



Explainable AI (XAI) for Soil Prediction Interpretability: Enhancing Trust and Understanding in Digital Soil Mapping Through Model Transparency

Dr. Harish Chandra ^{1*}, Deepika Rawat ²

¹ Associate Professor, Department of Computer Applications, Banaras Hindu University, Varanasi, India

² Assistant Professor, Department of Computer Applications, Banaras Hindu University, Varanasi, India

* Corresponding Author: Dr. Harish Chandra

Article Info

P-ISSN: 3051-3448

E-ISSN: 3051-3456

Volume: 06

Issue: 01

January - June 2025

Received: 09-01-2025

Accepted: 12-02-2025

Published: 19-03-2025

Page No: 19-25

Abstract

The widespread adoption of artificial intelligence in digital soil mapping has achieved remarkable predictive accuracy, but the "black box" nature of complex models limits their acceptance among soil scientists, agronomists, and land managers who require understanding of prediction rationale for decision-making. This study presents a comprehensive evaluation of Explainable AI (XAI) techniques for enhancing interpretability in soil property prediction models while maintaining predictive performance. We implemented and compared five XAI approaches: SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), Integrated Gradients, Attention Mechanisms, and a novel Soil-Specific Explanation Framework (SSEF) across 18,934 soil samples from diverse agricultural landscapes in North America and Europe. The models predicted soil organic carbon (SOC), pH, clay content, and available nitrogen using 68 environmental covariates including climate, topography, remote sensing, and geological variables. Model architectures included Random Forest, XGBoost, Neural Networks, and Transformer-based approaches to evaluate XAI effectiveness across different algorithm types. The SSEF approach achieved superior explanation quality with fidelity scores of 0.94 for SOC, 0.91 for pH, 0.89 for clay content, and 0.87 for nitrogen predictions. Feature importance rankings showed high consistency (Spearman correlation >0.85) across different XAI methods, with climate variables (precipitation, temperature) and topographic indices (elevation, slope) emerging as primary predictors. SHAP analysis revealed non-linear relationships and interaction effects previously undetected through traditional statistical approaches, including threshold effects of precipitation on organic carbon accumulation and complex terrain-climate interactions affecting soil pH. Local explanations through LIME successfully identified region-specific prediction drivers, with coastal areas prioritizing salinity-related variables and mountainous regions emphasizing elevation and slope factors. Attention mechanism visualization in transformer models revealed spatial patterns in feature importance that aligned with known soil formation processes. User evaluation studies with 47 soil scientists and agronomists demonstrated significant improvements in model trust (78% increase), decision confidence (65% increase), and practical adoption intentions (82% increase) when XAI explanations were provided. Computational overhead for explanation generation averaged 12-18% of prediction time, making real-time interpretability feasible for operational applications. The study establishes practical frameworks for implementing explainable AI in soil science applications, bridging the gap between predictive accuracy and scientific understanding while supporting evidence-based agricultural decision-making.

Keywords: Explainable Artificial Intelligence, Soil Prediction, Model Interpretability, Shap, Lime, Digital Soil Mapping, Feature Importance, Model Transparency, Trust in AI

Introduction

The integration of artificial intelligence in digital soil mapping has transformed our ability to predict soil properties at unprecedented scales and resolutions, enabling precision agriculture and sustainable land management practices ^[1]. However, the increasing complexity of AI models, particularly deep learning architectures, has created a significant barrier to adoption among soil scientists and agricultural practitioners who require understanding of model predictions to make informed decisions ^[2]. The "black box" nature of complex models undermines trust and limits their practical utility in domains where prediction rationale is as important as prediction accuracy.

Explainable AI (XAI) has emerged as a critical research area addressing the interpretability challenge in machine learning applications [3]. XAI encompasses techniques and methodologies designed to make AI model decisions transparent, interpretable, and trustworthy while maintaining predictive performance. For soil science applications, explainability is particularly crucial because soil formation processes are complex, spatially variable, and influenced by multiple interacting environmental factors that practitioners need to understand for effective management.

The need for interpretable soil prediction models extends beyond scientific curiosity to practical necessity. Agronomists making fertilizer recommendations need to understand why specific nutrient levels are predicted for particular fields. Environmental consultants assessing soil contamination require insight into which factors drive pollution predictions. Policy makers developing land use regulations need transparent evidence supporting soil quality assessments [4]. Without interpretable models, these critical decisions rely on blind trust in algorithmic predictions.

Traditional statistical approaches in soil science, such as multiple linear regression and geostatistical methods, provide inherent interpretability through parameter coefficients and variogram structures [5]. However, these methods often fail to capture the complex non-linear relationships and high-dimensional interactions characteristic of soil-environment systems. Modern machine learning approaches like ensemble methods and neural networks achieve superior predictive accuracy but sacrifice interpretability, creating a fundamental trade-off between performance and understanding.

The challenge of model interpretability in soil science is compounded by the heterogeneous nature of soil data and the diverse backgrounds of end-users. Soil scientists require explanations that align with pedological theory and soil formation factors (climate, organisms, relief, parent material, time). Agricultural practitioners need interpretations focused on management-relevant variables like nutrient availability and water retention. Environmental managers prioritize explanations related to contamination sources and remediation potential [6].

SHAP (SHapley Additive exPlanations) represents one of the most theoretically grounded approaches to model explanation, based on cooperative game theory principles [7]. SHAP provides consistent and efficient explanations by fairly attributing prediction contributions to individual features while satisfying mathematical axioms for explanation quality. The method works across different model types and provides both global and local explanations, making it versatile for diverse soil science applications.

LIME (Local Interpretable Model-agnostic Explanations) offers an alternative approach focused on explaining individual predictions through local linear approximations [8]. By perturbing input features and observing prediction changes, LIME generates interpretable explanations for specific instances. This approach is particularly valuable for understanding why models make specific predictions for individual soil samples or locations.

Integrated Gradients provides explanation methods specifically designed for neural networks, computing feature importance through integration of gradients along paths from baseline to input [9]. This approach addresses some limitations of simpler gradient-based methods and provides more stable explanations for deep learning models increasingly used in soil science applications.

Attention mechanisms in transformer-based models offer built-in interpretability by revealing which input features the model focuses on when making predictions [10]. For spatial applications like soil mapping, attention weights can reveal geographical patterns in feature importance that may align with known soil formation processes and landscape relationships.

The evaluation of explanation quality presents significant challenges, as there is often no ground truth for what constitutes a "correct" explanation. Fidelity measures assess how well explanations reflect actual model behavior, while stability evaluates consistency of explanations across similar inputs. User studies provide crucial validation by measuring whether explanations actually improve human understanding and decision-making [11].

The computational efficiency of explanation methods is critical for operational deployment in soil mapping applications. Real-time explanations enable interactive exploration of model predictions, while batch explanation generation supports systematic analysis of large-scale soil surveys. The trade-off between explanation quality and computational cost must be carefully balanced for practical implementation [12].

This study aims to develop and evaluate comprehensive XAI frameworks specifically designed for soil prediction applications, addressing both technical challenges and user needs in digital soil mapping. Specific objectives include: (1) comparing different XAI approaches for soil property prediction models, (2) evaluating explanation quality and consistency across different model types, (3) analyzing feature importance patterns and their alignment with soil science knowledge, (4) assessing user perception and trust in explanation-enhanced models, and (5) developing practical guidelines for implementing XAI in operational soil mapping systems [13].

The integration of XAI in soil science represents a paradigm shift toward transparent and trustworthy AI applications that combine predictive power with scientific understanding. This approach supports evidence-based decision-making while building user confidence in AI-assisted soil management and agricultural planning [14].

Materials and Methods

Study Areas and Data Collection

The research was conducted across six representative agricultural regions to ensure comprehensive evaluation of XAI methods under diverse environmental conditions. Study areas included: Corn Belt agriculture in Iowa, USA (42°15'N, 93°45'W), Prairie agriculture in Saskatchewan, Canada (52°30'N, 106°30'W), Mediterranean agriculture in Tuscany, Italy (43°30'N, 11°15'W), Subtropical agriculture in São Paulo, Brazil (23°30'S, 46°45'W), Temperate agriculture in Norfolk, England (52°45'N, 1°15'E), and Semi-arid agriculture in New South Wales, Australia (32°30'S, 147°30'E).

A total of 18,934 soil samples were collected using stratified random sampling design stratified by soil type, land use, and topographic position. Sampling density averaged 1.2 samples per km² with higher density in heterogeneous areas. Samples were collected from 0-20 cm depth during standardized conditions to minimize temporal variability.

Quality control procedures included duplicate sampling (10% of locations), split sample analysis, and inter-laboratory

calibration exercises to ensure data consistency across study regions.

Laboratory Analysis and Target Variables

Four soil properties were selected as prediction targets based on their importance for agricultural management: soil organic carbon (SOC) determined by dry combustion, pH measured in 1:2 soil-water suspension, clay content through laser diffraction particle size analysis, and available nitrogen via alkaline permanganate extraction.

All analyses followed standardized international protocols with quality assurance including certified reference materials, blank samples, and replicate analyses. Measurement uncertainties were quantified and incorporated into model uncertainty estimates.

Environmental Covariates and Feature Engineering

A comprehensive suite of 68 environmental covariates was compiled representing the five soil formation factors: climate (19 variables), organisms/vegetation (15 variables), relief/topography (18 variables), parent material/geology (8 variables), and time/age (8 variables).

Climate variables included temperature and precipitation statistics from WorldClim database, aridity indices, and growing degree days. Vegetation variables comprised NDVI time series, phenology metrics, and land cover classifications from MODIS and Landsat imagery. Topographic variables were derived from 30-meter SRTM DEM including elevation, slope, aspect, curvature, wetness index, and flow accumulation.

Geological variables incorporated parent material types, lithology, and age from global geological databases. Temporal variables included land use history, cultivation duration, and management intensity indices.

Feature engineering included multi-scale analysis through focal statistics at 100m, 500m, and 1km windows, temporal aggregation of vegetation indices, and derivation of climate-topography interaction terms.

Machine Learning Model Development

Four model architectures were implemented to evaluate XAI effectiveness across different algorithm types:

- **Random Forest (RF):** Ensemble of 500 decision trees with maximum depth of 15 and minimum samples per leaf of 5. Feature importance calculated through mean decrease in impurity and permutation importance.
- **XGBoost (XGB):** Gradient boosting with 1000 estimators, learning rate of 0.1, and maximum depth of 8. Early stopping based on validation loss with 50-round patience.
- **Neural Networks (NN):** Multi-layer perceptrons with 3-4 hidden layers, 128-256 neurons per layer, ReLU activation, and dropout regularization (0.2-0.3).
- **Transformer Models (TF):** Vision transformer architecture adapted for tabular data with self-attention mechanisms and positional encodings for spatial features.

All models employed 5-fold cross-validation with spatial blocking to account for spatial autocorrelation. Hyperparameter optimization used Bayesian optimization with 100 iterations per model type.

Explainable AI Implementation

Five XAI approaches were implemented and evaluated:

- **SHAP (SHapley Additive exPlanations):** Tree Explainer for tree-based models and Kernel Explainer for neural networks. Shapley values calculated for all features and samples, providing additive feature attribution with theoretical guarantees.
- **LIME (Local Interpretable Model-agnostic Explanations):** Local linear models fitted around individual predictions using perturbation-based sampling. Feature selection employed Lasso regularization with cross-validation for optimal sparsity.
- **Integrated Gradients (IG):** Gradient integration along linear paths from baseline to input features for neural network models. Baseline values set to feature means with 50 integration steps for stable attribution.
- **Attention Mechanisms (AM):** Native attention weights from transformer models visualized to show feature importance patterns. Multi-head attention averaged across heads and layers for comprehensive attribution.
- **Soil-Specific Explanation Framework (SSEF):** Novel approach incorporating pedological knowledge through hierarchical explanations organized by soil formation factors. Physical constraints ensure explanations align with soil science principles.

Explanation Quality Assessment

Multiple metrics evaluated explanation quality and consistency:

- **Fidelity:** Correlation between explanation scores and actual feature contributions measured through feature ablation studies. Higher fidelity indicates explanations accurately reflect model behavior.
- **Stability:** Consistency of explanations for similar inputs assessed through perturbation analysis. Stable explanations should remain consistent for minor input changes.
- **Consistency:** Agreement between different XAI methods measured through rank correlation of feature importance scores. High consistency suggests robust explanation quality.
- **Plausibility:** Expert evaluation of explanation alignment with soil science knowledge conducted through structured interviews with domain experts.

User Evaluation and Trust Assessment

Comprehensive user studies evaluated explanation effectiveness with 47 participants including soil scientists (n=18), agronomists (n=16), and land managers (n=13). Participants evaluated model predictions with and without explanations across multiple scenarios representing typical decision-making contexts.

Trust assessment employed validated psychological scales measuring cognitive trust, emotional trust, and behavioral intentions. Decision confidence was measured through self-reported certainty scores and decision time analysis.

Explanation preference surveys identified optimal explanation formats and information content for different user groups and application contexts.

Computational Efficiency Analysis

Computational overhead for explanation generation was measured across different methods and model types. Metrics included explanation generation time, memory usage, and scalability with dataset size and feature dimensionality. Optimization strategies were evaluated including approximation methods, parallel processing, and caching mechanisms to enable real-time explanation generation for interactive applications.

Results

Model Performance and XAI Method Comparison

All machine learning models achieved high predictive accuracy, with R² values ranging from 0.83-0.91 across soil properties (Table 1). The Transformer model achieved highest accuracy for most properties, while Random Forest provided good performance with inherent interpretability. XAI method evaluation revealed significant differences in explanation quality across approaches.

Table 1: Model performance and XAI method comparison across soil properties

Model Type	Soil Organic Carbon			Soil pH			Clay Content			Available Nitrogen		
	R ²	RMSE	Fidelity	R ²	RMSE	Fidelity	R ²	RMSE	Fidelity	R ²	RMSE	Fidelity
Random Forest												
SHAP	0.87	0.64	0.92	0.83	0.35	0.89	0.85	3.2	0.90	0.79	15.2	0.88
LIME	0.87	0.64	0.88	0.83	0.35	0.85	0.85	3.2	0.87	0.79	15.2	0.84
SSEF	0.87	0.64	0.94	0.83	0.35	0.91	0.85	3.2	0.89	0.79	15.2	0.87
XGBoost												
SHAP	0.89	0.59	0.93	0.85	0.33	0.90	0.87	3.0	0.91	0.81	14.5	0.89
LIME	0.89	0.59	0.86	0.85	0.33	0.83	0.87	3.0	0.85	0.81	14.5	0.82
SSEF	0.89	0.59	0.94	0.85	0.33	0.91	0.87	3.0	0.89	0.81	14.5	0.87
Neural Networks												
SHAP	0.88	0.61	0.90	0.84	0.34	0.87	0.86	3.1	0.88	0.80	14.8	0.86
LIME	0.88	0.61	0.85	0.84	0.34	0.82	0.86	3.1	0.84	0.80	14.8	0.81
Integrated Gradients	0.88	0.61	0.89	0.84	0.34	0.86	0.86	3.1	0.87	0.80	14.8	0.85
SSEF	0.88	0.61	0.92	0.84	0.34	0.89	0.86	3.1	0.87	0.80	14.8	0.85
Transformer												
Attention Mechanisms	0.91	0.53	0.88	0.87	0.30	0.85	0.89	2.7	0.86	0.83	13.6	0.84
SHAP	0.91	0.53	0.91	0.87	0.30	0.88	0.89	2.7	0.89	0.83	13.6	0.87
SSEF	0.91	0.53	0.94	0.87	0.30	0.91	0.89	2.7	0.89	0.83	13.6	0.87

The SSEF approach consistently achieved highest fidelity scores (0.87-0.94) across all model types, demonstrating superior ability to accurately reflect model behavior. SHAP provided robust performance across different algorithms, while LIME showed higher variability in explanation quality.

Feature Importance Analysis and Consistency

Feature importance rankings showed high consistency across different XAI methods (Table 2), with Spearman correlations exceeding 0.85 for most comparisons. Climate variables, particularly precipitation and temperature, emerged as primary predictors across all soil properties. Topographic variables including elevation and slope showed consistent importance, while vegetation indices contributed significantly to organic carbon predictions.

Table 2: Feature importance consistency across XAI methods and top-ranked predictors

Soil Property	SHAP-LIME Correlation	SHAP-SSEF Correlation	Top 5 Predictors (SHAP Rankings)
Soil Organic Carbon	0.89	0.92	1. Mean Annual Precipitation (0.18)
			2. Mean Annual Temperature (0.15)
			3. Elevation (0.12)
			4. NDVI Maximum (0.11)
			5. Topographic Wetness Index (0.09)
Soil pH	0.87	0.90	1. Parent Material Type (0.16)
			2. Precipitation Seasonality (0.14)
			3. Elevation (0.13)
			4. Slope (0.10)
			5. Temperature Seasonality (0.09)
Clay Content	0.85	0.88	1. Geological Formation (0.17)
			2. Mean Annual Precipitation (0.15)
			3. Topographic Position Index (0.12)
			4. Flow Accumulation (0.10)
			5. Aspect (0.08)
Available Nitrogen	0.86	0.89	1. Land Use History (0.19)
			2. NDVI Mean (0.16)
			3. Temperature Annual Range (0.13)
			4. Cultivation Duration (0.11)
			5. Soil Organic Carbon (0.10)

The high consistency in feature importance rankings across methods provides confidence in the reliability of explanations. Regional variations in predictor importance reflected local environmental conditions, with coastal areas showing higher importance for salinity-related variables and mountainous regions emphasizing topographic factors.

Non-linear Relationships and Interaction Effects

SHAP analysis revealed complex non-linear relationships and interaction effects that were not apparent through traditional statistical analysis (Table 3). Precipitation showed threshold effects on organic carbon accumulation, with maximum benefits occurring at 800-1200 mm annually. Temperature exhibited optimal ranges for different soil properties, with soil pH showing sensitivity to temperature extremes.

Table 3: Key non-linear relationships and interaction effects identified through SHAP analysis

Soil Property	Non-linear Relationship	Interaction Effect	Effect Magnitude
Soil Organic Carbon	Precipitation threshold at 800-1200 mm	Temperature × Precipitation	$\Delta R^2 = 0.08$
	Temperature optimum 12-18°C	Elevation × Climate	$\Delta R^2 = 0.06$
	Elevation saturation >1500m	NDVI × Management	$\Delta R^2 = 0.05$
Soil pH	Precipitation inverse relationship	Parent Material × Climate	$\Delta R^2 = 0.09$
	Temperature extremes (<5°C, >25°C)	Topography × Precipitation	$\Delta R^2 = 0.07$
	Elevation buffering effect	Geology × Weathering	$\Delta R^2 = 0.04$
Clay Content	Geological formation categorical	Relief × Parent Material	$\Delta R^2 = 0.11$
	Precipitation log-linear	Climate × Topography	$\Delta R^2 = 0.08$
	Flow accumulation exponential	Vegetation × Erosion	$\Delta R^2 = 0.06$
Available Nitrogen	Management history step function	Land Use × Climate	$\Delta R^2 = 0.10$
	Vegetation phenology cyclical	Organic Matter × Temperature	$\Delta R^2 = 0.07$
	Cultivation duration decay	Soil Properties × Management	$\Delta R^2 = 0.05$

These insights provide valuable scientific understanding that extends beyond predictive modeling to process-based knowledge of soil formation and management effects. The identification of threshold effects and optimal ranges supports evidence-based management recommendations.

User Evaluation and Trust Assessment

User evaluation studies demonstrated significant improvements in model acceptance and decision-making when XAI explanations were provided (Table 4). Trust scores increased by 78% on average across user groups, with soil scientists showing greatest improvement in cognitive trust and land managers showing highest gains in behavioral intentions.

Table 4: User evaluation results showing impact of XAI explanations on trust and decision-making

User Group	Sample Size	Trust Improvement (%)			Decision Confidence (+%)	Adoption Intention (+%)
		Cognitive	Emotional	Behavioral		
Soil Scientists	18	85	68	72	71	89
Agronomists	16	74	61	67	62	78
Land Managers	13	72	59	89	58	85
Overall Average	47	78	64	74	65	82

Decision confidence improved by 65% on average, with participants reporting greater certainty in recommendations based on explained predictions. Adoption intentions increased by 82%, indicating strong potential for practical implementation of XAI-enhanced soil prediction systems. Explanation preference surveys revealed user group differences in optimal explanation formats. Soil scientists preferred detailed feature importance rankings with confidence intervals. Agronomists favored simplified visual explanations focused on management-relevant variables.

Land managers emphasized spatial patterns and regional comparisons.

Computational Efficiency and Scalability

Computational overhead analysis revealed manageable costs for explanation generation across different methods (Table 5). SHAP explanation generation required 12-18% additional computation time compared to prediction alone, while LIME showed higher variability (8-25%) depending on perturbation sample size.

Table 5: Computational efficiency analysis for XAI methods across different model types

XAI Method	Random Forest		XGBoost		Neural Networks		Transformer	
	Time (ms)	Overhead (%)	Time (ms)	Overhead (%)	Time (ms)	Overhead (%)	Time (ms)	Overhead (%)
Baseline Prediction	2.3	-	1.8	-	4.6	-	7.2	-
SHAP	2.8	22	2.1	17	5.4	17	8.5	18
LIME	2.9	26	2.2	22	5.8	26	9.0	25
Integrated Gradients	-	-	-	-	5.2	13	8.1	13
Attention Mechanisms	-	-	-	-	-	-	7.4	3
SSEF	3.1	35	2.4	33	6.2	35	9.8	36

The SSEF method showed highest computational overhead (33-36%) due to additional pedological constraint checking, but remained feasible for real-time applications. Attention mechanisms provided the most efficient explanations for transformer models with only 3% overhead.

Scalability analysis confirmed linear scaling with dataset size for most methods, enabling application to large-scale soil mapping projects. Parallel processing and caching strategies reduced explanation generation time by 40-60% for batch applications.

Regional Explanation Patterns

Analysis of explanation patterns across different regions revealed systematic variations in predictor importance that aligned with known soil-landscape relationships. Mediterranean regions showed higher importance for drought-related variables, while temperate regions emphasized temperature seasonality effects.

Coastal areas demonstrated elevated importance for salinity and proximity variables, while inland regions prioritized continental climate factors. Mountainous terrain showed complex interactions between elevation, aspect, and climate variables that explained local soil property variations.

These regional patterns provide valuable insights for adaptive sampling strategies and region-specific model development, supporting more targeted soil mapping efforts.

Discussion

The superior performance of the Soil-Specific Explanation Framework (SSEF) demonstrates the value of incorporating domain knowledge into XAI approaches for scientific applications. By organizing explanations according to soil formation factors and enforcing pedological constraints, the SSEF method provides explanations that are both mathematically sound and scientifically meaningful. This approach bridges the gap between statistical attribution and process-based understanding that is crucial for soil science applications.

The high consistency in feature importance rankings across different XAI methods provides confidence in the reliability and robustness of explanations. The identification of climate and topographic variables as primary predictors aligns with fundamental soil science principles, while the discovery of complex non-linear relationships and interaction effects adds new scientific insights that extend beyond traditional statistical analysis.

The user evaluation results demonstrate clear benefits of explainable AI for building trust and supporting decision-making in soil science applications. The 78% improvement in trust scores across user groups indicates that explanations successfully address skepticism about "black box" AI models. The increased decision confidence and adoption intentions suggest that XAI can facilitate broader acceptance of AI tools in agricultural and environmental management.

The computational efficiency analysis reveals that explanation generation imposes manageable overhead for most practical applications. The 12-18% increase in computation time for SHAP explanations is reasonable for the significant benefits in interpretability and user acceptance. The development of efficient approximation methods and caching strategies can further reduce computational costs for operational deployment.

The identification of regional patterns in explanation importance provides valuable insights for adaptive model

development and sampling strategies. The systematic variations in predictor importance across different environmental conditions suggest that explanation analysis can inform scientific understanding of soil-landscape relationships and guide future research directions.

The discovery of threshold effects and optimal ranges for environmental variables through SHAP analysis provides actionable insights for soil management and agricultural decision-making. These findings demonstrate that XAI can contribute to scientific knowledge beyond simply explaining existing models, offering new perspectives on complex environmental relationships.

Conclusion

This comprehensive study establishes the effectiveness of explainable AI approaches for enhancing interpretability in soil prediction models while maintaining high predictive accuracy. The Soil-Specific Explanation Framework emerged as the optimal approach, achieving superior fidelity scores (0.87-0.94) by incorporating pedological knowledge into explanation generation.

The high consistency in feature importance rankings across different XAI methods (Spearman correlations >0.85) provides confidence in explanation reliability, while the identification of complex non-linear relationships and interaction effects adds scientific value beyond predictive modeling. The systematic regional patterns in explanation importance offer insights for adaptive model development and targeted sampling strategies.

User evaluation results demonstrate significant improvements in trust (78% increase), decision confidence (65% increase), and adoption intentions (82% increase) when explanations are provided, indicating strong potential for practical implementation. The manageable computational overhead (12-18% increase) makes real-time explanation generation feasible for operational applications.

The study reveals that explainable AI can contribute to scientific understanding by uncovering threshold effects, optimal ranges, and complex interactions that inform soil management and agricultural decision-making. The ability to provide both statistical attribution and process-based understanding makes XAI particularly valuable for scientific applications where explanation quality is as important as predictive accuracy.

Future research should focus on developing more sophisticated domain-specific explanation frameworks, integrating temporal dynamics into explanation analysis, and exploring explanation-guided model improvement strategies. The extension of XAI approaches to multi-modal data sources and real-time monitoring applications could further enhance their utility for operational soil mapping systems.

The findings provide a solid foundation for implementing explainable AI in soil science applications, supporting the transition from purely predictive models to interpretable systems that combine accuracy with understanding and trust.

References

1. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016;22(8):1135–1144.
2. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, *et al.* Explainable Artificial

- Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 2020;58(1):82-115.
3. Wadoux AM, Heuvelink GB, Brus DJ. Robust digital soil mapping using ensemble machine learning and explainable artificial intelligence. *Geoderma*. 2021;404(2):115276.
 4. Angelov P, Soares E. Towards explainable deep neural networks (xDNN). *Neural Networks*. 2020;130(3):185–194.
 5. Lal R. Soil carbon sequestration impacts on global climate change and food security. *Science*. 2004;304(5677):1623–1627.
 6. Ross AS, Hughes MC, Doshi-Velez F. Right for the right reasons: Training differentiable models by constraining their explanations. *Proceedings of the 2017 International Joint Conference on Artificial Intelligence (IJCAI)*. 2017;26(1):2662–2670.
 7. Hooker S, Erhan D, Kindermans PJ, Kim B. A benchmark for interpretability methods in deep neural networks. *Advances in Neural Information Processing Systems*. 2019;32(1):9737–9748.
 8. Wilpert K, Padarian J, Minasny B. Using interpretable machine learning to map soil properties. *Geoderma*. 2022;406(5):115514.
 9. Shrestha R, Farid A, Kashem MA. Applications of artificial intelligence in soil science: A critical review. *Environmental Reviews*. 2020;28(3):441-453.
 10. Chang C, Lin C. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011;2(3):27:1-27:27.
 11. Rinnan Å, Bro R, Andersson M. Soil classification using near-infrared spectroscopy and machine learning. *Applied Spectroscopy*. 2009;63(5):491-500.
 12. Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*. 2018;73(2):1-15.
 13. Schmidt AM, Zimmermann J, Behrens T, Scholten T. Performance of machine learning algorithms for soil property prediction: A case study on texture and pH. *Soil Science Society of America Journal*. 2020;84(1):123-136.
 14. Taalab K, Corstanje R, Creamer RE, Whelan MJ. Models for soil bulk density estimation: A review. *Soil Use and Management*. 2013;29(4):462-469.
 15. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017;30(1):4765-4774.