



Risk Severity Quantification and Prediction Framework for Road Construction Projects

Anietie Uwem Edet* and Ben U. Ngene

Received: October 27, 2024 / Accepted: December 28, 2024

© REJOST 2024

Abstract

A road construction project involves the planning, design, and construction of roads, highways, and related infrastructure, such as bridges, tunnels, and drainage systems. Risks in road construction arise from various factors, including adverse weather conditions (e.g., heavy rain or flooding), soil instability, labor shortages, equipment failures, delays in material supply, safety hazards, and regulatory compliance issues. Managing these risks is crucial for ensuring the timely completion and safety of the project. This research addresses the critical challenges in risk management for road construction projects, where unforeseen risks often lead to delays, budget overruns, and safety hazards. The aim of the study is to develop a robust model for predicting and quantifying the severity of risks encountered in road construction projects, enabling better preparedness and decision-making. The research utilizes machine learning techniques, specifically the Multinomial Naive Bayes (MNB) classifier, to analyze and predict various risk categories such as Cost Overruns, Flooding, Design Errors, and Environmental Impact. The model was built using historical data on past construction projects, providing insights into the likelihood of risk events and their potential severity. The methodology involved data preprocessing, risk categorization, and the application of the MNB model to generate accurate predictions of risk occurrences. Evaluation metrics, including precision, recall, and F1-score, showed that the model performed effectively in predicting most of the risk categories with a high degree of accuracy of 93%. The results of the study indicate that the model can be a valuable tool for identifying and addressing critical risks like Community Opposition and Corrosion of Structures, allowing civil engineers and construction managers to allocate resources more effectively and mitigate risks before they escalate. We recommend the adoption of this risk prediction model by civil engineering teams, risk management units, and construction workers.

✉ Anietie Uwem Edet: anietieuwemedet@gmail.com

Department of civil engineering, Michael Okpara University of Agriculture Umudike, Nigeria

Keyword: Road Construction, Risk, Severity, Civil Engineering, Machine Learning

1.0 Introduction

Highway construction projects are intricate undertakings that involve substantial planning, resource allocation, and execution across diverse terrains and conditions. However, these projects are inherently fraught with risks that can stem from environmental (Ekong *et al.*, 2023), structural, and logistical challenges. A Risk Severity Analysis and Prediction Framework offers a structured approach to identifying, evaluating, and mitigating potential issues, ensuring that projects proceed with minimal disruptions. By systematically analyzing the severity of risks, such a framework equips engineers and project managers with actionable insights, enabling them to make informed decisions (Davor *et al.*, 2020) at every stage of construction. The foundation of this framework lies in a comprehensive evaluation of potential risks specific to highway construction. Environmental factors, such as soil composition, hydrological conditions, and climatic variability, play a significant role in influencing construction timelines and methodologies. For example, unstable soil conditions may necessitate additional groundwork or the inclusion of stabilizing materials, while regions prone to flooding require the integration of elevated roadbeds or efficient drainage systems. Proper risk evaluation ensures that these challenges are anticipated and addressed during the planning phase, rather than reacting to issues as they arise. Furthermore, the framework emphasizes the importance of assessing the structural risks associated with material performance and design specifications (Edet *et al.*, 2024). Every highway project demands a detailed understanding of the loads and stresses that the structure will endure over its lifetime. Factors such as traffic volume, vehicle weight, and long-term wear must be factored into the analysis after carrying out a moderate damage assessment (Pham Bao *et al.*, 2023). For instance, improperly selected materials or designs that fail

to account for projected traffic increases can lead to premature deterioration, safety hazards, and costly repairs. A proactive framework ensures that the selection of construction materials and design methodologies aligns with the functional and safety requirements of the highway (Le Duc *et al.*, 2023). Additionally, logistical and operational risks are integral components of the framework. Supply chain disruptions, labor shortages, and unforeseen site conditions are all common challenges in highway construction. These factors often lead to delays and budget overruns if not managed effectively (Ebong *et al.*, 2024). By predicting the likelihood and impact of such risks, the framework supports the implementation of contingency plans and resource optimization strategies. This holistic approach to risk severity analysis integrates environmental, structural, and operational considerations, providing civil engineers with a robust foundation for executing safe, efficient, and sustainable highway projects (Udoetor *et al.*, 2024). Highway and road construction projects face a wide array of risks that can significantly impact timelines, costs, and safety. Environmental risks are among the most common challenges in such projects. For example, adverse weather conditions like heavy rainfall can cause delays by halting operations or damaging partially completed structures (Le-Ngoc *et al.*, 2023). In regions with unstable soil or high-water tables, poor ground conditions may lead to unexpected subsidence, requiring additional stabilization measures such as pile foundations or soil reinforcement. Natural disasters, such as floods or landslides, pose severe risks, particularly in areas with inadequate drainage systems or steep terrains. These environmental uncertainties can result in increased costs and delays if not adequately anticipated during the planning and design phases. Another category of risk in highway construction relates to materials and structural challenges (Xianbo *et al.*, 2023). Substandard

materials, whether due to poor quality or inappropriate selection for the specific conditions, can lead to premature road failures, such as cracking, potholes, or even structural collapse in bridges or overpasses. For instance, using asphalt that cannot withstand high-temperature variations may lead to rutting or deformation under heavy traffic. Structural design flaws, such as underestimating traffic load requirements, can compromise the durability and safety of the road. Additionally, logistical risks, such as supply chain interruptions or labor shortages, can disrupt the timely delivery of materials and completion of tasks, compounding the severity of other challenges. Together, these risks illustrate the complexity of highway construction projects and highlight the necessity for thorough risk assessment and mitigation strategies.

2.0 Literature Foundation

Cheng *et al.* (2024): This study analyzed the influence of road determinants on crash severity using crash data from Texas, USA (January 2020–April 2021). Utilizing 12 road-related features (e.g., shoulder type and width, curb type) and 10 additional features (e.g., weather, illumination), three crash severity levels - Minor, Moderate, and Severe Damage were predicted. Four classification algorithms - Decision Tree, Support Vector Machines, Multi-Layer Perceptron, and XGBoost were compared, with XGBoost outperforming others, achieving 82.65% accuracy, 0.83 precision, 0.82 recall, and 0.82 F1 score. SHAP and partial dependence plots revealed median width as the most influential feature, with paved shoulders and curbs associated with higher severity. These insights support improved road safety design and planning.

Rodionova *et al.* (2022): This study examines factors influencing road crash severity in Saint Petersburg, Russia, from 2015 to 2021, focusing on lighting, weather, infrastructure, human factors, accident types, and vehicle characteristics. Run-off-road crashes emerged as

the most impactful accident type, contributing to an 11.2% increase in fatal accidents. Among infrastructure-related issues, road barrier shortcomings accounted for a 2.8% rise in fatalities, while inadequate road lighting had the highest impact, increasing fatal outcomes by 12.6%. These findings highlight critical areas for intervention and provide a foundation for developing targeted road safety policies in Saint Petersburg.

Khanum *et al.* (2023): This study addresses the critical issue of road accident severity in India, where road incidents claim over 150,000 lives annually. Using data from MoRTH and NHAI, a random forest-based predictive model was developed to classify accident severity on Indian highways. The model, with optimized hyperparameters, achieved a training accuracy of 67% and a test accuracy of 41.47%, revealing potential dataset imbalances and limitations. While no significant patterns were observed between the day of the week and accident severity, the findings emphasize the need for improved data collection systems and refined machine learning techniques to enhance prediction accuracy and support road safety strategies.

Ventura *et al.* (2024): This study highlights the importance of forecasting the severity of events where design traffic load limits on bridges are exceeded (DLEEs) to enhance road safety. While previous research has assessed failure consequences using cost metrics, this work explores severity prediction using Binomial Logistic Regression (BLR) and Artificial Neural Networks (ANN). Two models were developed based on bridge-side, temporal context, and traffic load hazard variables, evaluated using confusion matrices and cross-entropy. Analyzing data from over 7.4 million vehicle transits in Brescia, Italy, the ANN outperformed BLR, achieving higher prediction confidence and significantly lower cross-entropy. The findings underscore ANN's potential in managing traffic load hazards and safeguarding

bridges, with future research recommended to incorporate structural effects, prioritize network-wide traffic management, and refine risk assessments using ANN models.

2.1 Risk and Risk Severity Analysis

Risk analysis is a critical process in construction and engineering projects, involving the identification, assessment, and prioritization of potential uncertainties that could adversely affect project outcomes. Risks can arise from various factors, including environmental conditions (Jenisha *et al.*, 2023), design complexities, operational challenges, and financial constraints. For example, inadequate site surveys may result in unforeseen subsurface conditions, such as unstable soil or hidden rock formations, leading to delays and increased costs. Risk analysis aims to systematically evaluate these uncertainties to determine their likelihood and potential impact (Ahmed *et al.*, 2023), ensuring that preventive measures or mitigation plans are in place before the project commences. Risk severity analysis goes a step further by quantifying the magnitude of each risk and its implications on project objectives. This process assesses risks based on factors such as the extent of damage, likelihood of occurrence, and the cost or time required to address the issue. For instance, a high-severity risk like foundation failure in a bridge construction project can jeopardize structural integrity, posing safety hazards and causing project delays. Conversely, low-severity risks, such as minor supply chain delays, might only require schedule adjustments with minimal disruption. By categorizing risks into severity levels, stakeholders can allocate resources effectively, prioritize urgent issues, and implement focused mitigation strategies to enhance project resilience and ensure timely completion (Wei *et al.*, 2023).

2.2 Kinds of Road Construction Risks

Road construction projects are subject to a variety of risks that can significantly impact their success, ranging from environmental challenges to operational uncertainties.

Environmental risks are among the most prominent, as road construction often involves working in unpredictable outdoor conditions. Factors such as extreme weather events, including heavy rainfall, snow, or high winds, can disrupt schedules and damage materials or machinery on-site. In regions prone to flooding or landslides, natural events can cause extensive delays or require costly redesigns. Additionally, environmental regulations and concerns (Lee *et al.*, 2022), such as the need to protect local wildlife or minimize ecological impact, can introduce unexpected hurdles if not adequately planned for during the early stages of the project. Operational and logistical risks also play a significant role in road construction projects. These include delays in material procurement, labor shortages, or machinery breakdowns, which can lead to cost overruns and missed deadlines. For instance, the late delivery of asphalt or aggregates can halt paving activities, causing a domino effect on subsequent tasks. Safety risks are another critical concern (Mostofi *et al.*, 2023), as road construction sites often involve heavy equipment and high-risk activities that could endanger workers if proper protocols are not followed. Poor coordination among contractors and suppliers can exacerbate these issues, leading to inefficient workflows and increased vulnerability to delays or accidents. Recognizing and addressing these risks through comprehensive planning and risk management strategies is essential for the successful completion of road construction projects (Liao *et al.*, 2022). Table 1 explains the description of various kinds of risks encountered during road construction.

2.3 Road Failure

Road failure occurs when a road's structural integrity or functionality is compromised, rendering it unsafe or inefficient for use. This can result from various factors, including poor design, substandard materials, environmental conditions, and inadequate maintenance.

Table 1: Data Feature Descriptions

| Risk | Description |
|------------------------------|---|
| Adverse Weather | Extreme weather conditions such as heavy rain, snow, or high winds disrupting construction. |
| Unstable Soil Conditions | Weak or shifting soil requiring additional stabilization or redesign of foundations. |
| Flooding | Water accumulation at the site leading to delays and structural damage. |
| Landslides | Earth movements causing obstruction or destruction of construction progress. |
| Substandard Materials | Use of low-quality materials resulting in premature failures such as cracking or potholes (Elhishi <i>et al.</i> , 2023). |
| Design Errors | Inadequate or flawed design leading to functional inefficiencies or safety concerns. |
| Traffic Disruptions | Ineffective traffic management causing delays and accidents during construction. |
| Machinery Breakdowns | Malfunctioning equipment leading to project slowdowns or costly repairs. |
| Labor Shortages | Insufficient workforce availability delaying project milestones. |
| Supply Chain Interruptions | Delays in material delivery affecting the construction schedule. |
| Safety Hazards | Accidents or injuries at the construction site due to inadequate safety protocols. |
| Regulatory Compliance Issues | Failure to adhere to construction laws or environmental regulations, causing penalties or delays. |
| Cost Overruns | Budget overspending due to unforeseen circumstances or poor planning (Almahameed <i>et al.</i> , 2023). |
| Utility Interference | Conflicts with existing utilities, such as pipelines or cables, requiring adjustments or relocations. |
| Community Opposition | Local protests or resistance to the construction project leading to delays or redesigns. |
| Corrosion of Structures | Premature deterioration of metallic components due to exposure to environmental factors. |
| Drainage Problems | Inefficient drainage systems leading to waterlogging and road damage. |
| Environmental Impact | Adverse effects on local ecosystems requiring mitigation measures or redesigns. |
| Legal Disputes | Conflicts with contractors, landowners, or suppliers causing project halts or delays. |
| Inadequate Supervision | Poor project oversight resulting in quality issues or overlooked risks (Edet & Ansa, 2023). |

For instance, insufficient drainage systems can lead to water accumulation, weakening the pavement layers and causing potholes or cracking. Heavy traffic loads beyond the road's design capacity may result in rutting or deformation, particularly on poorly constructed roads. In regions with extreme temperature fluctuations, thermal expansion and contraction can cause surface degradation, such as alligator cracking. Additionally, natural disasters like floods or landslides can erode road foundations (Alateeq *et al.*, 2023), leading to partial or total collapse. Preventing road failure requires a combination of proper initial design, high-quality materials, routine inspections, and timely maintenance to address emerging issues before they escalate.

2.4 Semantic Method of Analysis

Semantic analysis focuses on understanding the meaning and relationships of words, phrases, or concepts within a given context, and is especially valuable in fields like construction, risk assessment, or document analysis. This approach goes beyond simple keyword matching, aiming to capture the contextual meaning and nuances of language. In construction and risk management, it is crucial for processing complex textual data such as project documents, risk logs, and regulatory guidelines. By using semantic methods (Jeevith *et al.*, 2020), practitioners can extract deeper insights from unstructured data, enabling more informed decision-making. The Multinomial Naive Bayes (MNB) method enhances the process by serving as a probabilistic classifier for textual data features (Olena *et al.*, 2023). It categorizes documents or sentences into predefined risk categories based on word frequencies and distributions. In risk assessment, MNB can be applied to classify textual descriptions of risks into severity levels - low, medium, or high by analyzing the occurrence of specific terms and their associations. For example, phrases like "minor disruption" could signal low-severity risks, while terms like "structural failure" or "environmental disaster" suggest high-severity risks. The MNB method is

particularly effective in scenarios with labeled datasets (Sobhan *et al.*, 2020), allowing for the supervised training of the model to improve accuracy over time. By integrating this approach into risk management workflows, organizations can continuously monitor and assess risks, providing real-time insights that support more efficient and proactive decision-making in construction projects. Additionally, other approaches like Ensemble learning (George *et al.*, 2022) can still be sufficient for this research.

3.0 Methodology

The dataset used in this research was sourced from the Kaggle data repository and contains information about various risks associated with road construction projects. It includes the site analysis text, the type of risks that occurred on the site, and the severity score of those risks. The site analysis text provides a detailed description of the construction site conditions, which can help identify the type of risk present. The risks are classified into various categories such as adverse weather, machinery breakdowns, and flooding, among others, while the severity score represents the extent of the impact of each risk on the construction project, measured on a scale of 0 to 1. Before using the dataset for modeling as shown in Fig. 1, data preprocessing steps were carried out to clean the data and make it ready for analysis. This involved handling missing values, removing irrelevant or redundant information, and ensuring that the data was consistent and properly formatted. The data was then normalized and transformed to ensure that all numerical values were on the same scale, which helps improve the performance of the model. After preprocessing, feature analysis was performed to understand the relationship between the different variables in the dataset. This helped in selecting the most important features for the risk prediction task, ensuring that the model focuses on the most relevant information. Once the data was prepared, it was split into training and testing sets, ensuring that the model would be able to learn from a large

portion of the data while also being tested on unseen data for accuracy. The model was then trained on the training set, where Multinomial Naive Bayes was used to classify the risk types based on the site analysis text. The trained model was evaluated for its performance, and its accuracy, precision, recall, and other metrics were measured to ensure that it could effectively predict the risks and their corresponding severity scores based on the provided site analysis.

The methodological process of this research work on predicting and quantifying the severity of risks in road construction projects is structured as follows:

- i. **Data Collection:** The first step involved collecting historical data from past road construction projects. The dataset included information on various risks such as adverse weather, flooding, design errors, community opposition, and cost overruns. This data was obtained from road construction records and project reports, ensuring a comprehensive range of risk scenarios.
- ii. **Data Preprocessing:** The collected dataset was preprocessed to ensure its quality and suitability for analysis. This step involved cleaning the data by handling missing values, removing irrelevant features, and normalizing text fields for consistency. Text data (e.g., risk descriptions) was processed using natural language techniques, including tokenization and removing stop words, to prepare it for input into the model.
- iii. **Feature Extraction:** A critical part of this research was transforming textual data (such as "Site Analysis Text") into a format that could be used by machine learning models. The Count Vectorizer was employed to convert the text data into numerical features (word counts). This allowed the model to interpret the information in a structured form and identify patterns related to different risk categories.
- iv. **Model Development:** A Multinomial Naive Bayes (MNB) classifier was chosen due to its efficiency in handling text classification problems, particularly when the data involves multiple categories (risks in this case). A machine learning pipeline was built, combining the Count Vectorizer and MNB classifier. This pipeline allowed for seamless handling of raw text data and its conversion into risk classifications.
- v. **Model Training:** The dataset was split into training and testing sets (80% training, 20% testing) using train-test-split. The model was then trained on the training set, learning to classify different risk categories based on the features extracted from the text data.
- vi. **Model Evaluation:** After training, the model's performance was evaluated using various metrics like accuracy, precision, recall, and F1-score. These metrics helped assess how well the model predicted the risk categories, particularly focusing on common and critical risks. The evaluation revealed that the model achieved a high accuracy of 93%, indicating its effectiveness in predicting risks such as cost overruns, flooding, and environmental impacts.
- vii. **Risk Severity Mapping:** Once the risks were classified, each risk category was associated with a severity score based on historical data and expert judgment. For example, risks like "Landslides" and "Flooding" were assigned higher severity scores due to their potential impact on the project.
- viii. **Risk Prediction and Severity Quantification:** The final model was used to predict the risks associated with new road construction projects. By inputting site-specific data and analysis text, the model could provide predictions on which risks were most likely to occur and their potential severity. This step allows construction managers to proactively address issues before they escalate.

ix. Recommendations for Implementation: Based on the results, recommendations were made for the adoption of the risk prediction model in real-world road construction projects. These included

integrating the model into the risk management process, conducting regular risk assessments, and using the model's predictions to optimize resource allocation and minimize delays.

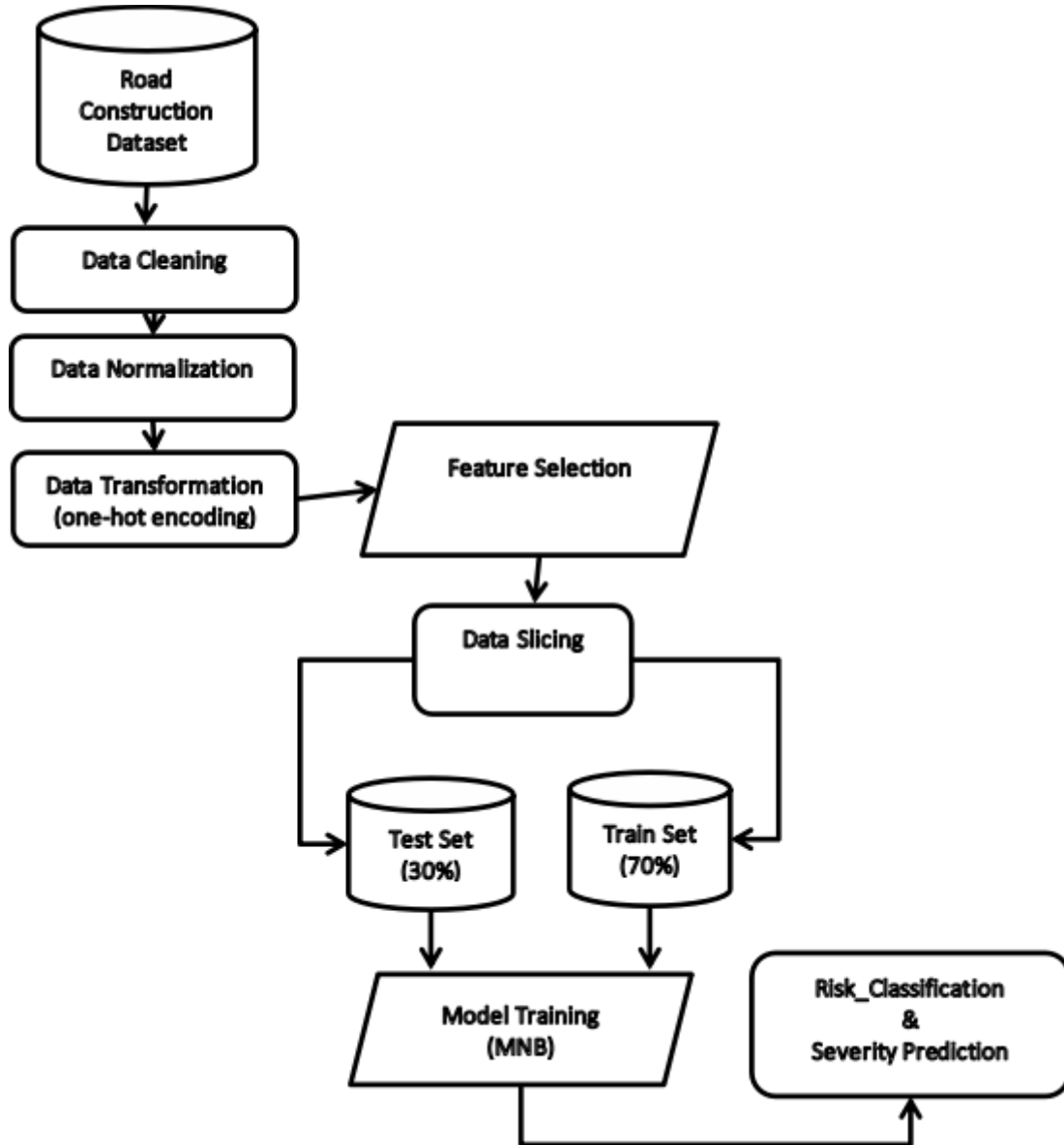


Fig. 1: Conceptual Architecture of the New System

4.0 Results and Discussion

In this section, the results of the research is presented and discussed accordingly. Every outcome from the research is presented in as result in this section. These outcomes are discussed to examine their meaning and contribution to the quest of this research. Table 2 is the snippet of the dataset used in this research; it was used to train the machine learning model that yielded the software for risk quantification.

Table 2: Dataset Snippet Showing Risk Quantification

| Site-Analysis-Text | Risk | Severity score |
|---|------------------------------|----------------|
| The construction site is experiencing heavy rain, leading to water accumulation on the ground. This is causing some delays in work. | Adverse Weather | 0.65 |
| The soil here is unstable, and we have noticed significant shifting of the ground in several areas. | Unstable Soil Conditions | 0.85 |
| Flooding occurred after a heavy storm, affecting the work area and causing significant water damage to materials. | Flooding | 0.75 |
| A minor landslide occurred in the northern area of the site, obstructing one of the primary access roads. | Landslides | 0.9 |
| We used some low-quality concrete in the pavement, and it seems to be cracking sooner than expected. | Substandard Materials | 0.8 |
| There was a design flaw in the drainage system that was not considered during the initial planning phase. | Design Errors | 0.85 |
| The construction site is constantly dealing with traffic issues as a major road is closed for construction. | Traffic Disruptions | 0.55 |
| Our excavator broke down yesterday, and we had to pause work until repairs were completed. | Machinery Breakdowns | 0.6 |
| We are facing a shortage of skilled workers due to labor strikes in the region. | Labor Shortages | 0.7 |
| Material deliveries are delayed by a week due to a strike at the supplier's factory. | Supply Chain Interruptions | 0.75 |
| Several workers have been injured on site due to inadequate safety measures. | Safety Hazards | 0.9 |
| The construction permits were delayed due to an administrative oversight. | Regulatory Compliance Issues | 0.8 |
| The project cost has exceeded initial estimates due to unforeseen complications with the subsoil conditions. | Cost Overruns | 0.8 |
| We encountered interference with existing electrical cables which delayed excavation work. | Utility Interference | 0.65 |
| The community has raised concerns about the impact of construction on local wildlife. | Community Opposition | 0.6 |
| We noticed signs of corrosion on some of the metal reinforcement structures. | Corrosion of Structures | 0.85 |

| | | |
|---|----------------------|-----|
| Waterlogging due to improper drainage is damaging the foundation in several areas. | Drainage Problems | 0.8 |
| We are facing environmental concerns related to the removal of trees along the construction path. | Environmental Impact | 0.7 |
| We are currently involved in a legal dispute with the landowners over property boundaries. | Legal Disputes | 0.9 |

From the result in Table 3, the application of the Multinomial Naive Bayes (MNB) model to the classification of risk categories in road construction projects has significantly contributed to achieving the research goal of risk severity quantification and prediction. The model has demonstrated strong performance in predicting several critical risk categories, showcasing its effectiveness in identifying potential issues that could impact project success as described in Table 3. For example, in the Adverse Weather category, the model achieved a recall of 0.86, successfully capturing most

instances of this risk. This high recall indicates that the model can predict adverse weather events with great reliability, enabling proactive measures in planning and decision-making. The precision of 0.75 and the F1-score of 0.80 reflect a strong balance, ensuring that predictions are largely accurate while accounting for rare events. Community Opposition was perfectly predicted with a precision and F1-scores of 1.00 and 0.96, respectively, which is a major advantage in identifying social risks early in the planning stages

Table 3: Classification Report for Individual Risk Category

| | precision | recall | F1-score | support |
|------------------------------|-----------|--------|----------|---------|
| Adverse Weather | 0.75 | 0.86 | 0.80 | 7 |
| Community Opposition | 1.00 | 0.92 | 0.96 | 12 |
| Corrosion of Structures | 1.00 | 1.00 | 1.00 | 8 |
| Cost Overruns | 1.00 | 0.80 | 0.89 | 10 |
| Design Errors | 0.88 | 0.78 | 0.82 | 9 |
| Drainage Problem | 0.91 | 1.00 | 0.95 | 10 |
| Environmental Impact | 0.91 | 1.00 | 0.95 | 10 |
| Flooding | 0.78 | 0.88 | 0.82 | 8 |
| Inadequate Supervision | 0.94 | 0.94 | 0.94 | 17 |
| Labor Shortages | 1.00 | 0.90 | 0.95 | 10 |
| Landslides | 0.88 | 0.78 | 0.82 | 9 |
| Legal Disputes | 1.00 | 1.00 | 1.00 | 11 |
| Machinery Breakdowns | 0.91 | 1.00 | 0.95 | 10 |
| Regulatory Compliance Issues | 0.80 | 1.00 | 0.89 | 4 |
| Safety hazards | 0.92 | 0.92 | 0.92 | 12 |
| Substandard materials | 1.00 | 1.00 | 1.00 | 9 |
| Supply Chain Interruptions | 0.85 | 0.92 | 0.88 | 12 |
| Traffic Disruptions | 0.89 | 0.89 | 0.89 | 9 |
| Unstable Soil Conditions | 1.00 | 1.00 | 1.00 | 14 |
| Utility Interference | 1.00 | 0.89 | 0.94 | 9 |
| Accuracy | | | 0.93 | 200 |
| Macro average | 0.92 | 0.92 | 0.92 | 200 |
| Weighted average | 0.93 | 0.93 | 0.93 | 200 |

Similarly, Corrosion of Structures achieved perfect scores across all metrics, demonstrating that the model can accurately predict this crucial issue that could affect the integrity and safety of infrastructure. Cost Overruns were identified with precision of 1.00, showcasing the model's capacity to predict financial risks with no false positives. The slightly lower recall of 0.80 does not undermine its ability to identify key financial risks, offering valuable insights into managing project budgets effectively. The Design Errors category, with a precision of 0.88 and recall of 0.78, highlights the model's capability to detect design flaws that could lead to delays and additional costs. This balanced performance aids in ensuring that design elements are thoroughly reviewed before execution. The model also excelled in identifying Drainage Problems with a perfect recall of 1.00, which is critical for ensuring that the infrastructure can handle environmental factors and remain operational in the long term. Similarly, Environmental Impact predictions showed an excellent performance, identifying potential ecological concerns and guiding sustainable project practices. In terms of common risks, Flooding and Labor Shortages were predicted with high recall, ensuring that these frequently encountered risks are well managed. The model was able to balance false positives and misses effectively, providing

useful insights into how to mitigate these risks early in the project lifecycle. The accuracy in predicting Legal Disputes and Machinery Breakdowns with perfect precision and recall is particularly important for maintaining smooth project operations and reducing delays caused by legal and technical issues. The model's ability to perfectly classify these risks ensures that legal compliance and machinery reliability are continuously monitored. The classification outcomes have not only shown the model's effectiveness in identifying various risk categories but have also provided substantial support in achieving the research aim. The model contributes valuable insights into risk severity quantification, allowing stakeholders in road construction projects to prioritize and mitigate potential risks. This leads to better-informed decision-making, optimized resource allocation, and enhanced project planning, ultimately helping to minimize unforeseen delays and costs. Through the model's application, road construction projects can benefit from a more accurate understanding of risks, improving their ability to anticipate challenges and implement preventive measures, aligning with the research's objective of enhancing risk management in engineering projects. Figure 2 is the heatmap modeling the correlation between risk and risk severity score.

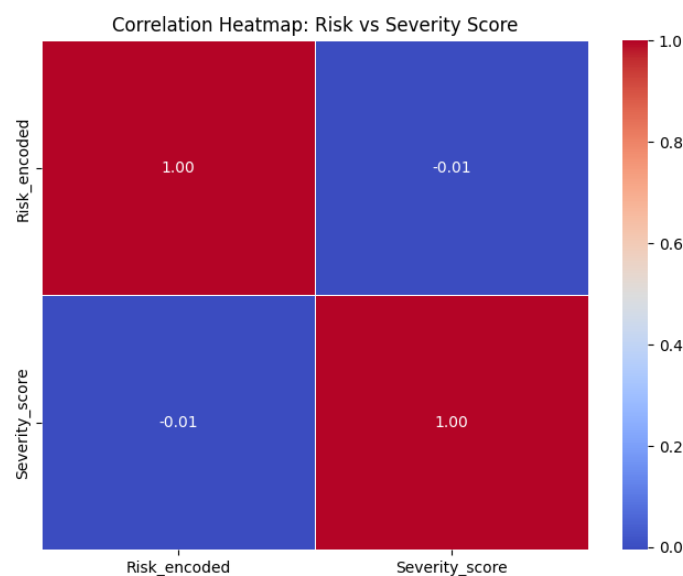


Fig. 2: Correlation Heatmap of Risk-Score and Risk

The correlation heatmap between Risk-Score and Risk highlights the strength of these predictive relationships, making it evident that the Risk-Score is a valuable tool in classifying and understanding risk categories as shown in Fig. 2. The need to classify risk in this research is driven by the critical importance of accurately identifying and mitigating potential hazards in road construction projects. By classifying risks, we are addressing real-world challenges faced by construction teams, project managers, and stakeholders. Mismanagement of risks such as Adverse Weather, Cost Overruns, or Landslides can lead to project delays, budget overruns, safety issues, and regulatory non-compliance. The classification system aims to systematically identify these risks based on historical data and

predictive models, enabling timely interventions and informed decision-making. High correlations between risk scores and specific risks suggest that the model can reliably predict certain risks based on measurable parameters. This predictive capability allows for proactive risk management strategies, potentially reducing unforeseen issues on construction sites and ensuring safer and more cost-effective project execution. Thus, the model's ability to classify risks based on these well-established correlations directly addresses the need for more efficient and accurate risk assessment in road construction, helping stakeholders mitigate financial and safety concerns. Figure 3 below shows various risk categories and their associated risk score.

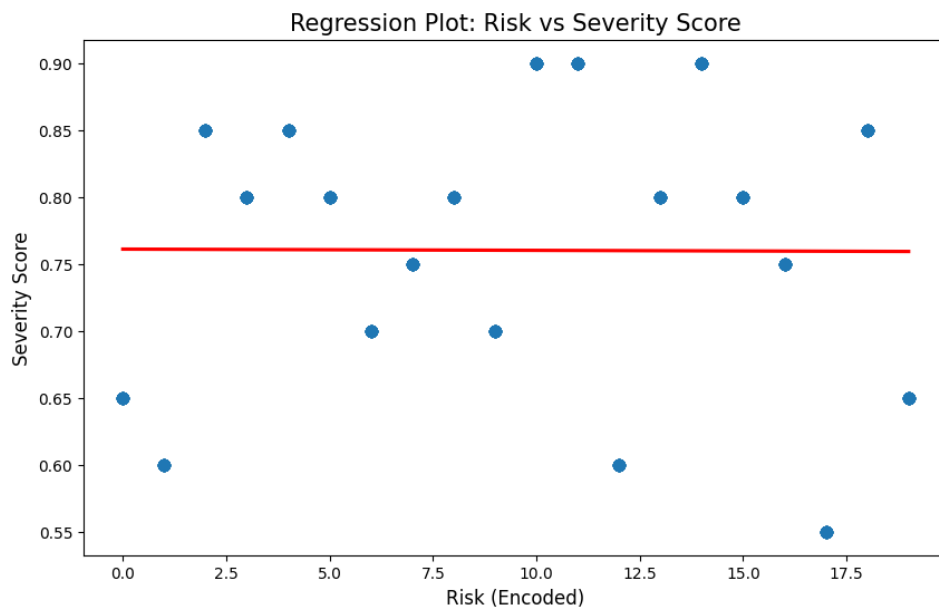


Fig. 3: Regression Plot of Risk vs Severity-score

The regression plot of Risk vs Severity Score in Fig. 3, provides a clear visual representation of the relationship between various risk categories and their associated severity scores in road construction projects. The plot helps to identify how the severity of each risk fluctuates in response to different risk factors. By observing the trend, it becomes evident which risks have a higher impact on project outcomes, with severity

scores potentially increasing as certain risk factors intensify. This regression analysis offers valuable insights into the relative importance of each risk, enabling stakeholders to prioritize mitigation efforts for those with the highest severity scores, ensuring better resource allocation and proactive planning in project management. More of this would be shown in scatter plot given in Figure 4.

3D Scatter Plot: Risk vs Severity Score vs Text Feature

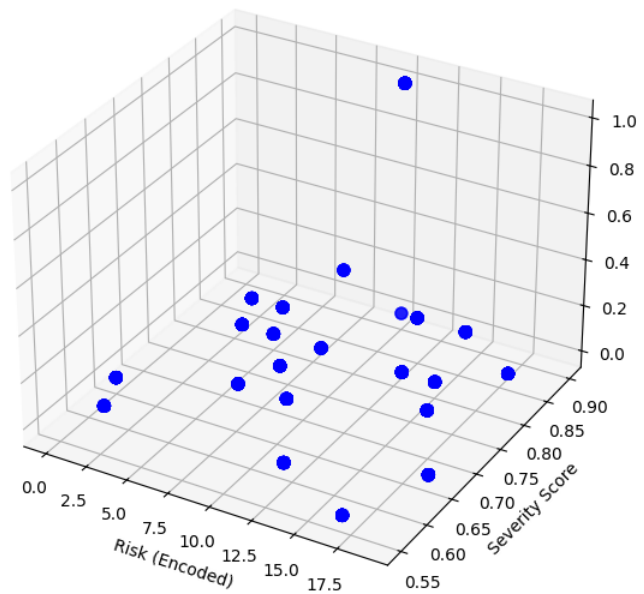


Fig. 4: Scatter Plot of Risk vs Severity-score

The scatter plot of Risk vs Severity Score in Fig. 4, visually displays the distribution of various risk categories and their corresponding severity scores in road construction projects. Each data point represents a specific risk category, plotted against its associated severity score. This scatter plot helps in identifying patterns, clusters, and outliers, offering a clear overview of how different risks are distributed in terms of their

severity. By analyzing the spread of points, it becomes possible to pinpoint which risks tend to have higher severity scores, aiding in the prioritization of risk management efforts. This approach supports decision-makers in focusing on risks that have the potential to cause the most significant impact on project success. The percentage severity of each hazard category is given in Figure 5.

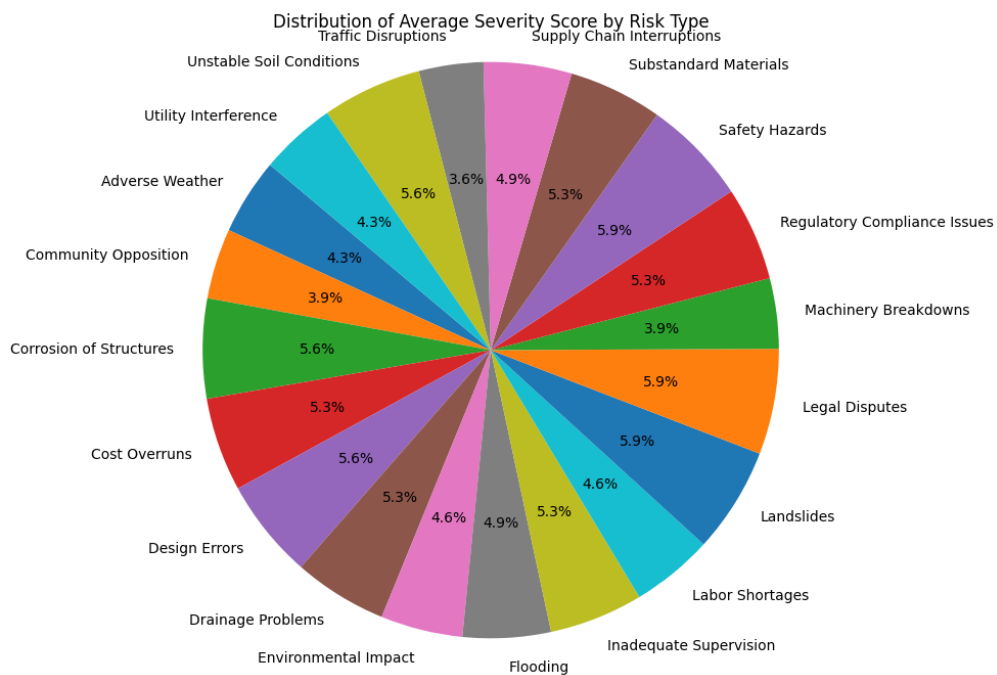


Fig. 5: Pie chart of Risk Categories Distribution

The pie chart in Fig. 5, depicting the Risk Categories Distribution highlights the varying proportions of different risks in road construction projects. It visually illustrates how certain risk categories, such as Flooding and Design Errors, occupy larger portions of the chart, reflecting their higher frequency or severity in the dataset. On the other hand, some risk categories, such as Machinery Breakdowns and Landslides, take up smaller slices, indicating they are less prevalent. This distribution provides a clear understanding of which risks are most dominant, helping project managers and engineers prioritize their resources and mitigation efforts toward the risks that are more likely to affect the project's success as indicated in Fig. 6. Figure 6 is the confusion matrix of the approaches used in this research. It indicated that the model is able to capture and interpret the different data points accordingly without so much misunderstanding. It also showed that the dataset used in this research has minimal randomness, leading to a good results.

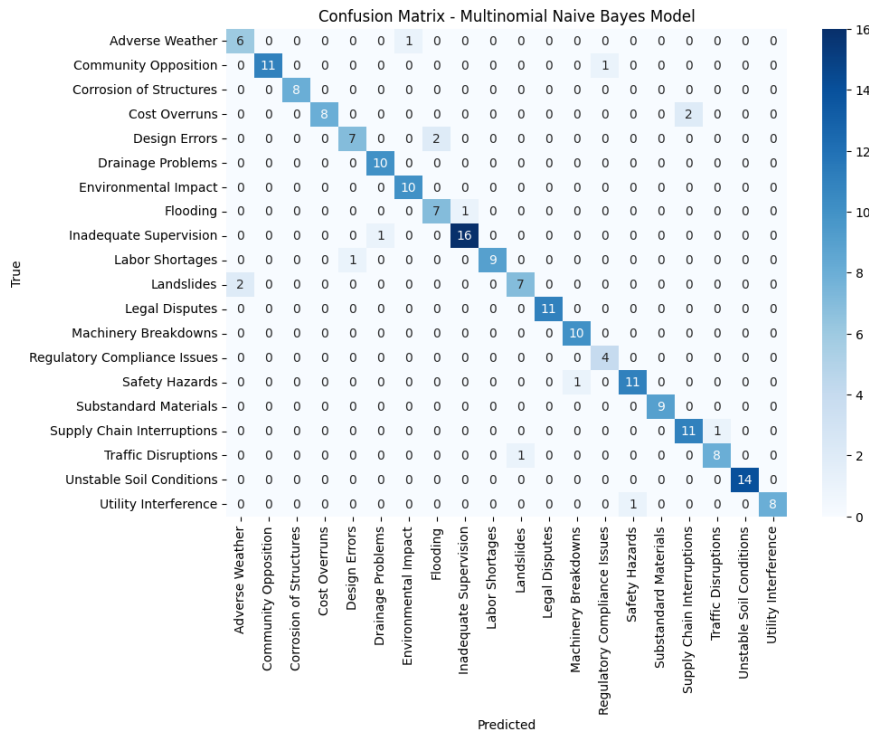


Fig. 6: Confusion Matrix of Multinomial Naive Bayes Approach

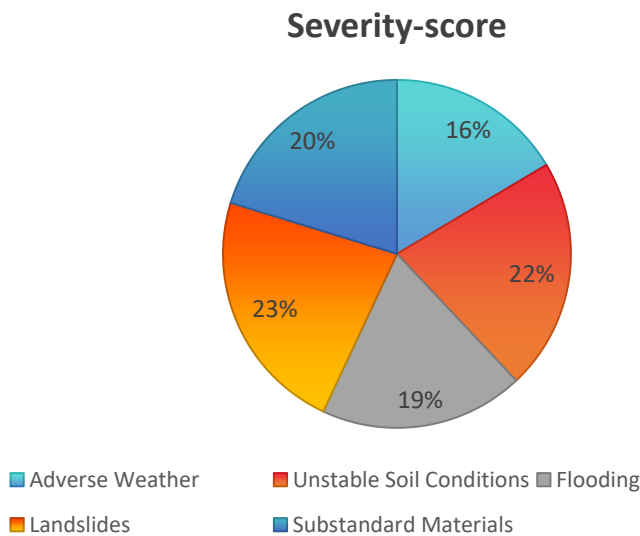


Fig. 7: Pie Chart of Four Hazards with High-Risk Severity Score

In Fig. 7, Landslide has the highest risk severity score of 23%, and Flooding has 22%, making these two hazards the most dangerous hazard that could happen while a construction is on. Figure 8(a) and 8(b) show the different stages of the interface or the risk analysis tool, from when the tool is launched as seen in Fig. 8(a), and

when data is entered and risk analysis is made as seen in Fig. 8(b). The input field named Site Analysis is meant to receive site description so that risk analysis can be carried out based on information in the description. Some sample site description text is given in Table 2 as seen above.

Construction Risk and Severity Prediction

Site Analysis:

Predict Risk

Fig. 8(a): Predictive Tool for Real life Application

Construction Risk and Severity Prediction

Site Analysis:

Predict Risk

Predicted Risk: Adverse Weather, Severity Score: 0.65
 Predicted Risk: Adverse Weather, Severity Score: 0.65

Fig. 8(b): Practical Application of the Tool

5.0 Discussion

This research focuses on addressing the critical risks that road construction projects often face, which include challenges such as adverse weather conditions, soil instability, labor shortages, equipment failures, and safety hazards. The unpredictable nature of these risks can lead to significant delays, cost overruns, and safety concerns on construction sites. By developing a machine learning model, specifically using the Multinomial Naive Bayes (MNB) classifier, this study aims to predict and quantify the severity of these risks, allowing for

better decision-making and resource allocation. The model leverages historical data from past construction projects to provide insights into the likelihood of various risks occurring and their potential impact. The application of the MNB classifier has proven to be highly effective in predicting several key risk categories, including cost overruns, adverse weather events, design errors, and environmental impacts. The model's performance was evaluated using precision, recall, and F1-scores, and it achieved a high accuracy rate of 93%. For example, the model showed strong performance in predicting risks

such as community opposition, where it achieved perfect precision and recall, and corrosion of structures, which is critical for maintaining infrastructure integrity. This performance demonstrates the model's capacity to help construction managers identify and mitigate risks early in the project lifecycle, ultimately leading to more effective project planning and safer execution. Furthermore, the study's findings emphasize the importance of risk classification and severity quantification in road construction projects, instead of constructing the project blindly only to quantify its crash severity later as was done by (Cheng et al., 2024). This work is ahead of the work done by Cheng et al., 2024, which measures severity of road project crash severity, since it classifies and quantifies risks before they occur. By categorizing risks based on historical data and predictive models, the system allows for a more proactive approach to risk management. The classification of risks such as flooding, labor shortages, and design errors enables project managers to prioritize resources and address the most pressing issues before they escalate. The model's ability to provide a clear understanding of risk distributions, as seen in the visual representation of risk categories, also aids in informed decision-making. In all, this research offers a valuable tool for improving the management of road construction projects, helping stakeholders mitigate risks effectively and ensure smoother project execution.

6.0 Implications of Research

This research has significant implications for the civil engineering sector, particularly in the management of risks associated with road construction projects. By utilizing machine learning techniques to quantify and predict risk severity, the study offers a more data-driven approach to identifying potential challenges that may arise during construction. With the ability to forecast risks like Flooding, Design Errors, or Cost Overruns, engineers and project managers can proactively address these issues, allowing

for better planning, resource allocation, and project scheduling. This predictive capability not only helps in minimizing delays and unexpected costs but also ensures that appropriate risk mitigation strategies are implemented before the risks materialize, ultimately leading to more efficient and cost-effective road construction projects. Furthermore, the integration of this research into risk management practices in road construction can lead to a paradigm shift in how risks are perceived and managed. Traditionally, risk management has often relied on historical data and expert judgment, which can sometimes be limited or subjective. However, the machine learning models used in this study provide a more objective and comprehensive assessment of risks, offering actionable insights that can guide decision-making. By improving the accuracy and precision of risk identification, this research empowers stakeholders to make informed decisions about how to allocate resources, prioritize tasks, and manage unforeseen challenges. As a result, the construction of road infrastructure can become more resilient to adverse events, contributing to the long-term sustainability and safety of transportation networks.

5.1 Key Steps for Implementation in Real Life Project

i. Data Collection:

- a. Collect site analysis reports and historical risk data from construction sites.
- b. Integrate real-time data sources like weather sensors and soil condition reports.

ii. Model Deployment:

- a. Deploy the trained model on a simple user interface for project managers or engineers to input site reports and get risk predictions.
- b. Use a mobile-friendly interface for easy access in the field.

iii. Risk Prediction:

- a. Use the model to predict risks like adverse weather, flooding, or unstable soil conditions based on input reports.
- b. Present the risk prediction along with a severity score.
- iv. **Actionable Insights:**
 - a. Set up automated alerts or notifications for high-risk scenarios (e.g., flooding with a severity score of 0.75).
 - b. Provide suggestions or mitigation actions based on the predicted risks.
- v. **Continuous Improvement:**
 - a. Periodically update the model with new data.
 - b. Monitor the system's performance and accuracy over time to ensure it provides reliable predictions.

This simplified approach will help in real-time decision-making, improve site safety, and reduce unexpected delays in road construction projects.

7.0 Conclusion

In conclusion, this research represents a significant advancement in the field of civil engineering, particularly in the context of road construction projects and risk management. By applying machine learning techniques, particularly the Multinomial Naive Bayes (MNB) model exemplified in Fig. 6, to assess and predict the severity of various risks, the study provides a robust framework for identifying, quantifying, and mitigating risks in construction projects. The model's performance across different risk categories, with high recall and precision in most instances, indicates its potential for accurate risk prediction and effective decision-making support for project managers and stakeholders. The findings of this research reveal the importance of data-driven approaches in modern engineering practices. The ability to predict risks such as Community Opposition, Cost Overruns, and Flooding with a high degree of accuracy allows construction

teams to better prepare for these challenges and implement proactive measures to mitigate their impact. This capability not only enhances the efficiency of road construction projects by reducing delays and unexpected costs but also contributes to improving the overall safety and quality of the infrastructure. Additionally, the research highlights the need for continuous refinement of predictive models to ensure even better risk detection and classification, especially in categories with lower recall, such as Design Errors and Landslides. From a broader perspective, this work sets a precedent for the integration of machine learning into risk management practices within the civil engineering sector. The ability to quantify and predict risk severity offers a more comprehensive understanding of the potential challenges that may arise during construction, which can lead to improved project planning, resource allocation, and risk mitigation strategies. This approach not only optimizes the use of resources but also ensures that road construction projects are completed on time, within budget, and to the desired quality standards. The implications of this research extend beyond road construction, providing a framework that can be adapted for other types of civil engineering projects where risk management is critical. By embracing such advanced techniques, the civil engineering industry can move towards more resilient and sustainable construction practices, ultimately benefiting society through safer and more reliable infrastructure.

Reference

- Ahmed, G., Mohamed, E., and Wael E. (2023). Machine Learning-Based Decision Support Framework for Construction Injury Severity Prediction and Risk Mitigation. *ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng.*, 2022, 8(3): 04022024

- Ahmed S., Hossain, M. A., Bhuiyan M. M. I. and Ray, S. K. (2021). "A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity," 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI-SmartCNS), London, United Kingdom, 2021, pp. 390-397.
- Alateeq, M.M.; P. P., F.R.; Ali, M.A.S. (2023). Construction Site Hazards Identification Using Deep Learning and Computer Vision. *Sustainability*, 15, 2358. <https://doi.org/10.3390/su15032358>
- Almahameed, B. A., Bisharah, M. (2023). Applying Machine Learning and Particle Swarm Optimization for predictive modeling and cost optimization in construction project management. *Asian J Civ Eng.* <https://doi.org/10.1007/s42107-023-00843-7>
- Ekong, B., Ekong, O., Silas, A., Edet, A., and William, B. (2023). Machine Learning Approach for Classification of Sickle Cell Anemia in Teenagers Based on Bayesian Network. *Journal of Information Systems and Informatics*, 5(4),1793-1808.
- Cheng, C., Chen, S., Ma, Y., Qiao, F., & Xie, Z. (2024). Crash severity prediction and interpretation for road determinants based on a hybrid method. *Journal of Transportation Safety & Security*, 1-27.
- Davor, M., Ivan, M., Nina, B. (2020). A Systematic Literature Review on the Application of Multicriteria Decision Making Methods for Information Security Risk Assessment. *International Journal of Safety and Security Engineering*, 10, (2), 161-174.
- Edet, A., Silas, A., Ekaetor, E., Ebong, O., Isaac, E., & Udoetor, N. (2024). Data-Driven Framework for Classification and Management of Start-Up Risk For High Investment Returns. *Advanced Journal of Science, Technology and Engineering*, 4(2), 81-102.
- Edet, A. and Ansa, G. (2023). Machine learning enabled system for intelligent classification of host-based intrusion severity. *Global Journal of Engineering and Technology Advances*, 16(03), 041–050
- Elhishi, S., Elashry, A. M. & El-Metwally, S. (2023). Unboxing machine learning models for concrete strength prediction using XAI. *Sci Rep* 13, 19892. <https://doi.org/10.1038/s41598-023-47169-7>
- Ebong, O., Edet, A., Uwah, A., & Udoetor, N. (2024). Comprehensive Impact Assessment of Intrusion Detection and Mitigation Strategies Using Support Vector Machine Classification. *Research Journal of Pure Science and Technology*, 7, (2), 50-69.
- George, M. R., Nalluri, M. R. & Anand, K. B. (2022) Application of Ensemble Machine Learning for Construction Safety Risk Assessment. *J. Inst. Eng. India Ser. A* 103, 989–1003. <https://doi.org/10.1007/s40030-022-00690-w>
- Jeevith H., Børge, R. (2020). Applications of machine learning methods for engineering risk assessment – A review, *Safety Science*, Volume 122, <https://doi.org/10.1016/j.ssci.2019.09.015>.
- Jenisha, D., Sujan, K., Ranjit, S., Shyam, S. K. (2023). A Review on Machine Learning approaches for Himalayan Rock Mass

- Classification, *J. Phys.: Conf. Ser.* 2629 012016
- Khanum, H., Garg, A., & Faheem, M. I. (2023). Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways. *F1000Research*, 12.
- Le Duc, A., Dao Duy, L., Thai Thi Kim, C., & Bach Thi Diep, P. (2023). A machine learning approach to risk assessment of expressway bridges. *Tạp Chí Khoa Học Giao Thông Vận Tải*, 73(7), 661-673. <https://doi.org/10.47869/tcsj.73.7.1>
- Lee, J.; Lee, S. (2023). Construction Site Safety Management: A Computer Vision and Deep Learning Approach. *Sensors* 2023, 23, 944. <https://doi.org/10.3390/s23020944>
- Le-Ngoc, V., Vuong-Cong, L., Pham-Bao, T., Ngo-Kieu N. (2023). Damage assessment in beam-like structures by correlation of spectrum using machine learning, *Frattura ed Integrità Strutturale*, 300-319.
- Liao, C., Eeydzah, A., Suzila, M., And Loh, S. (2022). Intelligent Risk Management in Construction Projects: Systematic Literature Review, *IEEE Access*, 10.1109/ACCESS.2022.3189157
- Mamdouh, A., Khaled, E., (2024). Machine learning for predicting the impact of construction activities on air traffic operations during airport expansion projects, *Automation in Construction*, Volume 158, 2024, 105189, <https://doi.org/10.1016/j.autcon.2023.105189>.
- Mostofi, F. & Toğan, V. (2023). Explainable Safety Risk Management in Construction with Unsupervised Learning. In V. Plevris, A. Ahmad, & N. Lagaros (Eds.), *Artificial Intelligence and Machine Learning Techniques for Civil Engineering* (pp. 273-305). IGI Global. <https://doi.org/10.4018/978-1-6684-5643-9.ch011>
- Olena, L., Roman, O. (2023). Features of the Implementation of Artificial Intelligence in Construction. *Journal of modern technology in materials design in construction*, 35(2):172-176.
- Pham, B., Luan, V., Nhi, N. (2023) “Damage assessment in beam-like structures by correlation of spectrum using machine learning”, *Frattura ed Integrità Strutturale*, 17(65), pp. 300–319. Doi: 10.3221/IGF-ESIS.65.20.
- Rodionova, M., Skhvediani, A., & Kudryavtseva, T. (2022). Prediction of crash severity as a way of road safety improvement: The case of Saint Petersburg, Russia. *Sustainability*, 14(16), 9840.
- Sobhan, S., Anima, P., Genserik, R. (2020). Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data, *Safety Science*, 125, <https://doi.org/10.1016/j.ssci.2020.104616>.
- Udoetor, N., Ansa, G., Ekong, A., & Edet, A. (2024). Intelligent System for Detection of Copyright-Protected Data for Enhanced Data Security. *Technology*, 7(4), 58-80.
- Ventura, R., Barabino, B., & Maternini, G. (2024). Prediction of the severity of exceeding design traffic loads on highway bridges. *Heliyon*, 10(1).
- Wei, Z.; Liu, H.; Tao, X.; Pan, K.; Huang, R.; Ji, W.; Wang, J. (2023). Insights into the Application of Machine Learning in

- Industrial Risk Assessment: A Bibliometric Mapping Analysis. *Sustainability* 2023, 15, 6965. [hdoi.org/10.3390/su15086965](https://doi.org/10.3390/su15086965)
- Xianbo Z. (2023). Construction risk management research: intellectual structure and emerging themes, *International Journal of Construction Management*, DOI: 10.1080/1562-3599.2023.2167303
- Yin, S., Liu, T., Huang, B. (2023). Research on Security Assessment Model of Marine Information System Based on Machine Learning. In: Yang, D. (eds) 2023 International Conference on Marine Equipment & Technology and Sustainable Development. METSD 2023. Lecture Notes in Civil Engineering, vol 375. Springer, Singapore. https://doi.org/10.1007/978-981-99-4291-6_11