Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-1593

Geo-ML: An online toolkit for probabilistic analysis of geotechnical assets using machine learning techniques

Surya Sarat Chandra Congress, Raja Jaladurgam, Anas Shaaban
Michigan State University, East Lansing, Michigan, USA, surya@msu.edu

ABSTRACT: Typical landslide risk management involves tasks such as hazard and consequence identification, risk analysis, calculation, evaluation, and acceptance, as well as risk mitigation, project execution, and monitoring. Accurate risk analysis and evaluation are essential to effective risk management and treatment, particularly in slope stability problems within the geotechnical domain. However, this process presents challenges in decision-making and in evaluating alternatives under conditions of uncertainty in field conditions, material properties, and analysis methods. These challenges are faced by field inspectors, geotechnical practitioners, designers, and researchers in selecting appropriate statistical parameters for uncertainty analysis, probability distributions, correlation coefficients, spatial variability parameters, suitable reliability methods, and acceptable risk levels. In this study, a comprehensive and logical framework, in the form of a web-based toolkit application, is developed from a compilation of past literature to provide guidelines and enhance the accuracy of risk estimation. This toolkit uses vision-based Large Language Models (LLMs) to parse borehole log data and to extract location information. The extracted borehole information is displayed on an interactive map within the web interface. The proposed toolkit is designed to serve as a practical resource for field inspectors, geotechnical practitioners, and designers, guiding them in selecting and quantifying uncertainty in soil and model parameters, and in assessing the advantages and limitations of various probability distributions for specific hazards. Furthermore, an inbuilt chatbot provides recommendations for suitable reliability methods and establishes standard acceptable risk levels for different hazard scenarios. This toolkit ensures robust risk assessment and facilitates the adoption of effective risk management strategies to achieve safety goals and optimize the design process using probabilistic analysis.

KEYWORDS: Slope stability, risk analysis, machine learning, LLMs.

1 INTRODUCTION

According to the Unified Global Landslide Database (UGLD), a total of 37,946 landslides and 185,753 fatalities were recorded across 161 countries between 1903 and 2020. Among these, rainfall-triggered landslides account for approximately 61%, followed by those caused by human triggers and earthquakes (Gómez et al. 2023). With rapid climate change, increased rainfall, and urbanization, the likelihood of landslides and their consequences is rising. To mitigate significant deaths, injuries, economic losses, and environmental degradation, it is essential to understand the mechanics of landslides and implement effective landslide risk management strategies. Landslide risk management for slope analysis involves tasks such as scope and problem identification, site characterization, risk analysis in terms of the likelihood or probability of its occurrence and the consequences of its occurrence, risk calculations, risk evaluation, risk mitigation, and risk monitoring (ISO, 2009; Wyllie, 2023). Due to project constraints and limited data, quantifying hazards and risk becomes challenging.

Geotechnical site characterization involves determining subsurface stratigraphy, representative design parameters, and groundwater conditions through a combination of field and laboratory investigations. It is a critical preliminary step for optimizing project decisions and conducting geotechnical slope stability analyses. Typical geotechnical investigation records include borehole logs, test pit logs, in-situ testing data (e.g., Standard Penetration Test (SPT) and Cone Penetration Test (CPT)), and groundwater monitoring data. In addition to quantitative data, field logs also capture key qualitative observations, such as ground elevation, location coordinates, local geological names, drilling conditions, the presence of slickensides, or other interbedded soil and rock lenses that can significantly influence geotechnical decision-making.

Traditionally, geotechnical engineers manually extract and compile this information into spreadsheets or project records, a process that is often time-consuming, labor-intensive, and susceptible to human error. However, with the integration of modern data management systems and advanced automation tools, particularly vision-enabled Large Language Models (LLMs) like OpenAI's GPT, the efficiency, accuracy, and reliability of geotechnical site characterization can be significantly enhanced. The development and application of such tools offer the potential to streamline project workflows, reduce time, and improve the precision of data interpretation. Since geotechnical field data are obtained at discrete locations, and subsurface conditions inherently vary spatially, the design parameters derived from limited field data carry significant variability and uncertainty.

The nature of uncertainty in geotechnical engineering falls into two major categories. The first is aleatory uncertainty (natural or inherent soil variability), which is random and can be modeled as temporal variability (variability over time at a single location) and spatial variability (variability over space at different locations at a single time). It also includes measurement and testing variability arising from limitations in instruments, sampling procedures, and laboratory/field testing methods. The second is epistemic uncertainty (knowledge uncertainty), which arises from limitations in data, analytical models, and scientific understanding of natural processes (National Research Council, 2000; Ang and Tang 1975). Each type of uncertainty influences the probability of failure; i.e., the higher the uncertainty, the more critical the safety of the slope. Aleatory and epistemic uncertainty associated with soil and model parameters can be described through statistical measures in risk and reliability analysis.

The most important of these measures are central tendency and variability of index parameters, strength properties, field measurements, and measurement errors, which can be defined using mean (μ), standard deviation (SD), coefficient of variation (COV), and scale of fluctuation (δ) in the vertical and horizontal directions (Phoon and Kulhawy 1999; Phoon et al. 1995; Harr, 1984; Lacasse and Nadim 1997; Duncan, 2000). When there is insufficient or limited field data, these parameters are used to quantify the variability of geotechnical soil properties and model parameters using probability distribution or a probability density function to describe the precise distribution of any natural phenomenon.

To fit probability distributions with high confidence intervals, sufficiently high-quality data are required. For large and high-risk projects, selecting the most appropriate probability distribution can be challenging when limited field data is available. Common distributions for engineering properties include normal, lognormal, exponential, and uniform distributions. Standard reliability methods such as first order reliability method (FORM), simulation-based Monte Carlo simulations (MCS), variance reduction techniques such as Latin Hypercube Sampling (LHS) and Subset Simulation (SS), and surrogative models such as Stochastic Response Surface Methods (SRSM) are used to evaluate the probability of failure. The choice of methods depends on model complexity, data availability, and precision.

Risk is a function of the probability of hazard occurrence and the consequences of these hazards. It can be evaluated either qualitatively or quantitatively. The qualitative approach involves subjective risk ranking in terms of a frequency rating scale for hazard occurrence and an impact rating scale for consequences. Dai et al. (2002) proposed that quantitative risk analysis relies less on qualitative estimates and judgment, and more on numerical values for the probability of event occurrence and monetary values for the consequences of these events.

In this paper, Geo-ML web interface with four modules is developed to assist in geotechnical asset management. Module 1 uses vision-based LLMs to parse borehole log data. Module 2 is a database that contains typical statistical measures for various soil types. Module 3 guides users into choosing an appropriate probability distribution based on the input soil property or model parameters. Module 4 feeds on the likelihood or probability of failure as input and calculates risk based on consequence scores for geotechnical projects applicable to embankments and natural slopes. A module-aware GPT-4 chatbot is also integrated to help users interpret results and guide parameter and distribution choices. The availability of these four modules in one place through the toolkit enables geotechnical engineers, consultants, and field inspectors to automate the interpretation of subsurface stratigraphy and geotechnical design parameters from available bore log data, and to select appropriate probabilistic parameters and probability distributions for risk assessment.

2 BACKGROUND

2.1 *Probabilistic parameters*

Natural and artificial slopes are constantly subjected to geomorphological changes, leading to inherent variability in soil properties. Accurate modeling of the highly complex, nonlinear behavior of soils using analytical and numerical models is cumbersome. To account for such highly complex behavior, common statistical parameters, such as the μ, SD, COV, and δ, are used. However, due to the limited availability of field data, practitioners often rely on historical information to estimate statistical parameters. The probabilistic characteristics of commonly used statistical parameters in risk and reliability analyses for fine-grained and sandy soils can be estimated from the literature. According to Phoon and Kulhawy (1999), the COV for undrained shear strength ranges from 11% to 49%, with a typical fluctuation scale of about 2.5 m. Unit weight exhibits lower variability, with COV values between 3% and 20% and a larger fluctuation scale of approximately 5.2 m. Effective friction angle shows moderate to high variability (10%–50%). Natural moisture content varies from 7% to 46%, with a fluctuation range of about 5.7 m. The SPT N-value also displays considerable variability, with COV values between 37% and 57%, while Young's modulus shows the widest range (30%–90%) of COV values.

2.2 *Probability distributions*

With limited field data, studies in literature can provide important guidance for selecting appropriate probability distributions. In geotechnical engineering, the normal distribution is the most used because many engineering variables conform to it, especially when they result from several random effects. Lumb (1970) highlighted that for soils with cohesive and frictional strength components, the natural variability often aligned more closely with beta distributions than with normal distribution. Different soil characteristics/properties follow different distributions, as shown in Table 1. Despite varying complexity and suitability, each distribution offers unique advantages for modeling geotechnical parameters under uncertainty.

Table 1. Typical probability distributions used for soil and model parameters in past studies (Baecher and Christian 2003).

ID	Parameter	Normal	Lognormal	Extreme value	Weibull	Beta	Uniform
1	Cohesion	✓	✓			✓	
2	Friction angle	✓	✓			✓	
3	Unit weight	✓	✓				✓
4	Tension crack depth	✓					✓
5	Water depth - tension crack						✓
6	Seismic acceleration ratio			✓	✓		
7	Length of a rock joint			✓	✓		

2.3 *Reliability methods*

The tools used for performing probabilistic analysis applicable to slope stability problems fall into three broad categories: direct reliability analysis, simulation-based techniques, and surrogate models (Baecher and Christian 2003). Direct reliability analysis propagates uncertainties in properties, geometries, loads, water levels, and other factors through analytical models/performance functions to derive probabilistic descriptions of a structure's or system's behavior. These include the First Order Second Moment (FOSM) method, which uses the first terms of a Taylor series expansion of the performance function to estimate its expected value and variance. It is called the second moment method because the variance, a second-order statistical moment, is the highest-order statistical result used in the analysis. For a system with N uncertain variables, this method requires either evaluating N partial derivatives of the performance function or performing a numerical approximation using evaluations at 2N points (Duncan, 2000).

Furthermore, the Second Order Second Moment (SOSM) utilizes the terms from the Taylor series up to second order. Computational difficulty is greater, and the improvement in accuracy may not always justify the extra effort. Rosenblueth (1975) introduced a simple and efficient approach to obtain performance function moments by evaluating it at selected discrete points. However, it requires 2N evaluations, making it impractical for many uncertain parameters. While the Rosenblueth method offers a straightforward framework, especially for systems with a small number of variables (e.g., 2 or 3), its computational demands increase exponentially with the number of variables, creating significant challenges (Hammah et al. 2009).

Simulation techniques include Monte Carlo Simulation (MCS), which is computationally intensive due to its randomization processes. In addition to the μ and SD of a variable, MCS also requires knowledge of the probability distribution of variables during randomization, obtained from prior experience or through assumptions. In this approach, the analyst generates many random parameter sets and computes the performance function for each. The results are used to compute the reliability index (β) or probability of failure (P_f). Though conceptually simple, MCS often requires multiple evaluations to achieve accurate results. Latin Hypercube Sampling (LHS) improves efficiency over MCS by reducing variance and achieving faster convergence with fewer samples (~1/5th of MCS), ensuring better stratification (Hoek, 2007). Subset Simulation (SS) is an efficient method for estimating rare event probabilities by breaking down rare event simulations into more probable conditional simulations, thereby reducing the total number of samples needed. However, it requires parameter tuning and relies on computationally complex Markov Chain Monte Carlo (MCMC) methods (Au and Beck, 2001, 2003). The Stochastic Response Surface Method (SRSM) reduces computational effort by using surrogate models, such as polynomial approximations (Yuan et al. 2024)

2.4 *Risk analysis*

The application of risk and reliability analyses to geotechnical slope stability problems involves scope and problem identification (mode of failure), soil and model parameter estimation, selecting an appropriate geotechnical analytical model and probability distribution, and uncertainty and risk analyses. Additionally, uncertainty analyses involve the evaluation of the probability of failure (p_f)/reliability index (β) using suitable probabilistic methods. Also, risk calculations consider additional data regarding the consequences of the failure event. Finally, risk assessment involves comparing the calculated risk with valuable judgments and acceptable or societal risk criteria to optimize risk-based decisions. Figure 1 illustrates a simple methodology for risk analysis, particularly applicable to slope stability problems.

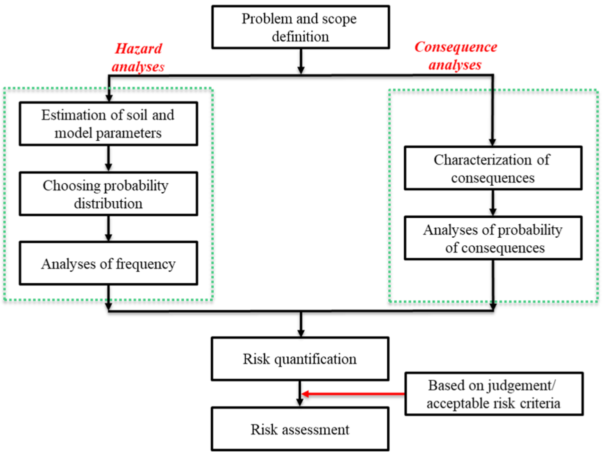


Figure 1. Methodology for risk analysis.

Duncan (2000) noted that reliability analyses require little effort beyond that involved in conventional geotechnical analyses. However, they become tedious and yield inaccurate results when performed during project development or preliminary designs. Additionally, it could be included during the planning or tendering stage for road and railway infrastructure by engineers, consultants, or field inspectors to prepare budgets, allocate funds, and develop preliminary designs to identify high-risk zones. Such designs and estimates are often based on limited field data. Inaccuracies in risk assessment could accumulate if applied to larger areas. Additionally, this could lead to an underestimation or overestimation of the risk involved. Therefore, proper planning and risk assessment are key challenges that must be addressed effectively.

2.5 *Large Language Models (LLMs)*

Large Language Models (LLMs) are deep learning models trained on large text corpora to understand and generate human language. According to Amazon Web Services, LLMs "understand and generate human language with a high degree of accuracy" (AWS, 2025). This makes them effective for tasks such as summarization, parsing, and classification. Models such as Meta's LLaMa and OpenAI's GPT have shown strong performance in automating previous manual tasks, including classifying tables from research papers and processing geotechnical data.

LLMs were considered in this study to create a chatbot for for their potential to automate the interpretation of borehole logs and organize complex geotechnical information across the Geo-ML workflow. OpenAI's GPT-4 API was selected for its strong natural-language performance, reliable cloud infrastructure, and ease of integration with the web interface. The chatbot is designed to be module-aware, drawing on probabilistic parameters, probability distributions, reliability methods, and risk analysis modules to answer queries about soil variability, explain risk calculations, and provide guidance on distribution and reliability method selection using natural language.

Geo-ML also lets users export chatbot conversations to PDF, enabling them to keep a written record of the AI's explanations for preliminary project notes, reporting, and quality assurance.

2.6 *Geosetta*

Geosetta is a web-based online application built to host historical subsurface/geotechnical data. It employs machine learning methods to train deep neural networks and analyze the datasets from auger boring, boring data, and foundation installation (Geosetta, 2022). It extracts data inputs such as soil strength, groundwater level, depth to rock, soil classification, rock classification, and rock quality. This information is used in conjunction with supervised learning algorithms to predict these details in locations where the information is unavailable. It uses tabular data neural network models, such as those from Fast AI or Random Forests, to predict Standard Penetration Test (SPT) data, rainfall estimates, pavement thickness, and more. Additionally, it provides visualization of topography and site features, as well as predictions of subsurface conditions.

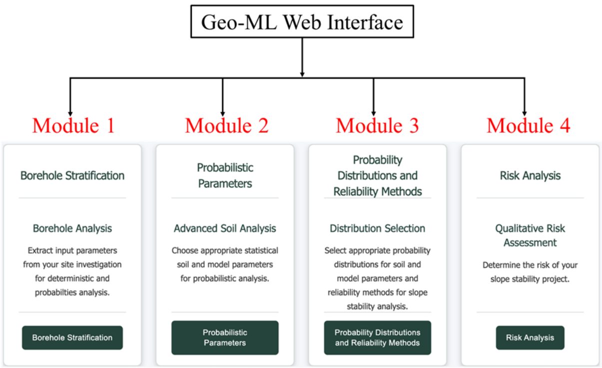
While Geosetta provides predictions of subsurface conditions, at some locations, it appears to overlook an additional step to verify the selected GPS locations against actual site-specific input data. As a result, the generated bore log stratification may not accurately reflect the actual subsurface conditions at the selected site. Furthermore, Geosetta does not provide information on reliability or risk analyses, which is key for transportation geotechnical asset management.

The current study developed a Geo-ML web interface that leverages advanced machine learning techniques to extract information from bore logs and support risk assessments, as discussed below.

3 METHODOLOGY

3.1 *Overview*

This toolkit is designed to streamline geotechnical data analysis and risk assessment through four modules, as illustrated in the Geo-ML web interface flowchart (Figure 2). Module 1, borehole stratification, enables comprehensive extraction of input parameters from site investigations. Module 2, probabilistic parameters, helps users select suitable statistical soil and model parameters for advanced probabilistic analyses. Module 3, probability distributions and reliability methods, guides the selection of appropriate probability distributions for geotechnical parameters and reliability methods to ensure accurate modeling using other commercial slope stability software. Finally, Module 4, risk analysis, builds on the output of the stability analysis, performs a risk assessment to evaluate potential risks associated with slope stability projects, and supports informed decision-making.



Extract input parameters from your site investigation for deterministic and probabilties analysis.

Choose appropriate statistical soil and model parameters for probabilistic analysis.

Select appropriate probability distributions for soil and model parameters and reliability methods for slope stability analysis.

Determine the risk of your slope stability project.

Figure 2. Flowchart of Geo-ML web interface.

3.2 *Module 1*

The first module extracts borehole data logs with the assistance of a vision-based LLM and displays all relevant information, including depth ranges for each soil type and SPT columns, in a user-friendly manner. Extracting borehole data from PDF reports posed several challenges due to the complex nature of geotechnical logs, particularly in handling depth ranges, lithology, and associated soil properties. Initially, the ExtractTable API was employed to extract tabular data by identifying and parsing each soil layer line by line based on depth ranges. This method enabled the creation of a structured Pandas DataFrame containing numerical values, such as SPT N-values, moisture content, and unit weights, but it could not interpret lithological boundaries that segment the borehole into meaningful layers.

To overcome this limitation, a hybrid extraction strategy was implemented. The ExtractTable API continued to be used for its precise numerical data-parsing capabilities. However, to accurately define depth ranges based on lithology, an additional step was introduced using a vision-enabled large language model (LLM). Borehole log images were converted to PNG format and submitted to OpenAI's o3 Vision API, along with a structured prompt that guided the model to visually identify lithological boundaries. The LLM interpreted the stratigraphic transitions and returned structured JSON objects specifying distinct soil layers and their respective depth intervals. This combination of tabular and visual extraction ensured both numerical precision and lithological accuracy. While it introduced some complexity in integrating outputs from different modalities, it significantly improved the overall quality of the data. It provided a more robust foundation for geotechnical interpretation, which is necessary for risk analysis.

After the borehole log is parsed and displayed, the toolkit stores the borehole data on a dedicated page and allows single-click exports to JSON and Excel format. During extraction, borehole coordinates are identified and used to display the extracted borehole and its associated information on an interactive map in the Geo-ML web interface. Selecting a borehole on the map allows you to view a summary of the borehole information. This feature eliminates the need to repeatedly upload borehole PDFs, enabling data reuse and providing an instant view of soil conditions.

The toolkit also includes 3D visualization of the subsurface-interpolated stratigraphy across multiple boreholes. The system automatically identifies matching soil layers using normalized classifications and generates linear interpolations rendered as trapezoids connecting corresponding strata between borehole pairs. The normalization algorithm recognizes standard descriptive terms, Unified Soil Classification System (USCS) codes (CH, CL, SP), and standard geotechnical abbreviations.

A unified stratification mode creates continuous soil layers spanning the entire borehole area as a cohesive block, employing Inverse Distance Weighting interpolation to generate tilted surfaces reflecting depth variations, and groups layers by generalized classifications (topsoil, clay, silt, sand, till, rock, etc.). Interactive controls include a borehole spread slider for adjusting horizontal separation, individual borehole toggles for selective visualization, and a polygon fill mode for three or more boreholes with matching layers.

To further enhance geotechnical analysis, the Geo-ML toolkit includes an advanced 3D visualization feature that enables users to select an area and interactively view it. The toolkit integrates publicly available LiDAR point cloud datasets. Users can select an area on the map, and the tile-selection algorithm converts the selected area's coordinates from geographic (WGS84) to Universal Transverse Mercator (UTM) and identifies the intersecting tiles using the appropriate LiDAR tile naming convention from the United States Geological Survey (USGS) database. The system then automatically downloads the LASzip file, a compressed file containing point-cloud data, and processes it using the Point Data Abstraction Library (PDAL) for cropping, filtering, and classification.

Coordinate transformations align the LiDAR coordinate system with the local 3D scene coordinates, ensuring that the surface topography and subsurface borehole data are accurately aligned. The processed point-cloud data are rendered in the web app using WebGL-based Three.js libraries, enabling real-time user interaction within the 3D virtual space. Borehole locations are precisely positioned using their GPS coordinates, and geological layers are represented as cylindrical segments at their respective depths below ground surface. This pipeline, from automated tile collection and PDAL processing to precise borehole placement, provides engineers with a quick, spatially accurate view of subsurface conditions that supports informed decisions during preliminary assessments.

It also incorporates satellite images overlaid on the point cloud. Users can interact with these soil layers via hover, where information about each soil layer, such as depth, SPT N-values, and soil classification, is displayed (Figure 3). By combining LiDAR point clouds, satellite imagery, and borehole data in a real-time 3D viewer, Geo-ML transforms geotechnical information extracted from borehole logs into an interactive 3D workspace.

3.3 *Module 2*

The second module is a database of probabilistic parameters derived from relevant literature, allowing the user to select among soil types such as gravel, sand, silt, and clay. Each soil type then has tabs for its respective ranges for basic, engineering, hydraulic, and settlement properties. Since geotechnical field datasets are typically collected at discrete locations and subsurface conditions vary spatially, the design parameters derived from such data inherently carry significant variability and uncertainty. This module addresses this by compiling and providing statistical measures, such as the published μ, range, SD, COV, and vertical/horizontal δ, for each property under a single umbrella. This directly addresses the challenges geotechnical practitioners face in quickly accessing information on aleatory and epistemic uncertainties for their preliminary risk analysis. Additionally, module 2 allows data to be adjusted and added to the tables.

3.4 *Module 3*

Module 3 guides the typical probability distribution and reliability method adopted for specific geotechnical soil properties in the past literature and serves as a guide for the user's selection. When field data is limited or of poor quality, selecting an appropriate probability distribution can be a significant challenge, especially for large or high-risk projects. This can lead to either underestimating or overestimating risk, thereby affecting the calculated probability of failure. The module also recommends reliability analysis methods, outlining their advantages and limitations. This allows users to select an appropriate reliability method for slope stability analysis in commercially available software, thereby improving project risk assessment outlined in the next module.

3.5 *Module 4*

Module 4 is a three-step activity that helps users calculate the slope stability risk. First, the user inputs both the probability of failure (e.g., the annual frequency of slope instability) and the consequences, including potential loss of life or property damage. The system then looks up these values in a probability table compiled from a literature review to assign a likelihood score (1–5) and a consequence score (1–5). It multiplies these scores to calculate the final risk score. This score is then categorized into clear qualitative levels, such as Very High Risk, Medium Risk, or Very Low Risk, which are displayed on a color-coded risk matrix. The result is displayed on a color-coded chart, accompanied by brief advice such as "monitor" or "take action immediately." The combination of these four modules leads to efficient characterization of the subsurface, selection of probability distributions, application of reliability methods, and preliminary risk assessment. The Geo-ML toolkit reduces manual data entry by using consistent numbers found in published research and presents the final risk in an easily understandable format.

Finally, a module-aware GPT-4 chatbot has been developed, leveraging probabilistic parameters, probability distributions, reliability methods, and risk analysis modules, enabling users to query risk calculations, soil variability interpretations, and distribution selection guidance using natural language. The chatbot also contains a feature to export the chat to a PDF document for project documentation.

4 CASE STUDY

In the present case study, borehole stratification and geospatial position for a random site location are extracted using the Geo-ML and Geosetta web interfaces. The predicted results are then compared with actual subsurface exploration data from SPT and boring logs, along with laboratory test results, to evaluate the accuracy and reliability of the data extraction procedures. For this comparative study, SPT boring log data were collected along Interstate Highway 94 in Minnesota at geographic coordinates 45°00'56.5 "N, 93°16'58.4 "W, with borehole ID 83616.

The Geosetta portal extracted the field location at 44°58'37.0 "N, 93°12'53.6 "W, which is approximately 7–8 miles away from the actual site. In contrast, the Geo-ML web interface with a multi-tier georeferencing check system accurately extracted location coordinates from the SPT boring log data and precisely plotted the field logging location where the actual boring was performed. These results indicate that Geo-ML's additional layer of checks for location accuracy plots the datasets in their correct site locations. Furthermore, the SPT boring log data indicate that from a ground elevation of 847.5 ft to a depth of 3.5 ft, the ground encountered loamy sand and gravel, with an average SPT N value of 30. This loamy sand layer is underlain by a slightly plastic sandy loam that extends down to 6 feet, with an average SPT N value of approximately 8. Below a ground elevation of 839.0 ft, extending to the end of the borehole at approximately 36.9 ft, the subsurface consists of weakly cemented sandstone, with SPT N values reported in a refusal state. This indicated that a very hard, dense layer exists below 6 ft of ground elevation. Furthermore, the boring log records the groundwater table at approximately 35 feet below ground elevation. Table 2 presents subsurface stratification based on actual bore log data, while Table 3 summarizes the automated extraction of boring log data from PDF files using Geo-ML and the Geosetta web interfaces.

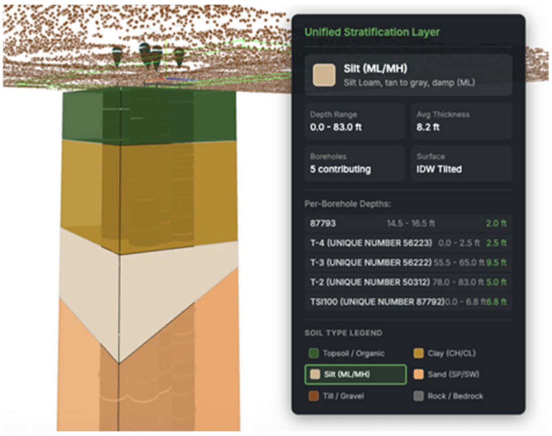
Table 2. Subsurface stratification from actual field boring log data.

Depth (ft)	**Elevation (ft)**	**Classification**	**SPT N**	**Moisture content (%)**
3.5	847.5	Loamy Sand and Gravel	30	5
6	844	Slightly plastic Sandy Loam	8	16
36.9	841.5	Sandstone	46	8

Table 3. Boring log data extraction from Geo-ML and Geosetta.

Geo-ML extracted data			Geosetta Predictions		
Classification	Depth (ft)	SPT N	Classification	Depth (ft)	SPT N
Loamy Sand and Gravel	3.5	30	Clay	20	1-9
Slightly Plastic Sandy Loam	6	8	Sand	40	10-25
Sandstone	36.9	46	Rock	60	50+

The Geosetta data extraction indicates clay up to 20 ft, sand up to 40 ft, and rock extending to 100 ft. At some locations, it does not provide the groundwater table depth information necessary for slope stability analyses. Additionally, Geosetta reports SPT N-values in broad ranges, such as 1–9 and 10–25. In contrast, the actual borehole data provides more precise values, such as 30 for loamy sand and 8 for slightly plastic sandy loam. The Geo-ML bore log data extraction aligns with the actual soil types, soil depth profiles, and other test results. It accurately identifies the soil classification at depths consistent with the actual bore log.



Unified Stratification Layer

Silt (ML/MH)
Silt Loam, tan to gray, damp (ML)

Depth Range: 0.0 - 83.0 ft
Avg Thickness: 8.2 ft
Boreholes: 5 contributing
Surface: IDW Tilted

Per-Borehole Depths:

87793	14.5 - 16.5 ft	2.0 ft
T-4 (UNIQUE NUMBER 56223)	0.0 - 2.5 ft	2.5 ft
T-3 (UNIQUE NUMBER 56222)	55.5 - 65.0 ft	9.5 ft
T-2 (UNIQUE NUMBER 50312)	78.0 - 83.0 ft	5.0 ft
TSI100 (UNIQUE NUMBER 87792)	0.0 - 6.8 ft	6.8 ft

Figure 3. 3D subsurface stratigraphic profile.

Additionally, Geo-ML provides information on ground elevation, geographic location coordinates, elevation relative to exploration depth, and groundwater table depth. Figure 3 shows a 3D visualization of the LiDAR point cloud of the borehole area integrated with bore log data in the Geo-ML Point Cloud Viewer. This interface allows users to interactively explore terrain geometry and visualize subsurface features, such as layer boundaries and ground elevations. Overall, Geo-ML provides a significantly detailed and accurate representation of subsurface conditions for the site considered in this study. In addition to providing accurate borehole stratification, Geo-ML offers advanced probabilistic analysis, distribution modeling, and risk assessment modules, making it a more powerful and comprehensive platform for risk-based decision making in slope stability problems.

5 CONCLUSIONS

This paper introduces Geo-ML, an online toolkit that automates and enhances geotechnical data analysis and preliminary slope stability risk assessment using machine learning. Geo-ML integrates four key modules: borehole stratification, probabilistic parameter selection, probability distribution and reliability method selection, and risk assessment. By leveraging vision-based LLMs, such as OpenAI's GPT, Geo-ML efficiently extracts borehole data from complex PDFs and visual logs. Beyond providing detailed stratigraphy, Geo-ML offers advanced probabilistic tools that support robust risk assessment. These include access to a database of probabilistic parameters, guidance on selecting appropriate probability distributions and reliability methods, and a dedicated module for calculating project-specific slope stability risk. A module-aware chatbot is also integrated to help users interpret results and support parameter and distribution choices through natural-language queries. Together, these features assist in a comprehensive and integrated approach to slope stability analysis and risk-informed decision-making.

6 REFERENCES

Amazon Web Services (AWS), 2025. What is a large language model?. [online] Amazon Web Services. Available at: https://aws.amazon.com/what-is/large-language-model/ [Accessed 21 July 2025].

Ang, A.H.-S. and Tang, W.H., 1975. Probability concepts in engineering planning and design, basic principles. Vol. 1. New York: Wiley.

Au, S.K. and Beck, J.L., 2001. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4), pp.263–277.

Au, S.K. and Beck, J.L., 2003. Subset simulation and its application to seismic risk based on dynamic analysis. *Journal of Engineering Mechanics*, 129(8), pp.901–917.

Baecher, G.B., and Christian, J.T. 2003. Reliability and statistics in geotechnical engineering. Chichester: John Wiley & Sons.

Dai, F.C., Lee, C.F. and Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1), pp.65-87.

Duncan, J.M., 2000. Factors of safety and reliability in geotechnical engineering. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(4), pp.307-316.

Geosetta, 2022. Geosetta user manual. Online user manual. Updated 29 March 2022. Maryland: Geosetta. Available at: https://geosetta.org [Accessed 23 July 2025].

Gómez, D., García, E.F. and Aristizábal, E., 2023. Spatial and temporal landslide distributions using global and open landslide databases. *Natural Hazards*, 117(1), pp.25-55.

Hoek, E., 2007. Practical rock engineering. Online edition. Toronto: Rocscience.

Harr, M.E., 1984. Reliability-based design in civil engineering. New York: McGraw-Hill.

Hammah, R.E., Yacoub, T.E., and Curran, J.H., 2009. Numerical modelling of slope uncertainty due to rock mass jointing. In: *Proceedings of the International Conference on Rock Joints and Jointed Rock Masses*, pp.7–8.

ISO (International Organization for Standardization). 2009. Risk management principles and guidelines; ISO 31000. Geneva: ISO.

Lacasse, S. and Nadim, F., 1997. Uncertainties in characterizing soil properties. Publication No. 201. Oslo: Norwegian Geotechnical Institute, pp.49-75.

Lumb P., 1970. Safety factors and the probability distribution of soil strength. *Canadian Geotechnical Journal*, 3(7): 225-242.

National Research Council, Commission on Geosciences, Resources, Water Science, Technology Board, and Committee on Risk-Based Analysis for Flood Damage Reduction, 2000. Risk analysis and uncertainty in flood damage reduction studies. Washington, D.C.: National Academies Press.

OpenAI. (2025). Security and privacy. OpenAI. Retrieved from https://openai.com/security-and-privacy/ [Accessed 21 July 2025].

Phoon, K.K. and Kulhawy, F.H., 1999. Evaluation of geotechnical property variability. *Canadian Geotechnical Journal*, 36(4), pp.625-639.

Phoon, K.K., Kulhawy, F.H. and Grigoriu, M.D., 1995. Reliability-based design of foundations for transmission line structures. Report TR-105000. Palo Alto: Electric Power Research Institute.

Rosenblueth, E., 1975. Point estimates for probability moments. Proceedings of the National Academy of Sciences, 72(10), pp.3812–3814.

Wyllie, D.C. 2023. Risk Management for Geotechnical Engineering: Hazard, Risks and Consequences. Boca Raton: CRC Press.

Army, U.S., 1997. Introduction to probability and reliability methods for use in geotechnical engineering. *Engineering Technical Letter*, 1110-2, 547.

Yuan, R., Jia, Z., Sun, J., Ou, J., Wang, W., He, Y. and Jiang, J., 2024. Probabilistic analysis of a three-dimensional slope based on limit analysis. *Natural Hazards Review*, 25(4), p.04024043.