Granite Residual Soil and Machine Learning in Geotechnical Engineering: A Survey

www.surveyx.cn

Abstract

This survey explores the integration of machine learning (ML) techniques in geotechnical engineering, focusing on granite residual soil and its shear strength parameters. Granite residual soil, formed through the weathering of granite rock, presents unique challenges due to its complex properties. ML methods, including neural networks and hybrid models, offer significant advancements over traditional approaches by accurately modeling the complex interactions within soil data. The emphasis on model interpretability ensures that predictions are reliable and comprehensible, crucial for practical engineering applications. The survey systematically reviews the role of ML in predicting shear strength parameters, addressing algorithm selection, data processing, and model transparency. It highlights case studies demonstrating successful ML applications in predicting soil behavior, emphasizing the importance of integrating geological factors with advanced ML techniques. Despite advancements, challenges remain, such as the need for high-quality datasets and computational efficiency. Future research should focus on hybrid models, iterative ML techniques, and the incorporation of missing data to enhance model robustness and applicability. The survey concludes that ML holds transformative potential for geotechnical engineering, promising more accurate, interpretable, and adaptable models that contribute to safer and more efficient engineering practices.

1 Introduction

1.1 Significance of Granite Residual Soil

Granite residual soil is pivotal in geotechnical engineering due to its unique formation processes and properties that influence construction stability and shear strength parameters. The weathering of granite leads to a soil type that presents distinct challenges for engineering applications, necessitating specialized management approaches to address its complex behavior [1]. Its mineral composition and structural attributes are crucial for understanding responses to various loading conditions and environmental factors [2]. The anisotropic behavior of granite residual soil, akin to the crystallographic texture of metallic materials, necessitates a comprehensive evaluation of its mechanical properties to ensure structural integrity [3]. Addressing these complexities requires an interdisciplinary approach, integrating insights from fields such as AI and software engineering to enhance predictive modeling and improve engineering outcomes [4].

1.2 Role of Machine Learning in Geotechnical Engineering

Machine learning significantly advances geotechnical engineering by providing sophisticated tools for predicting and analyzing shear strength parameters, essential for evaluating soil stability and ensuring structural safety. Techniques such as neural networks and hybrid models greatly improve upon traditional methods by accurately capturing material behavior under varying conditions [5]. This is particularly beneficial in modeling path-dependent behaviors, such as those seen in crystal plasticity, where conventional models often fall short [6].

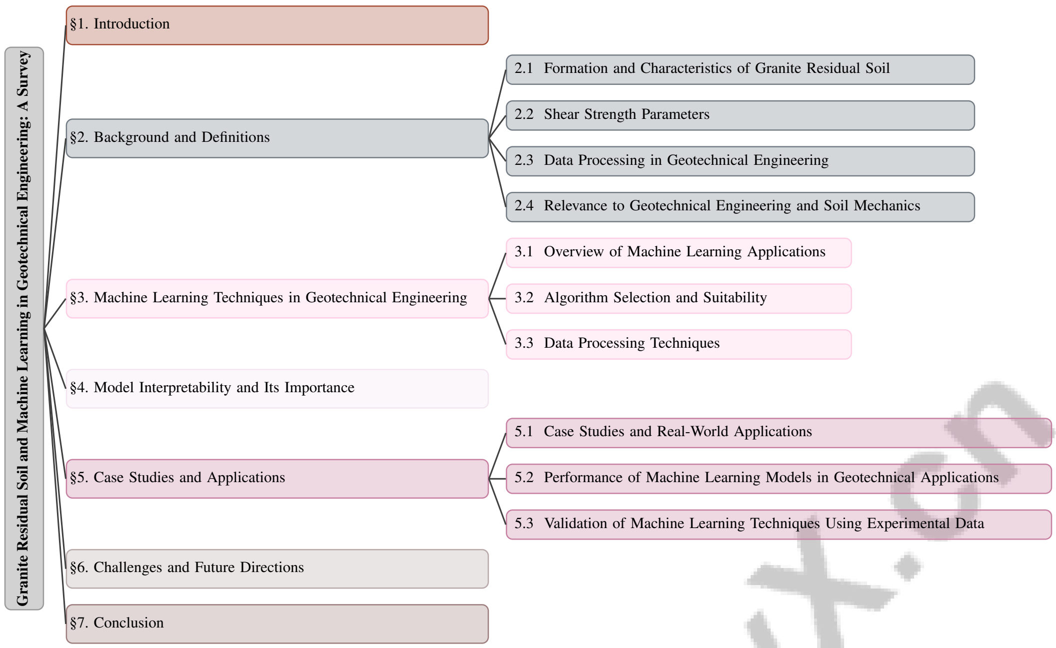


Figure 1: chapter structure

In geotechnical applications, machine learning models effectively predict critical soil properties by incorporating variables like sand content, saturation levels, and dry density, which are vital for understanding expansive soils' mechanical properties [7]. The iterative nature of machine learning enhances the adaptability and accuracy of predictive models, paralleling control systems [8]. Additionally, Convolutional Neural Networks (CNNs) have shown efficacy in classifying and predicting complex patterns within geotechnical data.

However, the reliance on black-box models in machine learning presents challenges regarding interpretability, crucial for practical engineering applications [9]. Developing glass-box models that balance predictive performance with transparency is essential to ensure predictions are reliable and comprehensible to engineers and stakeholders [10]. Moreover, integrating machine learning with satellite imagery analysis offers new opportunities for systematic classification of land cover types, enhancing understanding of environmental impacts on geotechnical properties [11].

Machine learning applications also extend to analyzing time series data, such as predicting water levels influenced by rainfall and climate variables, underscoring the versatility of these techniques in handling diverse data types [12]. As the field evolves, the synergy between machine learning and traditional geotechnical methods promises innovative solutions to complex engineering challenges and improved accuracy of uncertainty estimates [13].

1.3 Objectives and Importance of Model Interpretability

This paper aims to enhance the predictive accuracy of machine learning models in geotechnical engineering, focusing on granite residual soil. This objective involves integrating geological factors with advanced machine learning techniques to improve shear strength parameter predictions, thereby contributing to safer and more reliable engineering practices. A critical component of this objective is model interpretability, vital for ensuring predictions are accurate and understandable to engineers and stakeholders. This understanding fosters accountability in decision-making and enhances trust in machine learning models, particularly as the field evolves to incorporate interpretable models like generalized additive models (GAMs) that balance predictive performance with clarity [14, 15, 16].

Model interpretability serves as a bridge between complex algorithmic outputs and practical applications, enabling models to provide clear explanations for their decisions, which is crucial for reliability and transparency [10]. This is especially significant in geotechnical engineering, where understanding material behavior under varying conditions is vital for structural stability and safety

AI-generated, for reference only.

[13]. Additionally, interpretability aids in assessing prediction reliability, critical when incorporating innovative methodologies like Physics-informed Denoising Autoencoders (PI-DAE) [17].

The paper also evaluates recent advancements in machine learning applications for soil improvement, emphasizing the need to align these advancements with sustainable practices. Innovative approaches, such as developing automated detection pipelines and incorporating roughness effects into models, are explored, with interpretability playing a pivotal role in ensuring practical applicability [18]. Furthermore, optimal decision trees for model selection exemplify the balance between high accuracy and interpretability [9].

The pursuit of interpretability in machine learning models enhances decision-making in geotechnical engineering and improves overall trust and usability across scientific and engineering domains. This is underscored by the necessity of explaining machine learning model predictions through methods like counterfactual explanations, which bolster user trust and meet legal requirements [19]. Additionally, developing interpretable non-fuzzy rules for data representation and classification addresses existing methods' limitations, further reinforcing the importance of interpretability in machine learning. Techniques such as Split Integrated Gradients enhance attribution accuracy, contributing to more reliable and interpretable models [20].

1.4 Structure of the Survey

The survey is systematically organized to provide a comprehensive exploration of granite residual soil and the application of machine learning in geotechnical engineering. It begins by introducing the significance of granite residual soil and the pivotal role of machine learning in enhancing the understanding and prediction of shear strength parameters, emphasizing the importance of model interpretability [21]. Following the introduction, the survey delves into the background and definitions, offering a detailed explanation of granite residual soil, including its formation and characteristics, and defining key concepts such as shear strength parameters and relevant data processing techniques.

Subsequently, the paper reviews various machine learning techniques applied in geotechnical engineering, focusing on predictive modeling for shear strength parameters. This section discusses algorithm selection, data processing techniques, and the significance of model interpretability. The survey then presents case studies and real-world applications, illustrating successful implementations of machine learning in predicting shear strength parameters in granite residual soil.

The latter part addresses challenges and future directions in applying machine learning to geotechnical engineering, particularly concerning granite residual soil. This section identifies current challenges, explores innovative approaches in predictive modeling, and discusses advancements in algorithms and techniques. The survey concludes with a summary of key findings, reiterating the importance of integrating machine learning with geotechnical engineering to enhance the understanding and prediction of shear strength parameters while highlighting model interpretability's significance in ensuring practical applicability [22].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Formation and Characteristics of Granite Residual Soil

Granite residual soil results from the weathering of granite rock, involving mechanical disintegration and chemical decomposition, yielding distinct physical and chemical characteristics crucial for geotechnical assessments [7]. The soil's unique physical and mechanical properties significantly influence its behavior in engineering applications, impacting shear strength, compressibility, and stability [23]. Variations in mineral composition, texture, and structure, directly inherited from the parent granite, critically determine mechanical behavior, including load-bearing capacity and permeability, vital for foundation and slope design [24]. The heterogeneity of granite residual soil, marked by diverse particle sizes and moisture retention, affects shear strength and compressibility, necessitating detailed geotechnical evaluations to ensure structural stability and safety.

The interplay between mineralogical and mechanical properties underscores the need for comprehensive geotechnical evaluations to mitigate construction risks. Advanced data processing techniques enhance soil characterization by effectively managing data complexities. Understanding the mechanical and environmental properties of clayey soils, particularly with lime and waste marble powder enhancements, is crucial for optimizing construction practices, improving load-bearing capacity and permeability, and mitigating erosion risks. Data-driven approaches, including machine learning algorithms, further refine predictions of structural responses, enhancing construction integrity in challenging geological conditions [24, 25, 26].

2.2 Shear Strength Parameters

Shear strength parameters are crucial for stability analysis in soil structures, particularly granite residual soil, where unique characteristics significantly influence these parameters [22]. Defined by cohesion and internal friction angle, shear strength determines the soil's ability to resist shear stress without failure, critical for designing safe geotechnical structures like foundations and retaining walls [27]. Accurate determination of shear strength parameters is complicated by environmental factors such as air pressure, humidity, and temperature, which can obscure the soil's true mechanical properties [28]. This complexity parallels challenges in predicting mechanical properties in calcined clay cements, where inadequate understanding complicates empirical modeling [26]. Additionally, uncertainties in velocity structure models can bias fault activity predictions, underscoring the importance of precise shear strength assessments in soil mechanics [29].

The heterogeneity of granite residual soil, marked by variations in mineral composition and particle size distribution, necessitates advanced modeling techniques to accurately capture shear strength behavior. This need is similar to the precise measurement of mechanical properties in other materials, such as Na-SSEs, where shear modulus is vital for structural integrity [30]. Developing robust predictive models that incorporate these complexities is essential for reliable geotechnical assessments and optimizing engineering practices [24]. Existing methodologies often focus on shape differences in density functions without considering orientation distances and correlations, crucial for a comprehensive understanding of shear strength parameters [3]. Integrating additional features and contextual information into predictive models, as seen in recent advancements in benchmark datasets, can enhance the accuracy and reliability of shear strength evaluations [31]. These advancements emphasize the need to refine our understanding and modeling of shear strength parameters to ensure geotechnical structures' safety and stability.

2.3 Data Processing in Geotechnical Engineering

Data processing is crucial for accurately analyzing and modeling soil behaviors, particularly with extensive datasets from sources like satellite imagery, ground stations, and climate models, vital for understanding environmental impacts on geotechnical properties [11]. Advanced techniques, such as deep learning models like N2N-Seismic, reduce noise in seismic data using residual neural network techniques tailored for geophysical applications [32]. In seismic data processing, deep preconditioners address interpolation from sparsely sampled receiver arrays, reconstructing data onto a finely sampled grid to enhance geophysical interpretations [33]. The integration of pre-trained Transformer models facilitates the reconstruction of masked seismic data for various downstream geotechnical tasks [17].

Standardization techniques, such as Z-Normalization, ensure datasets are on a common scale, optimizing data for machine learning applications and enhancing predictive model accuracy [34]. As exascale machines become more prevalent, feature-driven methods for in situ data analysis and reduction are increasingly important for efficient data processing in high-performance computing environments [35]. Using Inline Coordinates to visualize n-dimensional data in two-dimensional space aids in interpreting complex geotechnical data, simplifying high-dimensional dataset analysis [36]. Machine learning algorithms applied to acceleration data from tri-axial accelerometers demonstrate the integration of traditional data with advanced modeling techniques, enhancing geotechnical dynamics understanding [37].

The selection of high-quality radar data and the application of inversion algorithms to estimate permittivity values exemplify the precision required in data processing for accurate geotechnical assessments [38]. These strategies collectively contribute to robust analysis and interpretation of geotechnical data, supporting the development of more accurate and reliable engineering models. Additionally, the survey of classical statistical methods and recent machine learning approaches for time series regression, such as dynamic regression and regARIMA, highlights the evolving landscape of data processing in geotechnical engineering [12].

2.4 Relevance to Geotechnical Engineering and Soil Mechanics

Granite residual soil is vital in geotechnical engineering and soil mechanics due to its distinct properties and associated challenges in engineering applications. Understanding geomorphological formation factors, such as sediment thickness and slope angles, is essential for comprehending the geological complexities of granite residual soil, which directly influence its mechanical behavior and stability. Detailed analysis of these factors is necessary to ensure reliable design and construction of geotechnical structures [39]. Integrating advanced data processing techniques, like the POCS algorithm, enhances the precision and reliability of geophysical data interpretation, crucial for assessing soil mechanical properties. This importance parallels the need for accurately determining shear strength parameters in geotechnical engineering, where precise data interpretation is vital for evaluating soil stability and performance [40]. Moreover, variability in measurement parameters and data processing methods across laboratories introduces significant uncertainties, highlighting the necessity for standardized approaches to ensure consistent and accurate evaluations of soil properties [41].

Machine learning and feature engineering are pivotal in addressing uncertainties associated with the dynamics and environmental interactions of granite residual soil. Developing innovative data-driven methods is necessary to improve decision-making processes in geotechnical engineering, particularly in predicting and managing soil behavior under varying conditions [42]. Additionally, challenges in generating accurate labels for geospatial imagery underscore the importance of high-quality datasets for training machine learning models, crucial for advancing geotechnical applications [43]. The complexities of simulating and predicting granular flow dynamics, especially under varying boundary conditions and geometries, further emphasize the need for robust modeling techniques in soil mechanics. These techniques are essential for capturing the complex behaviors of granite residual soil, ensuring the stability and safety of engineering structures [44]. Integrating AI and machine learning into geotechnical engineering addresses data accuracy and ethical implications while offering innovative solutions for enhancing the reliability and efficiency of soil mechanics analyses [4].

3 Machine Learning Techniques in Geotechnical Engineering

Category	Feature	Method
Overview of Machine Learning Applications	Performance Enhancement	OCSSM[24]
	Statistical Estimation	RF-OOB[13]
	Model Refinement	IML[8], AME[7], F-GMM[45]
Algorithm Selection and Suitability	Ensemble Methods	CB[46], ML-MWD[47], ML-UCPL[48]
	Data Enhancement Techniques	DP[33]
Data Processing Techniques	Feature and Knowledge Integration	MLFM[49], ML-SP[50], PIDDCM[5]
	Data Scaling	DPR-GUI[51]
	Numerical Techniques	RDLM[52], ML-UCS[53]

Table 1: This table provides a comprehensive summary of various machine learning methods applied in geotechnical engineering. It categorizes the methods into three main areas: machine learning applications, algorithm selection and suitability, and data processing techniques. Each category highlights specific features and methods, along with relevant references, demonstrating their contributions to enhancing predictive accuracy and decision-making in geotechnical contexts.

The integration of machine learning techniques has significantly transformed geotechnical engineering, enhancing traditional methodologies for sophisticated analyses and predictions. This section explores various machine learning applications, demonstrating their effectiveness in elucidating soil behavior and improving engineering practices. As illustrated in **??**, the hierarchical structure of machine learning techniques in geotechnical engineering categorizes these applications into predictive modeling, optimization, and advanced techniques. Table 4 presents a comparison of various machine learning methods utilized in geotechnical engineering, emphasizing their algorithm types, application focuses, and data processing techniques. Additionally, Table 1 offers a detailed summary of machine learning methods employed in geotechnical engineering, categorizing them into applications, algorithm selection, and data processing techniques. This figure highlights the selection and suitability of various algorithms and emphasizes the data processing techniques that facilitate accurate geotechnical analyses, thereby providing a comprehensive overview of how these innovative methods are reshaping the field.

3.1 Overview of Machine Learning Applications

Machine learning is pivotal in geotechnical engineering, enabling advanced predictive modeling and data analysis to deepen understanding of soil behavior and structural responses. Clustering algorithms enhance predictive modeling for shear strength parameters in granite residual soil, optimizing soil properties and ensuring structural stability [7]. Random Forest methods establish valid prediction intervals, enhancing point prediction reliability in geotechnical contexts [13]. These methods are crucial for forecasting geotechnical properties and integrating large datasets to extract meaningful insights.

Iterative machine learning (IML) techniques refine system models iteratively using kernel-based models, improving predictive model adaptability and accuracy for complex material behaviors [8]. Optimization of geotechnical properties is achieved through methods like Response Surface Methodology, which fine-tunes material mixtures to enhance construction practices and ensure project safety [24]. Hybrid models combine interpretable and black-box approaches, balancing accuracy and transparency, which is beneficial in comprehending soil behavior mechanisms [45].

Modern strategies for time series regression categorize research into classical statistical methods and contemporary machine learning approaches, aiding in selecting suitable methodologies for specific geotechnical challenges [12]. Machine learning algorithms effectively predict critical soil characteristics such as compressive strength, bearing capacity, and stress-strain behavior, essential for informed decision-making in soil improvement and construction practices. Advanced techniques like artificial neural networks and genetic programming set new standards for geotechnical modeling, promoting sustainability through green materials [54, 55, 53, 25].

As illustrated in Figure 2, this figure illustrates the hierarchical categorization of machine learning applications in geotechnical engineering, focusing on predictive modeling, optimization techniques, and time series regression. Each category highlights key methodologies and their respective contributions to enhancing predictive accuracy and decision-making in geotechnical contexts. The first image depicts a periodic waveform segmented into sections, useful for understanding cyclic patterns in geotechnical phenomena like soil stress-strain relationships. The second image explores the impact of a parameter, 1, on an implicit rule, emphasizing how varying values influence specific functions. These examples highlight machine learning's potential to provide deeper insights and robust solutions in geotechnical engineering [36, 56].

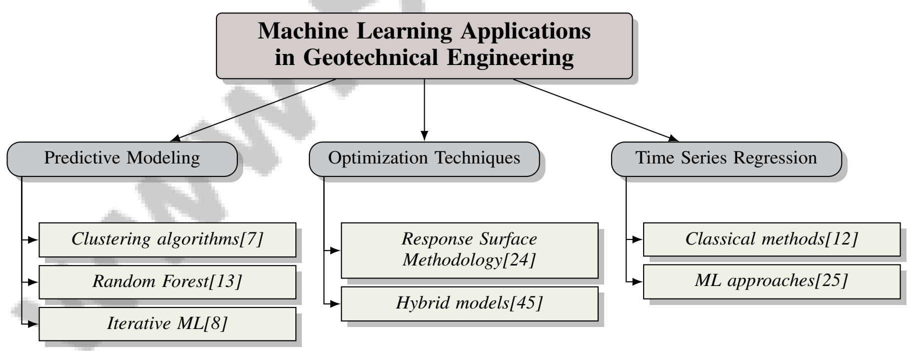


Figure 2: This figure illustrates the hierarchical categorization of machine learning applications in geotechnical engineering, focusing on predictive modeling, optimization techniques, and time series regression. Each category highlights key methodologies and their respective contributions to enhancing predictive accuracy and decision-making in geotechnical contexts.

3.2 Algorithm Selection and Suitability

Selecting appropriate machine learning algorithms is crucial for addressing the complex nature of geotechnical data, particularly with granite residual soil. Table 2 provides a comprehensive comparison of machine learning methods used for geotechnical data analysis, illustrating their suitability, data characteristics, and predictive accuracy. Figure 3 illustrates the categorization of machine learning algorithms suitable for geotechnical data analysis, highlighting neural networks, decision tree methods, and support vector machines with specific applications in seismic data

Method Name	Algorithm Suitability	Data Characteristics	Predictive Accuracy
DP[33]	Deep Neural Networks	Complex, Multi-dimensional	Improve Processing Tasks
ML-UCPL[48]	Street-by-street	Environmental Features	Superior Accuracy
CB[46]	Cart-Bagging Algorithm	Complex Data Relationships	Improve Prediction Accuracy
ML-MWD[47]	Svm, Gp, RF	Real-time Drilling	High Accuracy

Table 2: Comparison of Machine Learning Methods for Geotechnical Data Analysis: This table presents a detailed overview of various machine learning methods, highlighting their algorithm suitability, data characteristics, and predictive accuracy. The methods are evaluated for their applications in complex, multi-dimensional geotechnical datasets, emphasizing improvements in processing tasks and prediction accuracy.

processing, variability management, and high-dimensional data analysis. Neural networks, such as Convolutional Neural Networks (CNNs), excel in capturing intricate data patterns, enhancing inference accuracy in seismic data processing [57, 32]. AutoEncoder (AE) networks solve inverse problems in seismic contexts by mapping input seismic data to a latent manifold, with the trained decoder serving as a nonlinear preconditioner [33].

Decision tree-based methods, including Random Forests and gradient boosting algorithms like CatBoost, effectively manage variability and complexity in geotechnical datasets. These ensemble methods are valuable for robust predictions, such as path loss in urban environments, where diverse data integration is essential [48]. CART-Bagging enhances prediction accuracy by leveraging multiple decision trees in data-variable scenarios [46].

Support Vector Machines (SVMs) and Gaussian Processes (GPs) provide robust regression capabilities, with SVMs adept at handling high-dimensional data, making them suitable for analyzing complex soil behaviors [47]. Adversarial learning frameworks, such as MDA GAN, maintain spatial continuity and anisotropy, essential in geotechnical applications characterized by spatial heterogeneity [58].

Evaluating models like Fully Connected Neural Networks (FCNN) and specialized architectures such as SRGEN demonstrates effective denoising capabilities, essential for improving geotechnical data quality [59]. These models, along with frameworks for time series data analysis, underscore the diverse applicability of machine learning in geotechnical engineering.

Algorithm suitability should be guided by data characteristics and desired outcomes, balancing accuracy, interpretability, and computational efficiency. The evolution of machine learning methodologies enhances predictive accuracy and broadens practical applications, enabling effective analyses of critical soil characteristics. Recent reviews indicate that advanced algorithms, including artificial neural networks and gradient boosting techniques, optimize soil improvement processes through sustainable materials, providing valuable insights for civil and geotechnical engineers [54, 55, 60, 53].

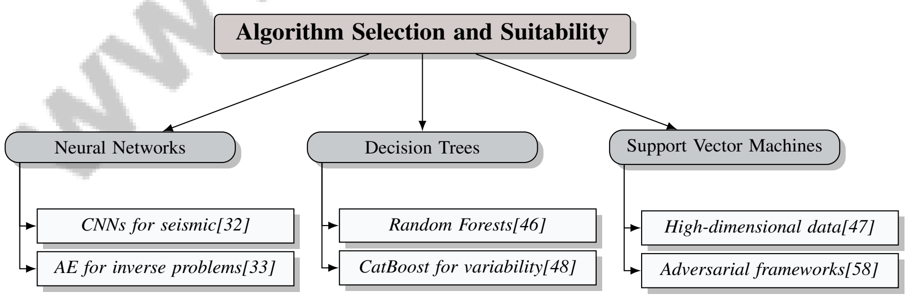


Figure 3: This figure illustrates the categorization of machine learning algorithms suitable for geotechnical data analysis, highlighting neural networks, decision tree methods, and support vector machines with specific applications in seismic data processing, variability management, and high-dimensional data analysis.

3.3 Data Processing Techniques

Data processing is fundamental for developing accurate machine learning models in geotechnical engineering, where soil data complexity necessitates robust methodologies. Advanced techniques

Method Name	Data Complexity	Model Integration	Domain Knowledge Integration
RDLM[52]	Exponential Complexity	Machine Learning Classification	Homotopy Continuation
ML-UCS[53]	Complex Relationships	Multiple ML Techniques	Soil Gradation
MLFM[49]	Complex Data Patterns	Feature Extraction, Training	Renewable Energy Forecasts
ML-SP[50]	-	Feature Generation Techniques	Human-designed Heuristics
PIDDCM[5]	Complex Mechanical Behavior	Combining Feature Extraction	Physics-informed Methods
DPR-GUI[51]	Uneven Distribution	Feature Extraction Integration	Field Knowledge Incorporation

Table 3: Comparison of data processing techniques in geotechnical engineering, highlighting the complexity of data, integration with machine learning models, and domain knowledge application. Each method demonstrates unique approaches to handling complex datasets, integrating feature extraction, and applying domain-specific knowledge for improved model accuracy.

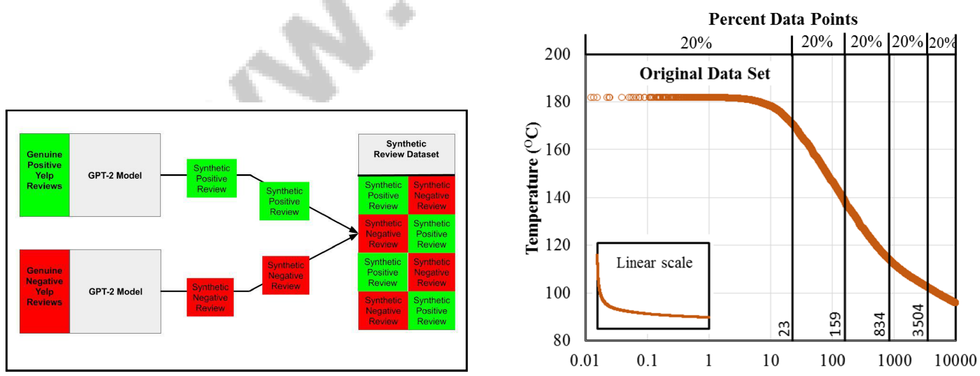
like the Real Discriminant Labeling Method (RDLM) enhance classification accuracy by sampling parameter space and using homotopy continuation to identify real solutions [52]. In soil-lime mixtures, effective data processing is critical for developing predictive models that accurately capture mechanical behavior [53].

Integrating feature extraction and model training is emphasized in machine learning forecasting methods predicting prices based on order book data and renewable energy forecasts [49]. Automating classifier training and testing through machine learning pipelines highlights the significance of data processing in determining optimal variable orderings for accurate model development [50].

Training surrogate models on limited datasets to capture stress-strain relationships of soft materials illustrates the critical role of data processing in physics-informed, data-driven discovery [5]. This approach integrates domain-specific knowledge into data processing for enhanced model interpretability and accuracy.

The methodologies discussed highlight the critical role of rigorous data processing and advanced machine learning techniques in geotechnical engineering. These approaches facilitate accurate predictions of soil characteristics and contribute to developing reliable predictive models. By systematically evaluating the strengths and weaknesses of different algorithms, practitioners can better integrate these technologies into their design processes, leading to improved outcomes in soil improvement and engineering practices [54, 55, 60, 53].

Table 3 provides a comprehensive overview of various data processing techniques employed in geotechnical engineering, emphasizing their data complexity, model integration strategies, and the role of domain knowledge.



(a) Synthetic Review Dataset Generation Using GPT-2 Model[61]

(b) Original Data Set[51]

Figure 4: Examples of Data Processing Techniques

As depicted in Figure 4, machine learning integration in geotechnical engineering has led to innovative data processing methods that enhance complex dataset analysis. The first figure illustrates synthetic dataset generation using advanced models like GPT-2, enriching the original dataset for robust analysis. The second figure presents a graph from an original dataset, highlighting the relationship between data points and temperature, emphasizing the importance of understanding data trends for interpreting geotechnical phenomena. These examples demonstrate machine learning's transformative potential in refining data processing techniques within geotechnical engineering, leading to more accurate and insightful analyses [61, 51].

Feature	Clustering Algorithms	Random Forest	Iterative Machine Learning
Algorithm Type	Clustering	Ensemble	Iterative
Application Focus	Shear Strength	Prediction Intervals	Model Adaptability
Data Processing Technique	Not Specified	Not Specified	Kernel-based Models

Table 4: This table provides a comparative analysis of three machine learning methods—Clustering Algorithms, Random Forest, and Iterative Machine Learning—highlighting their algorithm types, application focuses, and data processing techniques. The comparison underscores the distinct applications and processing approaches of each method, offering insights into their respective roles and adaptability in geotechnical engineering.

4 Model Interpretability and Its Importance

4.1 Model Interpretability and Transparency

Model interpretability is pivotal in the application of machine learning within geotechnical engineering, particularly for predicting shear strength parameters in granite residual soil [7]. The complex behavior of soils and inherent uncertainties in geotechnical data necessitate models that are not only accurate but also understandable and trustworthy [23]. Interpretability ensures that predictions are actionable, a critical requirement for practical engineering applications.

Techniques such as the CNN approach enhance interpretability without sacrificing accuracy, offering significant advantages in speed and precision while reducing outliers compared to traditional methods [57]. This is crucial in geotechnical contexts where parameter optimization demands a clear understanding of machine learning predictions [23]. Advanced architectures like the ANN-GMDH model exemplify the balance between accuracy and transparency, improving prediction precision while providing insights into model decisions [6]. Such integration is essential for accurately predicting shear strength by elucidating the influence of various soil properties. Additionally, techniques like Split Integrated Gradients enhance interpretability by highlighting feature importance through unsaturated gradients [20].

Iterative machine learning (IML) exemplifies adaptability in learning model uncertainties, emphasizing the importance of interpretability in machine learning applications [8]. This adaptability is crucial in geotechnical engineering, where models must handle diverse data types and complex interactions. Current research underscores the ability of machine learning models to capture complex, non-linear relationships and adapt to new data in real time, highlighting the need for models that are both interpretable and adaptable [12].

Incorporating self-attention mechanisms, as illustrated by StorSeismic, enables models to exploit relationships between seismic traces, enhancing generalization across various tasks while minimizing the need for extensive labeled data [17]. This approach demonstrates the potential of advanced architectures to improve the interpretability and applicability of machine learning models in geotechnical engineering.

Pursuing model interpretability in machine learning applications for geotechnical engineering bridges the gap between complex algorithmic outputs and practical engineering applications. By leveraging methodologies that prioritize transparency and accuracy, engineers can develop reliable predictive models that effectively forecast outcomes while elucidating underlying decision-making processes. This dual focus enhances model applicability across engineering disciplines, particularly as machine learning (ML) and artificial intelligence (AI) increasingly integrate into engineering education and practice. Hybrid predictive models that combine interpretable and black-box approaches exemplify advancements in model development, facilitating a deeper understanding of results while aligning with the United Nations Sustainable Development Goals by preparing future engineers to tackle complex challenges in fields like environmental engineering [62, 9].

5 Case Studies and Applications

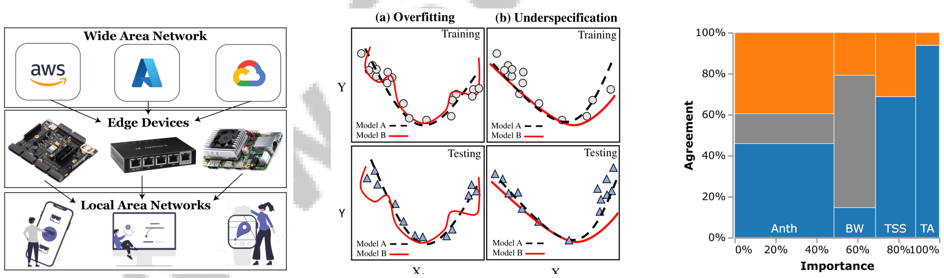
The integration of machine learning techniques in geotechnical engineering is increasingly recognized for its potential to enhance predictive accuracy and efficiency. This section delves into case studies that demonstrate the application of these methodologies in predicting shear strength parameters of granite residual soil. By examining real-world implementations, we highlight how machine learning not only improves predictive capabilities but also deepens our understanding of soil behavior across diverse contexts.

5.1 Case Studies and Real-World Applications

Case studies underscore the transformative role of machine learning in predicting shear strength parameters of granite residual soil, enhancing geotechnical engineering practices. Neural networks and hybrid models have notably advanced soil property predictions by incorporating variables such as sand content, saturation levels, and dry density, crucial for expansive soils [7]. This demonstrates machine learning's capability to capture complex interactions between soil components and environmental factors.

Iterative machine learning techniques refine predictive models by continuously incorporating new data, thus enhancing adaptability and accuracy, essential for managing granite residual soil's variability [8]. Convolutional neural networks (CNNs) effectively classify and predict intricate patterns within geotechnical data, establishing a robust framework for assessing soil stability and performance. These models improve predictive accuracy for soil characteristics and offer insights into optimizing soil improvement methodologies [54, 55, 25].

The development of glass-box models, balancing predictive performance with transparency, ensures machine learning predictions are both accurate and interpretable, crucial in geotechnical engineering [10]. These case studies illustrate machine learning's significant impact on geotechnical engineering, particularly in enhancing shear strength parameter predictions and overall reliability. Machine learning algorithms have effectively predicted various soil characteristics, often surpassing traditional methods, thus aiding civil and geotechnical engineers in optimizing soil properties through innovative methods, including green materials and amendments like lime and sand [54, 55, 53].



(a) Edge Computing and Edge Devices[62]

(b) Understanding Overfitting and Underspecification in Machine Learning Models[63]

(c) Importance vs. Agreement[64]

Figure 5: Examples of Case Studies and Real-World Applications

As shown in Figure 5, the examples offer diverse perspectives on practical applications of theoretical concepts. The first example highlights edge computing's role in enhancing data processing efficiency. The second example addresses overfitting and underspecification challenges in machine learning models, emphasizing the need for robust validation techniques. Lastly, the "Importance vs. Agreement" chart provides an analytical view of consensus factors, illustrating their relevance in real-world contexts [62, 63, 64].

5.2 Performance of Machine Learning Models in Geotechnical Applications

Evaluating machine learning models in geotechnical engineering involves assessing predictive accuracy and applicability, particularly concerning soil characteristics like compressive strength and stress-strain behavior. Studies show that algorithms such as neural networks and extreme gradient

Benchmark	Size	Domain	Task Format	Metric
GPT-4o[65]	1,000,000	Atmospheric Science	Climate Data Processing	Accuracy, F1-score
TAU-4[28]	1,244	Nuclear Physics	Half-life Measurement	T1/2, ASo
RBS-ANN[66]	500	Material Analysis	Spectra Analysis	Accuracy, Efficiency
GAM-Bench[15]	150,000	Predictive Analytics	Binary Classification	AUROC, RMSE
MOFSimplify[67]	5,311	Material Science	Stability Prediction	Accuracy, MAE
WCFZ-PB[68]	122	Mining Engineering	Regression	R2, MSE
RecBole[31]	41	Recommender Systems	Recommendation	Recall@10, MRR@10
ML-GPR[69]	3,472	Soil Analysis	Regression	Nugget-to-Sill Ratio, Mean Squared Error

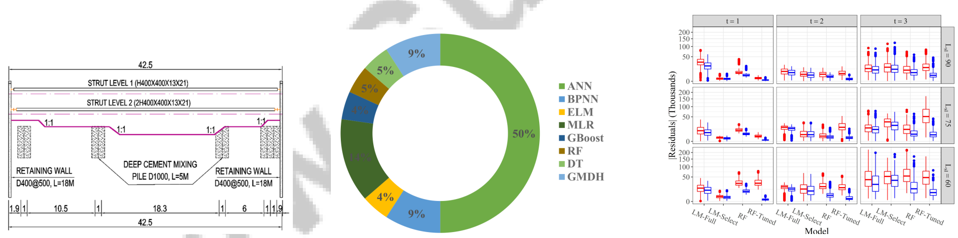
Table 5: This table presents a comprehensive overview of various benchmarks utilized in different scientific domains, including atmospheric science, nuclear physics, and material analysis. Each benchmark is characterized by its size, domain of application, task format, and performance metrics, highlighting the diverse methodologies and evaluation criteria employed in these fields.

boosting effectively model soil properties, providing valuable insights for engineers tackling complex challenges [54, 55, 53]. These models excel in capturing non-linear relationships and handling large datasets, critical for understanding diverse geotechnical properties.

Performance metrics like accuracy, precision, and recall are crucial for evaluating models' predictive capabilities across various soil conditions [23]. High accuracy ensures the safety and stability of engineering structures, making these evaluations essential. The adaptability of models to new data reflects their capability to generalize across different geotechnical scenarios [8].

Hybrid models, integrating machine learning with traditional methods, offer improved predictive performance and enhanced interpretability [45]. Advanced data processing techniques, such as feature extraction, enhance model performance by ensuring relevant data informs predictions [52]. Evaluating models also involves assessing robustness under varying environmental conditions, crucial for practical applicability in engineering projects [12].

Table 5 provides a detailed enumeration of benchmarks employed across multiple scientific disciplines, illustrating the scope and diversity of machine learning applications in geotechnical engineering and related fields.



(a) A structural diagram of a retaining wall and two structural levels with reinforcement details[25]

(b) Pie Chart of Machine Learning Algorithms[55]

(c) Box Plots of Residuals for Different Models at Different Time Points[63]

Figure 6: Examples of Performance of Machine Learning Models in Geotechnical Applications

Figure 6 illustrates the diverse applications and performance of machine learning models in geotechnical engineering through visualizations. The structural diagram of a retaining wall highlights reinforcement details crucial for stability. The pie chart categorizes machine learning algorithms, emphasizing their application in geotechnical studies. The box plots offer a comparative analysis of model performance, illustrating accuracy in predicting geotechnical outcomes [25, 55, 63].

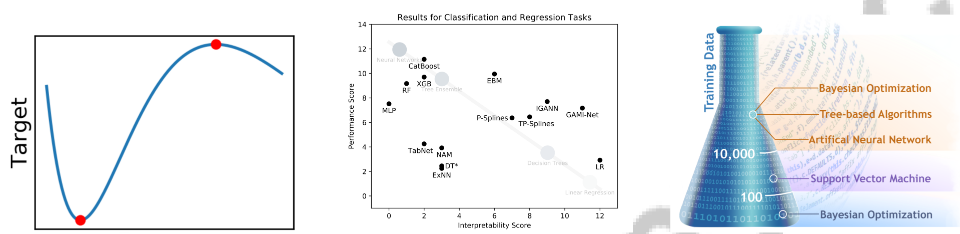
5.3 Validation of Machine Learning Techniques Using Experimental Data

Validating machine learning techniques in geotechnical applications ensures predictive models' reliability and accuracy. Studies demonstrate the effectiveness of algorithms in predicting key soil characteristics, emphasizing the integration of machine learning into geotechnical practices to enhance performance [60, 53, 54, 55, 27]. Experimental data benchmarks model performance, providing real-world context for assessing predictive capabilities. Incorporating experimental data identifies model strengths and weaknesses, facilitating algorithm refinement to capture complex geotechnical behaviors.

In geotechnical engineering, experimental data from laboratory tests and field observations offer insights into soil properties under various conditions, essential for calibrating models [23]. Comparing model predictions with experimental results evaluates accuracy and reliability, identifying areas for improvement.

Utilizing experimental data addresses challenges of data heterogeneity and variability in geotechnical applications. Diverse datasets capturing various soil conditions enable testing models for robustness and generalizability [8]. This ensures models are accurate and adaptable to dynamic geotechnical environments.

Experimental data also supports developing hybrid models combining machine learning with traditional methods, enhancing predictive accuracy and interpretability while maintaining empirical connections [45]. Validating hybrid models against experimental data underscores their practical applicability and effectiveness in geotechnical projects.



(a) The graph shows a curve with a target value at the top right corner.[70]

(b) Results for Classification and Regression Tasks[15]

(c) Bayesian Optimization and Machine Learning Algorithms[71]

Figure 7: Examples of Validation of Machine Learning Techniques Using Experimental Data

Figure 7 highlights diverse approaches for validating machine learning techniques using experimental data. The first graph represents a target value within a dynamic curve, crucial for understanding model outcomes. The second example compares performance and interpretability scores, addressing trade-offs in model deployment. The third example integrates Bayesian optimization with machine learning, emphasizing optimization techniques in refining model performance. These examples provide a comprehensive overview of leveraging experimental data to validate and enhance machine learning methodologies [70, 15, 71].

6 Challenges and Future Directions

6.1 Current Challenges in Machine Learning Applications

Machine learning applications in geotechnical engineering face several critical challenges that hinder their full potential. A significant issue is the dependence on high-quality, diverse datasets, which are vital for model accuracy, particularly in complex deformation scenarios [5]. The scarcity of comprehensive datasets for precise post-buyout land use assessments often leads to model overfitting, limiting applicability across various geotechnical contexts [11].

Moreover, the computational complexity of managing large datasets presents a substantial hurdle. Current techniques often struggle to efficiently process and reduce data while maintaining essential features during in situ analysis [35]. This complexity can impede result interpretability from sophisticated models, a crucial aspect in geotechnical engineering [6]. Additionally, reliance on gradients at saturated points in neural networks might lead to misleading attributions, complicating the reliability of model predictions [20].

Variability in soil composition, additive dosages, and curing conditions further complicate the consistency of machine learning model outcomes, highlighting the need for robust data collection and processing strategies [24]. The absence of reliable prediction intervals that accurately convey uncertainty, especially for Random Forest models, is a significant obstacle to model robustness [13].

Latency constraints and inefficiencies, such as the need for separate counterfactual searches for each input, exacerbate the complexity of applying machine learning in this field [18]. Addressing these challenges requires improved data collection techniques, expanded benchmark datasets, and more adaptable and efficient machine learning models. Proposals to minimize computational redundancy offer significant performance improvements as data characteristics evolve, crucial for advancing machine learning applications in geotechnical engineering [45].

6.2 Innovative Approaches in Predictive Modeling

Advancements in predictive modeling for geotechnical applications are vital for improving the accuracy and applicability of machine learning techniques. Refining regularization techniques to enhance model performance across varied geotechnical contexts is a promising direction. Incorporating user feedback can make models more interpretable, increasing their utility in practical engineering scenarios [56].

Active learning approaches focus on enhancing sampling methods near singularities, significantly improving the model's capacity to capture critical geotechnical features and resulting in more reliable predictions. Extending these methods to encompass a broader range of functions is a future research aim, enhancing their versatility and effectiveness in geotechnical applications [52].

Hybrid modeling techniques, such as those integrating Bayesian smoothing surfaces, provide a robust framework for managing irregularly spaced data and other linear features frequently encountered in geotechnical datasets. These models combine probabilistic and deterministic approaches, offering comprehensive solutions for complex geotechnical challenges [72]. By accommodating unique characteristics of geotechnical data, these innovative methods promise to enhance the predictive capabilities and practical applicability of machine learning models in the field.

6.3 Advancements in Algorithms and Techniques

Recent advancements in algorithms and techniques have significantly bolstered machine learning applications in geotechnical engineering, particularly concerning granite residual soil. The Inline Coordinates method offers a lossless data representation, enhancing the interpretability and efficiency of machine learning models [36]. This method enables high-dimensional geotechnical data visualization in two-dimensional space without sacrificing critical information, facilitating better understanding and analysis of complex datasets.

The dynamic learning principle of AICA allows machine learning models to adapt to shifts in data distribution [73]. This adaptability is crucial in geotechnical engineering, where soil properties and environmental conditions can vary significantly. By enabling models to dynamically adjust to new data, AICA enhances the robustness and accuracy of predictions, ensuring models remain relevant across diverse geotechnical scenarios.

The integration of Residual in Residual Dense Blocks within the autoencoder architecture represents a key innovation in feature learning and data reduction capabilities [35]. This approach markedly outperforms traditional methods by improving the model's ability to capture intricate patterns and features in geotechnical data. Enhanced feature extraction and data reduction capabilities are particularly beneficial for managing the complexity and variability inherent in soil behavior, leading to more accurate and reliable predictions.

Recent advancements in machine learning algorithms and techniques highlight their significant potential to transform geotechnical engineering by enabling the development of highly accurate, interpretable, and adaptable predictive models. These models can effectively forecast various soil properties, such as compressive strength and bearing capacity, while also addressing complex behaviors of soils like expansive types. Furthermore, integrating machine learning with experimental data enhances prediction precision and facilitates the optimization of soil treatment methods, thereby supporting civil and geotechnical engineers in making informed decisions and improving the sustainability of soil improvement practices [60, 53, 54, 55, 63]. As the field continues to evolve, integrating these innovative approaches promises to enhance the predictive capabilities of machine learning applications, ultimately contributing to safer and more efficient engineering practices.

6.4 Future Research Directions

Future research in machine learning for geotechnical applications should prioritize exploring hybrid models that effectively combine statistical inference with machine learning flexibility, thereby enhancing predictive capabilities and adaptability to diverse geotechnical scenarios [12]. Integrating advanced machine learning techniques for automating data processing could significantly improve the efficiency and accuracy of geotechnical assessments [7]. Moreover, research should focus on incorporating missing values into predictive models to enhance the coverage and accuracy of prediction intervals, critical for reliable geotechnical engineering applications [13].

The development of iterative machine learning (IML) techniques should also be emphasized, particularly in exploring alternative kernel functions that could improve model accuracy and inform advancements in geotechnical modeling [8]. Additionally, extending current methodologies to other deep learning architectures and exploring further optimizations could foster the development of more user-friendly implementations of machine learning algorithms, facilitating broader adoption in the field [45].

Future research should also investigate the long-term performance evaluations of treated soils, particularly in optimizing application parameters for a broader range of soil types and exploring the use of additional waste materials [24]. This inquiry is essential for developing sustainable geotechnical solutions that align with environmental and economic considerations.

Furthermore, exploring techniques to enhance the actionability of machine learning models, including applying CounteRGAN to new domains and investigating mutable and immutable features on counterfactual generation, represents a promising research avenue [18]. These efforts will contribute to developing more interpretable and actionable models, crucial for practical engineering applications.

7 Conclusion

The integration of machine learning into geotechnical engineering marks a significant advancement in predicting the shear strength parameters of granite residual soil. This survey underscores the superiority of machine learning techniques over traditional methods, particularly in estimating the unconfined compressive strength and elasticity modulus of expansive soils. These advancements lay the groundwork for more precise geotechnical evaluations. The adaptability of approaches like AICA highlights the potential for enhanced model accuracy and interpretability. The collaboration between machine learning and conventional geotechnical practices is pivotal in improving both interpretability and practical applicability. This is especially crucial in fields such as architectural heritage and geotechnical engineering, where understanding the interplay of complex variables is essential for making reliable predictions. Although smart sensors have progressed in compaction monitoring, their robustness still requires refinement. Machine learning offers a promising avenue to augment traditional methods, though extensive field validation remains necessary to ensure their dependability.

References

[1] Jiahang Li, Hitoshi Mikada, and Junichi Takekawa. Improved full-waveform inversion for seismic data in the presence of noise based on the k-support norm, 2023.

[2] Jonathan E. Kollmer and Karen E. Daniels. An experimental investigation of the force network ensemble, 2017.

[3] Tarek Iraki, Lukas Morand, Norbert Link, Stefan Sandfeld, and Dirk Helm. Accurate distances measures and machine learning of the texture-property relation for crystallographic textures represented by one-point statistics, 2023.

[4] Talia Crawford, Scott Duong, Richard Fueston, Ayorinde Lawani, Samuel Owoade, Abel Uzoka, Reza M. Parizi, and Abbas Yazdinejad. Ai in software engineering: A survey on project management applications, 2023.

[5] Kshitiz Upadhyay, Jan N. Fuhg, Nikolaos Bouklas, and K. T. Ramesh. Physics-informed data-driven discovery of constitutive models with application to strain-rate-sensitive soft materials, 2023.

[6] Yuqing He, Yousef Heider, and Bernd Markert. Embedding an ann-based crystal plasticity model into the finite element framework using an abaqus user-material subroutine, 2024.

[7] Meryam Chaieb and Mohamed Aymen Saied. Automate migration to microservices architecture using machine learning techniques, 2023.

[8] Santosh Devasia. Iterative machine learning for output tracking, 2017.

[9] Tong Wang and Qihang Lin. Hybrid predictive model: When an interpretable model collaborates with a black-box model, 2019.

[10] Zeynep Tuna Deger and Gulsen Taskin Kaya. Glass-box model representation of seismic failure mode prediction for conventional rc shear walls, 2021.

[11] Hakan T. Otal, Elyse Zavar, Sherri B. Binder, Alex Greer, and M. Abdullah Canbaz. Harnessing deep learning and satellite imagery for post-buyout land cover mapping, 2024.

[12] Stephanie Clark, Rob J Hyndman, Dan Pagendam, and Louise M Ryan. Modern strategies for time series regression, 2020.

[13] Burim Ramosaj. Interpretable machines: Constructing valid prediction intervals with random forests, 2021.

[14] Dylan Slack, Sorelle A. Friedler, Carlos Scheidegger, and Chitradeep Dutta Roy. Assessing the local interpretability of machine learning models, 2019.

[15] Sven Kruschel, Nico Hambauer, Sven Weinzierl, Sandra Zilker, Mathias Kraus, and Patrick Zschech. Challenging the performance-interpretability trade-off: An evaluation of interpretable machine learning models, 2024.

[16] John Mitros and Brian Mac Namee. A categorisation of post-hoc explanations for predictive models, 2019.

[17] Randy Harsuko and Tariq Alkhalifah. Storseismic: A new paradigm in deep learning for seismic processing, 2022.

[18] Daniel Nemirovsky, Nicolas Thiebaut, Ye Xu, and Abhishek Gupta. Countergan: Generating realistic counterfactuals with residual generative adversarial nets, 2021.

[19] Anton Persson, Niklas Dymne, and Fernando Alonso-Fernandez. Classification of ps and abs black plastics for weee recycling applications, 2021.

[20] Vivek Miglani, Narine Kokhlikyan, Bilal Alsallakh, Miguel Martin, and Orion Reblitz-Richardson. Investigating saturation effects in integrated gradients, 2020.

[21] Xiaolong Li, Federica B. Bianco, Gregory Dobler, Roee Partoush, Armin Rest, Tatiana Acero-Cuellar, Riley Clarke, Willow Fox Fortino, Somayeh Khakpash, and Ming Lian. Toward automated detection of light echoes in synoptic surveys: considerations on the application of the deep convolutional neural networks, 2022.

[22] Achmad Choiruddin, Aisah, Finola Trisnisa, and Nur Iriawan. Quantifying effect of geological factor on distribution of earthquake occurrences by inhomogeneous cox processes, 2021.

[23] Yudong Liang, Bin Wang, Jiaying Liu, Deyu Li, Yuhua Qian, and Wenqi Ren. Progressive residual learning for single image dehazing, 2021.

[24] Ghania Boukhatem, Messaouda Bencheikh, Mohammed Benzerara, SM Anas, Mohanad Muayad Sabri, and Hadee Mohammad Najm. Optimizing properties of clayey soil using lime and waste marble powder: a sustainable approach for engineering applications. *Frontiers in Materials*, 11:1392875, 2024.

[25] DinhHieu Tran, HongGiang Nguyen, YuRen Wang, KhacHai Phan, ThiTuyetNga Phu, Duy-Phuong Le, and TienThinh Nguyen. Analysis of artificial intelligence approaches to predict the wall deflection induced by deep excavation. *Open Geosciences*, 15(1):20220503, 2023.

[26] Yassine El Khessaimi, Youssef El Hafiane, Claire Peyratout, Karim Tamine, Samir Adly, Moulay Barkatou, and Agnès Smith. Towards accelerating the development of calcined clay cements: data-driven prediction of compressive strength exploiting machine learning algorithms. 2023.

[27] Lateef Bankole Adamolekun, Muyideen Alade Saliu, Abiodun Ismail Lawal, and Ismail Adeniyi Okewale. Investigating the competency of some selected soft computing techniques for modeling of lateritic soil strength based on index properties. 2024.

[28] E. N. Alexeev, Yu. M. Gavrilyuk, A. M. Gangapshev, A. M. Gezhaev, V. V. Kazalov, V. V. Kuzminov, S. I. Panasenko, O. D. Petrenko, and S. S. Ratkevich. Tau-4 installation intended for long-term monitoring of a half-life value of the ^{212}po, 2018.

[29] Ryoichiro Agata. Quantification of uncertainty and its propagation in seismic velocity structure and earthquake source inversion, 2025.

[30] Junho Jo, Eunseong Choi, Minseon Kim, and Kyoungmin Min. Machine learning aided materials design platform for predicting the mechanical properties of na-ion solid-state electrolytes, 2021.

[31] Lanling Xu, Zhen Tian, Gaowei Zhang, Lei Wang, Junjie Zhang, Bowen Zheng, Yifan Li, Yupeng Hou, Xingyu Pan, Yushuo Chen, Wayne Xin Zhao, Xu Chen, and Ji-Rong Wen. Recent advances in recbole: Extensions with more practical considerations, 2022.

[32] Xing Zhao, Ping Lu, Yanyan Zhang, Jianxiong Chen, and Xiaoyang Li. Attenuating random noise in seismic data by a deep learning approach, 2019.

[33] Matteo Ravasi. Deep preconditioners and their application to seismic wavefield processing, 2022.

[34] Usman Syed, Federico Cunico, Uzair Khan, Eros Radicchi, Francesco Setti, Adolfo Speghini, Paolo Marone, Filiberto Semenzin, and Marco Cristani. Material synthesis through simulations guided by machine learning: a position paper, 2024.

[35] Qun Liu, Subhashis Hazarika, John M. Patchett, James Paul Ahrens, and Ayan Biswas. Deep learning-based feature-aware data modeling for complex physics simulations, 2019.

[36] Boris Kovalerchuk and Hoang Phan. Full interpretable machine learning in 2d with inline coordinates, 2021.

[37] Nan Xu, Hassan Askari, Yanjun Huang, Jianfeng Zhou, and Amir Khajepour. Tire force estimation in intelligent tires using machine learning, 2020.

[38] S. E. Lauro, G. Gennarelli, E. Pettinelli, F. Soldovieri, F. Cantini, A. P. Rossi, and R. Orosei. A strategy for an accurate estimation of the basal permittivity in the martian north polar layered deposits, 2016.

[39] Polina Lemenkova. Factor analysis by r programming to assess variability among environmental determinants of the mariana trench, 2018.

[40] Matteo Ravasi and Nick Luiken. Why pocs works, and how to make it better, 2023.

[41] N. Castin, P. Klups, M. Konstantinovic, G. Bonny, M. I. Pascuet, M. Moody, and L. Malerba. How precisely are solute clusters in rpv steels characterized by atom probe experiments?, 2024.

[42] Francesco Salmaso, Mirko Trisolini, and Camilla Colombo. A machine learning and feature engineering approach for the prediction of the uncontrolled re-entry of space objects, 2023.

[43] Conrad M Albrecht, Fernando Marianno, and Levente J Klein. Autogeolabel: Automated label generation for geospatial machine learning, 2022.

[44] Yongjin Choi and Krishna Kumar. Graph neural network-based surrogate model for granular flows, 2023.

[45] Zhaoyue Chen, Nick Koudas, Zhe Zhang, and Xiaohui Yu. Efficient construction of nonlinear models over normalized data, 2021.

[46] Aurélie Fischer, Lucie Montuelle, Mathilde Mougeot, and Dominique Picard. Statistical learning for wind power : a modeling and stability study towards forecasting, 2018.

[47] Rami N Khushaba, Arman Melkumyan, and Andrew J Hill. A machine learning approach for material type logging and chemical assaying from autonomous measure-while-drilling (mwd) data, 2022.

[48] Ankit Gupta, Jinfeng Du, Dmitry Chizhik, Reinaldo A. Valenzuela, and Mathini Sellathurai. Machine learning-based urban canyon path loss prediction using 28 ghz manhattan measurements, 2022.

[49] Simon Schnürch and Andreas Wagner. Machine learning on epex order books: Insights and forecasts, 2019.

[50] Dorian Florescu and Matthew England. A machine learning based software pipeline to pick the variable ordering for algorithms with polynomial inputs, 2020.

[51] Palash Panja, Pranay Asai, Raul Velasco, and Milind Deo. Data processing protocol for regression of geothermal times series with uneven intervals, 2019.

[52] Edgar A. Bernal, Jonathan D. Hauenstein, Dhagash Mehta, Margaret H. Regan, and Tingting Tang. Machine learning the real discriminant locus, 2022.

[53] Xinghuang Guo, Cesar Garcia, Alexis Ivan Andrade Valle, Kennedy Onyelowe, Andrea Natali Zarate Villacres, Ahmed M Ebid, and Shadi Hanandeh. Modeling the influence of lime on the unconfined compressive strength of reconstituted graded soil using advanced machine learning approaches for subgrade and liner applications. *Plos one*, 19(4):e0301075, 2024.

[54] Ammar Alnmr, Haidar Hosamo Hosamo, Chuangxin Lyu, Richard Paul Ray, and Mounzer Omran Alzawi. Novel insights in soil mechanics: integrating experimental investigation with machine learning for unconfined compression parameter prediction of expansive soil. *Applied Sciences*, 14(11):4819, 2024.

[55] Ahmed Hassan Saad, Haslinda Nahazanan, Badronnisa Yusuf, Siti Fauziah Toha, Ahmed Alnuaim, Ahmed El-Mouchi, Mohamed Elseknidy, and Angham Ali Mohammed. A systematic review of machine learning techniques and applications in soil improvement using green materials. *Sustainability*, 15(12):9738, 2023.

[56] Andrew Slavin Ross. Training machine learning models by regularizing their explanations, 2018.

[57] Hao Liu, Yan Xu, Jiasheng Wang, Ju Jing, Chang Liu, Jason T. L. Wang, and Haimin Wang. Inferring vector magnetic fields from stokes profiles of gst/niris using a convolutional neural network, 2020.

[58] Yimin Dou, Kewen Li, Hongjie Duan, Timing Li, Lin Dong, and Zongchao Huang. Mda gan: Adversarial-learning-based 3-d seismic data interpolation and reconstruction for complex missing, 2023.

[59] Pablo M. Barros, Roosevelt de L. Sardinha, Giovanny A. M. Arboleda, Lessandro de S. S. Valente, Isabelle R. V. de Melo, Albino Aveleda, André Bulcão, Sergio L. Netto, and Alexandre G. Evsukoff. A real benchmark swell noise dataset for performing seismic data denoising via deep learning, 2024.

[60] Arnaud Mignan. Text classification of the precursory accelerating seismicity corpus: Inference on some theoretical trends in earthquake predictability research from 1988 to 2018, 2018.

[61] Dewayne Whitfield. Using gpt-2 to create synthetic data to improve the prediction performance of nlp machine learning classification models, 2021.

[62] Andrew Schulz, Suzanne Stathatos, Cassandra Shriver, and Roxanne Moore. Utilizing online and open-source machine learning toolkits to leverage the future of sustainable engineering, 2023.

[63] Mohsen Zaker Esteghamati, Brennan Bean, Henry V. Burton, and M. Z. Naser. Beyond development: Challenges in deploying machine learning models for structural engineering applications, 2024.

[64] Diego Rojo, Nyi Nyi Htun, Denis Parra, Robin De Croon, and Katrien Verbert. Ahmose: A knowledge-based visual support system for selecting regression machine learning models, 2021.

[65] Lujia Zhang, Hanzhe Cui, Yurong Song, Chenyue Li, Binhang Yuan, and Mengqian Lu. On the opportunities of (re)-exploring atmospheric science by foundation models: A case study, 2024.

[66] Renato da S. Guimarães, Tiago F. Silva, Cleber L. Rodrigues, Manfredo H. Tabacniks, Simon Bach, Vassily V. Burwitz, Paul Hiret, and Matej Mayer. Processing of massive rutherford back-scattering spectrometry data by artificial neural networks, 2022.

[67] A. Nandy, G. Terrones, N. Arunachalam, C. Duan, D. W. Kastner, and H. J. Kulik. Mofsimplify: Machine learning models with extracted stability data of three thousand metal-organic frameworks, 2021.

[68] Zhengyu Wu, Ying Chen, and Dayou Luo. Comparative study of five machine learning algorithms on prediction of the height of the water-conducting fractured zone in undersea mining. *Scientific Reports*, 14(1):21047, 2024.

[69] Chunlei Xu, Michael Pregesbauer, Naga Sravani Chilukuri, Daniel Windhager, Mahsa Yousefi, Pedro Julian, and Lothar Ratschbacher. Soil analysis with machine-learning-based processing of stepped-frequency gpr field measurements: Preliminary study, 2024.

[70] Tiago Botari, Rafael Izbicki, and Andre C. P. L. F. de Carvalho. Local interpretation methods to machine learning using the domain of the feature space, 2019.

[71] Sterling G. Baird, Marianne Liu, Hasan M. Sayeed, and Taylor D. Sparks. Data-driven materials discovery and synthesis using machine learning methods, 2022.

[72] Matthew Hofkes and Douglas Nychka. Hybrid bayesian smoothing on surfaces, 2024.

[73] Suze Nei Pereira Guiamarães and Valiya Mannathal Hamza. Avanços na caracterização das estruturas geológicas em subsuperfície da provincia uranífera lagoa real (ba) a partir de dados aerogeofísicos, 2011.

Disclaimer: