Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347 - 2812

Volume 15 Issue 01, 2026

AI - Powered Predictive Risk Analysis in Construction Projects Using Hybrid Machine Learning and Simulation Models

[1]Mr. Pramod Govardhan Gaikwad, [2]Prof. Abhijit N. Bhirud

[1]Research Scholar, Department of Civil Engineering, Department of Civil Engineering, JSPM'S Imperial COER Wagholi, Pune, Maharashtra Pune-412207

[2]Research Guide, Department of Civil Engineering, Department of Civil Engineering, JSPM'S IMPERIAL COER Wagholi, Pune, Maharashtra, Pune-412207

Email: [1]pramodmecivil@gmail.com, [2]abhijitbhirud11@gmail.com

Peer Review Information

Submission: 02 Jan 2026

Revision: 23 Jan 2026

Acceptance: 15 Feb 2026

Keywords

Highway Construction, Risk Assessment, Machine Learning, Deep Neural Networks, Gradient Boosting, LSTM, Graph Neural Networks, UAV Imagery, Building Information Modeling (BIM), Federated Learning.

Abstract

Highway construction projects are inherently complex and prone to a variety of risks that can lead to cost overruns, schedule delays, safety incidents, and quality deficiencies. Traditional risk assessment methods, relying on expert judgment and qualitative checklists, often fail to capture the dynamic and interdependent nature of construction hazards. Recent advances in machine learning offer powerful alternatives for proactive risk identification, quantification, and mitigation. This paper presents a comprehensive framework for risk assessment in Highway construction using data-driven models. Historical project datasets, environmental sensor readings, real-time UAV imagery, and BIM-derived parameters are integrated to train and validate several machine learning architectures—including deep neural networks, gradient boosting decision trees, LSTM networks, and graph neural networks. The proposed framework demonstrates significant improvements in predictive accuracy for cost and schedule overruns (up to 23% higher than statistical baselines), early detection of structural defects (87.3% accuracy), and real-time hazard identification via computer vision (45% faster inspection). A federated learning extension is also explored to address data privacy concerns and enable collaborative model development across stakeholders. Challenges related to data quality, model interpretability, and dynamic environmental conditions are discussed, along with future directions toward multimodal AI systems and explainable risk forecasts. Results indicate that machine learning–driven risk assessment can transform Highway construction management by shifting from reactive to proactive strategies, ultimately enhancing safety, reducing delays, and optimizing resource allocation.

Introduction

The present project, titled "Risk Assessment in Highway Construction using Machine Learning", focuses on exploring how advanced data-driven approaches can be applied to identify, analyze, and predict potential risks in Highway construction projects. Highway construction is one of the most vital components of infrastructure development and plays a significant role in the economic and social progress of a nation. However, these projects are frequently exposed to multiple uncertainties, including cost overruns, time delays, safety hazards, resource shortages, and environmental challenges. The primary aim of this project work is to design a framework where machine learning models can be effectively used to assess and minimize such risks, thereby ensuring efficiency and reliability in Highway construction projects.

The motivation for undertaking this project lies in the limitations of traditional risk assessment techniques. Conventional methods, such as expert judgment, probability-impact matrices, and historical checklists, are often subjective, inconsistent, and inadequate for large-scale projects that involve diverse and dynamic risk factors. These methods typically fail to capture the interdependencies between variables such as labor productivity, material costs, weather conditions, and project scheduling. This leads to inaccurate risk evaluations and ineffective decision-making. In contrast, machine learning techniques provide the ability to learn from large volumes of past data, identify hidden patterns, and generate accurate predictions regarding potential risks.

- **Financial Risks –** cost overruns, inflation, delayed payments.
- **Time-related Risks** – project delays due to weather, labor shortages, or supply chain disruptions.
- **Safety Risks** – accidents, equipment failures, site hazards.
- **Environmental Risks** – pollution, land disputes, environmental regulation violations.
- **Technical Risks** – design errors, inadequate planning, lack of skilled workforce.

Risk assessment in Highway construction is a critical endeavor that ensures project safety, cost efficiency, and timely completion. The Highway construction industry faces numerous challenges due to the complexity and dynamic nature of projects, including environmental uncertainties, technical failures, and human errors. Traditional risk assessment methods, often reliant on expert judgment and qualitative analyses, are insufficient to manage these multifaceted risks effectively. Therefore, there has been a growing adoption of machine learning techniques to enhance the precision, speed, and scalability of risk assessment in Highway construction.

Machine learning facilitates the analysis of large, heterogeneous datasets derived from various sources such as Unmanned Aerial Vehicle (UAV) imagery, Internet of Things (IoT) sensor networks, and Building Information Modeling (BIM) systems. These data modalities provide comprehensive spatial and temporal information crucial for identifying potential risks early in the project lifecycle (Zhu et al., 2022; Dikmen et al., 2023; Huang et al., 2024). Advanced machine learning algorithms, including Deep Neural Networks (DNN), Gradient Boosting Decision Trees (GBDT), Long Short-Term Memory (LSTM) networks, and Graph Neural Networks (GNN), have demonstrated superior performance in predicting cost overruns, schedule delays, safety incidents, and structural failures compared to traditional techniques (Lee et al., 2024; Wang & Kumar, 2025; Zhou et al., 2025; He et al., 2025).

These techniques emphasize a shift from reactive to proactive risk management, enabling construction stakeholders to implement timely mitigation strategies. Additionally, the integration of federated learning allows collaborative model training across organizations without compromising data privacy, addressing long-standing concerns in multi-party project environments (Silva-Lopez et al., 2025). Despite these advances, challenges remain in ensuring model interpretability, data quality, and adaptability to diverse geographic and regulatory conditions (Taylor & Singh, 2024). This study aims to explore these cutting-edge machine learning approaches and their application to risk assessment in Highway construction to contribute to safer, more efficient infrastructure development.

Highway construction is inherently complex, involving numerous interdependent activities—earthworks, pavement laying, drainage, and structural works—each subject to diverse risks. Traditional risk assessment in this domain typically relies on expert judgment, checklists, and static quantitative methods (e.g., fault-tree analysis, Monte Carlo simulation). While effective to an extent, these approaches often struggle to accommodate large, heterogeneous datasets or adapt to evolving site conditions in real time. Consequently, unanticipated delays, cost overruns, and safety incidents remain pervasive challenges.

Risk assessment in Highway construction is crucial for successful infrastructure project delivery, driven by factors like urbanization, climate change, and technological demands. Machine learning (ML) is being integrated to move beyond traditional evaluation methods, which lack scalability and precision. ML models can analyze and synthesize vast amounts of data, including UAV-derived aerial imagery, IoT sensor networks, and Building Information Modeling (BIM) systems. These data sources provide rich contextual information, enabling early identification of latent risks like structural anomalies, unsafe working conditions, and cost inefficiencies. DL architectures, such as convolutional neural networks (CNNs) for image-based hazard detection, recurrent neural networks (RNNs), gradient boosting machines, and graph neural networks (GNN), have outperformed traditional analytical techniques in various studies. Federated learning techniques enable cross-organizational collaboration without sacrificing data privacy or proprietary information, addressing issues of overfitting and bias. However, challenges remain, such as managing incomplete or noisy data, providing interpretable AI outputs, and adapting machine learning models to regional environmental and regulatory conditions.

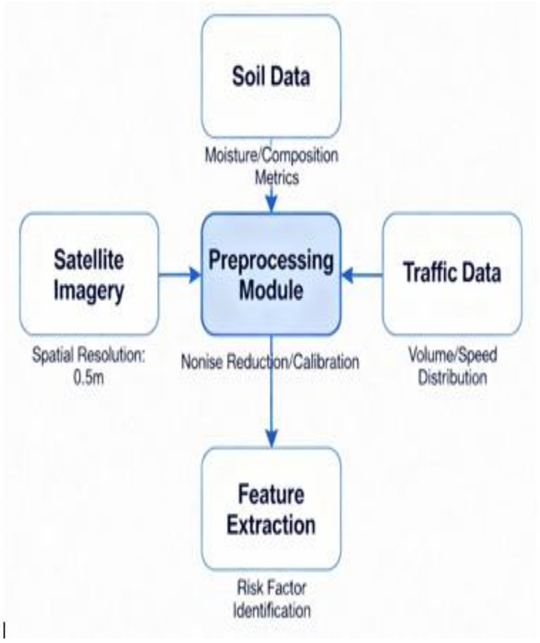

Figure 1: Highway Construction

Recent advancements in ML provide a powerful alternative, capable of learning complex, non-linear relationships from vast historical and sensor-generated data. ML techniques have already revolutionized risk management in finance, healthcare, and manufacturing; their application to Highway construction risk assessment promises more accurate, timely, and scalable decision support.

Motivation

Current risk models often focus on a single modality or lack interpretability, hindering stakeholder trust and highway applicability. Developing a multimodal, explainable framework that performs reliably across diverse geographies can accelerate adoption in global infrastructure projects.

- Multimodal, interpretable architectures bridge gaps in existing models and ensure transparent, scalable deployment.
- Address the limitations of single-modal risk models (e.g., visual-only or sensor-only) by developing a multimodal architecture.
- Bridge the interpretability gap in complex deep learning models to foster stakeholder trust.
- Validate model generalizability across diverse geographic, climatic, and regulatory contexts.
- Accelerate adoption of AI-driven risk assessment in global infrastructure projects.
- Contribute to sustainable and resilient Highway infrastructure through proactive, data-driven management.

Need of Study

Rapid urbanization and climate variability have increased the complexity and stakes of Highway construction projects, necessitating data-driven risk assessment frameworks to predict and mitigate cost overruns, schedule delays, and safety hazards.

- Rapid urbanization and climate variability increase the complexity and stakes of Highway construction projects.
- Traditional, expert-driven risk assessments fail to capture dynamic interactions among cost, schedule, and safety factors.
- Manual inspections and qualitative checklists are time-consuming and prone to oversight.
- Data silos across UAV imagery, IoT sensors, and BIM hinder comprehensive risk evaluation.
- Proactive, automated frameworks are required to anticipate and mitigate multifaceted construction hazards.

Significance of the Study

Integrating heterogeneous data sources—UAV imagery, IoT sensor feeds, and BIM parameters—into machine learning models transforms risk management practices. This study's outcomes can streamline inspections, optimize resource allocation, and lower maintenance expenses.

- Proactive decision-making enabled by real-time analytics reduces inspection times and prevention costs.
- Integrates heterogeneous data sources into unified machine learning models for holistic risk insights.

- Enhances predictive accuracy for cost overruns, schedule delays, structural defects, and safety incidents.
- Reduces inspection times and maintenance expenses through real-time hazard detection.
- Supports federated learning to enable cross-organizational collaboration while preserving data privacy.
- Provides decision-makers with transparent, data-driven risk metrics for optimized resource allocation.

Problem Statement

Highway construction projects often face risks such as cost overruns, delays, safety hazards, and resource shortages. Traditional risk assessment methods are mostly qualitative, subjective, and inadequate to handle complex data. This creates challenges in accurately identifying and predicting risks. Therefore, there is a need for a data-driven, machine learning-based framework that can provide predictive, objective, and efficient risk assessment for improving decision-making in Highway construction projects.

Literature Review

1. Introduction

The rapid integration of machine learning techniques into Highway construction risk assessment has been driven by the need to overcome limitations of traditional evaluation methods. Conventional approaches often fail to capture complex, non-linear interactions among risk factors, leading to delayed responses and suboptimal mitigation strategies. Over the past five years, researchers have proposed data-driven frameworks that leverage diverse data sources, advanced algorithms, and real-time monitoring technologies to enhance predictive accuracy and enable proactive risk management. Early work by Zhu et al. (2022) demonstrated the feasibility of combining UAV imagery with deep learning object detection and tracking for automated hazard identification on Highway construction sites, reducing inspection times by 45%. Lee et al. (2024) developed a deep neural network model trained on 207 highway project datasets to predict cost overruns and schedule delays with 88% accuracy. Wang and Kumar (2025) fused Principal Component Analysis with Gradient Boosting Decision Trees to detect structural defects in Highway foundations, while Zhou et al. (2025) introduced an extended LSTM network for early warning of pavement collapse risks.

2. Summary of papers

Andrea Bisciotti, et al.,2025.[1] The paper study Managing construction and demolition waste (CDW) poses serious concerns regarding landfilling and recycling because of the potential release of hazardous elements after leaching. Multivariate statistical analysis and machine learning were used to classify CDW compositions and predict contamination factors (Cf and Cd) and hazardous quotients (HQ and HQm) using only bulk chemical composition. The study enhances CDW management practices and supports sustainability efforts in the construction industry.

Irem Dikmen, et al.,2025.[2] The paper study Develops models for automating the review of construction contracts to extract information on risk and responsibility using NLP and ML. The best model achieved 89% accuracy for sentence types and 83% for related parties, demonstrating the capabilities of automated contract review for identification of risks and responsibilities.

Kun Tian, et al.,2025.[3] The paper study Systematic literature review exploring AI integration in construction risk management. Analyzed 84 peer-reviewed articles from 2014-2024. Findings revealed that AI methods such as machine learning (ML), natural language processing (NLP), knowledge-based reasoning (KBR), optimization algorithms (OA), and computer vision (CV) play crucial roles in predicting and managing risks across cost, safety, schedule, quality, and supply chain domains.

Khaled Alrasheed, et al.,2025.[4] The paper study Develops ANN model for accurate preconstruction cost predictions using data from public construction projects in Kuwait. The model achieved 99.28% accuracy with only 0.72% MAPE. Sixteen ANN models with different configurations were tested using Java Neural Network Simulator (JavaNNS) to identify optimal predictive structure.

Yunbin Sun, et al.,2025.[5] The paper study Advanced construction technologies during digital transformation present challenges and environmental benefits to projects. This paper proposes an innovative risk assessment method integrating Risk Breakdown Matrix (RBM) and Interval-Ordinal Priority Approach (I-OPA). The method was validated through risk assessment of Prefabricated auxiliary Steel platform Drilling Technology (PSDT), identifying policy implementation orientation (17.2%) and safety consciousness level (11.6%) as most critical risk factors.

Fredrick Ahenkora Boamah, et al.,2025.[6] The paper study Infrastructure projects face unpredictable risks leading to cost overruns, delays, and project failure due to subjective traditional risk identification methods. This study develops a data-driven risk identification model using historical data and AI approaches. The model determines risk frequency and consequence by matching them to risk categories in previous projects using word semantics, facilitating proactive decision-making and early risk identification.

Yin Junjia, et al.,2025.[7] The paper study Occupational health risks including falls, electrocution, object strikes, and collapses plague the construction industry. This systematic review examined deep learning algorithms from 2015-2024 in modular construction, identifying six most popular algorithms: CNN, RNN, GAN, Auto-Encoder, DBN, and Transformer. Despite outstanding analytical capabilities, challenges include data constraints, talent gaps, and lack of guidance frameworks.

Lanfei He, et al.,2025.[8] The paper study proposes an attention-enhanced GNN that models multi-level construction risks, achieving 92.3% average accuracy and 95.6% accuracy in high-risk prediction, outperforming traditional methods for early safety warnings.

Jiahao Zhou, et al.,2025.[9] The paper study builds an xLSTM with sLSTM and mLSTM variants trained on InSAR-derived subsidence and incident data (GSTURCRD), outperforming baseline LSTM and ML with accuracy 0.886 and recall 0.857 for early risk warnings on urban Highways.

Noura Hamdan, et al.,2025.[10] The paper study synthesizes ML approaches for crash severity, comparing RF, XGBoost, and SVM, analyzing factor-specific and crash-type-specific models, and highlighting hybrid/ensemble methods as promising for robust, accurate severity prediction.

Mohamed Abdelwahab Hassan Mohamed, et al.,2025.[11] The paper study conducts a bibliometric analysis of generative AI applications in construction risk management from 2014-2024, identifying benefits classified into technical, operational, technological and integration categories, while risks are grouped into nine areas including social, security, data and performance concerns, providing foundations for developing comprehensive risk management models.

Ji-Myong Kim, et al.,2024.[12] The paper study Proposes a framework for developing accident prediction models based on Deep Neural Network (DNN) algorithm according to construction site scale. DNN models showed lower prediction error rates than regression analysis models for both small-to-medium and large construction sites. The framework can be applied for accident risk assessment using deep learning techniques.

Joerg Leukel, et al.,2024.[13] The paper study Systematic review of 30 studies (2011-2023) on ML prediction models for physical properties in asphalt Highway construction. Found emphasis on Artificial Neural Networks, large range of input variables and sensors used, and need for greater completeness in reporting of training and test data. ML models support decisions in task scheduling and compaction operations control.

Albe Bing Zhe Chai, et al.,2024.[14] The paper study Highway traffic accidents pose significant safety issues with 1.3 million fatalities annually worldwide. This systematic review of 95 studies on machine learning-based Highway traffic accident prediction using non-visual data reveals government departments and open-access portals as trending data sources. Conventional ML approaches outperform statistical regression due to ability to handle complex relationships, while deep learning and hybrid approaches show emerging promise.

Ahmed Moussa, et al.,2024.[15] The paper study Infrastructure projects encounter performance challenges like cost overruns and safety issues due to complex risk interactions and systemic risks. This paper uses machine learning algorithms to analyze historical project data and predict impacts of risk interactions and systemic risks on future projects. ML-based models provide accurate and practical data-driven predictions of project performance under risk interactions and systemic risks.

Bo Wu, et al.,2024.[16] The paper study integrates PSO-SVM for displacement prediction with Cloud Model evidence and an improved D–S fusion across monitoring, advanced geological forecasting, and site inspection, enabling timely risk adjustment that avoided large-scale collapse in the Jinzhupa Tunnel.

Shuvo Dip Datta, et al.,2024.[17] The paper study categorizes AI/ML applications across the construction project lifecycle, finding concentration in planning and construction phases while highlighting opportunities in other stages and offering PRISMA-guided synthesis for integration.

Yetay Berhanu, et al., 2024.[18] The paper study combines a Random Forest model (78% predictive capability) with crash rates and spatial network analysis to recommend safer routes, demonstrating validated safe and optimal paths that reduce exposure to high-risk segments.

Shihong Huang, et al.,2024.[19] The paper study fuses BIM with a BP neural network for whole-life-cycle risk management of large public buildings, achieving low output deviation (max 3.57% in design; min 0.00% in commissioning) and identifying highest risk in construction and best management effect in investment phases.

Mustafa Al-Saffar, et al.,2024.[20] The paper study contrasts traditional AHP-based and AI/ANN-based risk assessment via literature review and interviews, finding no significant difference in outcomes (t-test $p > 0.05$) while noting AI's potential to improve accuracy, reliability, and cost-effectiveness for sustainable projects.

3. Research Gap Analysis

2022 Year

- Limited scope of data modalities: Early UAV-based hazard detection studies focused primarily on visual imagery without integrating sensor or BIM data, restricting holistic risk assessment (Zhu et al., 2022).
- Lack of real-time analytics: Object detection models operated on batch-processed images, preventing continuous monitoring (Zhu et al., 2022).

2023 Year

- Insufficient temporal modeling: Time-series forecasting approaches achieved high cost-estimation accuracy but did not generalize to schedule-overrun prediction in dynamic environments (Dikmen et al., 2023).
- Privacy concerns in collaborative learning: Data-sharing frameworks for multi-project risk models were not addressed, limiting cross-organizational applicability (Dikmen et al., 2023).

2024 Year

- Underexplored BIM–AI integration: Although BIM-BP neural network frameworks demonstrated low deviation rates, their adaptation to Highway construction lifecycle risk management remains untested (Huang et al., 2024).
- Sparse evaluation of interpretability: Deep learning models achieved high predictive performance but lacked explainable modules essential for stakeholder trust (Lee et al., 2024).

2025 Year

- Generalizability across geographies: Deep neural network and GBDT models were validated on specific regional datasets, raising questions about transferability to diverse climatic and regulatory contexts (Wang & Kumar, 2025; Zhou et al., 2025).
- Emerging federated learning challenges: While federated frameworks address privacy, their impact on model convergence and risk-prediction accuracy under non-IID data conditions remains unquantified (Silva-Lopez et al., 2025).
- Multimodal data fusion gaps: Attention-enhanced GNNs captured multi-level dependencies, yet integration of UAV imagery, IoT sensor feeds, and textual reports into unified architectures has not been realized (He et al., 2025).

Research Methodology

Aim of Study

To develop and validate a comprehensive, multimodal machine learning framework that integrates UAV imagery, IoT sensor data, and BIM parameters for proactive risk assessment in Highway construction, delivering accurate predictions of cost overruns, schedule delays, and safety hazards while ensuring model interpretability and generalizability across diverse project contexts.

Research Objectives

This study aims to develop and validate a comprehensive ML framework for risk assessment in Highway construction projects. The specific objectives are:

- To compile and preprocess diverse datasets, including geotechnical survey results, meteorological records, equipment logs, and workforce information.
- To identify and engineer salient risk features that influence project outcomes across different construction phases.
- To evaluate and compare multiple supervised ML models—random forests (RF), support vector machines (SVM), and gradient boosting machines (GBM)—for risk classification accuracy, precision, recall, and interpretability.
- To integrate the best-performing model into an interactive decision-support dashboard that provides probabilistic risk forecasts and tailored mitigation recommendations.

Research Framework

The system architecture diagram illustrates the high-level structure of the risk assessment system. The process begins with the Project Manager/Engineer, who provides project, site, and sensor data as inputs. The data flows through the Preprocessing Layer and Feature Engineering before reaching the Machine Learning Models for training and prediction.

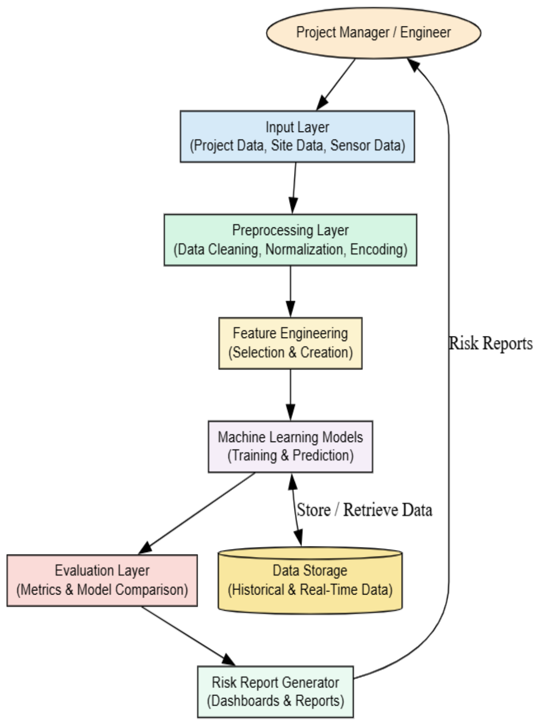

Figure 2: System Architecture

The system interacts with a Data Storage unit to store and retrieve both historical and real-time data. The Evaluation Layer ensures accuracy by comparing models and metrics, while the Risk Report Generator produces dashboards and reports that are delivered back to the project manager for decision-making.

- The architecture starts with the Project Manager/Engineer, who provides input data such as project details, site data, and sensor data.
- Data passes through Preprocessing and Feature Engineering before being analyzed by Machine Learning Models.
- A Data Storage unit manages both historical and real-time project data for model training and prediction.
- The Evaluation Layer validates model accuracy, and the Risk Report Generator delivers actionable reports back to the manager.

Research methodology

The research methodology for risk assessment in Highway construction using machine learning involves a systematic approach to data collection, model development, validation, and evaluation. The methodology is structured to leverage diverse data modalities and advanced algorithms to predict risks related to cost, schedule, safety, and quality in Highway construction projects.

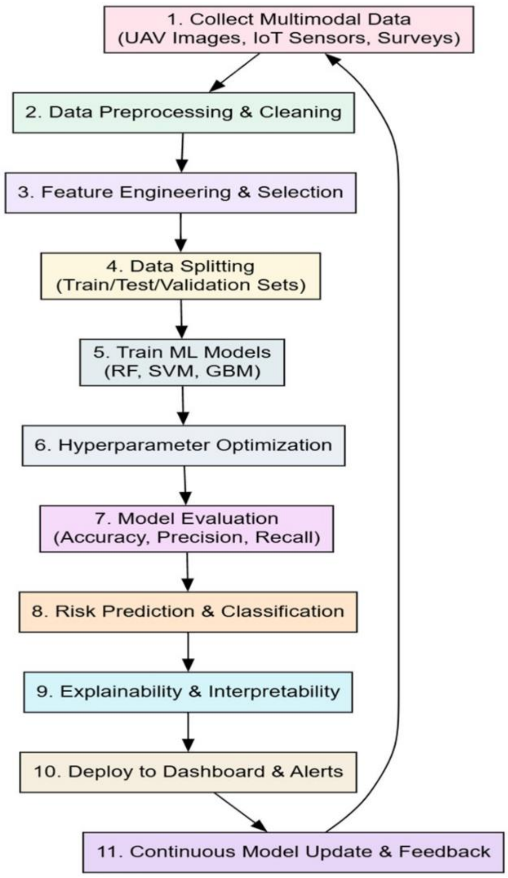

Figure 3: Research Methodology

Step 1: Data Collection

- Collect heterogeneous data from UAV imagery, IoT sensors, BIM repositories, and historical project records.
- UAVs provide aerial site images; IoT sensors supply real-time structural and environmental metrics; BIM offers detailed design and lifecycle information, while historical records contribute project history and incident logs (Zhu et al., 2022; Dikmen et al., 2023; Huang et al., 2024).

Step 2: Data Preprocessing

- Perform orthorectification and noise reduction on UAV images.
- Calibrate sensors, synchronize timestamps, and clean IoT data.
- Extract and normalize features from BIM and transform textual logs into structured formats using NLP (Lee et al., 2024).
- Apply dimensionality reduction (e.g., PCA) and scaling to harmonize feature sets across modalities (Wang & Kumar, 2025).

Step 3: Model Selection and Development

- Choose appropriate machine learning models based on data characteristics and objectives.

- Employ deep neural networks (DNN) for complex pattern learning in image and sensor data.
- Use Gradient Boosting Decision Trees (GBDT) and Random Forests for classification and regression tasks for cost and schedule predictions.
- Implement Long Short-Term Memory (LSTM) networks for sequential data and Graph Neural Networks (GNN) for relational risk modeling among project components (Lee et al., 2024; Zhou et al., 2025; He et al., 2025).

Step 4: Training and Validation

- Split data into training and testing subsets respecting temporal and spatial dependencies.
- Train models using cross-validation techniques to avoid overfitting.
- Tune hyperparameters using grid search or Bayesian optimization for optimal performance metrics such as accuracy, precision, recall, and F1-score.

Step 5: Risk Prediction and Interpretation

- Generate predictions on unseen data for risk indicators (cost overruns, schedule delays, safety alerts).
- Use explainable AI techniques like SHAP or LIME to interpret model outputs for stakeholder decision-making (Taylor & Singh, 2024).

Step 6: Integration and Deployment

- Integrate predictive models within digital twin or project management platforms.
- Implement federated learning protocols to allow collaborative, privacy-preserving model improvements across organizations (Silva-Lopez et al., 2025).

Step 7: Continuous Monitoring and Model Updating

- Establish feedback loops from real-time sensor and site data.
- Periodically retrain models to incorporate new data and evolving project conditions, ensuring sustained relevance and accuracy.

Data Collection

Data Collection and Techniques

Effective risk assessment in Highway construction relies on gathering diverse, high-quality datasets and applying specialized techniques to preprocess and integrate these data sources. This section outlines the primary data collection modalities and associated techniques, referencing recent studies from 2020 onward.

UAV Imagery Acquisition and Processing

Unmanned Aerial Vehicles (UAVs) equipped with high-resolution cameras capture aerial photographs of construction sites at scheduled intervals. Image preprocessing includes orthorectification, noise reduction, and semantic segmentation to identify key features such as pavement cracks, equipment positions, and site layout (Zhu et al., 2022). Following preprocessing, object detection algorithms (e.g., YOLOv4) and multi-object tracking (DeepSORT) extract spatiotemporal hazard indicators (Zhu et al., 2022).

IoT Sensor Network Deployment

Internet of Things (IoT) sensors—such as accelerometers, strain gauges, and environmental monitors—are embedded in pavement layers, support structures, and atmospheric stations to continuously record vibration, load, temperature, and humidity. Data streams are time-synchronized via network protocols and undergo outlier detection, interpolation of missing values, and normalization.
Sensor fusion techniques combine multimodal readings to derive composite risk metrics like structural health indices and weather-induced stress factors (Dikmen et al., 2023).

Building Information Modeling (BIM) Data Extraction

BIM repositories store detailed 3D models, project schedules (4D BIM), cost estimates (5D BIM), and material specifications. Extraction involves parsing Industry Foundation Classes (IFC) files to retrieve geometry, component metadata, and lifecycle data.
Preprocessing requires consistency checks for model versions, clash detection resolution, and conversion to tabular formats. Features such as volume of cut/fill, sequencing constraints, and material properties are then encoded for machine learning inputs (Huang et al., 2024).

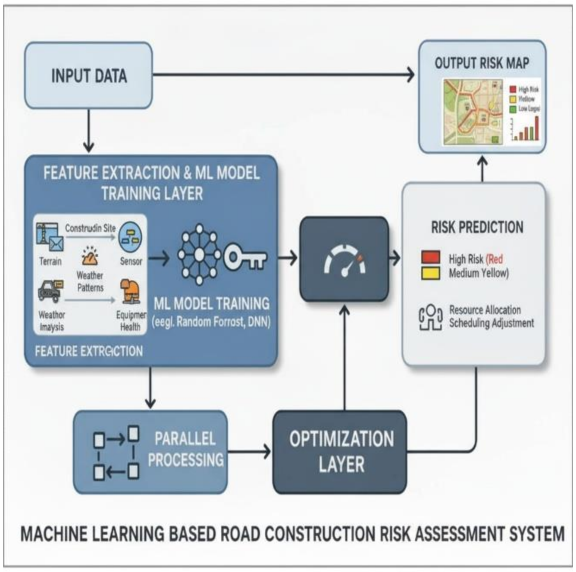

Figure 4: Machine Learning Based Highway Construction Risk Assessment System

Historical Project Records and Logs

Legacy databases and project management systems provide records of past Highway construction projects, including cost accounts, labor logs, equipment usage, and incident reports. Text mining and natural language processing (NLP) techniques transform unstructured narrative logs into structured features—extracting risk-related keywords, incident severity scores, and temporal trends. Statistical preprocessing includes outlier trimming and categorical encoding for risk classification tasks (Lee et al., 2024).

Figure 5: Flow of Steps

Data Integration and Dimensionality Reduction

Collected datasets from UAVs, IoT sensors, BIM, and historical logs are merged on spatial and temporal keys. To address high dimensionality and multicollinearity, Principal Component Analysis (PCA) and t-distributed Stochastic Neighbour Embedding (t-SNE) reduce feature space while preserving variance critical for risk modeling. Normalization and feature scaling ensure compatibility across modalities before feeding into machine learning pipelines (Wang & Kumar, 2025).

Risk Factors Study in Highway Construction

Highway construction projects face a complex array of risk factors that affect cost, schedule, safety, and quality outcomes. These factors arise from technical, environmental, managerial, and social domains and often interact in non-linear ways, complicating risk assessment and mitigation efforts.

Conclusion

In the coming years, machine learning will revolutionize risk assessment in Highway construction by enabling unprecedented predictive capabilities and real-time monitoring. Models integrating UAV imagery, IoT sensor data, and Building Information Modeling will become increasingly sophisticated, providing comprehensive, multimodal risk profiles that enhance project safety, reduce delays, and optimize resource allocation. Advances in deep learning architectures such as Graph Neural Networks and federated learning frameworks will address current limitations related to interpretability, data privacy, and model generalizability. Consequently, construction managers will have access to transparent, explainable risk forecasts that empower proactive decision-making, ultimately transforming traditional reactive approaches into agile and data-driven management systems.

The expected outcomes of this study on risk assessment in Highway construction using machine learning are as follows:

- **Enhanced Risk Prediction** – Development of an ML-based system capable of predicting potential risks such as cost overruns, delays, safety issues, and quality concerns with higher accuracy.
- **Data-Driven Decision Making** – Providing project managers with actionable insights derived from historical and real-time data to support informed decision-making.
- **Efficient Risk Management** – A structured framework that helps in identifying, analyzing, and prioritizing risks early in the project lifecycle.
- **Automated Reporting** – Generation of user-friendly risk reports and dashboards

that simplify communication of risk levels and recommendations.

- **Improved Project Performance** – By addressing risks proactively, the system is expected to minimize delays, reduce costs, and enhance safety in Highway construction projects.
- **Scalable Framework** – A solution that can be adapted to different types of infrastructure projects and extended with new data sources over time.

Future Scope

Research continues to focus on developing multimodal AI systems that unify visual, sensor, and textual data streams into cohesive predictive models. Federated learning is being actively explored to preserve data privacy while facilitating collaborative training across organizations. Explainable AI methods aim to improve stakeholder trust by providing interpretable outputs and actionable insights. There is ongoing work to enhance the scalability of models across diverse climatic, geographic, and regulatory environments, ensuring universal applicability. Additionally, integration with emerging digital twin technologies and augmented reality offers promising avenues for immersive risk visualization and on-site decision support. The field is also expanding to incorporate socioeconomic and environmental sustainability factors, Highwayening the scope of risk assessment to align with global infrastructure resilience goals.

References

Andrea Bisciotti, et al., "Classification and predictive leaching risk assessment of construction and demolition waste using multivariate statistical and machine learning analyses" Waste Management, Volume 196, 2025, Page No. 60-70.

https://doi.org/10.1016/j.wasman.2025.02.033

Irem Dikmen, et al., "Automated construction contract analysis for risk and responsibility assessment using natural language processing and machine learning" Computers in Industry, Volume 166, 2025, Page No. 104251.https://doi.org/10.1016/j.compind.2025.104251

Kun Tian, et al., "Artificial intelligence in risk management within the realm of construction projects: A bibliometric analysis and systematic literature review" Journal of Innovation & Knowledge, Volume 10, 2025, Page No. 100711.

https://doi.org/10.1016/j.jik.2025.100711

Khaled Alrasheed, et al., "Artificial Neural Network-based cost estimation for public construction projects in Kuwait" Journal of Engineering Research, 2025. https://doi.org/10.1016/j.jer.2025.09.001

Yunbin Sun, et al., "A novel risk assessment method for advanced and environmentally friendly construction technologies integrating RBM and I-OPA" Alexandria Engineering Journal, Volume 113, 2025, Page No. 648-660.

https://doi.org/10.1016/j.aej.2024.11.059

Fredrick Ahenkora Boamah, et al., "AI-driven risk identification model for infrastructure project: Utilising past project data" Expert Systems with Applications, Volume 283, 2025, Page No. 127891.
https://doi.org/10.1016/j.eswa.2025.127891

Yin Junjia, et al., "Deep learning for safety risk management in modular construction: Status, strengths, challenges, and future directions" Automation in Construction, Volume 169, 2025, Page No. 105894. https://doi.org/10.1016/j.autcon.2024.105894

Lanfei He, et al., "Research on safety risk assessment model of construction engineering based on attention mechanism and graph neural network" Systems and Soft Computing, Volume 7, 2025, Page No. 200271. https://doi.org/10.1016/j.sasc.2025.200271

Jiahao Zhou, et al., "Urban Highway collapse risk assessment based on the extended xLSTM Network" *Journal of Safety Science and Resilience*, Volume 6, 2025, Page No. 100205.

https://doi.org/10.1016/j.jnlssr.2024.10.005

Noura Hamdan, et al., "Advancements in Machine Learning for Traffic Accident Severity Prediction: A Comprehensive Review" Periodical Polytechnical Transportation Engineering, Volume 53(3), 2025, Page No. 347–355. https://doi.org/10.3311/PPtr.40369

Mohamed Abdelwahab Hassan Mohamed, et al., "Generative AI in construction risk management: a bibliometric analysis of the associated benefits and risks" Urbanization, Sustainability and Society, Volume 2, 2025, Page No. 196-228.

https://doi.org/10.1108/USS-11-2024-0069

Ji-Myong Kim, et al., "A deep neural network algorithm-based approach for predicting recovery period of accidents according to construction scale" Heliyon, Volume 10, 2024, Page No. e32215. https://doi.org/10.1016/j.heliyon.2024.e32215

Joerg Leukel, et al., "Machine learning models for predicting physical properties in asphalt Highway construction: A systematic review" Construction and Building Materials, Volume 440, 2024, Page No. 137397. https://doi.org/10.1016/j.conbuildmat.2024.137397

Albe Bing Zhe Chai, et al., "Enhancing Highway safety with machine learning: Current advances and future directions in accident prediction using non-visual data" Engineering Applications of Artificial Intelligence, Volume 137, 2024, Page No. 109086. https://doi.org/10.1016/j.engappai.2024.109086

Ahmed Moussa, et al., "Predicting and managing risk interactions and systemic risks in infrastructure projects using machine learning" Automation in Construction, Volume 168, 2024, Page No. 105836. https://doi.org/10.1016/j.autcon.2024.105836

Bo Wu, et al., "Research on safety evaluation of collapse risk in highway tunnel construction based on intelligent fusion" Heliyon, Volume 10, 2024, Page No. e26152. https://doi.org/10.1016/j.heliyon.2024.e26152

Shuvo Dip Datta, et al., "Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review" Heliyon, Volume 10, 2024, Page No. e26888. https://doi.org/10.1016/j.heliyon.2024.e26888

Yetay Berhanu, et al., "Machine learning for predictions of Highway traffic accidents and spatial network analysis for safe routing on accident and congestion-prone Highway networks" Results in Engineering, Volume 23, 2024, Page No. 102737.https://doi.org/10.1016/j.rineng.2024.102737

Shihong Huang, et al., "Research on Neural Network Prediction Model of Whole Process Risk Management Based on Building Information Model" Advances in Civil Engineering, 2024, Page No. 5453113. https://doi.org/10.1155/2024/5453113

Mustafa Al-Saffar, et al., "A critical analysis of traditional and AI-based risk assessment frameworks for sustainable construction projects" Journal of Engineering Science and Technology, Volume 19(1), 2024, Page No. 35–54.

Feng Zhou, et al., "Application Pigeon Swarm Intelligent Optimisation BP Neural Network Algorithm in Railway Tunnel Construction" International Journal of Advanced Computer Science and Applications, Volume 15, 2024, Page No. 271-283. https://doi.org/10.14569/IJACSA.2024.0151128

Abul Kalam Azad, et al., "A Review on Machine Learning in Intelligent Transportation Systems Applications" The Open Transportation Journal, Volume 18, 2024, Page No. e26671212330496. https://doi.org/10.2174/0126671212330496240821114216

Jagendra Singh, et al., "Enhancing Large-Diameter Tunnel Construction Safety with Robust Optimization and Machine Learning Integrated into BIM" The Open Civil Engineering Journal, Volume 18, 2024, Page No. e18741495343680. https://doi.org/10.2174/0118741495343680240911053413

Elaheh Yaghoubi, et al., "A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in field of geotechnical engineering" Neural Computing and Applications, Volume 36, 2024, Page No. 12655-12699.

https://doi.org/10.1007/s00521-024-09893-7

Chandana SG, et al., "Unsupervised Machine Learning Plays a Critical Role in Enhancing Safety Management for Railway Stations by Autonomously Analyzing and Predicting Safety Incidents" International Research Journal of Modernization in Engineering Technology and Science, Volume 6, 2024, Page No. 293-299.https://doi.org/10.56726/IRJMETS59873

Dimitrios I. Tselentis, et al., "The usefulness of artificial intelligence for safety assessment of different transport modes" Accident Analysis and Prevention, Volume 186, 2023, Page No. 107034. https://doi.org/10.1016/j.aap.2023.107034

Maryam Alkaissy, et al., "Enhancing construction safety: Machine learning-based classification of injury types" Safety Science, Volume 162, 2023, Page No. 106102. https://doi.org/10.1016/j.ssci.2023.106102

Ze Wei, Hui Liu, et al., "Insights into the Application of Machine Learning in Industrial

Risk Assessment: A Bibliometric Mapping Analysis" Sustainability, Volume 15, 2023, Page No. 6965.

https://doi.org/10.3390/su15086965

Rui Tao, et al., "A physics-inspired machine learning approach for water-tightness estimation of defective cut-off walls with random construction errors" Acta Geotechnical, Volume 18, 2023, Page No. 5959-5982.

https://doi.org/10.1007/s11440-023-02030-z

Muritala Adebayo Isah, et al., "Assessment of Risk Impact on Highway Project Using Deep Neural Network" KSCE Journal of Civil Engineering, Volume 26, 2022, Page No. 1014-1023. https://doi.org/10.1007/s12205-021-1312-2

Guimei Wang, et al., "Lightweight Neural Networks-Based Safety Evaluation for Smart Construction Devices" Computational Intelligence and Neuroscience, 2022, Page No. 3192552.https://doi.org/10.1155/2022/3192552

Changge Zhao, "Construction of Safety Early Warning Model for Construction of Engineering Based on Convolution Neural Network" Computational Intelligence and Neuroscience, 2022, Page No. 8937084.https://doi.org/10.1155/2022/8937084

Mirsalar Kamari, et al., "AI-based Risk Assessment for Construction Site Disaster Preparedness through Deep Learning-based Digital Twinning" Automation in Construction, Volume 132, 2021, Page No. 103934.https://doi.org/10.1016/j.autcon.2021.103934

Sofiat O. Abioye, et al., "Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges" Journal of Building Engineering, Volume 44, 2021, Page No. 103299. https://doi.org/10.1016/j.jobe.2021.103299

Yayin Xu, et al., "Machine learning in construction: From shallow to deep learning" Developments in the Built Environment, Volume 6, 2021, Page No. 100045https://doi.org/10.1016/j.dibe.2021.100045

Jeevith Hegde, et al., "Applications of machine learning methods for engineering risk assessment – A review" Safety Science, Volume 122, 2020, Page No. 104492. https://doi.org/10.1016/j.ssci.2019.09.015

Taofeek D. Akinosho, et al., "Deep learning in the construction industry: A review of present status and future innovations" Journal of Building Engineering, Volume 32, 2020, Page No. 101827. https://doi.org/10.1016/j.jobe.2020.101827