Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Monitoring-Based Assessment of Excavation Risk Using Data-Driven Models

Sachin Admane
sachinadmane@gmail.com

JSPM Group of Institutes Pune
Tejas Admane
JSPM University

Research Article

Keywords: Deep excavation, Excavation risk assessment, Monitoring data, Data-driven modeling, Machine learning, Geotechnical engineering, Construction safety

Posted Date: January 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8477200/v1

License: 🅭 🅯 This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Deep excavations involve complex interactions among soil properties, groundwater conditions, excavation geometry, support systems, and construction-induced responses, making reliable risk assessment essential for safe construction. With the increasing availability of monitoring data, there is a growing need for practical approaches that can systematically utilize such information for excavation risk evaluation. This study presents a monitoring-based assessment of excavation risk using data-driven modeling techniques suitable for routine geotechnical engineering practice. A comprehensive dataset comprising geotechnical, environmental, excavation-related, and structural response parameters is analyzed to classify excavation risk levels. Conventional machine learning models are employed to capture relationships among influencing factors, with emphasis placed on robustness and engineering interpretability rather than methodological complexity. Model performance is evaluated using standard classification metrics, and feature importance and sensitivity analyses are conducted to identify dominant contributors to excavation risk. Results indicate that excavation depth and deformation-related indicators govern risk classification, while environmental factors act as secondary modifiers under typical conditions. The findings demonstrate that data-driven interpretation of monitoring data can effectively complement traditional geotechnical assessment methods and support practical excavation risk evaluation.

1. Introduction

Deep excavations are an integral component of modern urban development, enabling the construction of basements, underground transportation systems, utility corridors, and foundation structures in space-constrained environments. The safety of such excavations depends on the complex interaction among subsurface conditions, groundwater regime, excavation geometry, support systems, and construction-induced responses. Uncertainty in any of these components may result in excessive deformation, damage to adjacent infrastructure, or instability of the excavation system. As a result, excavation risk assessment remains a fundamental concern in geotechnical engineering practice (Zhang et al., 2020; Peck 1969; Ou 2006).

Traditional approaches to excavation risk assessment are largely based on analytical formulations, numerical simulations, and empirical design charts (Finno et al.,2010). Limit equilibrium methods and finite element analyses are commonly used to evaluate stability and deformation, while the observational method emphasizes the use of field monitoring to guide construction decisions (Clough and O’Rourke 1990; Hashash and Whittle 2002). Although these approaches have significantly advanced understanding of excavation behavior, their practical application often relies on simplified assumptions regarding soil properties, boundary conditions, and construction sequences. Moreover, the increasing volume of monitoring data generated during excavation projects poses challenges for systematic interpretation using conventional tools (Finno and Calvello 2005; Long 2001).

In recent decades, the deployment of automated monitoring systems has become standard practice in deep excavation projects, particularly in dense urban areas. Measurements of wall displacement, ground settlement, pore water pressure, groundwater level, and structural response provide valuable information on excavation performance throughout construction (Zhang et al., 2021). Despite the availability of such data, excavation risk evaluation in practice frequently relies on threshold-based criteria or qualitative engineering judgment. While effective in many cases, these approaches may not fully capture the combined and nonlinear influence of multiple factors on excavation risk, especially under variable environmental and geological conditions (Ou and Hsieh 2011; Finno and Roboski 2006).

Data-driven modeling approaches provide an alternative means of extracting information from monitoring datasets by identifying patterns and relationships directly from observed data. Machine learning techniques have been increasingly explored in geotechnical engineering for applications such as deformation prediction, stability classification, and performance assessment of excavation systems (Goh 2015; Zhang and Goh 2016). These methods offer the ability to accommodate complex interactions among variables without requiring explicit constitutive modeling, which can be advantageous when dealing with heterogeneous subsurface conditions and incomplete information.

However, many existing data-driven studies emphasize methodological novelty or advanced algorithms, which may limit their acceptance and adoption in routine engineering practice. From a practitioner's perspective, excavation risk assessment tools should prioritize robustness, transparency, and interpretability over algorithmic sophistication. In particular, applied studies demonstrating how commonly available monitoring data can be integrated with well-established data-driven models to support practical risk evaluation remain limited in the literature (Baecher and Christian 2003; Zhang and Zhang 2014).

This study aims to address this gap by presenting a monitoring-based assessment of excavation risk using data-driven modeling approaches suitable for practical application. A comprehensive dataset comprising geotechnical properties, environmental conditions, excavation geometry, and structural response indicators is analyzed to classify excavation risk levels. Conventional machine learning models are employed, with emphasis placed on understanding the relative influence of key parameters rather than pursuing algorithmic complexity. Model performance is evaluated using standard classification metrics, and sensitivity analysis is conducted to identify dominant factors governing excavation risk.

The objective of this paper is not to replace traditional geotechnical analysis, but to demonstrate how data-driven interpretation of monitoring information can complement established methods and enhance excavation risk evaluation. By focusing on practical applicability and engineering insight, the proposed approach is intended to support decision-making in deep excavation projects and contribute to safer and more informed construction practice.

The remainder of the paper is organized as follows. Section 2 describes the monitoring dataset and variable characteristics. Section 3 outlines the data-driven modeling methodology. Section 4 presents the results of excavation risk classification and sensitivity analysis. Section 5 discusses the engineering implications of the findings, and Section 6 summarizes the main conclusions.

2. Dataset and Variable Description

2.1 Overview of the Monitoring Dataset

The dataset used in this study was compiled to represent a generalized deep excavation monitoring context, integrating geotechnical, environmental, excavation-related, and structural response parameters. Such multi-source datasets are increasingly common in contemporary excavation projects due to advances in instrumentation and data acquisition systems. The dataset reflects conditions typically encountered in urban excavations, where excavation performance is influenced by both subsurface properties and external environmental factors (Ou 2006; Long 2001).

The data consist of numerical and categorical variables describing soil characteristics, excavation geometry, support systems, environmental loading, and observed deformation responses. Each record corresponds to a specific excavation condition and is associated with a discrete excavation risk level. Although the dataset does not represent a single project site, its parameter ranges and distributions are consistent with those reported in monitoring-based excavation studies and design practice (Finno and Calvello 2005; Xu et al. 2018).

The use of a comprehensive dataset allows the analysis to capture interactions among multiple influencing factors, reflecting the complex and coupled nature of excavation behavior. Such integrated datasets have been shown to be particularly valuable for data-driven evaluation of excavation performance and risk (Goh et al. 2017; Zhang et al. 2019).

2.2 Description of Variables

The variables included in the dataset are summarized in Table 1, grouped according to their engineering relevance. Geotechnical variables describe inherent soil properties such as soil type, moisture content, shear strength, and bearing capacity. These parameters govern the fundamental resistance of the ground and are known to strongly influence excavation stability and deformation (Clough and O’Rourke 1990; Hashash and Whittle 2002).

Table 1
Description of variables used for excavation risk assessment

Variable	Category	Unit	Description
Soil_Type	Geotechnical	–	Soil classification
Soil_Moisture_%	Geotechnical	%	Soil moisture content
Shear_Strength_kPa	Geotechnical	kPa	Undrained shear strength
Bearing_Capacity_kPa	Geotechnical	kPa	Soil bearing capacity
Excavation_Depth_m	Excavation	m	Depth of excavation
Retaining_Wall_Type	Structural	–	Type of retaining wall
Support_System	Structural	–	Support system configuration
Rainfall_mm_day	Environmental	mm/day	Daily rainfall intensity
Temperature_C	Environmental	°C	Ambient temperature
Groundwater_Level_m	Environmental	m	Groundwater depth
Seismic_Activity	Environmental	–	Seismic activity indicator
Ground_Settlement_mm	Structural response	mm	Ground settlement
Wall_Displacement_mm	Structural response	mm	Lateral wall displacement
Deformation_mm	Structural response	mm	Overall deformation
Pore_Water_Pressure_kPa	Hydrological	kPa	Pore water pressure
Strain_Gauge	Structural response	–	Strain gauge reading
Risk_Level	Output	–	Excavation risk level (ordinal)

Excavation-related variables include excavation depth, retaining wall type, and support system configuration. Excavation depth serves as a key indicator of stress redistribution and geometric effects, while the retaining wall and support system reflect design choices that control deformation and stability during construction (Ou and Hsieh 2011; Kung et al. 2009).

Environmental variables, such as rainfall intensity, ambient temperature, groundwater level, and seismic activity indicator, are included to represent external influences that may alter soil behavior and loading conditions. Variations in groundwater level and rainfall have been widely reported as critical contributors to excavation-induced deformation and instability, particularly in soft ground and urban settings (Finno and Roboski 2006; Goh et al. 2017).

Structural response variables capture the observed performance of the excavation system, including ground settlement, wall displacement, overall deformation, pore water pressure, and strain gauge readings. These response indicators provide direct evidence of excavation behavior and are commonly used in monitoring-based evaluation and decision-making during construction (Finno and Calvello 2005; Long 2001).

The output variable represents excavation risk level, defined as an ordinal classification reflecting increasing degrees of excavation risk. Such classification-based representations are frequently adopted in applied risk assessment studies to facilitate interpretation and support decision-making (Baecher and Christian 2003; Zhang and Zhang 2014).

2.3 Statistical Characteristics of the Dataset

Descriptive statistics of the numerical variables are presented in Table 2. The dataset spans a wide range of excavation depths, groundwater levels, and deformation responses, indicating that both stable and potentially adverse excavation conditions are represented. The observed variability in geotechnical and environmental parameters reflects realistic site-to-site and stage-to-stage differences encountered in excavation projects.

Notably, deformation-related variables exhibit substantial dispersion, consistent with the inherently uncertain and nonlinear nature of excavation-induced ground movements. Similar variability has been reported in published excavation monitoring databases and case histories, underscoring the importance of probabilistic and data-driven interpretation approaches (Long 2001; Zhang et al. 2019).

The inclusion of variables with different physical meanings and scales necessitates appropriate preprocessing prior to modeling, as discussed in the following section. Nevertheless, the statistical characteristics of the dataset confirm its suitability for exploring relationships between monitoring parameters and excavation risk within a data-driven framework.

2.4 Engineering Relevance of the Dataset

From an engineering perspective, the dataset reflects the types of information commonly available to designers and site engineers during excavation projects. Rather than relying on specialized or difficult-to-obtain parameters, the analysis focuses on variables routinely measured or estimated in practice. This emphasis enhances the practical relevance of the study and supports the objective of developing an assessment approach that can complement existing geotechnical evaluation methods (Goh et al., 2018; Chen et al., 2022).

By integrating soil properties, environmental conditions, excavation geometry, and response measurements, the dataset provides a holistic representation of excavation behavior. Such integrated representations are increasingly recognized as essential for effective excavation risk assessment and monitoring-based decision support (Peck 1969; Wang et al. 2020, Zhou et al’, 2021).

3. Methodology

3.1 Overall Framework

The methodology adopted in this study follows a practical, monitoring-based risk assessment workflow designed for application in routine geotechnical engineering practice. Rather than developing new algorithms, the emphasis is placed on integrating commonly available monitoring data with established data-driven modeling techniques to support excavation risk evaluation. The overall framework consists of data preprocessing, model development, performance evaluation, and sensitivity analysis, as illustrated conceptually in Fig. 1.

This structure reflects the typical sequence of engineering decision-making in excavation projects, where monitoring information is progressively interpreted to assess performance and identify potential risk conditions (Peck 1969; Finno and Calvello 2005).

3.2 Data Preprocessing and Feature Handling

Prior to model development, the dataset was preprocessed to ensure consistency and suitability for data-driven analysis. Numerical variables were examined for missing values and outliers, while categorical variables representing soil type, retaining wall type, and support system configuration were encoded into numerical form. Such preprocessing steps are necessary to enable the application of machine learning models while preserving the engineering meaning of the original parameters (Zhang and Goh 2016; Goh 2015).

To avoid undue influence of variable scale, numerical features were normalized prior to model training. This step ensures that variables with large numerical ranges, such as pore water pressure or bearing capacity, do not dominate model learning at the expense of other relevant parameters. Similar preprocessing strategies have been widely adopted in monitoring-based geotechnical modeling studies (Leu et al., 2022; Sun et al. 2021; Zhang et al. 2019).

3.3 Selection of Data-Driven Models

Two conventional machine learning models were selected for excavation risk classification: logistic regression and random forest. These models were chosen due to their established performance, robustness, and interpretability in engineering applications.

Logistic regression provides a baseline probabilistic classification framework and offers transparency in the relationship between input variables and predicted outcomes. It has been widely used in geotechnical reliability and risk assessment studies due to its simplicity and ease of interpretation (Baecher and Christian 2003; Zhang and Zhang 2014).

Random forest, an ensemble-based decision tree model, was selected to capture nonlinear relationships and interactions among variables. Random forest models have demonstrated reliable performance in a range of geotechnical applications, including excavation risk assessment, due to their ability to handle heterogeneous datasets and complex variable interactions without excessive parameter tuning (Goh 2015; Zhang et al. 2019).

The use of these two complementary models allows comparison between a transparent baseline approach and a more flexible nonlinear model, while maintaining methodological simplicity appropriate for applied engineering studies.

3.4 Model Training and Validation

The dataset was divided into training and testing subsets to evaluate model performance objectively. Model training was conducted using the training subset, while predictive performance was assessed on unseen testing data. Standard classification metrics, including accuracy, F1-score, and balanced accuracy, were employed to evaluate model performance. These metrics provide complementary perspectives on classification quality, particularly in datasets where class distributions may be imbalanced (Sun et al. 2021).

Cross-validation was employed during model development to reduce sensitivity to data partitioning and enhance model robustness. Such validation strategies are commonly recommended in data-driven geotechnical studies to mitigate overfitting and improve generalization (Goh and Zhang 2016; Zhang et al. 2019).

3.5 Feature Importance and Sensitivity Analysis

To support engineering interpretation, feature importance analysis was conducted to identify variables that most strongly influence excavation risk classification. For the random forest model, feature importance was evaluated based on the contribution of individual variables to model decision-making across the ensemble of trees. This approach provides insight into dominant risk drivers while remaining computationally efficient and easy to interpret (Breiman 2001; Goh 2015).

In addition to global feature importance, sensitivity analysis was performed by systematically varying selected key parameters while holding others constant. This analysis allows examination of how changes in excavation depth, groundwater level, and deformation-related indicators affect predicted risk. Sensitivity analysis has been widely recognized as an effective tool for linking data-driven predictions to physical understanding in geotechnical applications (Baecher and Christian 2003; Zhang and Zhang 2014).

3.6 Engineering Interpretation and Practical Emphasis

Throughout the methodology, emphasis is placed on maintaining a clear connection between model outputs and engineering interpretation. Rather than treating machine learning predictions as black-box results, the analysis focuses on understanding relative parameter influence and response trends that can inform practical decision-making.

By combining conventional data-driven models with straightforward preprocessing, validation, and sensitivity analysis, the proposed methodology is designed to complement traditional excavation assessment methods. This approach aligns with the broader objective of enhancing the practical use of monitoring data in excavation risk evaluation without introducing unnecessary methodological complexity (Peck 1969; Ou 2006).

4. Results and Analysis

4.1 Descriptive Characteristics of the Dataset

The descriptive statistics of the numerical variables used for excavation risk assessment are summarized in Table 2. The dataset spans a wide range of geotechnical, environmental, and structural response conditions, reflecting the diversity typically encountered in deep excavation projects. Excavation depth varies from shallow to deep conditions, while groundwater levels and deformation indicators exhibit substantial dispersion, indicating the presence of both stable and potentially adverse excavation states.

Table 2
Descriptive statistics of numerical variables

Variable	Mean	Std. Dev.	Min	Max
Soil_Moisture_%	17.49	7.3	5.12	29.99
Shear_Strength_kPa	174.54	71.57	50.39	299.59
Bearing_Capacity_kPa	350.7	144.65	100.01	599.45
Excavation_Depth_m	16.1	7.73	3	29.99
Deformation_mm	24.72	14.41	0.07	49.97
Rainfall_mm_day	97.28	57.37	0.05	199.78
Temperature_C	16.52	15.57	−9.99	44.78
Groundwater_Level_m	4.97	2.8	0	9.99
Seismic_Activity	0.51	0.5	0	1
Ground_Settlement_mm	51.17	28.49	0.21	99.97
Wall_Displacement_mm	25.12	14.73	0.01	49.97
Pore_Water_Pressure_kPa	271.88	131.32	50.54	499.48
Strain_Gauge	49.35	29.29	0.14	99.83
Risk_Level	1.79	0.45	0	2

Geotechnical parameters such as shear strength and bearing capacity show considerable variability, consistent with heterogeneous subsurface conditions reported in excavation case histories (Long 2001; Ou 2006). Environmental variables, including rainfall and temperature, also display wide ranges, supporting their inclusion as potential secondary contributors to excavation risk. Structural response variables, particularly ground settlement, wall displacement, and pore water pressure, exhibit high standard deviations, highlighting the inherently uncertain and nonlinear nature of excavation-induced responses (Finno and Calvello 2005; Hashash and Whittle 2002).

Overall, the statistical characteristics confirm that the dataset captures a broad spectrum of excavation conditions, providing a suitable basis for data-driven risk classification.

4.2 Distribution of Excavation Risk Levels

The distribution of excavation risk levels is presented in Fig. 2. The majority of samples are classified as high risk, while moderate- and low-risk cases constitute smaller proportions of the dataset. This distribution is consistent with monitoring datasets derived from conservative engineering practice, where data are often collected intensively during conditions of heightened concern or advanced excavation stages (Ou and Hsieh 2011; Zhang et al. 2019).

Although the class distribution is imbalanced, such imbalance reflects realistic excavation monitoring scenarios rather than sampling bias. In practice, risk assessment tools must be capable of handling skewed data distributions, particularly when high-risk conditions dominate monitoring records. The use of balanced accuracy and F1-score in subsequent analysis therefore provides a more reliable evaluation of model performance than accuracy alone (Sun et al. 2021).

4.3 Model Performance Comparison

Model performance results are summarized in Table 3, while the corresponding ROC–AUC values are shown in Fig. 3. Both logistic regression and random forest models demonstrate satisfactory predictive capability, indicating that excavation risk can be reasonably classified using monitoring data and conventional data-driven techniques.

Table 3
Performance comparison of excavation risk classification models

Model	Accuracy	F1-score	Balanced accuracy
Logistic Regression	0.876	0.879	0.746
Random Forest	0.952	0.941	0.612

The random forest model achieves higher overall accuracy and F1-score, reflecting its ability to capture nonlinear interactions among variables. Logistic regression, while slightly less accurate, still provides robust performance and serves as a transparent baseline model. The ROC–AUC values further confirm strong discriminatory ability for both models, with the random forest model exhibiting near-perfect separation capability.

These results are consistent with previous geotechnical studies reporting improved classification performance of ensemble-based models for excavation-related risk assessment, while simpler probabilistic models remain valuable for baseline evaluation and interpretability (Goh 2015; Zhang and Goh 2016; Zhang et al. 2019).

4.4 Identification of Dominant Risk Drivers

The global feature importance results derived from the random forest model are presented in Table 4 and illustrated in Fig. 4. Excavation depth emerges as the most influential variable, underscoring the dominant role of excavation geometry and stress redistribution in governing excavation risk. This finding aligns well with classical excavation behavior, where deeper excavation stages are associated with increased deformation demand and reduced stability margins (Peck 1969; Ou 2006).

Table 4
Global feature importance derived from the random forest model

Feature	Importance
Excavation_Depth_m	0.222
Seismic_Activity	0.199
Ground_Settlement_mm	0.149
Wall_Displacement_mm	0.146
Temperature_C	0.039
Groundwater_Level_m	0.035
Soil_Moisture_%	0.033
Bearing_Capacity_kPa	0.029
Shear_Strength_kPa	0.028
Strain_Gauge	0.027

Seismic activity and deformation-related indicators, including ground settlement and wall displacement, also exhibit high importance. These variables directly reflect external disturbances and system response, reinforcing the role of monitoring data in capturing evolving excavation conditions. The importance of these parameters has been widely reported in excavation monitoring studies and observational databases (Long 2001; Finno and Roboski 2006).

Environmental and soil-related variables, such as temperature, groundwater level, soil moisture, and shear strength, show moderate importance. While not dominant individually, their collective influence highlights the coupled nature of excavation behavior, where environmental conditions may modulate response under certain circumstances (Goh et al. 2017; Xu et al. 2018).

4.5 Sensitivity of Excavation Risk to Groundwater Level

The sensitivity of predicted excavation risk to groundwater level variation is examined in Fig. 5 and summarized in Table 5. Within the analyzed groundwater range, predicted risk probability remains nearly constant, and the corresponding risk gradient is effectively zero. This behavior indicates that, under the selected baseline conditions, excavation risk is not highly sensitive to small groundwater fluctuations.

Table 5
Sensitivity of predicted excavation risk to groundwater level variation

Groundwater level (m)	Predicted risk probability
0.578	0.843
0.627	0.843
0.675	0.843
0.723	0.843
0.771	0.843

From an engineering perspective, this result suggests that groundwater effects may become critical only beyond certain thresholds or in combination with other adverse conditions, such as increased excavation depth or excessive deformation. Similar threshold-dependent groundwater effects have been reported in both numerical and monitoring-based excavation studies (Hashash and Whittle 2002; Xu et al. 2018).

The feature-wise sensitivity analysis further supports this interpretation. Excavation depth exhibits the largest response variance, followed by deformation, while groundwater level shows negligible variance under the examined perturbations. These findings reinforce the conclusion that geometric and deformation-related factors dominate excavation risk in the analyzed dataset, while groundwater effects act as secondary modifiers under typical conditions.

4.6 Engineering Interpretation

Taken together, the results demonstrate that excavation risk classification using monitoring data is both feasible and informative when combined with straightforward data-driven models. The dominance of excavation depth and deformation indicators highlights the importance of geometric control and system response in excavation stability, consistent with established geotechnical understanding.

At the same time, the relatively stable response to groundwater perturbation emphasizes the need for contextual interpretation of sensitivity results. Rather than treating individual parameters in isolation, excavation risk should be evaluated considering combined effects and potential threshold behavior. The presented analysis illustrates how data-driven tools can complement traditional engineering judgment by quantifying parameter influence and response trends using routinely available monitoring data.

5. Discussion

5.1 Interpretation of Model Performance in an Engineering Context

The results presented in Section 4 indicate that excavation risk can be effectively classified using monitoring data and conventional data-driven models. The strong performance of both logistic regression and random forest models suggests that the selected monitoring variables contain sufficient information to distinguish between different excavation risk levels. While the random forest model achieves higher predictive performance, the logistic regression model also demonstrates satisfactory accuracy and discrimination capability, confirming that excavation risk classification does not necessarily require complex modeling frameworks.

From an engineering standpoint, this finding is significant. In practice, robustness and consistency of prediction are often more important than marginal gains in accuracy. The strong performance of a relatively simple probabilistic model reinforces the feasibility of applying data-driven risk assessment approaches in routine excavation projects, where transparency and ease of implementation are valued (Baecher and Christian 2003; Zhang and Zhang 2014).

The near-perfect AUC observed for the random forest model should be interpreted with caution. Rather than indicating model superiority alone, it reflects the strong association between excavation risk level and a subset of dominant parameters in the dataset, particularly excavation depth and deformation-related indicators. Similar observations have been reported in previous excavation risk studies, where monitoring variables directly reflecting system response dominate predictive outcomes (Long 2001; Finno and Calvello 2005).

5.2 Dominance of Excavation Depth and Deformation Indicators

The feature importance analysis clearly identifies excavation depth as the primary driver of excavation risk, followed by seismic activity, ground settlement, and wall displacement. This hierarchy aligns well with established geotechnical understanding, where excavation depth governs stress redistribution and structural demand, while deformation indicators provide direct evidence of system response (Peck 1969; Ou 2006).

The prominence of deformation-related variables underscores the value of monitoring data in excavation risk assessment. Ground settlement and wall displacement are routinely monitored parameters and have long been recognized as key indicators of excavation performance. Their high importance in the data-driven model confirms that these measurements capture critical aspects of excavation behavior and can serve as reliable inputs for practical risk evaluation (Finno and Roboski 2006; Ou and Hsieh 2011).

Seismic activity also emerges as an influential factor, reflecting the sensitivity of excavation systems to external disturbances. Although seismic effects may be intermittent or site-specific, their inclusion in the dataset allows the model to account for transient loading conditions that can amplify excavation risk. This observation is consistent with monitoring-based studies emphasizing the interaction between construction-induced vibrations, environmental loading, and excavation stability (Goh et al. 2017; Wang et al. 2020).

5.3 Groundwater Sensitivity and Threshold Effects

The sensitivity analysis reveals that predicted excavation risk remains relatively insensitive to small variations in groundwater level under the examined baseline conditions. The near-zero risk gradient suggests that groundwater fluctuations within this range do not significantly alter risk classification when excavation geometry and deformation response are held constant.

This result does not imply that groundwater effects are unimportant in excavation engineering. Instead, it highlights the threshold-dependent nature of groundwater influence. Numerous studies have shown that groundwater impacts on excavation stability often become pronounced only beyond certain drawdown levels or when combined with unfavorable soil conditions and advanced excavation stages (Hashash and Whittle 2002; Xu et al. 2018). The present findings suggest that, for the analyzed dataset, groundwater acts as a secondary modifier rather than a primary driver of risk.

The feature-wise sensitivity comparison further supports this interpretation, showing that excavation depth and deformation exhibit higher response variance than groundwater level. Such differentiation among parameters is valuable for practical decision-making, as it helps engineers prioritize monitoring and mitigation efforts toward factors with the greatest influence on excavation risk.

5.4 Implications for Monitoring-Based Excavation Risk Assessment

The results of this study have several practical implications. First, they demonstrate that commonly collected monitoring data can be systematically integrated into a quantitative risk assessment framework without reliance on complex or site-specific numerical modeling. This capability is particularly valuable in urban excavation projects, where rapid interpretation of monitoring information is essential for construction control.

Second, the dominance of excavation depth and deformation indicators suggests that risk assessment frameworks should emphasize geometric progression and response trends rather than isolated parameter thresholds. Monitoring strategies that focus on deformation evolution relative to excavation depth may provide more reliable insight into excavation stability than static limit values alone.

Finally, the applied nature of the proposed methodology supports its potential use as a complementary tool alongside traditional geotechnical analysis and the observational method. Rather than replacing engineering judgment, data-driven assessment can enhance situational awareness and provide an additional layer of quantitative support for decision-making (Peck 1969; Finno and Calvello 2005).

5.5 Limitations and Scope of Applicability

Several limitations of the present study should be acknowledged. The dataset represents a generalized excavation monitoring context rather than a single site-specific case history. While the parameter ranges and trends are consistent with reported excavation behavior, direct application to individual projects may require calibration using site-specific data.

In addition, the sensitivity analysis explores parameter variation within limited ranges around selected baseline conditions. More extreme conditions or combined perturbations may lead to different risk responses. Future studies incorporating time-series monitoring data and explicit excavation sequencing could further enhance understanding of excavation risk evolution.

Despite these limitations, the study provides a practical demonstration of how monitoring data and data-driven models can be integrated for excavation risk assessment in engineering practice.

6. Conclusions

This study presented a monitoring-based assessment of excavation risk using data-driven modeling approaches suitable for practical geotechnical engineering applications. By integrating geotechnical properties, environmental conditions, excavation geometry, and structural response indicators, the proposed framework demonstrates how routinely collected monitoring data can be systematically utilized to support excavation risk evaluation.

The results indicate that conventional machine learning models are capable of classifying excavation risk with high reliability when applied to comprehensive monitoring datasets. Among the analyzed parameters, excavation depth and deformation-related indicators emerge as the dominant contributors to excavation risk, highlighting the governing role of excavation geometry and system response in stability assessment. Environmental and soil-related variables act as secondary modifiers, influencing risk under specific conditions rather than serving as primary drivers.

Sensitivity analysis further shows that excavation risk response is strongly dependent on geometric and deformation parameters, while groundwater effects exhibit threshold-dependent behavior within the examined range. These findings are consistent with established geotechnical understanding and reinforce the importance of interpreting monitoring data in an integrated and contextual manner.

From an engineering perspective, the proposed approach does not seek to replace traditional analytical or numerical methods. Instead, it provides a complementary, data-driven means of interpreting monitoring information and identifying dominant risk factors during excavation. The simplicity and transparency of the adopted methodology support its potential application in routine excavation projects, particularly where rapid interpretation of monitoring data is required for construction control.

Overall, the study demonstrates that data-driven analysis of monitoring data can enhance excavation risk assessment by providing quantitative insight into parameter influence and response trends. The presented framework offers a practical foundation for incorporating monitoring-based intelligence into excavation safety management and decision-making processes.

Declarations

Ethics approval and consent to participate

This study does not involve human participants or animals. The analysis is based on secondary data and computational modeling. Ethics approval and consent to participate are therefore not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The dataset used in this study is derived from an openly available excavation risk monitoring dataset. The processed data and analysis scripts supporting the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors’ contributions

Corresponding author: Conceptualization, methodology development, modeling design, analysis and interpretation of results, writing – original draft, and supervision.
Co-author: Data preparation, model implementation, validation, visualization, and writing – review and editing.
All authors have read and approved the final manuscript.

Acknowledgements

The authors acknowledge the availability of open-access datasets and computational tools that facilitated this research.

References

1. Baecher, G.B., Christian, J.T. (2003). *Reliability and Statistics in Geotechnical Engineering*. Wiley, New York.
2. Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32.
3. Chen, Z., Zhang, Y., Wang, J. (2022). Monitoring-based excavation risk prediction using hybrid data-driven approaches. *Tunnelling and Underground Space Technology* **121**, 104338.
4. Clough, G.W., O'Rourke, T.D. (1990). Construction-induced movements of in-situ walls. In: *Design and Performance of Earth Retaining Structures*. ASCE, New York, pp. 439–470.
5. Finno, R.J., Blackburn, J.T., Roboski, J.F. (2010). Three-dimensional effects for supported excavations in clay. *Journal of Geotechnical and Geoenvironmental Engineering* **136**, 1473–1486.
6. Finno, R.J., Calvello, M. (2005). Supported excavation monitoring. *Journal of Geotechnical and Geoenvironmental Engineering* **131**, 403–412.
7. Finno, R.J., Roboski, J.F. (2006). Effects of consolidation and excavation sequencing on deformation. *Journal of Geotechnical and Geoenvironmental Engineering* **132**, 1221–1233.
8. Goh, A.T.C. (2015). Support vector machines in geotechnical engineering: State-of-the-art and future trends. *Computers and Geotechnics* **65**, 28–40.
9. Goh, A.T.C., Zhang, W., Zhang, Y. (2018). Reliability assessment of excavation-induced deformation using field monitoring data. *Computers and Geotechnics* **100**, 20–33.
10. Goh, A.T.C., Zhang, W., Zhang, Y., Xiao, Y. (2017). Monitoring-based evaluation of excavation-induced deformation. *Tunnelling and Underground Space Technology* **63**, 45–57.
11. Hashash, Y.M.A., Whittle, A.J. (2002). Ground movement prediction for excavation support systems. *Journal of Geotechnical and Geoenvironmental Engineering* **128**, 462–478.
12. Kung, G.T.C., Juang, C.H., Hsiao, E.C.L., Hashash, Y.M.A. (2009). Simplified model for wall deflection prediction of deep excavations. *Journal of Geotechnical and Geoenvironmental Engineering* **135**, 134–145.
13. Liu, Y., Wang, X., Chen, Z. (2022). Risk analysis of deep excavation considering soil–water interaction using data-driven models. *Computers and Geotechnics* **149**, 104708.
14. Long, M. (2001). Database for retaining wall movements. *Journal of Geotechnical and Geoenvironmental Engineering* **127**, 203–224.
15. Ou, C.Y. (2006). *Deep Excavation: Theory and Practice*. Taylor & Francis, London.
16. Ou, C.Y., Hsieh, P.G. (2011). Wall deformation induced by deep excavation. *Computers and Geotechnics* **38**, 287–298.
17. Peck, R.B. (1969). Advantages and limitations of the observational method in applied soil mechanics. *Journal of Geotechnical and Geoenvironmental Engineering* **95**, 171–187.

18. Sun, Y., Goh, A.T.C., Zhang, W. (2021). Progressive risk assessment of excavation using monitoring data. *Computers and Geotechnics* **135**, 104155.
19. Wang, J., Zhang, D., Li, Y. (2020). Monitoring-based risk evaluation for deep excavations. *Tunnelling and Underground Space Technology* **96**, 103207.
20. Xu, C., Shi, C., Liu, H. (2018). Time-dependent deformation behavior of deep excavations. *Tunnelling and Underground Space Technology* **77**, 55–67.
21. Zhang, D., Li, M., Zhou, W. (2020). Monitoring-based risk evaluation of urban deep excavations. *Tunnelling and Underground Space Technology* **96**, 103192.
22. Zhang, J., Zhang, L.M. (2014). Reliability-based design of excavation systems. *Journal of Geotechnical and Geoenvironmental Engineering* **140**, 04014005.
23. Zhang, Y., Goh, A.T.C. (2016). Multivariate adaptive regression splines for analysis of geotechnical engineering problems. *Computers and Geotechnics* **72**, 66–79.
24. Zhang, Y., Goh, A.T.C., Zhang, W. (2019). Risk assessment of deep excavation using machine learning techniques. *Tunnelling and Underground Space Technology* **83**, 281–293.
25. Zhang, Y., Sun, Y., Goh, A.T.C. (2021). Physics-guided machine learning in geotechnical engineering: State-of-the-art and future directions. *Computers and Geotechnics* **130**, 103884.
26. Zhou, Y., Li, X., Wu, Y. (2021). Data-driven safety assessment of deep excavations using monitoring indicators. *Safety Science* **141**, 105335.

Figures

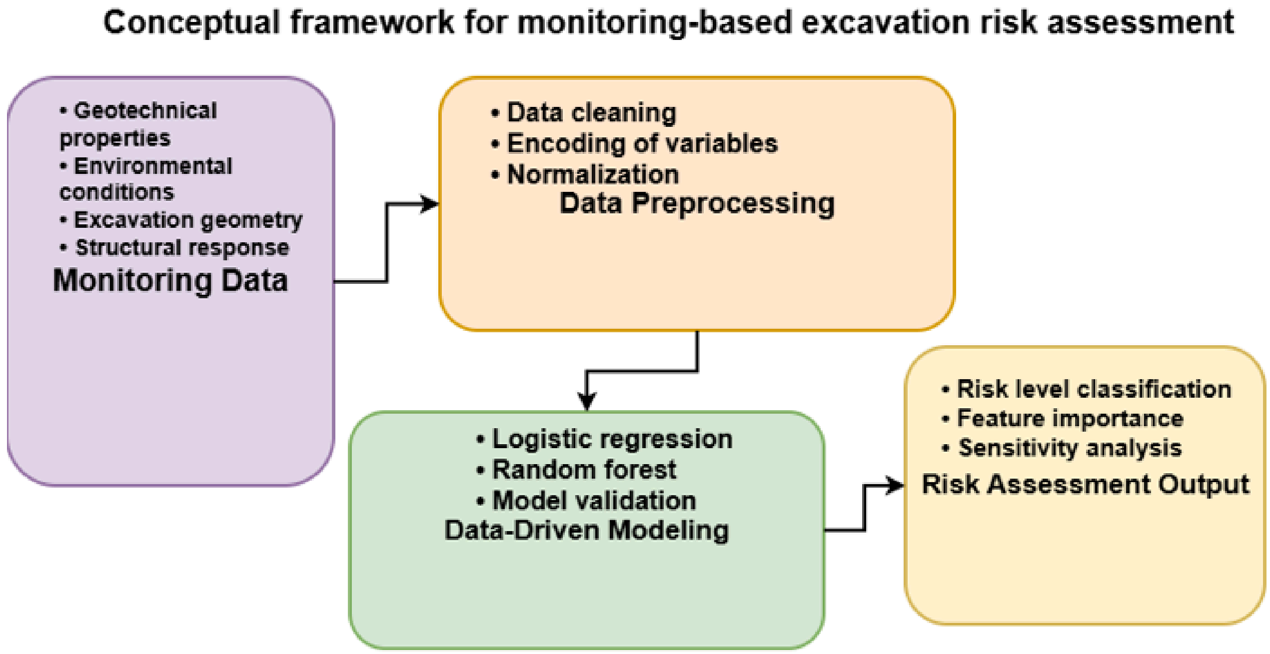

Figure 1

Conceptual framework for monitoring-based excavation risk assessment

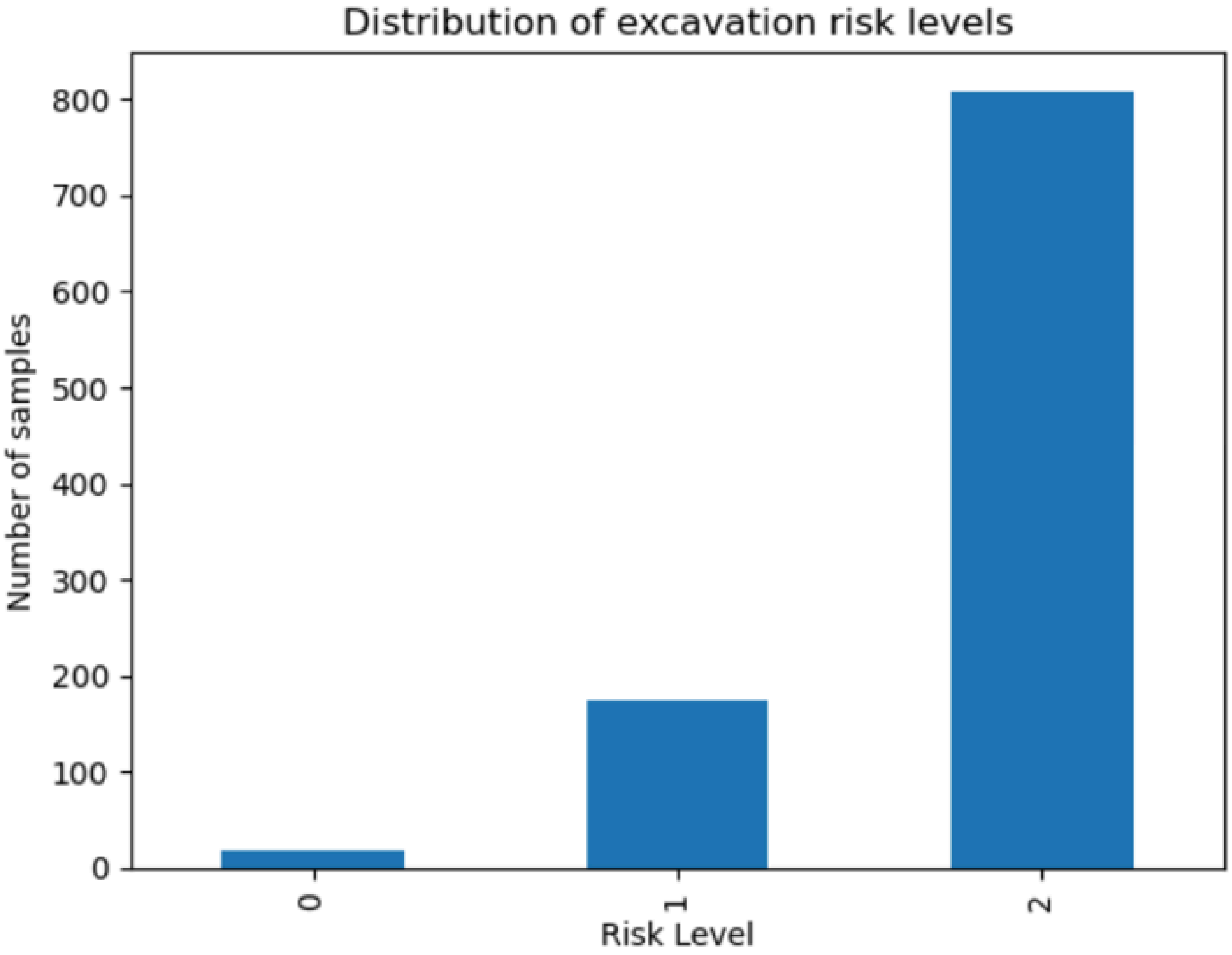

Figure 2

Distribution of excavation risk levels

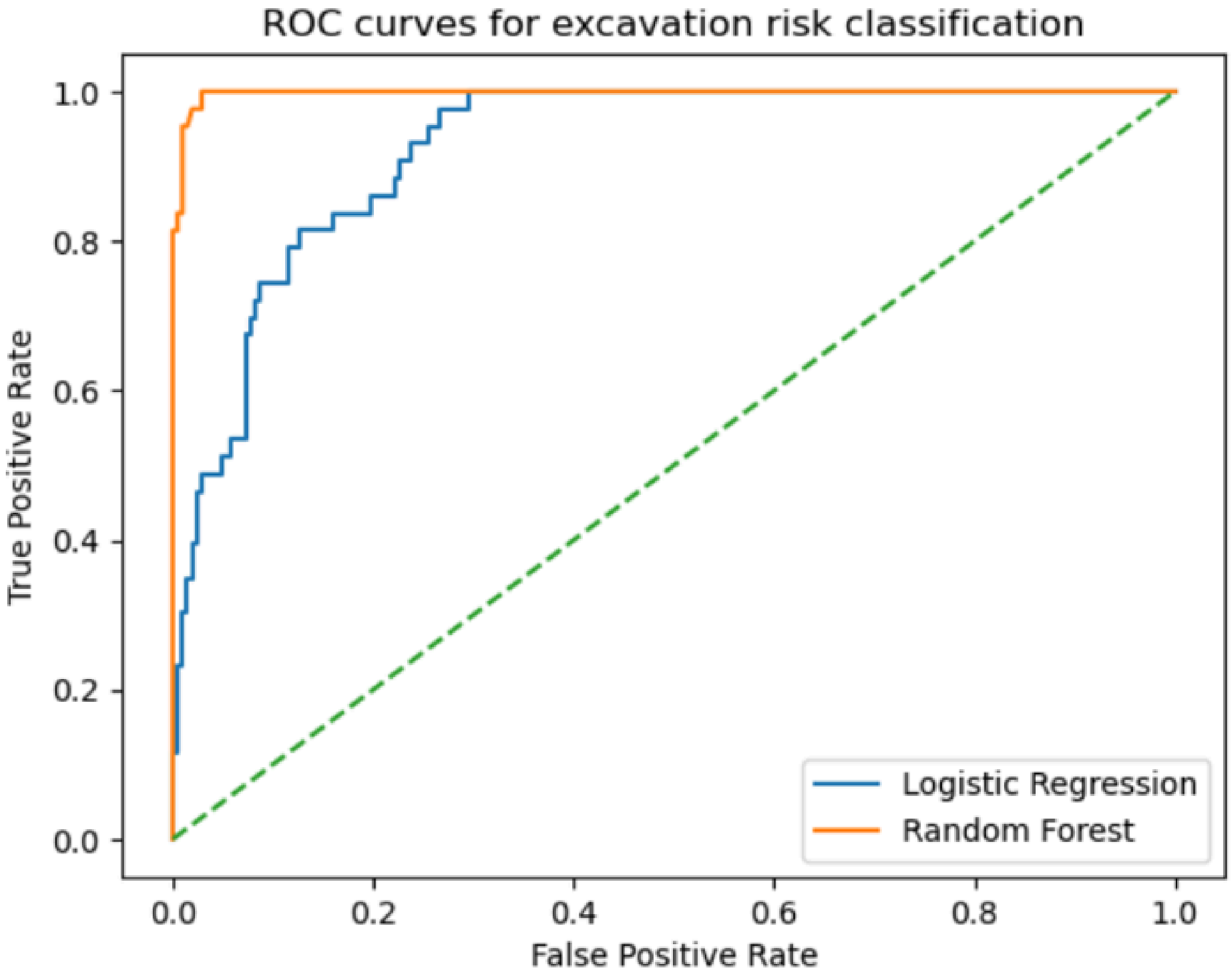

Figure 3

ROC curves for excavation risk classification models

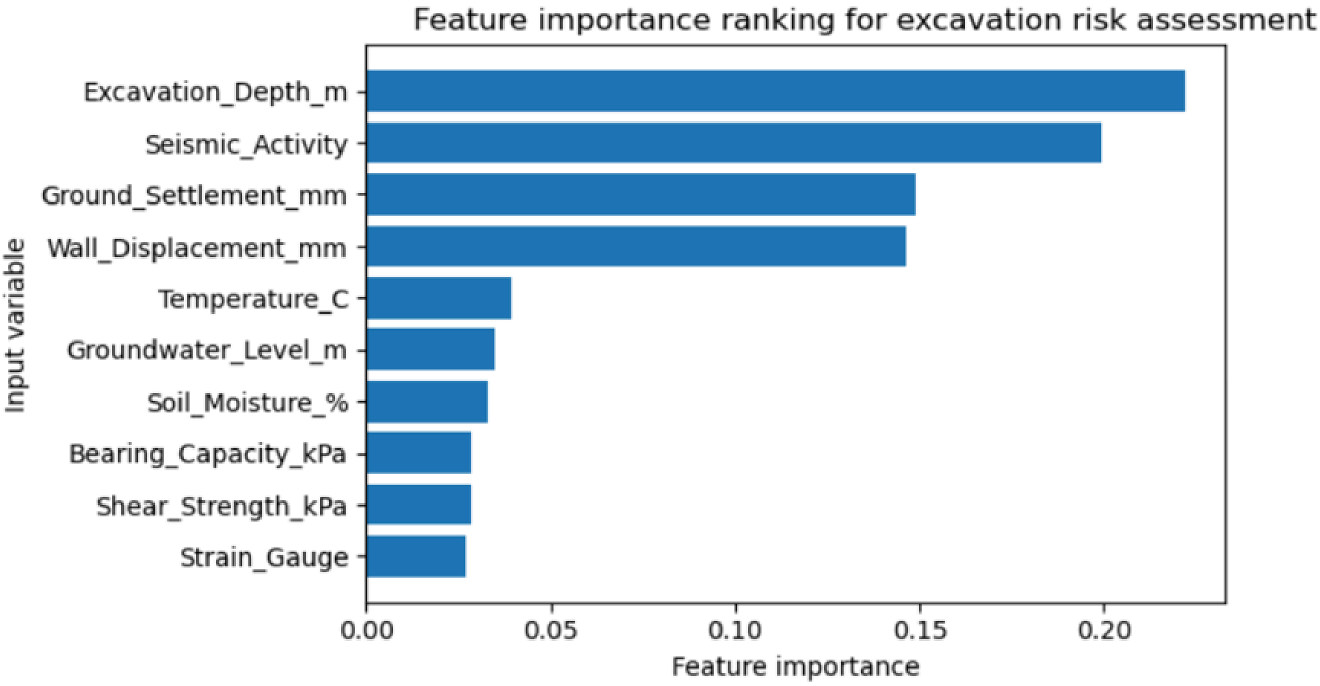

Figure 4

Feature importance ranking for excavation risk assessment

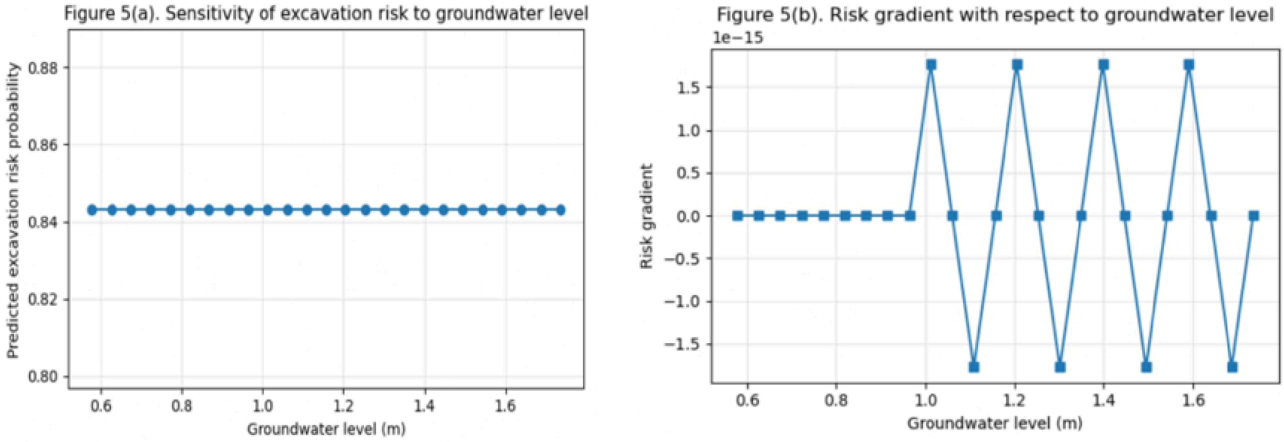

Figure 5

Sensitivity of excavation risk to groundwater level variation