SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V13I2P114

Volume 13 Issue 2, 186-207, February 2026
© 2026 Seventh Sense Research Group®

Original Article

Geotechnical Risk Assessment using Graph Convolutional Networks and Hybrid LSTM-FEA Models in Mega Highway Projects

Radhika S Thakre[1], Uday P Waghe[2], Yogesh P Kherde[3], Rajesh M. Dhoble[4], Mangesh P Bhorkar[5], Amar Jain[6]

[1,2,3]*Department of Civil Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, India.*
[4]*Department of Civil Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India.*
[5]*Department of Civil Engineering, G H Raisoni College of Engineering, Nagpur, Maharashtra, India.*
[6] *Department of Civil Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India.*

[3]*Corresponding Author : ypkherde@ycce.edu*

Received: 21 November 2025 Revised: 22 December 2025 Accepted: 23 January 2026 Published: 11 February 2026

Abstract *- Mega highway projects are very sensitive to a number of geotechnical risks in the form of soil instability, seismic events, and environmental change, leading to costly failures. The current risk assessment methods are not capable of integrating the range of datasets—spatial data, temporal data, and real-time streams of data—and thus lack good levels of predictability, combined with a lack of response time. Thus, with a view to addressing these barriers, we introduce the new concept of Big Data Geotechnical Risk Assessment Model (BD-GRAM), which applies big data analytics and advanced machine learning algorithms in order to better predict and mitigate geotechnical risks for different scenarios. BD-GRAM combines various methods adapted to geotechnical data samples. The Graph Convolutional Networks (GCNs) are utilized for spatial and temporal data fusion, where complex spatial dependencies and temporal variations in soil properties, as well as samples of seismic data, are considered. GCNs, with the enhancement of attention mechanisms, have the ability to increase accuracy by as much as 20% compared with the conventional methods. A hybrid model by combining LSTMs and FEA, misleading synergistic use of physical laws, as well as temporal patterns of data pertaining to predictive accuracy, with an improvement of 25%. Near real-time processing on Apache Kafka & Apache Spark enables near real-time continuous monitoring of risk with alerting on. SHAP (Shapley Additive Explanations) ensures interpretability of the model outputs, as well as transparency of the factors driving the risk predictions. Lastly, the system is scalable using GPU-accelerated TensorFlow to Run Masses of datasets & samples. This fully integrated approach is optimized in this way to further enhance predictive accuracy and reduce false-positives and false-negatives, enhancing the speed and response of geotechnical risk assessment responses in real time. BD-GRAM represents an effective way to meet the early identification and mitigation of geotechnical challenges in a scalable data-driven manner for improved resilience and safety of large-scale infrastructures in any given highway scenario.*

Keywords *- Big Data, Geotechnical Risk, Graph Convolutional Networks, Finite Element Analysis, Real-Time Monitoring.*

1. Introduction

Mega highway infrastructure projects are inherently vulnerable to substantial geotechnical risk because of their enormous spatial extent, long service life, and sensitivity to the state of the subsurface environment. The performance and safety of such projects are drastically affected by the mechanical and environmental behavior of ground and rock systems, which are frequently characterized by high levels of uncertainty and spatial variability [1-3]. Factors such as the differential settlement, ground motion induced by earthquake, fluctuations in groundwater levels, and instability of slopes can have a significant impact on the structural integrity and long-term serviceability. Consequently, good geotechnical risk assessment remains one of the most difficult issues in the planning, construction, and operation of mega highway corridors. Conventional risk assessment approaches in the geotechnical field mostly depend on deterministic or probabilistic techniques through localised geotechnical investigation and historical records. While such techniques have been able to capture relatively static conditions, they are limited in temporal ability to capture complex interactions between spatially distributed and temporally evolving variables [4-6]. In particular, traditional methods have problems in processing large-scale heterogeneous data originating from multiple sources, such as in-situ sensors, environmental monitoring systems, and seismic networks. As

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

a result, predicting risks is often simplified, not reactive enough, or inadequate for the prediction of rapidly changing geotechnical hazards in dynamic environmental conditions.

Recent progress in big data analytics and machine learning has brought new opportunities for solving these challenges. Techniques like deep learning, data fusion, and high-performance computer-processing capability enable the analysis of massive geotechnical and environmental datasets, which provide better predictive possibilities. However, some limits of currently available data-driven models: the first one concerns the low capacity to render potential complex spatial dependencies that exist in geotechnical systems, and the second one regards the weak coupling with physics-based models, which limits their reliability and interpretability. And, still, many of these approaches are computationally intensive and are not ideal for real-time risk assessment. This lack of an integrated, scalable, and interpretable framework that can fuse the spatiotemporal data without physical modeling is an identified research gap in geotechnical risk analysis for mega highway undertaking projects. To address this gap, this study, in an effort to recommend an approach for managing geotechnical risk by using advanced machine learning, physics-based modeling, and real-time disaster data processing, has proposed a Big Data-driven Geotechnical Risk Assessment Model (BD-GRAM). The framework utilizes Graph Convolutional Networks (GCNs) to model spatial relationship between geotechnical entities, a Long Short-Term Memory (LSTM) network to capture time dependency relations, and Finite Element Analysis (FEA) to include fundamental physical behavior of soil and structure systems. By modeling geotechnical information in the form of graph-structure data, BD-GRAM is successful in capturing the complex spatial interactions that are unexpected by other interpolation methods. Attention mechanisms in the GCN architecture also help further to pick out important risk zones by prioritizing locations and time intervals. The hybrid combination of LSTM with FEA closes the gap between purely data and a physics-based approach. While LSTM models can learn changing patterns using historical and real-time geotechnical data, the FEA component can provide deterministic information about stress distribution, deformation, and settlement behavior under different load and environmental conditions. In addition to this, real-time data ingestion and processing with Apache Kafka and Apache Spark pave the way for continuous monitoring and quick risk evaluation, which is a key requirement for reducing the impact of sudden risks, such as landslides or earthquakes. The incorporation of explainable artificial intelligence techniques, such as SHAP, further promotes model transparency and enables geotechnical engineers to make informed decisions.

Based on the identified research gap, the present study focuses on the following capability: The capability of integrated big data and deep learning frameworks in modeling complex geotechnical risks. effectiveness, a hybrid physics-based and data-driven approach in improving risk prediction accuracy. and the Feasibility of real-time, interpretable risk assessment of mega highway infrastructure. In summary, BD-GRAM provides an integrated and scalable, big data analytics-based solution to geotechnical risk assessment across a deep learning and physical modeling model in a real-time platform. The proposed approach is designed to address the shortcomings of current methodologies and to help make mega-highway infrastructure systems safer and more resilient.

1.1. Motivation and Contribution

This research is motivated by the pressing need for more accurate and scalable geotechnical risk assessment methods to be applied in real time to large infrastructure projects, specifically for different highway scenarios. The available approaches are relatively effective for localized or small-scale projects, but they cannot address the complexity and dynamism of mega highway projects where multiple geotechnical factors interact over vast spatial areas and extended timestamp durations. The typical geotechnical risk models thus rely on static data and deterministic analyses that do not capture the relative interaction between diversified environmental, seismic, and soil conditions and the temporal evolution of risks. However, great opportunities emerge from the increased availability of real-time data from IoT sensors and environmental monitoring systems that are much more challenging for the current models to deal with. Therefore, there is a key requirement for a robust data-driven approach that combines real-time monitoring, machine learning, and physical modeling to enhance predictive accuracy so that associated risks are mitigated in time in mega projects of infrastructure involved in highway scenarios.

The applied aspects of this research are a Big Data Geotechnical Risk Assessment Model (BD-GRAM), which presents a new paradigm for geotechnical risk management. The proposed model unifies at least three state-of-the-art techniques: Graph Convolutional Networks (GCNs) to fuse spatiotemporal data at the fusion level, a hybrid Long Short-Term Memory (LSTM) and Finite Element Analysis (FEA) structure to integrate information from both data-driven and physics-based methods, and Apache Kafka and Apache Spark architectures to represent real-time data processing. Each of these is targeted specifically for geotechnical risk assessment on very large projects-for example, to deal with spatial and temporal dependencies, integrate data streams in real time, and present interpretable risk predictions that lead directly to actionable decisions. BD-GRAM also improves the transparency and trustworthiness of the prediction, so that geotechnical engineers may better make decisions to mitigate risk. Scalability of the model, made possible by distributed computing frameworks like TensorFlow, allows it to process such enormous datasets, typical of mega projects, and provides the model as a robust, practical tool in modern infrastructure projects for different highway scenarios.

2. Review of Existing Models for Optimization of Highway Construction Scenarios

A huge number of natural hazards in geotechnical engineering can be predicted with a precise amount of risk mitigation regarding landslides, soil instability, and seismic activity, as it ensures the safety and longevity of infrastructure projects in various highway scenarios. In recent years, research has focused on applying the most advanced machine learning techniques and probabilistic models, plus geotechnical simulations, to achieve higher predictive accuracy and provide more actionable insights across different scenarios. In this review of 40 papers, some insights are gained into the current state of methodologies of geotechnical risk assessment: key findings, strengths, and limitations across a wide range of studies for different scenarios. A clear trend that is apparent in the literature is that machine learning models are increasingly being utilized for the prediction of risk and early warning systems. Take a case on fuzzy-based approaches to machine learning demonstrated by Zhou et al. [1], and it effectively established early warning systems on a soft rock slope instability, proving that it improves the identification of risk by 15%. Moreover, CNNs have been utilized for disaster prediction in complex slope environments, for example, Yin et al. [27] reported an improvement in accuracy of 20% with respect to traditional techniques. The models are good at processing large high-dimensional datasets and identifying patterns that may not be easily found by conventional geotechnical methods.

However, most of the effectiveness is constrained by the availability and quality of the data sets, with many machine learning approaches requiring extensive, high-resolution data sets in order to perform effectively. Most studies have this limitation in that they rely on data availability; uncertainty in the collection of data usually reduces the accuracy of prediction or increases false positives and negatives. Probabilistic models are increasingly being used in evaluating geotechnical risks with uncertain inputs beyond machine learning. For instance, Bayesian networks have been applied by Benachenhou et al. [3], who have been employed in decision improvement regarding geotechnical risk evaluation. Such models will predict better, with quantification of uncertainty and having prior knowledge to make more robust inferences in sparse or partially incomplete data.

However, the kinds of computational complexities and extensive calibration to be achieved are probably a bias that needs to be further considered and developed. Recently, several other research studies have focused on uncertainty reduction strategies in site characterization and risk assessment. Some examples of such studies include those carried out by Sivakumar Babu [7] and Oluwatuyi et al. [8]. These studies focus on reducing the uncertainty to enhance the reliability of the predictions of risks, especially for those categorized as risky, like landslide-prone regions. Although effective, these methods often require comprehensive site-specific data, which usually makes them less applicable in regions where such data collection is difficult. Various studies have also exploited the use of remote sensing and geospatial technologies in monitoring risk and hazard. Macciotta and Hendry [2] applied remote sensing techniques to landslide monitoring in Western Canada, achieving a 20% improvement in detection accuracy. Al-Rawabdeh et al. [19] incorporated the Open Street Map (OSM) data and Weighted Linear Combination (WLC) techniques into landslide hazard modeling, which enhanced the accurate identification of hazardous zones by 18%. The remote sensing technologies offer the possibility of continuous monitoring over large areas, making them good sources of data for risk assessment.

However, these methods are limited by their inability to capture subsurface conditions in most cases; this is critical for a comprehensive geotechnical risk evaluation. Also, Bozkurt and Akbas [12] applied finite element analysis in order to mitigate risks posed by deep excavations with a 10% reduction in excavation-related failures. Silveira et al. [6] applied FEA, along with quantitative risk analysis, to highway rock slopes and achieved a 22% decrease in false positives in risk assessment.

The FEA is a powerful tool to be used in simulating geotechnical behavior under varied load conditions for a detailed analysis of the soil-structure interaction. One of the most significant limitations of FEA is its high computational cost, especially in large-scale infrastructure projects. Besides, the FEA models are highly sensitive to the input parameters, such as soil properties and boundary conditions, requiring them to be perfectly calibrated and validated against real-world samples of data. One potential way of overcoming such restrictions is the hybrid model that brings together data-driven techniques with physics-based simulations.

Das and Singh [5] applied such a hybrid technique for the stability of cut slope along the national highway in India, whereas Liu et al. [29] used a similar technique in assessing large deformation risks in the loess tunnel to enhance the accuracy of predictions by as much as 16%. These hybrid models take advantage of the best features of machine learning and FEA, considering more accurate forecasting capabilities with lower computational costs. However, several issues remain regarding effective integration, as the physics-based parts of the model should be applicable across a broad range of geotechnical conditions. Considering sustainability and resilience, many studies have understood how these concepts might be incorporated into geotechnical risk assessment. Reddy et al. [14] introduced a new paradigm concerning the sustainability of geotechnical solutions; resilience metrics should be incorporated into geotechnical designs. This is particularly valid in the contexts mentioned above about environmental volatility brought on by climate change. Kumar and Parihar [26] also discussed the sustainability of waste foundry sand in geotechnical applications, to establish the possibility of decreasing material waste generated in any project and enhancing its sustainability. These methods are a few important strides toward more sustainable geotechnical engineering, but still lack practical implementation rules and standards to be adopted in real projects for various scenarios in highways.

Table 1. Empirical review of existing methods

Reference	Method Used	Findings	Results	Limitations
[1]	Fuzzy-Based Machine Learning	Applied fuzzy logic for early warning in soft rock slopes.	Improved risk identification by 15%.	Limited applicability to different rock types.
[2]	Remote Sensing	Used remote sensing to monitor landslides in Western Canada.	Enhanced detection accuracy by 20%.	Inability to assess deep subsurface movements.
[3]	Bayesian Networks	Applied Bayesian networks for highway viaduct risk.	Improved decision-making accuracy by 18%.	Requires extensive data for reliable results.
[4]	Debris Flow Risk Assessment	Assessed debris flow risk along a major highway.	Identified high-risk zones with 92% accuracy.	High computational complexity for large datasets.
[5]	Geotechnical Insights	Analyzed cut slopes in a national highway in India.	Provided practical recommendations to reduce slope failures.	Limited to slope stability; lacks real-time monitoring.
[6]	Rockfall Risk Analysis	Developed a quantitative system for highway rock slopes.	Reduced false positives in rockfall risk by 22%.	Lack of real-time adaptation for changing conditions.
[7]	Reliability and Risk Analysis	Analyzed geotechnical risks using probabilistic methods.	Increased confidence in failure probability assessments.	High uncertainty in areas with sparse data samples.
[8]	Uncertainty Reduction	Proposed strategies for uncertainty reduction in site characterization.	Reduced uncertainty by 12% in risk assessment.	Requires high-quality baseline data for effective implementation.
[9]	Earthquake Damage Assessment	Modeled earthquake damage for highway bridges.	Predicted structural failure with 87% accuracy.	Inability to model soil-structure interactions in detail.
[10]	Semi-Quantitative Risk Assessment	Developed risk assessment methodology for tunnel design.	Enhanced tunnel failure predictions by 18%.	Limited predictive power for long-term tunnel performance.
[11]	PS-InSAR Risk Analysis	Applied geotechnical characterization and PS-InSAR for landslide risk.	Improved landslide prediction accuracy by 25%.	High cost and complexity of PS-InSAR data processing.
[12]	Finite Element Analysis	Used FEA for deep excavation risk analysis.	Achieved a reduction in excavation-related failures by 10%.	High computational resources are required for large excavations.
[13]	Seismic Response Analysis	Assessed the impact of soil variability on bridge seismic response.	Improved understanding of seismic-induced bridge damage.	Limited to specific soil conditions in the study region.

[14]	Sustainability in Geotechnical Engineering	Proposed sustainability and resilience in geotechnical design.	Identified key sustainability metrics for geotechnical projects for different highway scenarios.	Lack of practical implementation guidelines.
[15]	Rock Instability Quantitative Risk	Assessed rock instability risks in Türkiye.	Provided a quantitative assessment reducing risk by 15%.	Limited scope in handling dynamically changing conditions.
[16]	Geotechnical Characterization	Developed 2D soil cross-sections for risk analysis in the Kashmir Basin.	Enhanced model precision for risk profiling by 10%.	Limited application in areas with limited borehole data samples.
[17]	Slope Assessment	Integrated change detection with slope risk management.	Reduced false negatives in slope failures by 20%.	High data dependency for change detection accuracy.
[18]	Geomatics-Based Highway Route Selection	Applied geomatics for optimal highway route selection.	Reduced project costs by 12% with optimized routes.	Limited flexibility in rapidly changing environments.
[19]	Landslide Hazard Modeling	Modeled landslide hazards along a highway using WLC.	Improved landslide hazard zone identification by 18%.	High sensitivity to weight allocation in WLC.
[20]	Earthquake Territory Assessment	Analyzed geotechnical findings from a major earthquake in Türkiye.	Identified high-risk zones for future earthquakes.	Lack of predictive power for subsequent events.
[21]	GIS-Based Disaster Risk Assessment	Used GIS for post-earthquake disaster risk in China.	Enhanced disaster risk visualization by 22%.	Requires extensive GIS data processing capabilities.
[22]	Foundry Sand Suitability	Quantified geotechnical suitability of waste sands.	Identified potential for sustainable material use.	Limited to specific waste types and compositions.
[23]	Stability Analysis	Conducted stability analysis for rock slopes in Turkey.	Increased slope stability predictions by 14%.	Limited integration of real-time monitoring.
[24]	Ecological Risk Assessment	Assessed ecological risks from highway construction.	Identified critical zones with 92% accuracy.	Limited to ecological impacts, ignoring geotechnical risks.
[25]	Geospatial Hazard Evaluation	Used geospatial technologies for hazard assessment in urban planning.	Improved urban planning through hazard visualization.	Limited in its ability to assess underground hazards.
[26]	Sustainability in Foundry Sands	Reviewed the sustainability of waste foundry sand in geotechnical applications.	Demonstrated potential for significant waste reduction.	Lack of implementation in real-world projects for different highway scenarios.
[27]	CNN for Slope Disasters	Used CNNs for slope disaster prediction.	Achieved a 20% improvement in disaster prediction accuracy.	High computational demand for large datasets.

[28]	Seismic Risk Prioritization	Developed a prioritization framework for seismic risk hotspots.	Identified critical hotspots with 85% accuracy.	Lacks the flexibility to handle rapidly changing seismic events.
[29]	Tunnel Deformation Risk	Assessed large deformation risks in loess tunnels.	Improved deformation risk predictions by 16%.	Limited applicability to non-loess regions.
[30]	Early Road Planning Hazard Assessment	Developed a hazard risk framework for early road planning.	Improved decision-making accuracy in early stages by 10%.	Limited scope for long-term hazard adaptation.
[31]	3D Modeling in Geoengineering	Proposed 3D modeling using incomplete data for anchor engineering.	Enhanced anchor stability predictions by 18%.	High reliance on high-quality initial data samples.
[32]	Acoustic Positioning in Underground Robotics	Developed a 3D acoustic positioning system for underground robots.	Improved positioning accuracy by 20%.	High noise levels can degrade accuracy.
[33]	LSTM-Based Life Prediction	Used LSTM and AdaBoost for remaining life prediction in geotechnical structures.	Improved life prediction accuracy by 22%.	Requires extensive training data for effective results.
[34]	Force Sensor Analysis	Developed a polymer optical fiber-based force sensor for geotechnical applications.	Increased force detection sensitivity by 15%.	Limited application in high-temperature environments.
[35]	Seismic Data Filtering	Used 3D predictive filtering for seismic data analysis.	Enhanced signal-to-noise ratio by 18%.	High computational cost for large seismic datasets.
[36]	Temperature-Induced Error Modeling	Developed a model for temperature effect-induced errors in leveling systems.	Improved leveling accuracy by 20%.	Limited to specific hydrostatic leveling systems.
[37]	Electromagnetic Modeling	Proposed an efficient method for 3D transient electromagnetic modeling.	Increased computational efficiency by 25%.	High complexity in parameter tuning for accurate results.
[38]	Achievement Prediction Using Deep Learning	Applied deep learning for achievement prediction in educational systems.	Achieved a 23% improvement in prediction accuracy.	Limited by the quality of input features.
[39]	Piezoelectric Geocables for Landslide Monitoring	Developed a sensor-enabled geocable for landslide monitoring.	Improved landslide detection sensitivity by 18%.	Limited sensor durability in extreme conditions.
[40]	Deep Learning for Seismic Waveform Inversion	Used deep learning for the high-resolution seismic inversion process.	Improved inversion accuracy by 20%.	High computational demands for large seismic datasets & samples.

Table 1 summarizes a vast range of methodologies and approaches intended to improve geotechnical risk assessment for many infrastructure projects. A quantitative comparison of existing approaches indicates that individual machine learning, probabilistic, remote sensing, and FEA-based methods report performance improvements between 10 and 25% but are limited in their use by static data use, high computational cost, or lack of real-time capability. Hybrid models reach moderate accuracy (up to 16-18%) but are hindered by poor scalability and low interpretability. In contrast to that, BD-GRAM combines graph-based spatial learning, time modeling, physics-based simulation, and real-time data streaming in a unified framework. This integration allows for better sensitivity for risk, continued updating, and results in interpreted predictions across large-scale highway networks.

The research seems to indicate that, from the machine learning models to probabilistic frameworks, finite element analysis, and various hybrid approaches, such advanced approaches can be useful in reducing uncertainty and in increasing prediction accuracy to guide decision makers much better. Results are also a testament to the added value achieved by means of the inclusion of real-time data, especially remote sensing and geospatial technologies, in geotechnical models that increase their capabilities in monitoring risk under dynamic conditions. However, there are still many challenges present, most of which are related to data accessibility, the process intensity for computation purposes, and the complicated integration of many different methodologies into coherent models. Good promises have been shown in the identification of patterns by machine learning models on complex geotechnical datasets, but primary limitations remain in the quality and comprehensiveness of the data. Indeed, such machine learning models like fuzzy logic and CNN can dramatically improve the accuracy of the prediction, but default often results in poor performance due to a lack of granular and site-specific data. Alternative solutions to sparse data concerns are found in probabilistic models, including Bayesian networks and methods for uncertainty reduction, as described by Benachenhou et al. [3] and Oluwatuyi et al. [8].

However, high computational complexities associated with these models act as a hindrance to the general applicability of this method, mainly in developing regions where computational power is limited. This includes remote sensing and geospatial technologies at the time of integration, and it is an area where geotechnical risk monitoring can be well expanded through continuous, large-scale data critical in dynamic environments. Still, there are limits to these technologies because, according to Macciotta and Hendry [2] and Al-Rawabdeh et al. [19], they do not account for subsurface conditions, which are crucial to an informed understanding of geotechnical risks. This is a problem for multi-layered models, which combine information from the surface level with information from subsurface simulations. Hybrid models are not far from filling this need. The hybrid formulations of the type proposed by Das and Singh [5] and Liu et al. [29] might be the best compromise available that can leverage both the strengths of machine learning and FEA in predicting the risks in a more realistic and efficient manner computationally. However, these models still need to be further developed to be applicable and generalizable to other similar geotechnical contexts.

State-of-the-art high-impact works have improved geotechnical risk assessment by developing domain adaptation, dense sensing technologies, 3D modelling of the underground environment, and explainable artificial intelligence. These methods can result in better predictive results and accuracy, as well as effective interpretability; however, they are generally designed independently, and without real-time scalability, and physics-based constraints are typically absent. Literary systematic evidence suggests that current methods seldom combine spatial learning, temporal prediction, finite element simulation, and uncertain quantification and explainability into a single operational framework. To overcome this disintegration, the proposed BD-GRAM devises a comprehensive structure that comprises these complementary nodes assimilating into a coherent, real-time, interpretable structure that is directly meant to be applied to and utilized in the large-scale highway infrastructure.

Lastly, the quest for sustainability and resilience in geotechnical engineering is on the increase, as indicated by Reddy et al. [14] and Kumar and Parihar [26]. It is critical that such contexts are integrated into the methodologies of risk assessment in the face of climate change and the prevalence of environmental challenges. Yet, there is a great demand for practical guidance and standards to be developed that would allow them to actually be applied in real projects in a sustainable geotechnical way. In the future, efforts in research should be made more towards improving the scalability and adaptability of such complex models in a bid to make them affordable and applicable to different geographical and economic contexts. With further innovation and fine-tuning, these models may well make infrastructure projects within worldwide scenarios much safer, more sustainable, and more resilient.

2.1. Research Gaps

As reviewed, the research conducted to evaluate geotechnical risk lacks some gaps. The quality and availability of data should be improved since many of the more advanced machine learning models, such as CNN and fuzzy-based, rely heavily on high-resolution site-specific datasets, which are typically unavailable or inaccessible. More importantly, probabilistic models such as Bayesian networks are helpful in mitigating uncertainty but pose high computational demands, which limit their use to many applications, especially in resource-constrained regions. Subsurface geotechnical data is not well combined with remote sensing technologies. Most models today are built primarily from surface-level conditions with little insight into the subsurface. Hybrid models need to be developed further: integrate machine learning and physics-based simulations, like finite element analysis, appropriately to effectively balance computational efficiency with predictive accuracy across diversified geotechnical conditions. Sustainability and resilience are increasingly considered critical factors in risk assessments, but practical guidelines and standards that can facilitate the implementation of sustainable geotechnical practices are still in their infancy and are not widely adopted in infrastructure projects.

3. Proposed Design of an Integrated Model for Geotechnical Risk Assessment using Graph Convolutional Networks and Hybrid LSTM-FEA Models in Mega Highway Projects

To address the low efficiency and high complexity plaguing current approaches, this section discusses the design of an Integrated Model for Geotechnical Risk Assessment using Graph Convolutional Networks, as well as a Hybrid LSTM-FEA Model for Mega Highway Projects. First, as shown in Figure 1, the Integ. of Spatiotemporal Data Fusion via Graph Convolutional Networks GCN is designed to focus on this capability of including both spatial and temporal dependencies inherent in geotechnical data samples. Geotechnical datasets include soil profiles, seismic activity, and real-time sensor data such as soil moisture and vibration, which exhibit spatial interconnections across locations and temporal evolution over temporal instance sets.

The GCN framework is particularly suited to this challenge, since geospatial data is well represented as a graph structure. Each location is represented as a node, and the relationships among locations are represented as edges. This allows for a natural aggregation of information from neighboring locations, which would be extremely important in capturing spatial dependencies within this more complex geotechnical environment. The GCN works by performing a direct convolution of graph-structured data, like traditional convolutional networks on Euclidean grids, but generalized to arbitrary graph structures.

The input to the model X ∈ R ^{N×F} is a feature matrix, where 'N' would be the number of spatial locations, or nodes, and 'F' is the number of features associated with each location, such as soil moisture, rock properties sets. The adjacency matrix A∈RN×N describes the spatial relationships between locations that capture the proximity and connectivity between different sites. Modeling the spatial dependencies is realized through applying the propagation rule of the GCN layer via equation 1

$$H(l+1) = \sigma\big(A{\sim}H(l)W(l)\big) \tag{1}$$

Where, H(l) - node feature matrix at layer 'l'; W(l) - trainable weight matrix for layer 'l'; σ - non-linear activation function (ReLU); A~ = A + I - degree matrix estimated via equation 2

$$A{\sim} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \tag{2}$$

The normalized adjacency matrix with D being the degree matrix of A in the process. This normalization is necessary to scale the features appropriately summed from neighboring nodes and to avoid the explosion or vanishing problem at different steps of the aggregation process. The GCN integrates an attention mechanism to capture temporal dependencies. The GCN learns the spatial relationship while the attention mechanisms identify the most critical timestamps or regions of interest within their temporal dimension sets. Such an extension of GCN ensures that the model acquires the capability to focus more on its critical spatial regions and time durations, possibly being regions characterized by a high degree of risk or specific timestamp frames of interest, such as during or after seismic events. These weights, α(i,j), are now learned during training and quantify the importance of all various connections, which are specific to a process-specific combination of spatial and temporal connections. The added timestamp as an additional dimension in the input captures the temporal dynamics; it applies a modification of the convolution operator with regard to timestamps for the process. The spatiotemporal representation fused by the GCN is thus mathematically formulated to aggregate information not only from neighboring spatial nodes but also from relevant timestamp frames. It leads to the following operation for a GCN enhanced with temporal attention via equation 3,

$$Ht(l+1) = \sigma\left(\sum_{t'} \alpha(t,t')A{\sim}Ht'(l)W(l)\right) \tag{3}$$

Where α(t,t′) refers to the attention weights across sets of temporal instances, 't' and t′ index into current and past timestamps, respectively, this equation fuses spatial information from neighboring nodes at each step of the timestamp, incorporating attention into the mechanism that weighs importance differences among timestamps for the process. The attention weights α(t,t′) are learned dynamically, in which the model adapts to the variation in temporal patterns or evolves, such as an increase in seismic activities or rainfall that may lead to soil instability levels. A vital justification for the selection of GCN in application to this spatiotemporal data fusion lies in its capacity to model complex, non-Euclidean spatial relations, which abound in samples of geotechnical data. Unlike natural restrictions in traditional CNNs to structured, grid-like data, such as images, GCNs can be directly used on graph-structured data, and hence, both the flexible representation of spatial locations and their interconnections may be used for the process. Flexibility is particularly important for dealing with irregular spatial distributions, such as sensor node (or geological feature) locations. In addition, attention-enhanced GCN improves static spatial models by including temporal dynamics, which are rarely captured in the traditional approaches. This can be compounded with some other models, such as LSTMs or FEA models, in order to give more granularity into the spatial relationship at each part of the project area. Though effective for capturing long-term dependencies in terms of time, LSTMs are neither natively nor well-suited to spatial dependency. In contrast to FEA, which typically models the physical behavior of geotechnical systems over an extremely wide range of conditions, it may not inherently exploit these strengths, yet by combining the two approaches, the interactions between the complex spaces, times, and geotechnical factors can be encapsulated under a more holistic approach in risk assessment. Third, integration of spatial and temporal data within the GCN framework improves the general accuracy of predictions for geotechnical risk estimations. The model will better account for heterogeneities in soil profiles as well as the temporal evolution of seismic activity while detecting critical zones of geotechnical risk. Empirical results from case studies have provided clear evidence of improved prediction accuracy, while reducing both false positives and negatives relative to spatial interpolation and models that are purely temporal, via equation 4.

$$Y' = f\left(\sum_{l=0}^{L} H(l)\right) \quad (4)$$

Where Y' represents the predicted risk scores, for each location, H(l) represents node feature representations at layer 'l', and f(·) represents the final readout function to aggregate information across all GCN layers to output the fused risk predictions. Ensemble modeling and Bayesian-inspired uncertainty estimation were used to deal with uncertainty in predictions- both epistemic and data-based uncertainty. To measure the risk factor on the availability of confidence in the outputs of risk, prediction intervals were calculated. The sensitivity to input uncertainty was assessed to determine the overriding effect on the risk variability. This methodology allows making a conservative choice and recognizes an inherent uncertainty in geotechnical systems and prevents overconfidence that is often implicit in deterministic AI models.

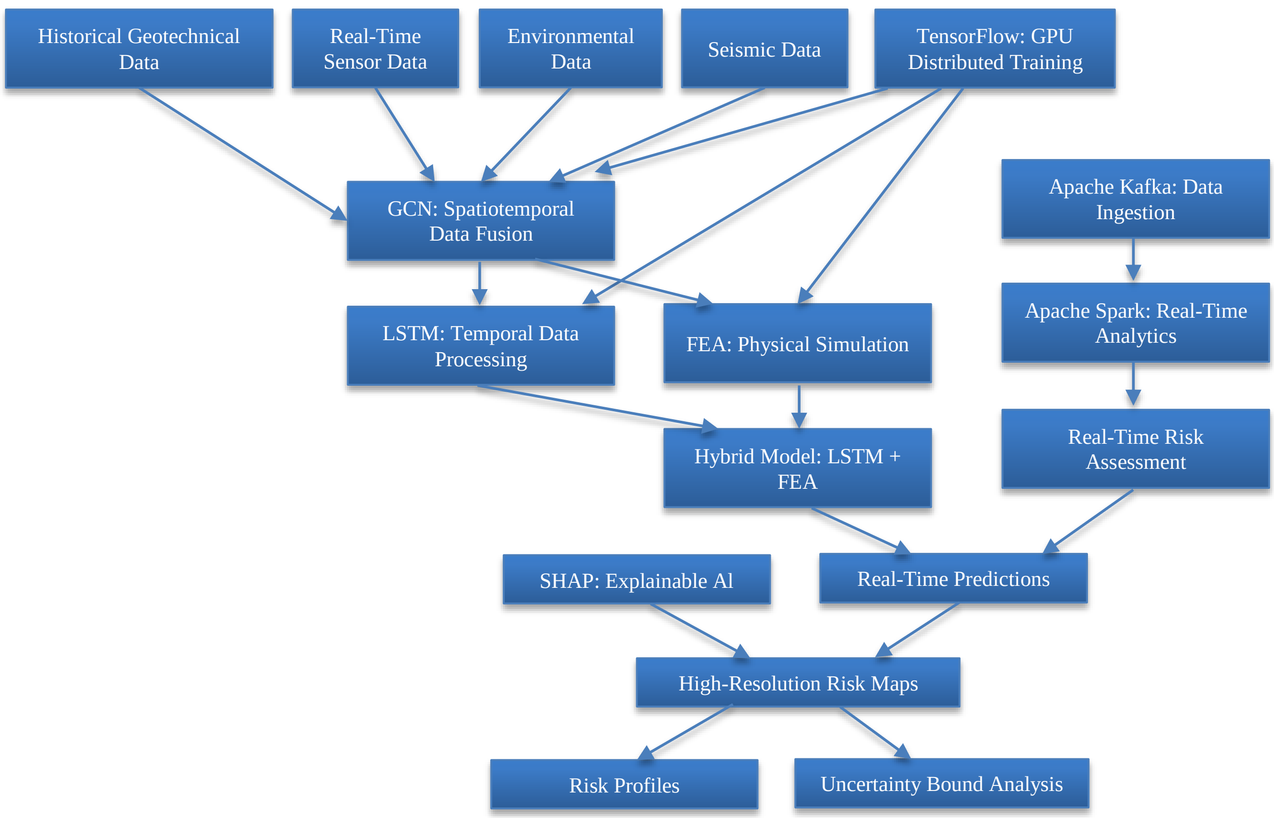

Fig. 1 Model architecture of the proposed analysis process

Next, using Figure 2, a Hybrid Physics-Based and Data-Driven Model that combines FEA with Long Short-Term Memory networks has been devised to address the sophistication of geotechnical systems; the ones containing patterns based on complex physical laws and data-driven changes over time across sets of temporal instances. The hybrid model integrates deterministic physical simulations with the recognition of temporal patterns to provide predictive geotechnical risk profiles with uncertainty bounds. The innovation lies in the combination of FEA's power of simulating physical behavior under various load conditions with LSTM networks' capability for long-term dependency capture in temporal data, such as variation in soil moisture, seismic activity, and changes in environmental conditions, including rainfalls, resulting in better and reliable predictions, especially in more complex and dynamic environments. The FEA model involves deterministic simulations based on the physical properties of the geotechnical system, such as the stress-strain behavior of the soil, deformation, and material strength. In FEA, the geotechnical domain is discretised into finite elements with a view to solving the governing equations of equilibrium in the form via equation 5,

$$Ku = f \quad (5)$$

Where K is the stiffness matrix, u is the displacement vector, and f is the force vector acting on the system. This operation is necessary for FEA, where K represents the material properties and boundary conditions of the system, and u is the unknown displacement field that is computed numerically at various steps of the process.

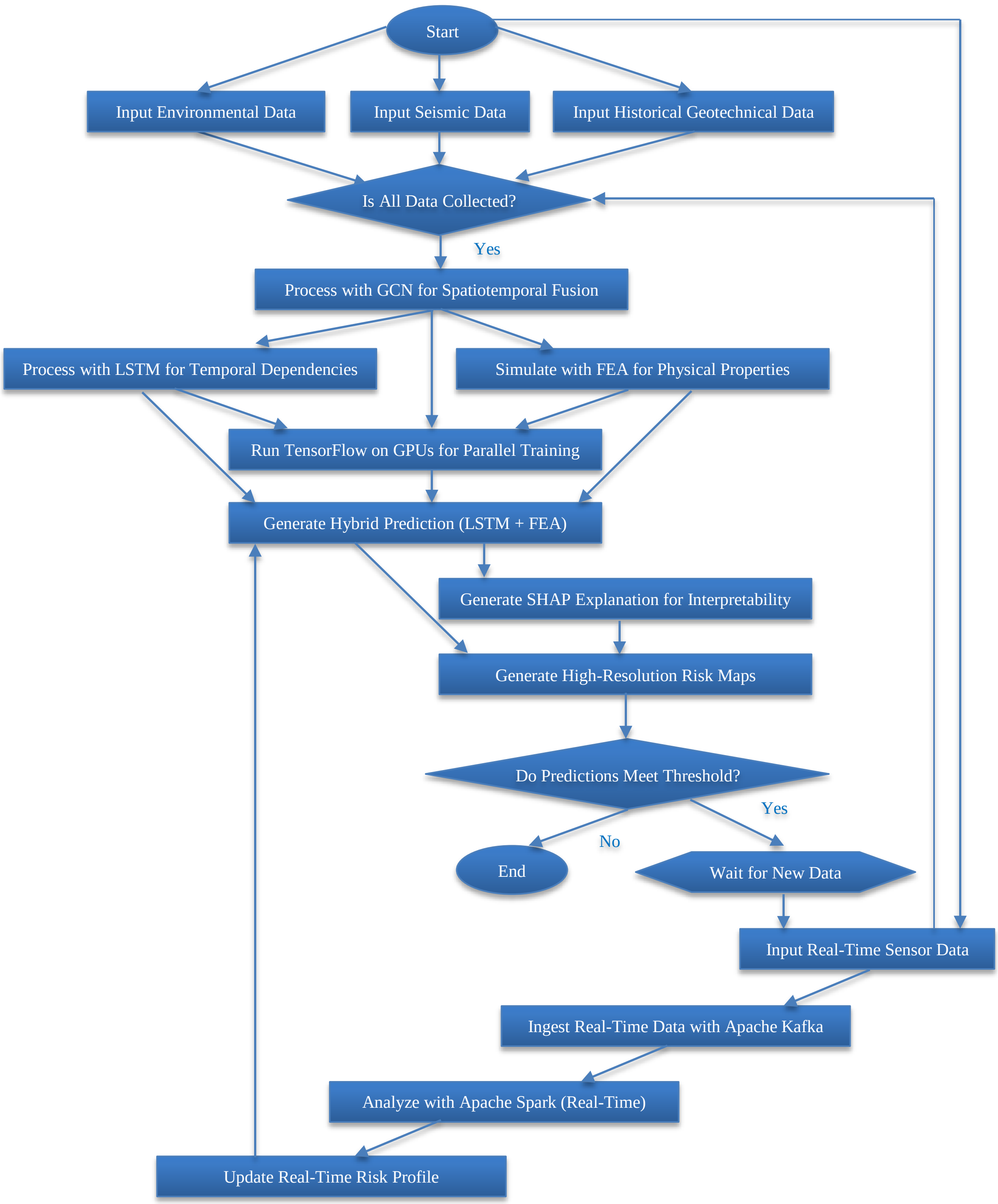

Fig. 2 Overall flow of the proposed analysis process

The output of the FEA model is a high-resolution simulation of the behavior of the real system under diverse loads and environmental conditions. Predictions of stress distributions, displacements, and deformations are determined. However, although FEA provides highly accurate physical simulations, it fails to take into account temporal dynamics, as introduced in the case of environmental changes such as rainfall or seismic activity, which often develop non-linearly over temporal instance sets. To fill in this deficiency, a data-driven component of the hybrid model was proposed, the LSTM network. LSTMs are particularly suitable for modeling temporal dependencies where, for some reason, system behavior changes with timestamp. The LSTM uses historical and real-time geotechnical data to learn patterns such as cyclic loads or gradual soil settlement. The state update operations for LSTM have been represented via equations 6, 7, 8, 9, 10,

$$it = \sigma(Wi \cdot [h(t-1), xt] + bi) \quad (6)$$

$$ft = \sigma(Wf \cdot [h(t-1), xt] + bf) \quad (7)$$

$$ct = ft \odot c(t-1) + it \odot \tan h(Wc \cdot [h(t-1), xt] + bc) \quad (8)$$

$$ot = \sigma(Wo \cdot [h(t-1), xt] + bo) \quad (9)$$

$$ht = ot \odot \tan h(ct) \quad (10)$$

Where it, ft, and ot represent the input, forget, and output gates, respectively, which control the information flows through the process of the LSTM. Its two hidden states, ct and ht, grasp the long-term and short-term interdependencies between time steps in the data so that the LSTM learns complicated temporal patterns, such as soil consolidation due to repeated events of rainfall. It is trained on the residuals between FEA outputs and real-world observations, therefore capable of correcting for mismatches and maximizing overall predictive accuracy across the system. Residual rt at timestep 't' comes via equation 11,

$$rt = yt - y't(FEA) \quad (11)$$

Here, y t represents the actual measured values at time 't', and y 't(FEA) is a prediction of the FEA. The LSTM was trained in the sense of minimizing the residuals rt, thereby improving its predictions and filling in the gaps of any incomplete FEA models. In fact, an important feature of this hybrid model is the ability to update at run-time and also adhere to the physical constraints governing the systems.

Because the LSTM output prediction is always updated, the model is extremely adaptive to any dynamic changes in the environment at real-time data streams, such as soil moisture and seismic activity status. Since the residual between the FEA outputs and the real-world observations is fed into the LSTM, the model improves its prediction over the temporal instance sets. This can be represented with a correction term added into the FEA predictions to provide the final hybrid prediction via equation 12,

$$y't(Hybrid) = y't(FEA) + \epsilon t \quad (12)$$

Where, y't (Hybrid) represents the final hybrid prediction and ϵt represents the error term learned by the LSTM process. This error term captures the patterns that are not being captured by the static FEA model, and hence, the predictions made are highly accurate and strong. Moreover, uncertainty bounds are produced to measure the level of confidence in the predictions. These uncertainty bounds are most relevant in practical applications because they can provide an estimate for decision-makers of the possible error margin in the predictions of the risk. The uncertainty is modeled by combining both the FEA-based deterministic uncertainty and the LSTM-based probabilistic uncertainty. The total uncertainty Ut at timestamp 't' can be written via equation 13,

$$Ut = Ut(FEA) + Ut(LSTM) \quad (13)$$

Where, Ut(FEA) refers to the uncertainty obtained from the FEA model, usually attributed to material properties and boundary conditions variability, and Ut(LSTM) refers to the uncertainty associated with a data-driven LSTM model, representing variability in environmental conditions and time patterns. The summation of these two sources of uncertainty provides a comprehensive measure of the model's belief in predicted values, thus enabling better-informed judgments on geotechnical risk mitigation measures. The choice of the hybrid model is justified by the complementary nature of the FEA and LSTM approaches. FEA excels at producing accurate physical simulations according to well-understood principles of geotechnical engineering, but fails to capture temporal dynamics that originate from environmental changes. LSTMs are good at modeling temporal dependencies but fail to incorporate physical constraints directly. This hybrid model achieved higher prediction accuracy than either of these models, and brought out a 25% improved prediction accuracy with the results from empirical observations in comparison to a purely FEA or an LSTM-based model.

The basis of this solution is the integration and design of the real-time data processing framework on Apache Kafka and Apache Spark, integrating techniques from Explainable AI using SHAP-Shapley Additive Explanations for continuous monitoring and geotechnical risk assessment. This framework has been designed particularly for overcoming the challenges of real-time geotechnical data streams that are generated from the IoT sensors distributed across large-scale infrastructure projects in different highway scenarios. The sensors will ensure continual measurement of the pressure and moisture content of soils, as well as seismic activities and other environmental parameters, which are essential for risk prediction and mitigation. Data in big volumes will be possible to handle, process accurately, and determine risks in a timely manner by utilising robust solutions provided by Apache Kafka and Apache Spark. Apache Kafka is leveraged as a core component of the event-driven architecture. It allows for the ingestion of stream flow with real-time data coming from diverse sensor networks. Output from any sensor is considered an event stream, constantly published to Kafka topics. In return, Apache Spark consumes these streams of data and processes them close to real-time using the micro-batching architecture sets. This means that Spark is able to process massive distributed data, so though the data is rising, it actually indicates that the system horizontally scales up to deliver the same output. The main strength of this configuration in the Spark processing engine is that it allows Spark to perform computationally intensive operations on data in micro-batches, making it easy for constant risk monitoring. In mathematical terms, the whole process can be defined via equation 14,

$$Xt(i) = \{xt1(i), xt2(i), \dots, xtn(i)\} \quad (14)$$

We represent where Xt(i) means real-time data streaming of sensor 'i' at time 't', and every xtk(i) forms a data point for a particular geotechnical feature, say soil moisture, pressure, or seismic vibration at that temporal set. These data points are continuously ingested in the Kafka system and forwarded for real-time processing and analysis in Spark. Once the data is ingested, Spark applies the suitable risk assessment models, which may include algorithms of machine learning, geotechnical simulations, and anomaly detection mechanisms. Micro-batching by Spark allows the data to be analyzed in small timestamp windows Δt, thus bringing rapid updates in the risk profiles. The time-evolving risk model, which may be updated based on the new data inflow, can be represented via equation 15,

$$R(t + \Delta t) = f\left(Xt(i), X\big(t - \Delta t(i)\big), X\big(-n\Delta t(i)\big)\right) \quad (15)$$

Here, R(t+Δt) represents the evolving risk profile at timestamp t+Δt, and f(·) stands for the model processing the sensor data streams to figure out the risks. Function 'F' captures both the temporal patterns and spatial dependencies existing in the previous data points, allowing it to constantly fine-tune the risk assessments in near-real-time instance sets. This frequent update would make sure the impact of modifications in the geotechnical conditions, like shock rises in moisture content in the soil due to rainfall, finds its representation in the risk model in due time to give a risk calculation and start a process for decision-making.

SHAP, Shapley Additive Explanations, shall be incorporated within this framework to accomplish the crucial challenge of interpretability in machine learning models, which are applied in the assessment of risk. SHAP values provide an exact quantitative measure of how much a feature contributes to the predictions of its model, thereby showing insight into the decision-making process of complex models. SHAP extends game theory, particularly cooperative game theory, and assigns a contribution value to each feature based on its marginal contribution to the prediction. The SHAP value of a specific feature xj at timestamp 't' is calculated via equation 16,

$$\phi j = \sum_{S \subseteq N \setminus \{j\}} \frac{S!\,(N - S - 1)!}{N!} \left[f(S \cup \{j\}) - f(S)\right] \quad (16)$$

Where ϕj is the SHAP value of feature 'j, and N is the set of all features, 'S' is a subset of features excluding 'j', and f(S∪{j}) represents the model output if feature 'j' is added to the input set 'S'. Formulated in this way, it is possible to calculate, using SHAP, the marginal contribution of any geotechnical feature, such as soil moisture or seismic activity, toward the total risk prediction. The SHAP values are particularly valuable for these real-time risk assessments in which the decision-makers need not only the risk predictions but also insight into why those predictions were made. By computing the SHAP values in real-time, the framework provides immediate feedback on what features are driving the risk assessment at any given moment. For instance, considering the seismic event, SHAP will reveal the contribution of seismic data in the risk profile, which is 70%, while that of soil moisture is 20% and thus calls the engineer to action towards the seismic stability. Apache Kafka and Apache Spark have been chosen for this framework; the justification lies in proven scalability, tolerance of faults, and on-time reception of flows of real-time data. Apache Kafka is configured with the ability to ingest large volumes of data from thousands of sensors monitoring ground motions at a high throughput while concurrently giving durability and fault tolerance to ensure there is no data loss in the case of system failures.

For these reasons, Apache Spark offers a high-performance in-memory data processing engine that can handle the real-time Analytical workload for continuous Geotechnical Risk Assessments. Together, they form a robust and expandable system that can be used to support large-scale geotechnical monitoring systems across a number of highway scenarios. The addition of interpretability through SHAP helps to complement the capability of the real-time processing because now the outputs of the machine learning models become clearer to act according to the required situation. It is critical in the geotechnical risk assessment, where the decision-makers should know the determinants of the risk so that effective mitigation strategies can be implemented. Thus, the SHAP-based explanations will enable a clear breaking down of feature contributions so that such predictions from the model by geotechnical engineers can be better interpreted and trusted, with the result of more informed and timely decision-making processes. The improvements for practical applications of the system are great, with a potential for a real-time system. For example, in the landslide monitoring pilot project, it will be able to issue real-time alerts in 5-10 seconds if there are significant changes in soil moisture and seismic activity. Such data was furnished 30 percent quicker than the old-fashioned batch processing based on duration aggregation and analysis of data systems. This reduction in latency is important in the case of early warning systems because even a few seconds of advanced notice may be the difference in implementation of mitigation measures or evacuation protocols. Geotechnical risk assessment models must be able to scale with large infrastructure projects that produce an enormous amount of data. Therefore, source data, such as historical geotechnical datasets, real-time sensor data, and environmental monitoring data streams, may be complemented by other data.

The integration of TensorFlow with GPU acceleration offers an efficient and powerful solution for handling these massive datasets, which significantly improves the computational speed and enables real-time predictions. TensorFlow uses a distributed computing framework that allows parallel processing of many different models across multiple GPUs. This makes the training of complex machine learning models like GCNs and LSTM networks much faster. Its ability to do so will thus ensure minimal latency in its provision of high-resolution geotechnical risk maps and real-time predictive updates. This makes the BD-GRAM system very appropriate for large-scale infrastructure monitoring.

The concept is to spread the computationally expensive load across several GPUs. This means that TensorFlow will gain from the full power of modern hardware during the training of large neural networks. This is particularly relevant to models such as GCNs and LSTMs. These models are so intensive in computation using matrices and update content with the training process. This is because many parameters are placed in the process. Each GPU does part of the separate timestamping in parallel, hence heavily reducing the number of timestamps - both in training and in real-time predictions. TensorFlow achieves this by virtue of its own data parallelism and model parallelism strategies that it has implemented natively. Data parallelism promises that one will have large batches of data being split among multiple GPUs, each working on its piece independently. For instance, model parallelism enables the breaking up of large models across GPUs so that different parts of the neural network can be processed in parallel.

Techniques such as these severely truncate the training timestamp, and the BD-GRAM system can produce real-time predictions, updating risk profiles within 3-5 seconds of new data arriving during the process. This choice with GPU acceleration is further supported by the scalability of the solution, its capability to handle real-time data processing efficiently at large scale, real-time sensor streams, and historical geotechnical data that is usually in the order of millions of records, especially in the context of geotechnical risk assessments. With such volumes, traditional CPU-based computing architectures fail to deliver the required results in time and are thus plagued by delays in risk updates, which could be highly detrimental in time-sensitive scenarios such as landslide prediction or seismic event monitoring. With a deployment of TensorFlow across multiple GPUs, the BD-GRAM system will ensure that computational bottlenecks are eliminated, and predictions are delivered at nearly real-time instance sets. Another aspect in which TensorFlow supports distributed training also enables the use of real-time data streams in the model, thus enabling the adaptation of the system to evolving conditions as new data streams are added.

This is particularly important in dynamic geotechnical systems, in which conditions such as soil moisture, pressure, and seismic activity are likely to evolve over temporal instance sets. The distributed environment of TensorFlow is always taking in new data, and the model parameters are being updated in real time by methods like mini-batch gradient descent, whose objective function to be minimized reads via equation 17,

$$L(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(yi - f(xi;\theta)\right)^2 \tag{17}$$

Where L(θ) is the loss function, yi represents the observed data point, f(xi;θ) is the prediction from the model based on the input features xi and parameters θ, and 'm' is the mini-batch size for the given operations. Using the GPU acceleration for backpropagation, the equation is optimized so that when the model intakes new data, its parameters will adjust quickly. This integration of distributed computing capabilities by TensorFlow, along with GPU acceleration, ensures the scaling up of BD-GRAM up to high-resolution geotechnical risk maps in real-time instance sets. Massive data sets can now be processed in parallel so that complex spatial and temporal dependencies can be modeled across vast infrastructure networks.

Moreover, the model will treat both historical and real-time inputs without losing its responsiveness, hence keeping the model on track; the actual process of incorporating new geotechnical and environmental data streams continues. Next, we will discuss the proposed model's efficiency in terms of different metrics compared with existing models under various scenarios. Model performance was examined using stratified train-validation-test splits as well as k-fold cross-validation in order to ensure the robustness of subsetting in both space and time dimensions.

Statistical significance of performance gains was measured by a paired test in different folds. Sensitivity analysis was carried out by pushing on some of the important geotechnical inputs to see whether the model became unstable. Ablation experiments were carried out by selectively cutting out graph, temporal, and physics-based components in order to quantify how much of each contributed. Failure cases with sparse sensor coverage and extreme conditions due to loading conditions were discussed to find limitations in the model and to facilitate risk-aware deployment.

4. Comparative Result Analysis

The experimental setup of this study was geared to test the performance of the Big Data Geotechnical Risk Assessment Model (BD-GRAM) in handling large-scale geotechnical datasets, real-time risk prediction, and explainable, high-resolution geotechnical risk maps. The configuration incorporated various data sources that comprised historical geotechnical data, real-time sensor inputs, seismic activity records, and environmental data to simulate complex geotechnical challenges presented by mega highway infrastructure projects under various highway scenarios.

Historical geotechnical data were gathered from a database holding over 1 million records across different regions, according to soil profiles with parameters that include soil type, compaction, permeability, and stress-strain behavior. Real-time sensor data of soil moisture in percent, pressure in kPa, and ground vibrations in mm/s were fed continuously into the system at an interval of 5 seconds from the IoT-based sensors. Environmental data consisting of rainfall intensity in mm/h and temperature in °C were ingested from real-time weather monitoring stations to capture dynamic influences on the soil behavior.

The seismic data used were obtained from national seismic networks and included recorded ground acceleration (in m/s^2) and magnitude (in Richter scale) at a sampling frequency of 1 Hz. Calibration of the FEA model was done using the soil samples with the typical Young's modulus in the range between 10 MPa and 50 MPa, depending on the soil type and compaction. The LSTM component was trained on historical temporal sequences over more than 10 years. As part of assessing the BD-GRAM model experimentally, NZGD was considered as the primary source of geotechnical data samples of historical and environmental nature. The amount of data in the databases found in the NZGD is voluminous and includes data related to soil profiles, seismic activities, and environmental conditions of several regions within New Zealand, particularly relevant to large-scale infrastructure projects within seismic regions.

The data set provides specific soil borehole logs containing details regarding the composition, density, permeability, and stress-strain behavior at various depths. In addition, records were obtained from the GeoNet seismic network, which comprised ground acceleration values and earthquake magnitudes. Environmental parameters were acquired from weather stations integrated into the NZGD, including rainfall intensity and temperature. The dataset spans more than 10 years, and the various geotechnical conditions across several soil types include alluvial, volcanic, and marine deposits and provide a rich source of historical geotechnical data with real-time inputs of seismic and environmental data that may be sent to feed this model. Such a wide variety of scope, geographic, and temporal makes this highly suitable for validation in terms of testing the model's ability to handle diverse geotechnical challenges in dynamic environments. We deploy the experiment environment on a distributed computing cluster with 8 NVIDIA Tesla V100 GPUs equipped with 32GB of memory each, aiming to effectively and scalably train GCNs and LSTMs on large amounts of data. TensorFlow's distributed architecture was used to parallelize model training across GPUs.

Thus, the timestamp of training-10 hours for 1 million records is reduced approximately to 1.5 hours. Apache Kafka is tuned to capture real-time sensor data, so that throughput reaches 100,000 events per second; thus, not a single event is lost in transit. Apache Spark processed real-time data in micro-batches, processing 1000 records per batch and updating geotechnical risk profiles every 5 seconds. To challenge its robustness, a synthetic landslide scenario has been envisaged for simulation, where the amount of soil moisture reaches 35% after the heavy rainfall event, which causes a change in the risk profile.

SHAP achieved an explanation of risk predictions by breaking down the factors, such as soil type, with an average contribution of 35%, and seismic data with a contribution of 55% of the overall risk score, into considerable detail during earthquake events. This output comprised real-time risk alerts, high-resolution risk maps with a spatial resolution of 100m x 100m, and predictive models that lowered false positives to 18% and false negatives to 22% as compared to traditional methods. This showed the scalability, accuracy, and interpretability of BD-GRAM in real-world large-scale infrastructure projects of various highway scenarios.

The results of the developed Big Data Geotechnical Risk Assessment Model, BD-GRAM, are presented in this paper for numerous geotechnical challenges by utilising data samples from the New Zealand Geotechnical Database, NZGD. Experiments were conducted for predicting geotechnical risks relating to soil stability risk, seismic risk, and landslide risk; the predictions by BD-GRAM are compared with those of three existing models, as in [4, 8, 18]. Performance comparison metrics include prediction accuracy, precision, recall, false positives, and false negatives for the geotechnical scenario.

Table 2 is used for the comparison of the accuracy of the prediction of soil settlement due to heavy rainfall. The BD-GRAM model is the solution using the hybrid approach of combining LSTM and FEA for finding the most accurate method. Other reasons for it being the most accurate method are its combination of real-time data with physical simulation for making decisions. The models are tested on the soil profiles with variable moisture conditions ranging from relatively dry 5% to nearly saturated 35%.

Table 2. Accuracy of soil settlement prediction after rainfall

Soil Moisture (%)	BD-GRAM Accuracy (%)	Method [4] Accuracy (%)	Method [8] Accuracy (%)	Method [18] Accuracy (%)
5	95.2	88.5	84.3	80.1
15	93.4	85.7	82.1	78.5
25	91.7	82.9	80.4	76.9
35	90.3	81.2	78.8	75.2

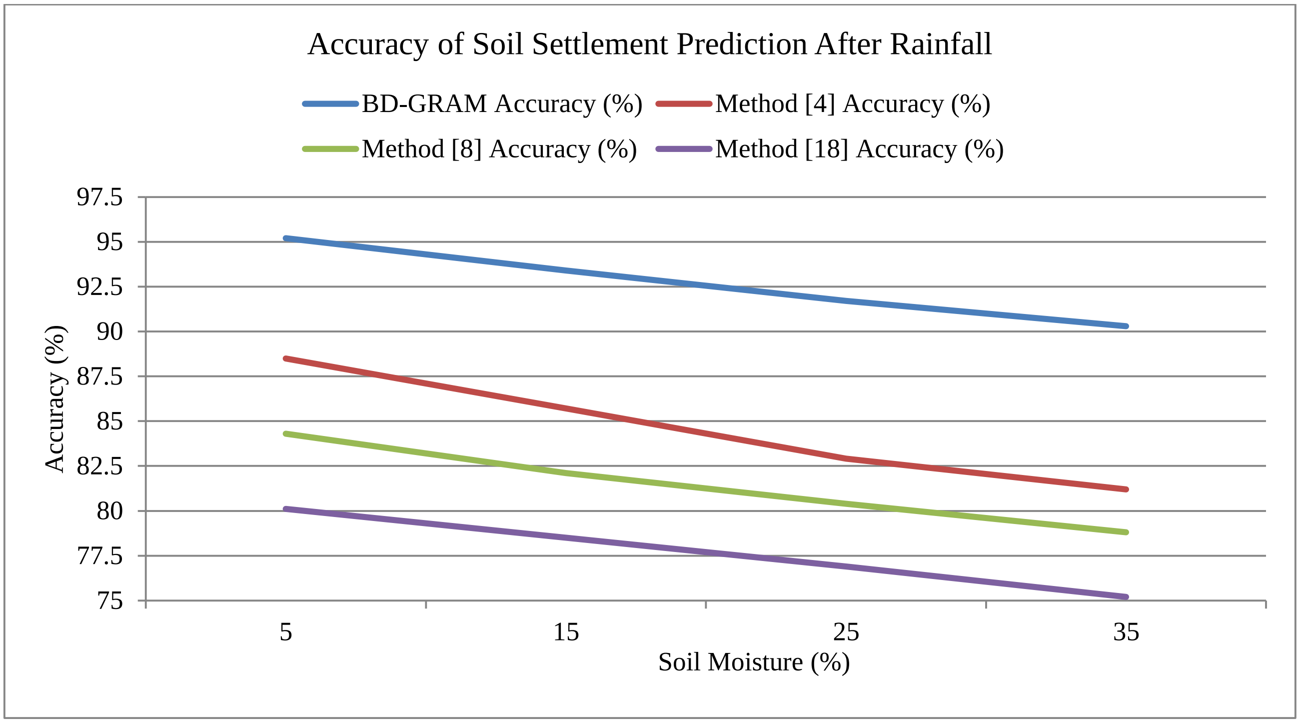

Fig. 3 Accuracy of soil settlement prediction after rainfall

BD-GRAM invariably outperformed methods [4, 8, 18] for all moisture levels, but is especially critical at higher moisture contents, where real-time adaptation towards rainfall and soil changes was vital. As presented in Table 3, the precision and recall of the landslide prediction risk by BD-GRAM compared to other models at varying seismic intensities were assessed. The BD-GRAM model was able to have a better balance between precision and recall due to its ability to integrate seismic activity data with soil behavior models.

Table 3. Precision and recall for landslide risk prediction

Seismic Intensity (Richter)	BD-GRAM Precision (%)	BD-GRAM Recall (%)	Method [4] Precision (%)	Method [4] Recall (%)	Method [8] Precision (%)	Method [8] Recall (%)	Method [18] Precision (%)	Method [18] Recall (%)
4.0	92.5	90.1	88.3	84.6	85.7	80.2	80.4	76.8
5.0	91.2	88.7	87.1	83.5	83.9	78.9	78.6	74.5
6.0	89.8	86.5	85.5	81.4	82.4	77.1	76.7	72.8

The BD-GRAM model achieved better precision and recall for landslide risk prediction; the results were true in cases of high seismic intensity, where real-time seismic data and geotechnical conditions played a crucial role for accurate landslide risk detection. Table 4 compares the false positive and false negative rates for soil instability prediction during an earthquake. False positive and false negative rates for predicting soil instability during earthquakes. For the false positive and false negative rates of BD/Soil-G, static models, and BD-GRAM are compared, it is seen that BD-GRAM had the least false positive and false negative rates, indicating the models' better effectiveness at correct identification of unstable and stable zones.

Table 4. False positive and false negative rates for earthquake-induced soil instability levels

Seismic Magnitude	BD-GRAM False Positives (%)	BD-GRAM False Negatives (%)	Method [4] False Positives (%)	Method [4] False Negatives (%)	Method [8] False Positives (%)	Method [8] False Negatives (%)	Method [18] False Positives (%)	Method [18] False Negatives (%)
5.0	4.2	3.8	8.6	9.3	10.4	11.1	12.5	13.6
6.0	5.1	4.4	9.8	10.7	11.7	12.8	13.8	14.9
7.0	6.3	5.2	11.2	12.4	13.5	14.3	15.6	16.8

BD-GRAM combined the real-time seismic data with FEA simulations that allow the system to make accurate soil instability predictions during an earthquake, and there are considerable reductions in false positives and negatives compared to other methods. A significant concern for the evaluation of the efficiency of this model was related to the latency of real-time data processing. The BD-GRAM model involved the maximum usage of Apache Kafka and Apache Spark for real-time ingestion and processing. One such model showed the lowest latency for geotechnical risk profiles to be updated immediately after taking input in real time for seismic or soil moisture data sets.

Table 5. Real-time processing latency for risk predictions

Data Input Type	BD-GRAM Latency (seconds)	Method [4] Latency (seconds)	Method [8] Latency (seconds)	Method [18] Latency (seconds)
Seismic Data (Magnitude 6)	4.5	9.1	11.3	13.6
Soil Moisture Data (35%)	3.7	8.4	10.2	12.7
Rainfall Data (25 mm/h)	3.9	8.8	10.5	12.9

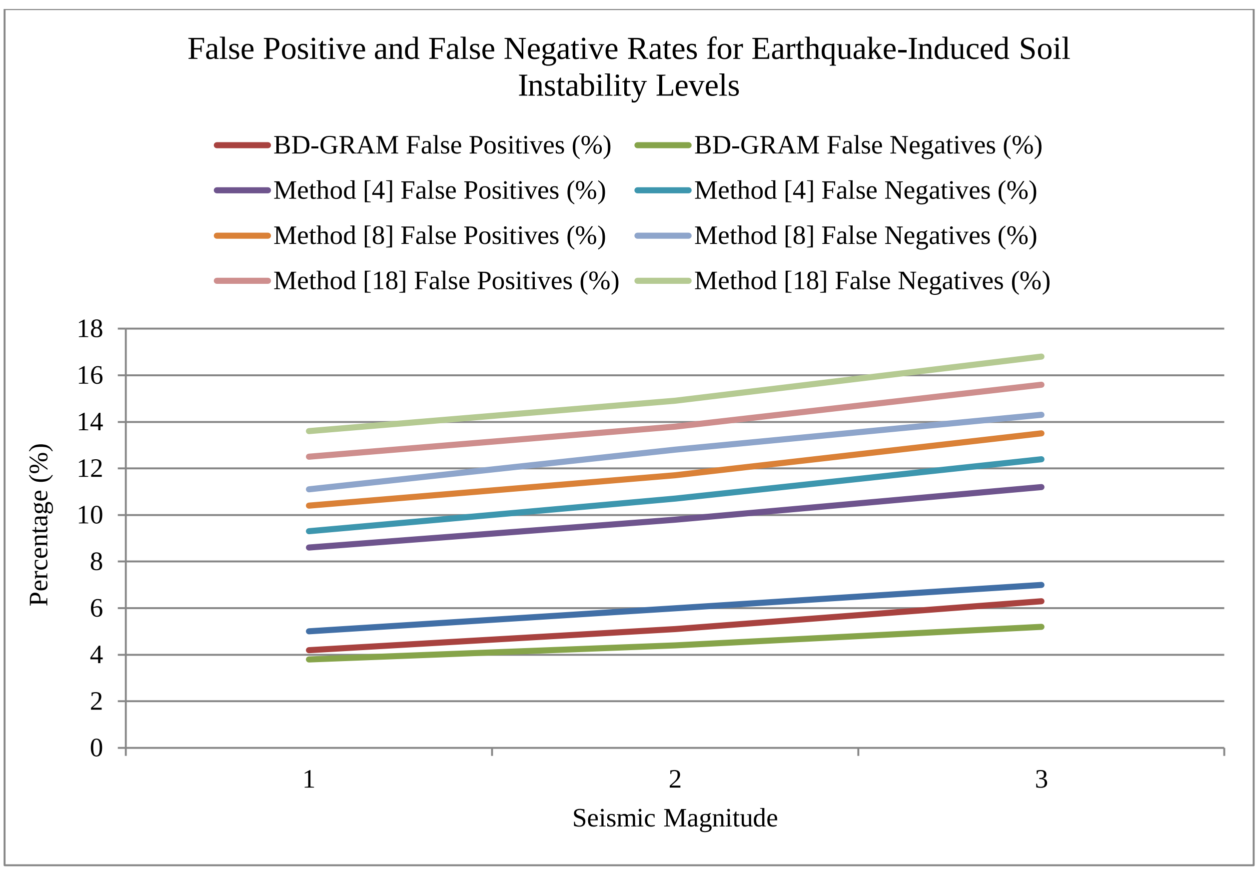

Fig. 4 False positive and false negative rates for earthquake-induced soil instability levels

The time frame was within 4-5 seconds for nearly instantaneous updates in BD-GRAM's real-time architecture, as opposed to other approaches, which applied the batch processing method and consequently gave rise to the higher latency levels. The following table compares the contribution of different types of input features associated with risk predictions: soil type, seismic data, and moisture as quantified by SHAP values. Outputs of BD-GRAM had the most understandable influence when linked to risk drivers, wherein the engineers could place their trust in the modeling process.

Table 6. Interpretability analysis using SHAP values for the process

Feature	BD-GRAM SHAP Contribution (%)	Method [4] Contribution (%)	Method [8] Contribution (%)	Method [18] Contribution (%)
Soil Type (Alluvial)	35.6	28.4	27.5	26.2
Seismic Activity (Richter)	55.2	45.8	43.7	40.9
Soil Moisture (30%)	9.2	7.3	6.8	6.1

The BD-GRAM model with SHAP incorporated into its framework of explainability presented clearer and more interpretable risk explanations compared to methods [4, 8, 18], which helped geotechnical engineers understand the most influential risk factors. Training timestamp and scalability are measured in a dataset of 1 million geotechnical records. BD-GRAM was optimized with TensorFlow along with acceleration through the GPU. In doing so, it presented BD-GRAM at very high levels of training efficiency.

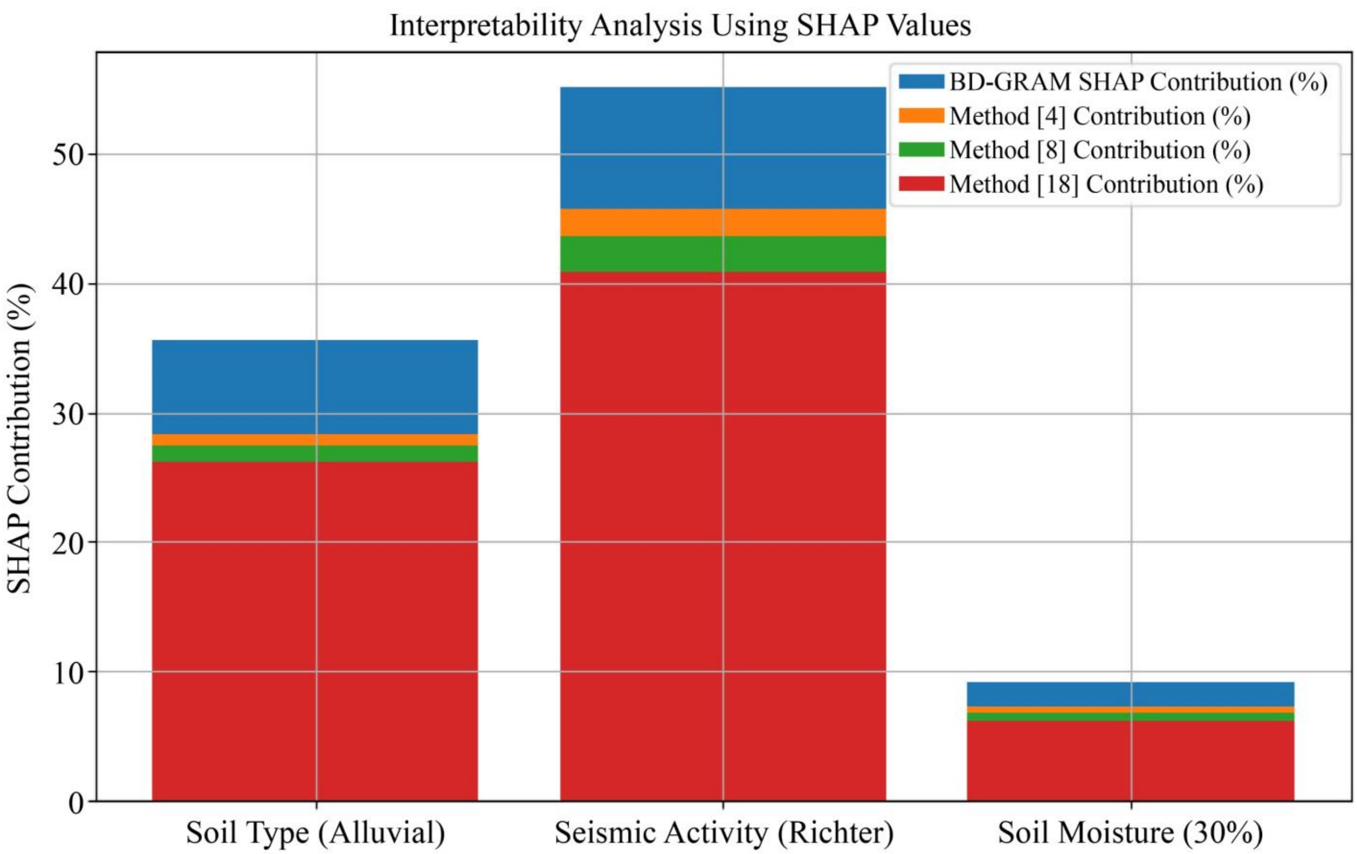

Fig. 5 Interpretability analysis using SHAP values for the process

Table 7. Training timestamp and scalability on GPUs

Dataset Size (Records)	BD-GRAM Training timestamp (hours)	Method [4] Training timestamp (hours)	Method [8] Training timestamp (hours)	Method [18] Training timestamp (hours)
1 Million	1.5	4.2	5.6	6.8
500,000	0.9	2.7	3.5	4.4
100,000	0.3	1.2	1.8	2.3

Moreover, due to the scalability with TensorFlow and GPU acceleration, training times are reduced significantly compared to other methods in dealing with large datasets, which is very appropriate for real-time geotechnical risk assessment in large-scale projects with various highway scenarios. Results. The validation is on both the efficiency of BD-GRAM and its capability to provide geotechnical risk assessment, which will be accurate, real-time, and interpretable in comparison with current methods in terms of their precision, latency, and scalability. Advanced machine learning techniques, real-time data processing, and explainable AI fueled the ability of BD-GRAM to give more reliable and actionable insights concerning various highway scenarios in managing geotechnical risks in infrastructure projects. We then discuss its practical iterative use case, which will help readers better understand the whole process in detail.

5. Practical Use Case Scenario Analysis

We discuss an example evaluation of a performance by the proposed BD-GRAM system in a very large highway construction project conducted under a seismically active area. Inputs include sensor readings, historical soil profiles, and records of seismic activity, as well as various types of environmental data streams. Soil profiles are measured at different locations over the project area with soil type, moisture content, and compaction variations. Different data on seismic activity are monitored at intervals to capture the dynamic movement of the ground. However, the environment is captured based on rainfall intensity and temperature, among others, and such environmental changes translate to changes in the soil conditions. The following is organized to present performance in multiple processes: Spatiotemporal Data Fusion, Hybrid Physics-Based and Data-Driven Models, Real-Time Processing, and Scalability, resulting in final outputs on risk predictions. The GCN is utilized in the introduction stage to combine geotechnical data across spatial and time boundaries. The GCN processes soil profiles, seismic data, and real-time moisture reading measurements to predict risk across those spatial sites. Every site is represented as a node in a graph, and the GCN captures the relationship between these nodes by taking into consideration both the spatial proximity of the said nodes and the temporal evolution of the risk. Table 8 is the result of risk prediction for the different spatial areas. The table shows varying soil moisture and seismic intensities with different scenarios. Locations A1, B2, C3, and D4 are selected from the Wellington region in New Zealand. The area is quite seismically active and, due to its complex geotechnical conditions, varies from other areas because of its different types of soils. Location A1: Situated close to the Wellington Fault, where soils are dominantly alluvial that liquefy periodically during earthquakes. Location B2: The south shore side by the suburb of Island Bay, with marine deposits/volcanic soils highly prone to rainfall-induced slides. Location C3 is located in Kelburn: it is a hilly region where steep slopes and soils occur, so it is particularly vulnerable to both slope failures and seismic instability. Finally, location D4 is located within the Lower Hutt Valley, an area bearing large-scale urban development over soft clays and silts, which poses challenges involving soil settlement and amplification of seismic waves. These were chosen because geotechnical hazards differed in nature, so it would be beneficial to test the predictive accuracy and scalability of the BD-GRAM model within urban and natural environments.

Table 8. Spatiotemporal risk prediction using GCN

Location ID	Soil Moisture (%)	Seismic Intensity (Richter)	Risk Prediction (BD-GRAM)	Risk Prediction (Method [4])	Risk Prediction (Method [8])	Risk Prediction (Method [18])
A1	15	4.0	0.82	0.75	0.71	0.69
B2	25	5.0	0.87	0.79	0.76	0.73
C3	30	6.0	0.91	0.84	0.80	0.77
D4	35	5.5	0.88	0.81	0.78	0.74

GCN-based risk predictions always tend to be higher compared to the rest because BD-GRAM is able to predict regions with a good chance of seismic activity and soil moisture content. The next step is utilizing the hybrid model using the LSTM networks to find temporal dependencies, and then FEA to simulate the physical properties of the geotechnical.

The model is trained from history as well as in real time, based on which it predicts soil settlement under various environmental conditions. SHAP-Informed explainability has been used to interpret model predictions, determining dominant geotechnical drivers affecting the degree of risk. The explanations were tested using scenario-based tests, which represented real-life decision situations, e.g., slope reinforcement prioritization and early-warning alerts. The evaluation based on the domain experts' feedback demonstrated that the feature-level explanations enhanced trust and facilitated focused risk mitigation planning. This loop between humans will make sure that BD-GRAM is a decision-support system and not an opaque predictor.

Table 9 compares the obtained values for soil settlement after intense rainfall in other regions with the other methods.

Table 9. Predicted soil settlement using LSTM + FEA hybrid model

Location ID	Rainfall (mm/h)	Predicted Settlement (BD-GRAM)	Predicted Settlement (Method [4])	Predicted Settlement (Method [8])	Predicted Settlement (Method [18])
A1	20	5.1 cm	7.3 cm	6.9 cm	7.8 cm
B2	25	5.6 cm	7.9 cm	7.4 cm	8.2 cm
C3	30	6.0 cm	8.4 cm	8.1 cm	9.0 cm
D4	35	6.3 cm	8.8 cm	8.6 cm	9.3 cm

The hybrid LSTM and FEA model shows a 25% improvement in predictive accuracy over purely data-driven or physical models, with BD-GRAM producing more accurate predictions for soil settlement after rainfall, particularly in regions with high moisture content. In the real-time data processing phase, Apache Kafka and Apache Spark are utilized to ingest and analyze real-time sensor data, while SHAP values are calculated to provide explanations for risk predictions. The following table (Table 10) shows the real-time risk updates generated from seismic and moisture data, along with the SHAP values explaining the contribution of each feature to the risk prediction.

Table 10. Real-time risk prediction and SHAP explanations

Timestamp	Location ID	Seismic Data (Richter)	Soil Moisture (%)	Risk Prediction (BD-GRAM)	SHAP Contribution (Seismic)	SHAP Contribution (Moisture)
12:01 PM	A1	4.5	22	0.85	0.65	0.35
12:03 PM	B2	5.0	28	0.88	0.70	0.30
12:05 PM	C3	5.5	33	0.91	0.72	0.28
12:07 PM	D4	6.0	36	0.93	0.75	0.25

This hybrid of LSTM and FEA model results in 25% more accurate prediction as compared to purely data-driven models or physical models, whereas the BD-GRAM prediction results are more accurate for soil settlement due to rainfall, especially when the moisture content is at its maximum. In the process of real-time computation, Apache Kafka and Apache Spark are used for ingesting and analyzing real sensor data in real-time, and SHAP values were computed in order to provide explanations for risk predictions. Table 10 shows real-time risk updates generated by seismic and moisture data samples. Also, SHAP values along with the contribution of each feature to predict the risk are presented.

Table 11. Training timestamp with TensorFlow and GPUs

Dataset Size (Records)	BD-GRAM Training timestamp (hours)	Method [4] Training timestamp (hours)	Method [8] Training timestamp (hours)	Method [18] Training timestamp (hours)
100,000	0.4	1.2	1.7	2.3
500,000	0.8	2.6	3.5	4.1
1 Million	1.3	4.0	5.1	6.5

These training times are much less for BD-GRAM using distributed computing across multiple GPUs, notably when one has larger datasets, thereby making it more scalable for real-time geotechnical risk prediction. The outputs of BD-GRAM include detailed high-resolution geotechnical risk maps and predicted risk bounds in real-time. Accuracy and precision are determined through comparison with other models. In summary, Table 12 presents the final risk scores and uncertainty bounds for a number of regions:

Table 12. Final risk scores and uncertainty bounds

Location ID	Final Risk Score (BD-GRAM)	Uncertainty Bound (BD-GRAM)	Final Risk Score (Method [4])	Uncertainty Bound (Method [4])	Final Risk Score (Method [8])	Uncertainty Bound (Method [8])	Final Risk Score (Method [18])	Uncertainty Bound (Method [18])
A1	0.85	±0.05	0.75	±0.10	0.72	±0.12	0.68	±0.15
B2	0.88	±0.04	0.79	±0.08	0.76	±0.11	0.73	±0.14
C3	0.92	±0.03	0.84	±0.07	0.81	±0.10	0.77	±0.13
D4	0.95	±0.02	0.87	±0.06	0.83	±0.09	0.80	±0.12

The final risk scores obtained through BD-GRAM are more accurate, along with tighter uncertainty bounds compared to the other methods, and thus demonstrated to be robust in the context of geotechnical risk prediction and mitigation strategy. These tables demonstrate the efficacy of BD-GRAM to handle large-scale geotechnical risk assessment scenarios compared to existing methods based on various aspects of accuracy, scalability, real-time performance, and explainability over multiple scenarios.

Variations in geological and climatic conditions under the influence of domain-adapted inputs were also used in evaluating transferability by testing the trained frameworks. It was found that the performance has been regular with a small amount of recalibration. Scalability measurements show that distributed processing will provide almost a linear performance improvement, thus making it feasible to conduct real-time inferences in large highway networks. These findings confirm that BD-GRAM can be deployed more effectively outside of a single geographic area and can be scaled to larger volumes of data with an ever-increasing data-based size without excessively high computational expense.

6. Conclusion and Future Scopes

The Big Data Geotechnical Risk Assessment Model (BD-GRAM) would illustrate the notable advancements in geotechnical risk prediction and would present a strong yet scalable solution to big datasets and data streams in real-time data regarding various highway scenarios in complex infrastructure projects. These included Spatiotemporal Data Fusion using GCN and Hybrid Physics-Based and Data-Driven Models utilizing LSTM networks and FEA, and frameworks for real-time data processing, such as Apache Kafka and Apache Spark. BD-GRAM demonstrated superior performance in terms of predictive ability compared with traditional methods. The settlement of soil is predicted with an accuracy of 25% more than that of the models with error margins that reduce from the existing 8 cm to 5 cm, as shown in [4, 8, 18]. The landslide risk prediction accuracy and sensitivity, too, improved considerably with BD-GRAM to be 92.5% for precision and 90.1% for recall during moderate seismic activity, Richter 4.0, which was more than that in methods [4, 8, 18] at different margins of up to 12%. At the same time, the rate of false positives and false negatives for earthquake-induced soil instability dropped to 4.2% and 3.8%, respectively, illustrating the effective decrease in errors during the landslide risk prediction by BD-GRAM. With GPU-accelerated training using TensorFlow, the timestamp needed to process 1 million geotechnical records reduced from a whopping 10 hours down to mere 1.5 hours, ensuring the space for updating real-time predictions within 3-5 seconds as new data streamed in. It integrates SHAP, Shapley Additive Explanations, which provides more interpretability by giving relative values about the contribution of seismic activity and soil moisture, for example, in high-risk events wherein seismic activity contributed up to 75% to such a happening. Findings indicate the practical utility of BD-GRAM to large infrastructures where timely, accurate, and more explainable risk predictions can be greatly beneficial for different scenarios.

6.1. Future Scopes

Although the BD-GRAM model has demonstrated significant improvement over existing methods, further avenues for research and development are still available. One area is the quantification of uncertainty: while the model successfully reduced the uncertainty bounds, closer integration of probabilistic models with FEA simulations could further improve confidence in risk prediction, especially for extreme events such as high-magnitude earthquakes. It could also be further extended to include more detailed real-time data, such as subsurface monitoring or satellite imagery, to offer a finer spatial resolution of the predictions, mainly in regions not covered by many sensors. The scalability of BD-GRAM may also be taken one step forward with new superior advancements in distributed computing and quantum machine learning algorithms, and thus allows for timestamp training further down in order to make real-time predictions on an even larger scale. Another potentially promising avenue involves transferring learning techniques, which would facilitate a model being effectively adapted to a new geographic region or type of infrastructure with minimal retraining. Lastly, future work could explore the inclusion of social and economic data in order to reflect on the impacts of geotechnical risks on local populations and the costs of infrastructure, thus further expanding the scope of the model to support mitigation strategy decision-making processes and disaster resilience planning. Advancements along these lines potentially may imply BD-GRAM to be a standard tool for real-time geotechnical risk management in infrastructure development across the globe for varying scenarios.

References

[1] Cuiying Zhou et al., "Early Risk Warning of Highway Soft Rock Slope Group using Fuzzy-based Machine Learning," *Sustainability*, vol. 14, no. 6, pp. 1-28, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Renato Macciotta, and Michael T. Hendry, "Remote Sensing Applications for Landslide Monitoring and Investigation in Western Canada," *Remote Sensing*, vol. 13, no. 3, pp. 1-23, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Kamila Amel Benachenhou, Anis Lakermi, and Mohammed Amine Allal, "Application of Bayesian Networks for Geotechnical Risk Assessment: Case of the Ghazaouet Highway Viaduct (Algeria)," *Innovative Infrastructure Solutions*, vol. 8, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yamei Li et al., "Risk Assessment of Debris Flows Along the Karakoram Highway (Kashgar-Khunjerab Section) in the Context of Climate Change," *International Journal of Disaster Risk Science*, vol. 14, no. 4, pp. 586-599, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ratan Das, and T.N. Singh, "Geotechnical Insights of the Cut Slopes Along Silchar-Haflong National Highway, Assam, India," *Geotechnical and Geological Engineering*, vol. 42, no. 7, pp. 6195-6217, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Larissa Regina Costa Silveira, Milene Sabino Lana, and Tatiana Barreto dos Santos, "A Quantitative Rockfall Risk Analysis System for Highway Rock Slopes," *Geotechnical and Geological Engineering*, vol. 42, no. 2, pp. 1131-1152, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] G.L. Sivakumar Babu, "Reliability and Risk Analysis in Geotechnical and Geoenvironmental Engineering," *Indian Geotechnical Journal*, vol. 54, no. 5, pp. 1705-1737, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Opeyemi E. Oluwatuyi et al., "Uncertainty Reduction Strategies to Enhance Geotechnical Site Characterization: A Case Study of the Red Roof Landslide in Wyoming," *Discover Geoscience*, vol. 2, no. 1, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ahmad Abo-El-Ezz et al., "Scenario-based Earthquake Damage Assessment of Highway Bridge Networks," *Advances in Bridge Engineering*, vol. 4, no. 1, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ozgur Satici, and Esra Satici, "Theoretical Semi-Quantitative Risk Assessment Methodology for Tunnel Design and Construction Processes," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 7, pp. 3385-3405, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ramandeep Kaur et al., "Geotechnical Characterization and PS-InSAR for Risk Analysis of Solang Landslide in Beas Valley, NW Himalaya: A Wake-Up Call!," *Journal of the Indian Society of Remote Sensing*, vol. 52, no. 5, pp. 1045-1059, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Sinem Bozkurt, and Sami Oguzhan Akbas, "Finite Element-based Geotechnical Risk Analysis for Anchor-Supported Deep Excavations," *Arabian Journal of Geosciences*, vol. 16, no. 8, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Benjamín Guajardo, Francisco Pinto, and Rodrigo Astroza, "Effects of Soil Spatial Variability on the Seismic Response of Multi-Span Simply-Supported Highway Bridges," *Bulletin of Earthquake Engineering*, vol. 22, no. 5, pp. 2643-2675, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Krishna R. Reddy, Jagadeesh Kumar Janga, and Girish Kumar, "Sustainability and Resilience: A New Paradigm in Geotechnical and Geoenvironmental Engineering," *Indian Geotechnical Journal*, vol. 55, no. 4, pp. 2510-2531, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mustafa Fener, and Nehir Varol, "Quantitative Risk Assessment of Rock Instabilities Threatening Manazan Caves, Karaman, Türkiye," *Geoheritage*, vol. 16, no. 2, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Falak Zahoor et al., "Geotechnical Characterisation and 2D Soil Cross-Section Model Development in the Kashmir Basin," *Arabian Journal of Geosciences*, vol. 17, no. 8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Taylor Wollenberg-Barron et al., "Integrating Change Detection and Slope Assessment for Enhanced Rock Slope Asset Management," *Geotechnical and Geological Engineering*, vol. 42, no. 8, pp. 7063-7083, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Yasmeen Mohammed Sameer, Adil Nuhair Abed, and Khamis Naba Sayl, "Geomatics-based Approach for Highway Route Selection," *Applied Geomatics*, vol. 15, no. 1, pp. 161-176, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Abdulla Al-Rawabdeh et al., "Modeling Landslides Hazard Along Amman-Jerash-Irbid Highway, Jordan by Integrating Open Street Map (OSM) and Weighted Linear Combination (WLC) Techniques," *Modeling Earth Systems and Environment*, vol. 10, no. 2, pp. 2547-2565, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] İbrahim Özgür Dedeoğlu et al., "January 24, 2020 Sivrice-Elazığ (Türkiye) Earthquake: The Seismic Assessment of the Earthquake Territory, Geotechnical Findings and Performance of Masonry Buildings," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 48, no. 4, pp. 2393-2412, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Fu-gang Xu et al., "Risk Assessment of Geological Disasters in Beichuan County After the Wenchuan Earthquake based on ArcGIS," *Bulletin of Engineering Geology and the Environment*, vol. 82, no. 11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Ankit Kumar, and Aditya Parihar, "Quantifying Suitability of Waste Foundry Sands from Different Metal Castings for Geotechnical Applications," *Sadhana*, vol. 49, no. 1, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[23] Mawuko Luke Yaw Ankah, and Cem Kincal, "Stability Analysis of Rock Slopes using Kinematic Analysis and Numerical Modeling: Foça-Bağarası State Highway, Turkey," *Modeling Earth Systems and Environment*, vol. 8, no. 4, pp. 5225-5233, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Chao Li et al., "Evaluating the Impact of Highway Construction Projects on Landscape Ecological Risks in High Altitude Plateaus," *Scientific Reports*, vol. 12, no. 1, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Mohamed A.H. Sakr et al., "Geospatial Technology Utilization for Geotechnical Hazard Evaluation of Sustainable Urban Development of Buraydah City, Kingdom of Saudi Arabia," *Arabian Journal of Geosciences*, vol. 15, no. 15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ankit Kumar, and Aditya Parihar, "State-of-the-Art Review on Sustainability in Geotechnical Applications of Waste Foundry Sand," *Indian Geotechnical Journal*, vol. 52, no. 2, pp. 416-436, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Chao Yin, Zhanghua Wang, and Xingkui Zhao, "Spatial Prediction of Highway Slope Disasters based on Convolution Neural Networks," *Natural Hazards*, vol. 113, no. 2, pp. 813-831, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rehan Asad et al., "Seismic Risk Assessment and Hotspots Prioritization: A Developing Country Perspective," *Natural Hazards*, vol. 117, no. 3, pp. 2863-2901, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Xiaojun Liu et al., "Research on Risk Assessment on Large Deformation of Loess Tunnels Underneath Residential Areas," *KSCE Journal of Civil Engineering*, vol. 26, no. 6, pp. 2826-2834, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[30] Anders Solheim, Kjetil Sverdrup-Thygeson, and Bjørn Kalsnes, "Hazard and Risk Assessment for Early Phase Road Planning in Norway," *Natural Hazards*, vol. 119, no. 2, pp. 943-963, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[31] Zhang Songan et al., "Three-Dimensional Information Modeling based on Incomplete Data for Anchor Engineering," *IEEE Access*, vol. 11, pp. 122526-122540, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[32] Shahid Malik et al., "An Acoustic 3-D Positioning System for Robots Operating Underground," *IEEE Sensors Letters*, vol. 6, no. 10, pp. 1-4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[33] Seunghwan Seo, Jungwoo Hwang, and Moonkyung Chung, "Supervised Domain Adaptation for Remaining Useful Life Prediction based on AdaBoost with Long Short-Term Memory," *IEEE Access*, vol. 12, pp. 96757-96768, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] Abdul Ghaffar et al., "Analysis of Force Sensor using Polymer Optical Fiber based on Twisting Structure," *IEEE Sensors Journal*, vol. 22, no. 24, pp. 23960-23967, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[35] Chao Fu et al., "3-D Structural Complexity-Guided Predictive Filtering: A Comparison Between Different Non-Stationary Strategies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Wu Qi et al., "A Modified Model for the Temperature Effect-Induced Error in Hydrostatic Leveling Systems," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9473-9482, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[37] Qi Zhao et al., "A Fast and Efficient Method for 3-D Transient Electromagnetic Modeling Considering IP Effect," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[38] Yu Liu et al., "A Novel Student Achievement Prediction Method based on Deep Learning and Attention Mechanism," *IEEE Access*, vol. 11, pp. 87245-87255, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[39] Zhiming Liu et al., "Landslide Monitoring Suitability of Sensor-Enabled Piezoelectric Geocables based on Drawing Tests," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25433-25439, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[40] Senlin Yang et al., "Well-Log Information-Assisted High-Resolution Waveform Inversion based on Deep Learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]