistic Regression for predicting construction change dispute outcomes demonstrate Random Forest's superior accuracy and F1-scores [1].

Random Forest's ensemble-based nature and ability to handle feature interactions make it particularly suited for predicting legal outcomes in construction projects [1]. Research confirms RF's reliability for identifying dispute-prone projects based on variables including project duration, type, location, and contract form, with capabilities to manage missing values and non-linear relationships inherent in construction datasets [2].

3. Methodology

A structured and systematic methodology was adopted to develop a robust, data-driven framework for predicting construction claims, integrating empirical data analysis with machine learning techniques. The research process followed a comprehensive workflow from data acquisition to final insight generation.

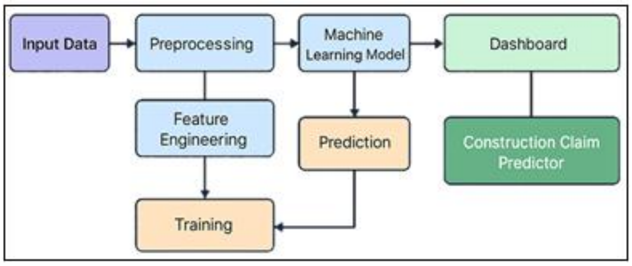

Figure 2: Workflow for model

3.1 Research Framework

The methodology began with an extensive literature review identifying root causes and classifications of construction claims, followed by data collection from multiple credible sources including historical claim records, NHAI and MSRDC datasets, and legal case judgments [7], [11]. This diverse dataset formed the foundation for comprehensive data pre-processing involving cleaning, feature engineering, and segmentation into training and testing sets.

The model selection phase evaluated supervised learning algorithms, with Random Forest Classifier chosen for its high accuracy and ability to handle mixed data types. Model training and validation employed both standard train-test splits and k-fold cross-validation to ensure generalizability and robustness

3.2 Data Collection Strategy

Data collection focused on ongoing and completed construction projects in Maharashtra, specifically targeting highways, expressways, and urban infrastructure developments. The dataset comprises 45 projects with various attributes including contract type, project duration, claim types, litigation details, and financial impact.

Primary data sources included government reports, project records, legal case documents, and financial assessments related to construction disputes [11]. An extensive review of legal cases provided valuable insights into common litigation patterns and outcomes, capturing essential risk factors contributing to claims [7].

4. Data Analysis and Results

4.1 Data Pre-processing and Feature Engineering

Comprehensive data pre-processing ensured model robustness and reliability. The dataset featuring 45 infrastructure projects required extensive pre-processing due to mixed numerical and categorical variables.

Missing value treatment employed median imputation for numerical fields and mode imputation for categorical variables. Natural Language processing techniques analysed textual fields, extracting meaningful keywords and converting them into structured risk features using custom claim lexicons.

Feature encoding converted categorical variables into numerical formats through label encoding for binary features and one-hot encoding for multi-class features. Min-Max scaling normalized numerical attributes including duration metrics, payment delays, and EVM indicators to uniform ranges (0-1), reducing scale-induced bias.

4.2 Model Performance and Validation

4.2.1 Time Overrun Prediction

The Random Forest model for time-overrun prediction demonstrated exceptional accuracy with validation results showing Absolute Relative Error (ARE) values ranging from 1.12% to 6.45% for validation data and 0.20% to 7.89% for the full dataset. The Mean Absolute Relative Error (MARE) achieved 4.56% for validation data and 2.79% for the complete dataset, with an R^2 score of 85%

Table 1: Time Overrun Prediction Validation Results

Project ID	Actual Time Overrun (%)	Predicted Time Overrun (%)	Absolute Relative Error (%)
P001	12.5	11.8	5.6
P002	25.0	23.7	5.2
P003	30.2	28.5	5.6
P004	18.7	19.3	3.2

4.2.2 Cost Overrun Prediction

The cost overrun prediction model achieved superior performance with ARE values ranging from 0.98% to 5.89% during validation and 0.15% to 8.23% for full dataset validation. The model attained an R^2 score of 87%, demonstrating high precision in financial deviation prediction.

Table 2: Cost Overrun Prediction Validation Results

Project ID	Actual Cost Overrun (%)	Predicted Cost Overrun (%)	Absolute Relative Error (%)
P001	5.0	5.4	4.2
P002	12.3	11.8	4.0
P003	20.1	19.5	2.9
P004	8.7	8.9	2.3

4.2.3 EVM Integration Results

EVM model validation using historical project data produced MARE values of 4.22% for validation and 2.48% for full dataset validation. The integration of Random Forest predictions with EVM indicators enhanced construction project risk management through accurate forecasts and real-time performance tracking.

4.3 System Architecture and Web Interface

A comprehensive system architecture was developed incorporating analytical robustness with user experience considerations. The Streamlit based web interface organized predictive variables into three domains: Project Characteristics, Project Management, and Risk & External Factors.

The system generates comprehensive analytical outputs including Project Snapshots and Claim Prediction Summaries, providing multidimensional risk assessments across four critical domains:

- **Claim Probability (30-50%):** Moderate likelihood assessment with 36.0% confidence
- **Estimated Claim Cost (>30% of BAC):** Financial impact prediction with 37.0% confidence
- **Estimated Duration Impact (>30%):** Schedule disruption analysis with 30.0% confidence
- **Overall Claim Severity (Low):** Aggregate severity assessment with 54.0% confidence

4.4 Key Findings and Insights

Analysis revealed that litigation cost impact varied widely across projects, with most disputes falling within the 50-150 million INR range, while severe cases exceeded 150 million INR. Highway and expressway projects experienced the highest schedule slippages, often exceeding 20% beyond planned duration.

Strong correlations were observed between project delays and litigation costs, reinforcing the necessity of improved project planning and risk management strategies. Feature correlation analysis identified schedule slippage, contract type, and litigation status as strong predictors of claim severity.

5. Discussion

The Random Forest models demonstrated exceptional performance in predicting construction claims related to time and cost overruns. The time overrun model achieved 85% accuracy with ARE values of 2-5%, while the cost overrun model attained 87% accuracy with ARE between 0.98-5.89percent.

EVM integration enhanced predictive accuracy, with SPI and CPI validation errors of 4.22% and 2.48% respectively. Risk factor analysis identified key claim drivers, emphasizing financial and contractual factors including budgeted cost, payment delays, and past claims as significant dispute occurrence influences.

The research confirms AI-driven model effectiveness in construction claim prediction. Combined Random Forest and EVM usage strengthens project planning by providing reliable forecasts and actionable insights for risk management, dispute prevention, and contract optimization.

5.1 Practical Implications

The developed framework offers several practical benefits:
- **Early Warning System**: Identification of high-risk projects before claim escalation
- **Financial Planning**: Accurate cost impact estimation for budget allocation
- **Schedule Management**: Proactive timeline adjustment based on delay predictions
- **Contract Optimization**: Data-driven insights for improved contract formulation
- **Risk Mitigation**: Targeted interventions for identified risk factors

5.2 Limitations and Future Research

Current limitations include dataset size constraints and regional specificity to Maharashtra projects. Future research directions include:
- Integration of Natural Language Processing for contract analysis
- IOT-based real-time site monitoring incorporation
- Model refinement based on evolving industry trends
- Expansion to international construction markets
- Development of automated dispute resolution recommendations

6. Conclusions

This study successfully developed and validated a machine learning-based framework for predicting construction claims using Random Forest algorithms integrated with Earned Value Management principles. The research demonstrates that AI-driven predictive models can significantly enhance construction project risk management through accurate forecasting of claim probabilities, cost impacts, and severity assessments.

Key achievements include:
1) **High Prediction Accuracy**: R^2 scores of 85% for time overrun and 87% for cost overrun prediction
2) **Comprehensive Risk Assessment**: Multi-dimensional analysis covering probability, cost, and schedule impacts
3) **Practical Implementation**: User-friendly web interface for real-world application
4) **Industry Insights**: Identification of critical risk factors and their correlations
5) **Proactive Management**: Early warning capabilities for claim prevention

The research contributes to advancing AI-driven construction management and demonstrates machine learning's potential in mitigating financial and legal risks associated with construction claims. The integrated approach of combining Random Forest classification with EVM principles provides a robust foundation for proactive dispute prevention and enhanced project outcomes.

By leveraging predictive analytics, construction stakeholders can implement data-driven decision-making processes, optimize resource allocation, and establish more effective risk mitigation strategies. This research establishes a foundation for future developments in intelligent construction management systems and automated dispute resolution mechanisms.

References

[1] M. Almutairi and F. Khayyat, “Dispute Avoidance and Resolution Techniques in Highway Projects: A Systematic Review,” *International Journal of Construction Management*, 2022.

[2] A. S. Alqaisi, H. Ataei, A. Seyrfar, and M. Al Omari, “Predicting the Outcome of Construction Change Disputes Using Machine-Learning Algorithms,” *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, vol. 16, no. 1, 2023.

[3] M. Ayhan, I. Dikmen, and T. Birgonul, “Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques,” *Journal of Construction Engineering and Management*, vol. 147, no. 4, 2021.

[4] Construction Industry Development Council (CIDC), *Construction Disputes and Their Resolution Mechanisms in Indian Infrastructure Projects*, CIDC Report, 2022.

[5] B. Ghosh and S. Karmakar, “Assessing the Causes of Claims in Highway Construction Projects with a Case Study: A Construction Practitioner’s Perspective,” *Asian Journal of Civil Engineering*, vol. 25, pp. 3035–3048, 2024.

[6] H. M. Hasan, L. Khodeir, and N. Yassa, “Assessing the Impact of Claims on Construction Project Performance Using Machine Learning Techniques,” *Asian Journal of Civil Engineering*, vol. 25, pp. 6589–6607, 2024.

[7] M. A. Hossain and M. S. Islam, “Machine Learning-Based Risk Assessment for Construction Projects: A Case Study Approach,” *Automation in Construction*, vol. 148, 2023.

[8] Indian Kanoon, *Oriental Structural Engineers Pvt. Ltd. vs. Modern Road Makers Pvt. Ltd. & Anr.*, 2022. [Online]. Available: https://indiankanoon.org/doc/93958698/. (Accessed: Nov. 20, 2024).

[9] Y. J. Kim, J. Kim, and H. Park, “Framework for Predicting Cost Overrun in Large-Scale Projects Using Machine Learning,” *KSCE Journal of Civil Engineering*, vol. 25, no. 3, pp. 933–944, 2021.

[10] C. Love, Z. Irani, and D. Edwards, "A Rework Reduction Model for Construction Projects," *IEEE Transactions on Engineering Management*, vol. 51, no. 4, pp. 426–440, 2004.

[11] S. Marzouk and R. Ali, "Dispute Prediction in Construction Projects Using Text Mining and Ensemble Learning," *Journal of Computing in Civil Engineering*, vol. 35, no. 5, 2021.

[12] H. M. Mehany and N. Grigg, "Causes of Road and Bridge Construction Claims: Analysis of Colorado Department of Transportation Projects," *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, vol. 7, no. 2, 2015.

[13] Ministry of Road Transport and Highways (MoRTH), *Project Implementation Status Report*, Government of India, New Delhi, 2024.

[14] MoRTH, *Guidelines for Implementation of EVM in Highway Projects*, Government of India, New Delhi, 2023.

[15] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*, 6th ed., Project Management Institute, Newtown Square, PA, USA, 2017.

[16] J. S. Russell, A. M. Rush, and A. Shrestha, "Dispute Causation in Highway Construction Projects," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2268, pp. 104–112, 2012.

[17] R. G. Virle and S. Y. Mhaske, "Monitoring of Construction Projects Using EVM and ESM Tools," *International Journal of Structural and Civil Engineering Research*, vol. 2, no. 4, pp. 128–135, 2013.