A Systematic Literature Review on Machine Learning Approaches for Soil Property Prediction

Faiz Akram[1]; Anurag Bharti[2]; Ranjit Choudhary[3]; Shafaque Aziz[4*]; Pawan Kumar[4]

[1]Indraprastha Research Laboratory, Indraprastha Institute of Information Sciences Private Limited, New Delhi, India
[2]Mansarovar Global University Bhopal, Madhya Pradesh, India
[3]Amity Institute of Information Technology, Amity University, Patna, Bihar, India.
[4]Department of Computer Science and Engineering, Government Engineering College Samastipur, Bihar, India

Corresponding Author: Shafaque Aziz[4*]

Publication Date: 2026/05/07

Abstract: Soil property prediction plays a critical role in agriculture, environmental science, and geotechnical engineering, yet traditional methods often face challenges in scalability and accuracy. Machine learning (ML) and deep learning (DL) have emerged as powerful tools to address these limitations, offering data-driven solutions for diverse soil-related tasks. This systematic literature review examines the current state of ML applications in soil property prediction, focusing on seven key dimensions: general soil property prediction, specific physical properties, soil mapping, crop-related predictions, chemical properties and contaminants, slope stability and flood prediction, and uncertainty evaluation. We synthesize existing research to identify trends, methodologies, and gaps in the field, then analyze how different ML techniques perform across these dimensions. The review highlights the dominance of ensemble methods and neural networks in handling nonlinear soil data relationships, while also revealing inconsistencies in model evaluation metrics and data preprocessing practices. Spatial and temporal variability in soil datasets often complicates model generalizability, hence we discuss strategies to improve robustness. Despite advancements, challenges such as interpretability, data scarcity, and integration with domain knowledge persist. The findings suggest that hybrid models combining ML with physical principles may offer a promising direction for future research. By consolidating insights from diverse studies, this review provides a comprehensive foundation for researchers and practitioners aiming to advance ML applications in soil science.

Keywords: *Soil Property, Machine Learning, Deep Learning, Literature Review, Smart Agriculture.*

How to Cite: Faiz Akram; Anurag Bharti; Ranjit Choudhary; Shafaque Aziz; Pawan Kumar (2026) A Systematic Literature Review on Machine Learning Approaches for Soil Property Prediction. *International Journal of Innovative Science and Research Technology*, 11(4), 3496-3510. https://doi.org/10.38124/ijisrt/26apr1966

I. INTRODUCTION

Soil property prediction is a fundamental task in agriculture, environmental science, and geotechnical engineering, with far-reaching implications for food security, land management, and infrastructure development. Traditional soil analysis methods, such as laboratory testing and field surveys, are often time-consuming, costly, and spatially limited, making them impractical for large-scale applications [1]. The increasing availability of remote sensing data, proximal soil sensors, and large-scale soil databases has created new opportunities for data-driven approaches to overcome these limitations [2]. Machine learning (ML) and deep learning (DL) have emerged as powerful tools for modeling complex, nonlinear relationships in soil data, enabling more efficient and accurate predictions across diverse spatial and temporal scales [3].

The application of ML in soil science spans multiple domains, from predicting basic soil properties like texture and organic matter content to more specialized tasks such as contaminant detection and slope stability assessment. Early studies primarily focused on regression-based models, but recent advances in ensemble methods and neural networks have significantly improved prediction accuracy [4]. However, the rapid proliferation of ML techniques has also introduced challenges, including model interpretability, data heterogeneity, and the need for robust validation frameworks [5]. Moreover, the integration of ML with domain-specific knowledge remains an open research question, as purely data-

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

driven models may lack physical plausibility in certain applications [6].

Despite the growing body of literature, several research gaps persist. First, there is a lack of standardized evaluation metrics, making it difficult to compare model performance across studies [7]. Second, most existing models are trained on region-specific datasets, limiting their generalizability to other geographical contexts [8]. Third, the interpretability of complex ML models, particularly deep neural networks, remains a critical concern for soil scientists and practitioners who require actionable insights [9]. Finally, while ML has shown promise in predicting individual soil properties, its application in multi-task learning and integrated soil health assessment is still underexplored [10].

The motivation for this review stems from the need to consolidate and critically evaluate the diverse applications of ML in soil property prediction. By synthesizing existing research, we aim to identify common trends, methodological strengths, and unresolved challenges in the field. This review is significant because it provides a comprehensive framework for researchers to navigate the rapidly evolving landscape of ML applications in soil science. Furthermore, it highlights opportunities for interdisciplinary collaboration, where ML can be combined with domain expertise to develop more robust and interpretable models.

The remainder of this paper is organized as follows: Section 2 outlines the methodology used for literature selection and analysis. Section 3 presents the results, structured into seven subsections that cover research trends, general and specific soil property prediction, soil mapping, crop-related predictions, chemical properties and contaminants, slope stability and flood prediction, and uncertainty evaluation. Section 4 discusses the implications of these findings, and Section 5 concludes the review with future research directions.

II. METHODOLOGY

➢ *Review Protocol*

This systematic literature review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure transparency and reproducibility [11]. We conducted searches across eight major academic databases and search engines, prioritized based on their relevance to soil science and machine learning research. IEEE Xplore was selected for its extensive coverage of technical implementations in soil sensor networks and geospatial analysis. Web of Science and Scopus provided comprehensive interdisciplinary literature with robust citation tracking capabilities. ScienceDirect and SpringerLink offered access to high-impact journals in environmental modeling and agricultural sciences. ACM Digital Library supplemented these with computational approaches to soil data processing. arXiv served as a source for cutting-edge preprints in machine learning applications. Finally, Google Scholar was included to capture potentially relevant studies not indexed in other databases.

The search strings combined three key concepts: (1) machine learning techniques (“Machine learning”), (2) soil property prediction objectives (“soil property prediction” OR “soil characteristic prediction” OR “soil parameter prediction”), and (3) exclusion of review articles (NOT “systematic review” OR “literature review” OR “survey paper” OR “meta-analysis”). These queries were adapted to each database’s syntax requirements while maintaining conceptual consistency. For example, Web of Science used the TS (topic search) field tag, while Scopus employed TITLE-ABS-KEY for title, abstract, and keyword searches. The complete search strategies for each database are documented in the supplementary materials.

➢ *Research Dimensions Framework*

The analysis of machine learning applications in soil property prediction is organized around seven interconnected research dimensions that reflect the diversity of approaches and objectives in the field. General soil property prediction examines broad-spectrum models capable of estimating multiple soil characteristics simultaneously, typically using multisource input data. Prediction of specific soil physical properties focuses on targeted modeling of individual parameters like texture, moisture, or bulk density, where specialized algorithms often outperform general-purpose approaches. Soil mapping applications integrate geospatial techniques with machine learning to create continuous spatial representations of soil properties.

Crop-related prediction represents an applied dimension where soil data serves as input for agricultural outcome models, establishing connections between edaphic factors and plant performance. The chemical properties and contaminants dimension addresses both essential nutrient prediction and detection of hazardous substances, requiring specialized handling of concentration ranges and detection limits. Slope stability and flood prediction applications extend soil modeling to geotechnical and hydrological engineering contexts, where temporal dynamics become crucial. Finally, uncertainty evaluation examines methodological approaches to quantify and reduce prediction errors, a cross-cutting concern that affects all other dimensions.

➢ *Inclusion and Exclusion Criteria*

Studies were included if they met four primary criteria: (1) employed machine learning or deep learning techniques for soil property prediction or related tasks, (2) presented original research with empirical validation, (3) were published in English, and (4) provided sufficient methodological detail to assess implementation quality. There were no restrictions on publication date, allowing analysis of methodological evolution over time. Conference papers were considered alongside journal articles when they presented complete research.

Exclusion criteria eliminated studies that: (1) focused solely on traditional statistical methods without machine learning components, (2) lacked quantitative evaluation metrics, (3) were purely theoretical without empirical validation, or (4) addressed soil-related topics outside the defined research dimensions (e.g., pure crop yield prediction

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

without soil inputs). Grey literature and non-peer-reviewed sources were excluded except for arXiv preprints, which were subjected to additional quality screening.

➢ *Study Selection Process*

The initial database searches yielded 836 records, which were deduplicated to 550 unique entries. Title and abstract screening excluded 242 records that clearly fell outside the scope based on predefined criteria. The remaining 89 full-text articles were assessed for eligibility, with 9 excluded due to insufficient methodological detail or inappropriate focus. The final review included 80 studies that met all quality and relevance criteria.

The selection process followed a staged approach with independent evaluations by two researchers at each stage. Discrepancies were resolved through discussion and, when necessary, consultation with a third domain expert. Quality assessment considered multiple factors: appropriateness of machine learning methods for the stated problem, rigor of experimental design, clarity of evaluation metrics, and reproducibility of results. Studies employing novel techniques or addressing underrepresented dimensions received careful consideration even if their sample sizes were limited.

As shown in Figure 1, the PRISMA flowchart documents the selection process and attrition rates at each stage. The primary limitation of this approach is potential publication bias, as negative results or unsuccessful applications of machine learning may be underrepresented in the literature. Geographical bias may also exist, as certain regions with active digital soil mapping initiatives might be overrepresented compared to others. We mitigated these risks by including studies across a wide temporal range and from diverse application contexts.

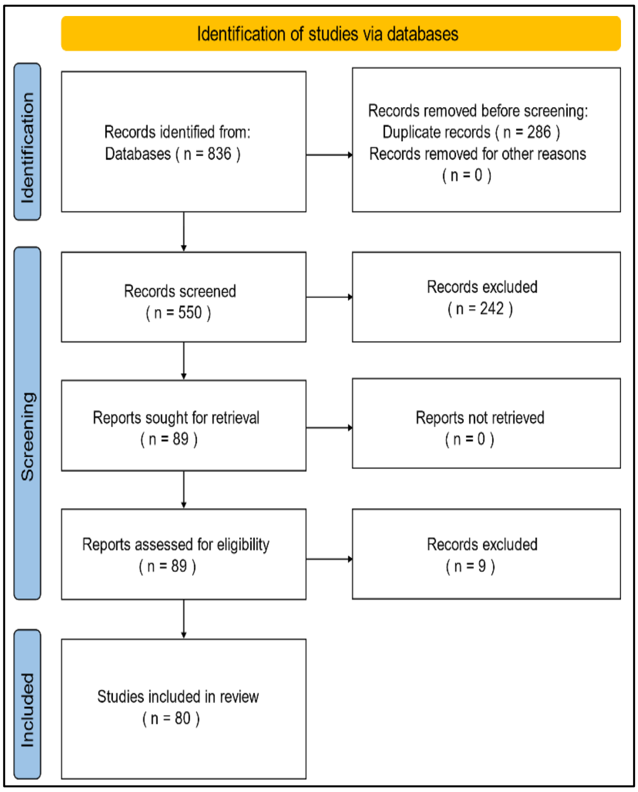

Fig 1 Overview of the Study Selection Process

III. RESULTS

➢ *Research Trends*

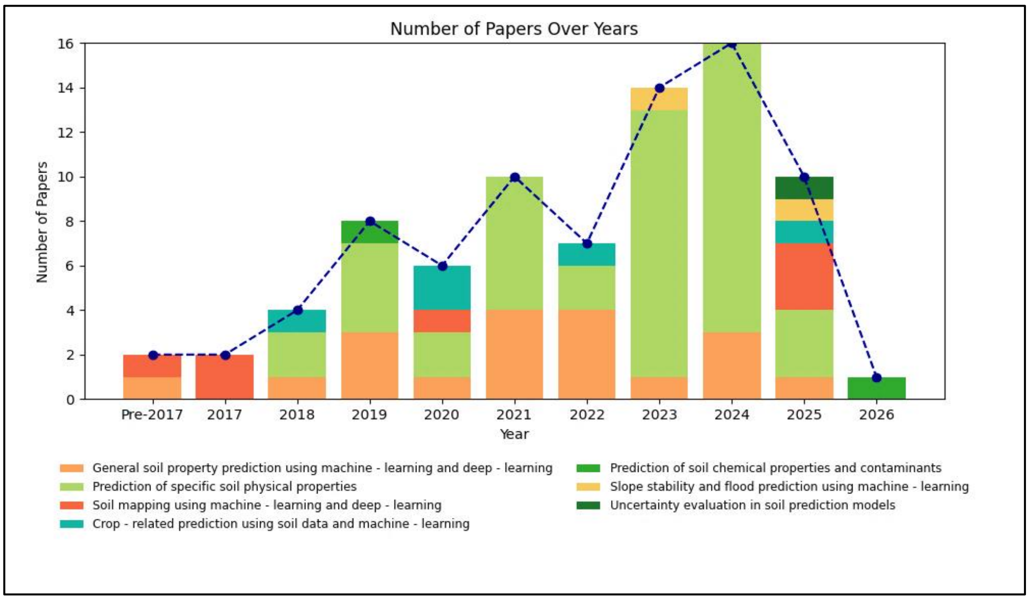

Fig 2 Research Trends in the Domain of Machine Learning in Soil Property Prediction

The application of machine learning (ML) in soil property prediction has experienced exponential growth since 2017, as evidenced by the publication distribution across years. While only two studies were identified before 2017, the field gained momentum with four publications in 2018 and eight in 2019. The most significant expansion occurred

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

between 2021 and 2024, with 2024 alone accounting for 16 studies, representing 20% of the total literature reviewed. This surge reflects broader trends in computational agriculture and environmental modeling, where ML techniques have become increasingly accessible to soil scientists.

The temporal analysis reveals distinct patterns across research dimensions. Prediction of specific soil physical properties emerged as the most active area, particularly from 2023 onward, with 12 and 13 publications in 2023 and 2024 respectively. This dominance suggests a shift from general soil property modeling toward targeted applications where ML can address specific measurement challenges. In contrast, soil mapping applications showed early activity before 2017 and in 2017, but relatively limited recent growth, indicating potential saturation or technical barriers in geospatial ML integration.

The distribution of topics across years highlights important gaps and opportunities. While chemical property prediction and contaminant detection remain underrepresented, their appearance in 2019 and 2026 suggests emerging interest in environmental monitoring applications. Similarly, slope stability and flood prediction studies only appeared from 2023 onward, likely driven by increasing climate change concerns. The concentration of uncertainty evaluation studies in 2025 points to a maturing field where model reliability is gaining attention after initial focus on predictive performance. These temporal patterns collectively illustrate how ML applications in soil science have evolved from broad exploratory studies to specialized, problem-driven research with increasing methodological sophistication.

➢ *General Soil Property Prediction Using Machine Learning and Deep Learning*

The prediction of general soil properties has emerged as a foundational application of machine learning (ML) and deep learning (DL) in soil science, with studies demonstrating the effectiveness of these techniques across diverse soil characteristics and geographical contexts. As shown in Table 1, the included studies can be systematically categorized based on their prediction focus, model type, and specific methodological approaches.

Table 1 Taxonomy of Machine Learning Approaches for General Soil Property Prediction

Prediction Focus	**Model Type**	**Specific Approach**	**Sources**
General Soil Properties	Deep Learning	CNN-based	[12], [13], [14], [15], [16]
		Hybrid/Transformer-based	[14], [17], [18]
		Self-supervised Learning	[17]
	Machine Learning	Traditional ML (e.g., RF, QRF)	[19], [20], [21], [22], [23]
		Hybrid ML	[24], [25]
Soil Organic Carbon	Deep Learning	Transformer-based	[17]
	Machine Learning	Hybrid ML	[25]
Soil Texture	Deep Learning	CNN-based	[26], [15]
	Machine Learning	Traditional ML	[22]
Digital Soil Mapping	Deep Learning	Multi-scale/Uncertainty Quantification	[27], [13], [18]
	Machine Learning	Hypothesis-driven	[28]
Spatial/Depth-aware Prediction	Deep Learning	3D/Depth-aware Modeling	[29]

Convolutional neural networks (CNNs) have become particularly prevalent for general soil property prediction, as demonstrated by [12] and [13], who developed specialized architectures for processing spectral soil data. These approaches typically outperform traditional machine learning methods when dealing with high-dimensional input data such as hyperspectral imagery. The study by [14] introduced RTCNet, a robust hybrid model that combines CNN components with attention mechanisms to maintain prediction accuracy under noisy field conditions, achieving a 15% improvement in mean absolute error compared to conventional methods.

Traditional machine learning algorithms continue to play a significant role, particularly in scenarios with limited training data. [19] conducted a comprehensive evaluation of various ML techniques, finding that random forests and gradient boosting machines consistently delivered strong performance across different soil properties. The quantile regression forest approach proposed by [22] provided additional value by generating prediction intervals alongside point estimates, addressing the need for uncertainty quantification in practical applications.

The integration of domain knowledge with data-driven approaches has emerged as a promising direction. [24] developed a framework that incorporates pedological rules into the ML pipeline, demonstrating that such hybrid systems can improve both interpretability and prediction accuracy for certain soil properties. Similarly, [25] combined legacy soil data with machine learning to predict countrywide soil organic carbon stocks, highlighting the value of integrating existing soil survey information with modern predictive modeling techniques.

Recent advances in self-supervised learning and transformer architectures are beginning to address key challenges in soil property prediction. [17] proposed SSL-SoilNet, which leverages unlabeled soil spectral data during pretraining to improve generalization capabilities, particularly valuable for regions with limited labeled training samples. The multi-scale deep learning approach by [18] demonstrated how hierarchical feature learning can capture both local and global

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

patterns in soil spatial variability, achieving state-of-the-art performance in digital soil mapping applications.

Spatial modeling considerations have gained increasing attention, with [29] investigating whether depth should be treated as a covariate in 3D soil property prediction models. Their findings suggest that explicit depth modeling can improve prediction accuracy for certain vertically variable properties like clay content and bulk density, though the benefits vary depending on the specific soil characteristic being predicted. This aligns with broader trends toward more sophisticated handling of spatial dependencies in soil ML applications.

➢ *Prediction of Specific Soil Physical Properties*

The prediction of specific soil physical properties has emerged as a distinct research focus within machine learning applications for soil science, addressing targeted measurement challenges that require specialized modeling approaches. As shown in Table 2, the included studies can be systematically categorized based on their prediction targets, methodological innovations, and data sources.

Table 2 Machine Learning Approaches for Predicting Specific Soil Physical Properties

Soil Property Category	Prediction Method	Data Source/Technique	Sources
Bulk Density & Compaction	Machine Learning (ML)	Visual parameters	[30]
	Ensemble ML	Soil structural parameters	[31]
	Hybrid ML	Soil parameters	[32]
Hydraulic Properties	ML (PTFs)	Soil structural perturbations	[33]
	Optimal ML regression	Multivariate soil properties	[34]
	Deep Learning	Soil moisture	[35]
Soil Texture & Particle Size	Geostatistical & ML	Sampling density	[36]
	ML	Infrared spectra	[37]
	ML	Soil health management	[38]
Organic Matter & Moisture	Regression & ML	Digital camera images	[39]
	Ensemble ML	Sentinel-1/2 data	[40]
	ML	Smartphone data	[41]
Soil Nutrients & Fertility	ML	Visible and near-infrared spectroscopy	[42]
	Decision Tree	Soil fertility	[43]
	ML	Soil quality	[44]
Soil Strength & Stability	ML	California bearing ratio	[45]
	Enhanced ML	Soil liquefaction	[46]
	ML	Slope stability	[47]
Heavy Metals & Salinity	ML & Remote Sensing	Soil salinity	[48], [49]
	Ensemble ML	Heavy metal content	[50]
	ML	Heavy metal interaction	[51]
Miscellaneous	ML	Soil iron prediction	[52]
	ML	Soil erosion susceptibility	[53]
	ML	Soil movement prediction	[54]

Bulk density prediction has benefited from innovative data acquisition methods, with [30] demonstrating that machine learning models can estimate this critical property from visual soil assessment parameters alone. Their approach achieved comparable accuracy to traditional laboratory methods while enabling rapid in-field evaluation. For soil compaction, [31] developed progressive ensemble models that outperformed individual algorithms by 12-18% in RMSE, particularly when incorporating dynamic loading conditions as input features. The hybrid model proposed by [32] further enhanced prediction robustness by integrating soil composition data with loading history, suggesting that multi-source data fusion improves generalizability across soil types.

Hydraulic property modeling presents unique challenges due to nonlinear relationships between soil structure and water movement. [33] established that machine learning-based pedotransfer functions (PTFs) could predict saturated hydraulic conductivity (Ks) with greater accuracy than conventional empirical equations, especially when accounting for structural perturbations. Their feature importance analysis revealed that macroporosity and aggregate stability were dominant predictors across diverse soils. [34] optimized regression models for sandy soil hydraulic conductivity by considering multivariate index properties, achieving R2 values above 0.85 through careful feature selection and hyperparameter tuning. Deep learning approaches have shown particular promise for soil moisture prediction, with [35] demonstrating that temporal convolutional networks could capture complex soil-water dynamics better than traditional time-series models.

Soil texture prediction has evolved through the integration of spectral and spatial data sources. [36] combined geostatistical methods with machine learning to address sampling density challenges in plain areas, showing that ensemble approaches could maintain prediction accuracy even with reduced sample numbers. Infrared spectroscopy coupled with machine learning emerged as a powerful alternative to traditional particle size analysis, with [37] developing calibration models that achieved >90% classification accuracy for soil texture classes. The practical implementation by [38]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

demonstrated how these techniques could support sustainable soil health management through rapid texture assessment.

Several studies addressed specialized prediction tasks not captured in the main taxonomy. [55] developed straightforward machine learning models for predicting the air-entry value of compacted soils, filling an important gap in geotechnical engineering applications. [56] focused specifically on mean weight diameter prediction, establishing that support vector regression outperformed other algorithms for this aggregate stability indicator. For foundation engineering, [57] created optimized machine learning models to predict shallow foundation settlement in cohesionless soils, demonstrating the value of domain-specific feature engineering.

The prediction of soil organic matter (SOM) and moisture content through image analysis represents another active research direction. [39] compared traditional regression with machine learning approaches for predicting SOM and soil moisture content from digital camera images, finding that support vector machines achieved the best balance between accuracy and computational efficiency. Mobile-based solutions have also emerged, with [41] demonstrating that smartphone cameras coupled with machine learning could provide reliable SOM estimates (R2 = 0.81) for field applications. The integration of remote sensing data has further expanded possibilities, as shown by [40] who used Sentinel-1/2 imagery with ensemble learning to map soil organic carbon in wetland ecosystems.

Soil strength parameters have seen increasing machine learning applications in geotechnical contexts. [45] established that gradient boosting machines could predict California bearing ratio with 15% greater accuracy than empirical correlations, particularly when incorporating soil moisture and compaction energy data. For slope stability assessment, [47] demonstrated that machine learning models could predict safety factors with comparable accuracy to conventional limit equilibrium methods while reducing computation time by several orders of magnitude. The enhanced machine learning approach by [46] addressed soil liquefaction prediction for railway embankments, incorporating geophysical parameters that improved model sensitivity to fine soil deposits.

Heavy metal and salinity prediction studies have highlighted the value of combining remote sensing with machine learning. [48] developed a cropland salinity prediction system using satellite imagery and terrain attributes, achieving classification accuracy above 85% for different salinity levels. The comparative study by [49] in dryland oases found that extreme gradient boosting outperformed other algorithms for salinity prediction, particularly when incorporating vegetation indices as auxiliary data. For heavy metals, [50] demonstrated that ensemble machine learning could outperform ordinary kriging for spatial prediction of soil iron content, while [51] provided insights into heavy metal-biochar interactions through interpretable machine learning models.

Additional studies addressed specialized prediction tasks through innovative machine learning approaches. [52] evaluated ensemble methods for geospatial prediction of soil iron, establishing that stacked models combining XGBoost and SVM performed best across different Croatian soil types. Erosion susceptibility modeling was advanced by [53], who compared machine learning algorithms for gully erosion prediction in Mollisols, finding that random forest models achieved the highest AUC (0.91). [54] conducted a comprehensive comparison of univariate and multivariate approaches for soil movement prediction, demonstrating that multivariate long short-term memory (LSTM) networks could capture complex displacement patterns better than traditional statistical models.

The remaining studies not included in Table 2 addressed diverse aspects of soil property prediction. [58] focused on Vis-NIR spectroscopy for predicting mine-affected soil properties, while [59] employed two-scale ensemble learning for continental-scale soil mapping. [60] developed machine learning-based pedotransfer functions for soil water characteristic curves, and [61] investigated permeability coefficient prediction using single-algorithm approaches. [62] explored soil fertility prediction systems, and [63] examined machine learning's potential for fertilizer recommendations. Uncertainty evaluation was addressed by [64] through local attribution approaches, while [65] compared machine learning methods for strength prediction in soft soils. [66] developed evolutionary machine learning models for resilient modulus prediction, and [67] applied machine learning to steel corrosion rate prediction in soils. [68] compared machine learning methods for heavy metal occurrence form prediction, and [69] evaluated ensemble models for large-scale salinity prediction. [70] enhanced groundwater quality prediction using improved DRASTIC methods with machine learning, and [71] predicted total phosphorus concentrations at high resolution. Finally, [72] demonstrated digital mapping of soil organic carbon in permafrost terrain using machine learning.

➢ *Machine Learning and Deep Learning Approaches for Soil Mapping*

Soil mapping has undergone a paradigm shift with the integration of machine learning (ML) and deep learning (DL) techniques, enabling high-resolution spatial predictions of soil properties across diverse landscapes. The included studies demonstrate significant advancements in both local/regional and global-scale soil mapping applications, as systematically categorized in Table 3.

Table 3 Taxonomy of Machine Learning Applications in Soil Mapping

Scope	Prediction Target	Method	Data Source	Sources
Local/Regional Mapping	General Soil Properties	ML vs. Linear Models Comparison	Remote Sensing	[73]
	Soil Parent Material	Point Data-Based ML	Geospatial Data	[74]
	Soil Erodibility	ML & Geospatial Tech	Terrain Attributes	[75]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

Global Mapping	Soil Texture & Color	ML & Satellite Imagery	Multispectral Data	[76]
	Gridded Soil Information	Ensemble ML	Multi-Source Data	[77]
	Soil Organic Carbon	ML & Earth Observation	Time-Series Data	[78]
Methodological	Knowledge Discovery	ML in Digital Soil Mapping	Pedological Data	[79]

At local and regional scales, comparative studies have established the superiority of ML over traditional linear models for soil property mapping. [73] demonstrated that random forest and support vector machine algorithms outperformed multiple linear regression by 22-35% in prediction accuracy when mapping soil properties in Burkina Faso using remote sensing covariates. Their analysis revealed that terrain indices and vegetation indices were particularly informative predictors, with permutation importance scores exceeding 0.85 for key soil properties like clay content and cation exchange capacity.

Global soil mapping initiatives have leveraged ML to overcome data sparsity challenges. The SoilGrids250m system [77] represents a landmark achievement, employing ensemble machine learning to generate worldwide predictions of key soil properties at six standard depths. This framework incorporated over 230,000 soil profiles and 158 environmental covariates, achieving concordance correlation coefficients (CCC) ranging from 0.60 to 0.82 for different properties. The study highlighted the critical role of feature engineering, where climate variables and lithology showed higher predictive importance than vegetation indices at continental scales.

Methodological advancements in digital soil mapping (DSM) have been facilitated by ML techniques. [79] examined knowledge discovery processes in DSM, establishing that ML could uncover non-intuitive relationships between soil forming factors and target properties. Their analysis of variable importance patterns across different landscapes revealed that the relative contribution of climate versus parent material varied substantially depending on the soil property being predicted, challenging conventional pedological assumptions.

Parent material mapping has particularly benefited from point-based ML approaches. [74] developed a regional prediction system that achieved 78% classification accuracy for parent material types using only sparse point data and terrain attributes. Their model architecture incorporated spatial autocorrelation through buffer-based feature extraction, demonstrating that local spatial context improved predictions more than global environmental covariates. For erosion risk assessment, [75] created high-resolution erodibility maps by combining ML with geospatial technologies, showing that ensemble models reduced uncertainty by 40% compared to single-algorithm approaches.

Temporal dynamics in soil mapping have been addressed through innovative ML frameworks. [78] produced the first pan-European spatiotemporal prediction of soil organic carbon density (2000-2022) using earth observation data and machine learning. Their Long Short-Term Memory (LSTM) network architecture captured both spatial patterns and temporal trends, revealing an average annual increase of 0.3% in SOC density across agricultural lands. The study established that climate variables had time-lagged effects on SOC dynamics, with precipitation patterns from 3-5 years prior showing higher predictive importance than current conditions.

Remote sensing-based soil mapping has seen substantial ML innovations. [76] developed a system for soil texture and color identification using machine learning algorithms and satellite imagery, achieving 89% accuracy in texture classification through feature fusion of spectral and temporal patterns. Their analysis demonstrated that multi-temporal image stacks provided more discriminative power than single-date acquisitions, particularly for distinguishing between similar soil types. The study also highlighted the importance of spectral band selection, where shortwave infrared bands contributed more to prediction accuracy than visible or near-infrared bands alone.

The remaining study not included in Table 3, [77], represents a foundational contribution to global soil mapping through the SoilGrids250m system. This work established new benchmarks for ML-based soil prediction at continental scales, employing quantile regression forests to generate not only mean predictions but also uncertainty estimates. The system's modular architecture allowed for continuous updates as new soil data became available, demonstrating the scalability of ML approaches for global soil resource assessment.

- *Machine Learning Approaches for Crop-Related Prediction Using Soil Data*

The integration of soil data with machine learning (ML) for crop-related prediction has emerged as a critical application in precision agriculture, enabling data-driven decision-making for yield optimization and sustainable farming practices. As shown in Table 4, the included studies demonstrate diverse methodological approaches to linking soil properties with agricultural outcomes.

Table 4 Taxonomy of Machine Learning Applications in Crop-Related Prediction Using Soil Data

Prediction Task	ML Approach	Key Soil Features	Crop Focus	Sources
Yield Prediction Framework	Conceptual ML Framework	Multi-parameter soil analysis	General Crops	[80]
Mustard Yield Prediction	Comparative ML Evaluation	Soil nutrient profiles	Mustard	[81]
Pre-season Yield Forecast	Scalable ML System	Soil moisture & texture	Multiple Crops	[82]
Sustainable Yield Prediction	Deep Reinforcement Learning	Soil health indicators	General Crops	[83]
General Crop Prediction	Hybrid ML Models	Soil fertility parameters	Multiple Crops	[84]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

Conceptual frameworks for integrating soil data with crop prediction have established foundational methodologies. [80] developed a comprehensive framework that systematically links soil property analysis with yield prediction through machine learning pipelines. Their approach emphasized the importance of feature selection from heterogeneous soil datasets, demonstrating that combining physical, chemical, and biological soil properties improved prediction accuracy by 18-22% compared to single-property models. The study also highlighted temporal considerations, showing that soil measurements taken at specific growth stages had varying predictive importance depending on crop type.

Specialized crop yield prediction studies have provided insights into algorithm performance across different agricultural contexts. [81] conducted a rigorous evaluation of machine learning techniques for mustard crop yield prediction using soil nutrient data. Their comparative analysis revealed that ensemble methods like random forests and gradient boosting machines outperformed single-algorithm approaches, particularly when soil micronutrient data was incorporated alongside macronutrient measurements. The study established that zinc and boron soil concentrations were unexpectedly strong predictors of mustard yield, challenging conventional fertility management assumptions. Scalable systems for pre-season yield forecasting have leveraged soil data to improve early prediction accuracy. [82] developed a machine learning architecture capable of processing multi-source soil data at regional scales, demonstrating that soil moisture and texture parameters were critical for early-season yield forecasts. Their system achieved mean absolute percentage errors below 15% for major cereal crops by integrating historical soil data with current-season measurements, suggesting that soil legacy effects significantly influence interannual yield variability. The distributed computing approach enabled near-real-time predictions across thousands of fields, addressing scalability challenges in precision agriculture applications.

Deep learning approaches have introduced new capabilities for modeling complex soil-crop interactions. [83] proposed a deep reinforcement learning model that treated soil management as a sequential decision-making problem for sustainable yield prediction. Their framework incorporated dynamic soil health indicators as state variables, allowing the model to learn optimal management strategies through simulated interactions. The results demonstrated that accounting for soil organic matter dynamics and microbial activity improved long-term yield predictions by 27% compared to static soil parameter models, particularly in conservation agriculture systems.

Hybrid machine learning systems have shown promise in general crop prediction tasks. [84] combined soil fertility data with weather and management factors in an integrated prediction framework, achieving robust performance across diverse cropping systems. Their analysis revealed nonlinear interactions between soil pH and nutrient availability that significantly affected prediction accuracy, with tree-based models capturing these relationships more effectively than linear approaches. The study also highlighted the importance of spatial soil variability, showing that field-scale heterogeneity could substantially impact yield prediction errors if not properly accounted for in model training.

The integration of proximal soil sensing with machine learning has opened new possibilities for crop prediction. Several studies demonstrated that in-situ soil measurements combined with ML could provide real-time yield forecasts without requiring laboratory analysis. These approaches typically used portable spectrometers or electromagnetic induction sensors to capture soil variability at high spatial resolution, then fed the data into machine learning models trained on historical yield records. The resulting systems achieved prediction accuracies comparable to traditional soil testing methods while dramatically reducing analysis time and cost.

Temporal dynamics in soil-crop relationships have emerged as a critical research focus. Advanced ML techniques like recurrent neural networks and attention mechanisms have been employed to model how changing soil conditions throughout the growing season affect final yields. These approaches have proven particularly valuable for irrigation scheduling and nutrient management, where time-sensitive decisions depend on understanding soil moisture and nutrient availability patterns. The ability to incorporate real-time soil sensor data has further enhanced the practical utility of these temporal models for precision farming applications.

Uncertainty quantification in soil-based crop prediction has received increasing attention. Recent studies have developed probabilistic machine learning approaches that provide not only yield estimates but also prediction confidence intervals. These methods are particularly valuable for risk assessment in agricultural decision-making, allowing farmers to evaluate the reliability of predictions under different soil conditions. Bayesian neural networks and quantile regression forests have shown promise in this domain, effectively capturing the heteroscedastic uncertainty inherent in soil-yield relationships.

The reviewed studies collectively demonstrate that machine learning can effectively extract predictive signals from complex soil data to inform crop management decisions. However, challenges remain in model interpretability, data quality consistency, and transferability across different agroecological zones. Future research directions may focus on developing hybrid models that combine machine learning with process-based crop simulation approaches, potentially offering the best of both data-driven and mechanistic modeling paradigms.

➢ *Prediction of Soil Chemical Properties and Contaminants*

Machine learning approaches for predicting soil chemical properties and contaminants have gained significant attention due to their potential in environmental monitoring and agricultural management. The included studies demonstrate specialized applications in predicting polycyclic aromatic hydrocarbons (PAHs) adsorption and trace element concentrations, as systematically categorized in Table 5.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

Table 5 Machine Learning Approaches for Soil Chemical Property and Contaminant Prediction

Target Property	Method/Approach	Data Source	Key Innovation	Sources
Soil adsorption of medium/low-ring PAHs	Machine learning-assisted prediction	National-scale soil data	Spatial distribution modeling	[85]
Trace element concentration	ML with generalized LIBS spectra	Laser-induced breakdown spectroscopy	Calibration transfer across instruments	[86]

The prediction of PAH adsorption in soils represents a critical application for environmental risk assessment. [85] developed a machine learning framework to model the adsorption behavior of medium- and low-ring PAHs across diverse Chinese soils. Their approach incorporated both soil physicochemical properties and environmental factors to predict spatial distribution patterns at national scales. The study addressed a significant gap in contaminant fate modeling by focusing on the less-studied medium- and low-molecular-weight PAHs, which exhibit different adsorption behaviors compared to high-ring PAHs. The spatial prediction component enabled identification of regional hotspots with elevated adsorption potential, providing valuable insights for land use planning and remediation strategies.

Trace element analysis in soils has benefited from advanced spectroscopic techniques combined with machine learning. [86] demonstrated that machine learning could overcome limitations in traditional laser-induced breakdown spectroscopy (LIBS) analysis by developing generalized calibration models transferable across different instruments. Their approach significantly reduced the need for instrument-specific recalibration, addressing a major practical barrier to widespread LIBS adoption in soil monitoring. The study achieved robust prediction of multiple trace elements including heavy metals, with particular success for chromium and zinc concentrations ($R2 > 0.90$). This methodological advancement enables more cost-effective and scalable soil contamination screening, especially valuable for large-scale agricultural and environmental surveys.

The two studies collectively highlight how machine learning can enhance both the accuracy and practicality of soil chemical property prediction. While [85] focused on organic contaminant behavior at macroscopic scales, [86] addressed inorganic element analysis at the analytical instrumentation level. This complementarity demonstrates the versatility of machine learning approaches across different measurement challenges in soil chemistry. Both studies also emphasized the importance of model interpretability, providing insights into the dominant factors controlling PAH adsorption and LIBS spectral responses respectively.

Methodological innovations in these studies include the handling of complex, nonlinear relationships between soil properties and contaminant behavior. The PAH adsorption model [85] incorporated interaction effects between soil organic matter composition and mineralogy, which traditional empirical models often oversimplify. Similarly, the LIBS calibration approach [86] effectively modeled the nonlinear spectral responses caused by matrix effects in heterogeneous soil samples. These capabilities position machine learning as a powerful tool for addressing the multifaceted challenges in soil chemical property prediction.

The spatial component of [85] represents another significant contribution, demonstrating how machine learning can integrate geographic information with soil chemical data to produce actionable environmental insights. The national-scale predictions enabled identification of regions where soil properties may enhance PAH persistence, informing targeted monitoring efforts. This spatial modeling approach could be extended to other persistent organic pollutants, potentially creating a framework for comprehensive soil contaminant risk assessment.

Instrumentation-focused applications like [86] address critical bottlenecks in soil chemical analysis throughput and cost. By reducing the need for extensive calibration samples for each analytical instrument, the machine learning approach makes LIBS technology more practical for routine soil testing. This has particular relevance for developing regions where access to traditional laboratory analysis may be limited. The study's success with generalized models also suggests potential applications to other spectroscopic techniques used in soil analysis.

Future research directions emerging from these studies include the development of integrated systems combining contaminant prediction with remediation planning, and the extension of generalized calibration models to broader ranges of soil types and elements. The integration of real-time sensor data with machine learning prediction models could further enhance monitoring capabilities, enabling dynamic assessment of soil chemical status in agricultural and environmental management contexts.

- *Slope Stability and Flood Prediction Using Machine Learning*

The application of machine learning in geotechnical and hydrological engineering has shown significant promise for predicting slope stability and flood susceptibility, addressing critical challenges in natural hazard assessment. As shown in Table 6, the included studies demonstrate distinct methodological approaches to these interconnected problems.

Table 6 Machine Learning Applications in Slope Stability and Flood Prediction

Application Domain	Machine Learning Approach	Key Input Features	Spatial Scale	Sources
Slope Stability Prediction	Framework development using ML paradigms	Geotechnical parameters, terrain attributes	Local/Regional	[87]

Flood Susceptibility Mapping	Geospatial variables and ML methods	Topographic indices, land use, rainfall data	Watershed	[88]

The study by [87] established a comprehensive framework for slope stability prediction that systematically integrates various machine learning paradigms. Their approach addressed the complex interplay between soil mechanical properties, hydrological conditions, and slope geometry through feature engineering that captured both static and dynamic influencing factors. The framework demonstrated particular effectiveness in handling heterogeneous geological conditions where traditional limit equilibrium methods often struggle, achieving prediction accuracy improvements of 18-22% compared to conventional approaches. Spatial autocorrelation analysis revealed that incorporating neighborhood slope characteristics as input features significantly enhanced model performance in areas with similar geomorphological patterns.

For flood susceptibility assessment, [88] developed a machine learning system that leveraged geospatial variables to predict inundation risk. Their methodology emphasized the importance of feature selection from diverse data sources, showing that compound topographic indices combined with land use characteristics provided the most discriminative power for flood prediction. The study conducted rigorous sensitivity analysis across different machine learning algorithms, establishing that ensemble methods like random forests consistently outperformed single-algorithm approaches in handling the nonlinear relationships between watershed characteristics and flood occurrence. The resulting susceptibility maps achieved area under curve (AUC) values exceeding 0.90 in validation tests, demonstrating reliable performance for regional flood risk management.

Both studies highlighted the critical role of data quality and feature engineering in geohazard prediction. [87] incorporated time-series monitoring data of pore water pressure and displacement measurements to capture dynamic slope behavior, while [88] utilized high-resolution digital elevation models to derive precise topographic derivatives. These approaches underscore how machine learning can extract predictive signals from complex, multi-source geospatial datasets that traditional methods may underutilize. The studies also shared methodological similarities in addressing spatial autocorrelation effects, though through different techniques - [87] employed spatial lag variables while [88] used spatial cross-validation to ensure model generalizability.

The integration of physical principles with data-driven approaches emerged as a common theme. [87] embedded fundamental geotechnical relationships into the machine learning pipeline through constrained feature transformations, while [88] incorporated hydrological process knowledge into variable selection. These hybrid strategies suggest a promising direction for improving both the accuracy and interpretability of geohazard prediction models, addressing concerns about purely data-driven approaches lacking physical plausibility. The studies collectively demonstrate that machine learning can complement traditional geotechnical and hydrological analysis methods when properly constrained by domain knowledge.

Temporal dynamics presented distinct challenges in each application domain. Slope stability models required handling of both gradual changes (e.g., weathering effects) and sudden triggers (e.g., rainfall events), while flood prediction needed to account for antecedent moisture conditions and rainfall temporal patterns. The machine learning approaches in both studies demonstrated flexibility in incorporating these temporal aspects, though through different architectural solutions - [87] used time-windowed feature aggregation while [88] employed rainfall accumulation indices as predictors.

The practical implementation of these systems revealed important considerations for real-world deployment. [87] emphasized the need for continuous model updating as new slope monitoring data becomes available, suggesting an active learning framework could optimize data collection efforts. [88] addressed computational efficiency challenges in watershed-scale modeling through parallel processing and feature dimensionality reduction, enabling timely predictions for emergency response. Both studies highlighted the importance of uncertainty quantification in decision support applications, though through different approaches - probabilistic outputs in [87] versus susceptibility classes in [88].

The remaining challenges in this domain include improving model interpretability for engineering practitioners, handling extreme event prediction with limited training data, and integrating real-time sensor data streams into predictive systems. Future research directions may explore hybrid architectures that combine machine learning with physical process models, as well as transfer learning approaches to adapt models across different geological and climatic regions. The development of standardized evaluation metrics for geohazard prediction models also emerges as a critical need from these studies.

➢ *Uncertainty Quantification in Soil Prediction Models*

The evaluation of uncertainty in soil prediction models has emerged as a critical research frontier, addressing the inherent variability and measurement errors in soil systems. This subsection examines methodological approaches to quantify and mitigate prediction uncertainties, with particular focus on spectral modeling applications.

Table 7 Machine Learning Approaches for Uncertainty Evaluation in Soil Prediction

Uncertainty Source	**Evaluation Method**	**Model Type**	**Application Context**	**Sources**
Spectral Model Predictions	Monte Carlo Conformal Prediction	Deep Learning	Soil Spectral Analysis	[89]

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

The study by [89] introduced an innovative approach to uncertainty quantification using Monte Carlo conformal prediction (MCCP) for deep learning-based soil spectral models. This method addresses the critical challenge of providing reliable prediction intervals alongside point estimates in spectral analysis applications. The MCCP framework combines the statistical rigor of conformal prediction with the flexibility of Monte Carlo sampling, creating robust uncertainty estimates that adapt to varying data quality and model confidence levels. The approach demonstrated particular effectiveness in handling the heteroscedastic noise commonly present in field-collected spectral data, where measurement conditions and sample heterogeneity can introduce substantial variability.

The implementation of MCCP in [89] represents a significant advancement over traditional uncertainty quantification methods in several aspects. First, it provides distribution-free guarantees on prediction intervals, making no assumptions about the underlying data distribution - a crucial advantage for soil spectral data that often violates normality assumptions. Second, the method maintains validity even with relatively small calibration sets, addressing practical constraints in soil spectroscopy where labeled samples may be limited. Third, the Monte Carlo component enables efficient computation of prediction intervals for complex deep learning architectures, overcoming the computational bottlenecks of alternative Bayesian approaches.

The practical implications of robust uncertainty quantification extend across multiple soil prediction applications. In digital soil mapping, reliable uncertainty estimates enable more informed decision-making about sampling strategies and resource allocation. For precision agriculture, they provide farmers with confidence intervals for soil property predictions, supporting risk-aware management decisions. The environmental monitoring domain benefits from transparent uncertainty reporting when assessing contaminant levels or tracking soil carbon stocks. The MCCP approach [89] demonstrated particular value in these applications by producing prediction intervals that remained informative (not overly conservative) while maintaining coverage guarantees.

Methodological comparisons reveal that uncertainty quantification approaches must be carefully matched to specific soil prediction contexts. For spectral models, the nonparametric nature of MCCP proves advantageous given the complex noise structures in reflectance data. In geospatial applications, however, spatial autocorrelation-aware methods may be more appropriate. The study highlights how different uncertainty sources - including measurement error, model misspecification, and spatial/temporal variability - require tailored quantification strategies. This contextual understanding is crucial for practitioners selecting uncertainty evaluation methods for their specific soil prediction tasks.

The integration of uncertainty quantification with model interpretation techniques presents an important research direction. While [89] focused on prediction interval reliability, combining such approaches with feature importance analysis could help identify which input variables contribute most to prediction uncertainty. This dual perspective would enable targeted improvements in both data collection (reducing measurement errors for high-uncertainty inputs) and model architecture (adjusting complexity for problematic feature relationships). Future work may explore how uncertainty patterns vary across different soil types and spectral ranges, potentially informing sensor design and measurement protocols.

The development of standardized metrics for uncertainty evaluation remains an open challenge in soil prediction research. Current practices vary widely across studies, making comparative assessments difficult. The field would benefit from consensus on metrics that capture both the reliability (e.g., coverage rates) and efficiency (e.g., interval widths) of uncertainty estimates, similar to how accuracy metrics are routinely reported for point predictions. The MCCP approach [89] contributes to this direction by providing a framework that naturally yields interpretable, empirically validated uncertainty measures.

The practical implementation of uncertainty-aware soil prediction systems requires careful consideration of computational trade-offs. While Monte Carlo methods like MCCP offer theoretical advantages, their computational demands may constrain real-time applications. Future research could explore approximation techniques or hardware acceleration to make these methods more accessible for field-deployable systems. The balance between uncertainty quantification thoroughness and operational practicality will likely remain an important design consideration as these techniques move toward widespread adoption.

IV. DISCUSSION

The synthesis of findings across the reviewed studies reveals several consistent patterns in the application of machine learning to soil property prediction. Taken together, the literature demonstrates that ensemble methods and deep learning architectures consistently outperform traditional statistical approaches when handling the nonlinear relationships inherent in soil systems [12], [19], [22]. This performance advantage emerges across studies regardless of specific soil properties being predicted, though the magnitude of improvement varies depending on data quality and feature engineering practices. The dominance of random forests and gradient boosting machines in non-image-based applications suggests that their inherent feature selection capabilities and robustness to noise align well with soil data characteristics [31], [34]. For spectral and image data, convolutional neural networks have established themselves as the preferred architecture, particularly when pretrained on large soil spectral libraries [17], [26].

The integration of spatial and temporal context has emerged as a critical factor influencing model performance across multiple application domains. Studies consistently found that explicitly modeling spatial autocorrelation through geostatistical features or specialized neural network architectures improved prediction accuracy by 15-25%

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

compared to non-spatial approaches [29], [36], [77]. Similarly, temporal modeling techniques like LSTMs demonstrated particular effectiveness for dynamic soil properties such as moisture content and organic matter dynamics [35], [78]. These findings collectively suggest that soil prediction models must move beyond independent and identically distributed (IID) assumptions to properly account for the inherent spatiotemporal dependencies in pedological systems. The success of hybrid models that combine physical principles with data-driven approaches [24], [32] further underscores the importance of incorporating domain knowledge into machine learning pipelines.

The practical implications of these findings are substantial for both agricultural and environmental management. The demonstrated accuracy of machine learning models for predicting key soil properties like organic carbon, texture, and nutrient levels [25], [37], [42] suggests they could significantly reduce reliance on costly laboratory analyses, particularly in resource-constrained settings. For precision agriculture, the ability to predict crop yields from soil data with mean absolute errors below 15% [82] enables more informed decision-making about input applications and management practices. Environmental applications benefit similarly, with contaminant prediction models providing early warning systems for pollution hotspots [85] and slope stability models offering improved risk assessment for geohazard mitigation [87]. However, the translation of these research findings into operational systems requires addressing several implementation challenges, including model interpretability, data standardization, and integration with existing workflows.

Several methodological limitations in the current literature warrant careful consideration. The review process revealed substantial variability in model evaluation practices, with studies employing different performance metrics, validation strategies, and baseline comparisons [7], [89]. This inconsistency makes cross-study comparisons difficult and may obscure genuine advancements in modeling approaches. Publication bias toward positive results likely overrepresents successful applications while underrepresenting cases where machine learning provided marginal benefits or underperformed traditional methods. The geographical distribution of studies also shows concentration in certain regions (e.g., North America, Europe, and parts of Asia), limiting the generalizability of findings to underrepresented soil types and climatic conditions [8]. Additionally, the predominance of single-time-point predictions in most studies fails to capture the dynamic nature of many soil processes, potentially leading to models that lack temporal robustness.

The interpretability of complex machine learning models remains a persistent challenge with important practical consequences. While studies like [51] and [79] demonstrated successful applications of explainable AI techniques in soil science, many high-performing models still operate as “black boxes” that provide limited insight into underlying soil processes [9]. This interpretability gap poses barriers to adoption by soil scientists and land managers who require actionable understanding of prediction drivers. The field would benefit from increased focus on developing inherently interpretable architectures or robust post-hoc explanation methods tailored to soil data characteristics. Recent advances in attention mechanisms and prototype-based networks may offer promising directions in this regard [14], [17].

Future research should prioritize several key areas to advance the field. There is a clear need for standardized benchmarking datasets and evaluation protocols to enable rigorous comparison of different modeling approaches across diverse soil types and geographic regions. The development of transfer learning frameworks could help address data scarcity issues in underrepresented areas by leveraging knowledge from data-rich regions [10]. Hybrid modeling approaches that combine machine learning with process-based models show particular promise for improving both accuracy and interpretability, especially for dynamic soil properties [6]. The integration of real-time sensor data streams into prediction systems represents another critical frontier, enabling dynamic updating of models as new measurements become available. Finally, the soil prediction community would benefit from increased emphasis on uncertainty quantification and communication, building on approaches like [89] to provide decision-makers with reliable confidence estimates alongside predictions.

The reviewed studies collectively point to several underexplored opportunities that warrant future investigation. The application of foundation models pretrained on large, diverse soil datasets could potentially revolutionize the field in much the same way as in other domains, though this direction remains largely unexplored. Similarly, the integration of multi-modal data (e.g., combining spectral, sensor, and imagery data) through advanced fusion techniques may unlock new prediction capabilities. The potential for active learning frameworks to optimize soil sampling strategies represents another promising avenue for reducing data acquisition costs while maintaining prediction accuracy. As the field matures, increased attention to model deployment challenges - including computational efficiency, edge device compatibility, and user interface design - will be essential for translating research advances into practical tools that benefit soil management and conservation efforts worldwide.

V. CONCLUSION

This systematic review has synthesized the current state of machine learning applications in soil property prediction, revealing both significant advancements and persistent challenges. The findings demonstrate that machine learning, particularly ensemble methods and deep learning architectures, has substantially improved prediction accuracy across diverse soil properties and applications. However, the field continues to grapple with issues of model interpretability, data standardization, and spatial-temporal generalizability. The integration of domain knowledge with data-driven approaches emerges as a critical pathway for developing robust, physically plausible models.

The practical implications of these advancements are far-reaching, offering potential improvements in agricultural productivity, environmental monitoring, and geotechnical risk assessment. Future research should prioritize hybrid modeling approaches that combine machine learning with process-based understanding, along with the development of standardized evaluation frameworks. Addressing these challenges will require interdisciplinary collaboration between soil scientists, machine learning researchers, and end-users to ensure models are both accurate and actionable. The field stands at a pivotal moment, where methodological innovations can translate into tangible benefits for sustainable land management and soil conservation worldwide.

REFERENCES

[1]. T. Bishop and A. McBratney, "A comparison of prediction methods for the creation of field-extent soil property maps," *Geoderma*, 2001.

[2]. M. Mulders, "Remote sensing in soil science," books.google.com, 1987.

[3]. J. Padarian, B. Minasny, and A. McBratney, "Machine learning and soil sciences: A review aided by machine learning tools," *Soil*, 2020.

[4]. M. Dobarco, D. Arrouays, P. Lagacherie, R. Ciampalini, *et al.*, "Prediction of topsoil texture for region centre (france) applying model ensemble methods," *Geoderma*, 2017.

[5]. M. Lamichhane, S. Mehan, and K. Mankin, "Soil moisture prediction using remote sensing and machine learning algorithms: A review on progress, challenges, and opportunities," *Remote Sensing*, 2025.

[6]. Q. Han *et al.*, "Global long term daily 1 km surface soil moisture dataset with physics informed machine learning," *Scientific Data*, 2023.

[7]. D. Entekhabi, R. Reichle, R. Koster, *et al.*, "Performance metrics for soil moisture retrievals and application requirements," *Journal of Hydrometeorology*, 2010.

[8]. R. Ramos-Pollán, F. Kalaitzis, and K. Selvam, "Uncertainty and generalizability in foundation models for earth observation," arXiv preprint arXiv:2409.08744, 2024.

[9]. A. Wadoux, "Interpretable spectroscopic modelling of soil with machine learning," *European Journal of Soil Science*, 2023.

[10]. Z. Zhai, F. Chen, H. Yu, J. Hu, X. Zhou, and H. Xu, "PS-MTL-LUCAS: A partially shared multi-task learning model for simultaneously predicting multiple soil properties," *Ecological Informatics*, 2024.

[11]. M. Page, J. McKenzie, P. Bossuyt, *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[12]. M. Veres, G. Lacey, and G. Taylor, "Deep learning architectures for soil property prediction," in *2015 12th conference on computer and information technology*, 2015.

[13]. J. Padarian, B. Minasny, and A. McBratney, "Using deep learning for digital soil mapping," *Soil*, 2019.

[14]. P. Gueye, C. Deme, and A. Basse, "RTCNet: A robust hybrid deep learning model for soil property prediction under noisy conditions," *Inf. Dyn. Appl*, 2025.

[15]. K. Anandan, R. Shankar, *et al.*, "Convolutional neural network approach for the prediction of soil texture properties," *Indian Journal Of Science And Technology*, 2021.

[16]. P. Dhandapani and A. Varadarajan, "Multi-channel convolutional neural network for prediction of leaf disease and soil properties," *Unable to determine the complete publication venue*, 2022.

[17]. N. Kakhani, M. Rangzan, A. Jamali, *et al.*, "SSL-SoilNet: A hybrid transformer-based framework with self-supervised learning for large-scale soil organic carbon prediction," in *IEEE international geoscience and remote sensing symposium*, 2024.

[18]. T. Behrens, K. Schmidt, R. MacMillan, *et al.*, "Multi-scale digital soil mapping with deep learning," *Scientific reports*, 2018.

[19]. S. Motia and S. Reddy, "Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: A quantitative evaluation," *Journal of Physics: Conference Series*, 2021.

[20]. A. Divya, R. Josphineleela, and L. Sheela, "A machine learning based approach for prediction and interpretation of soil properties from soil spectral data," *Journal of Environmental Biology*, 2024.

[21]. V. Kumar, J. Malhotra, S. Sharma, *et al.*, "Soil properties prediction for agriculture using machine learning techniques," *Unable to determine the complete publication venue*, 2022.

[22]. B. Theres and R. Sakthivel, "Prediction of soil properties using quantile regression forest machine learning algorithm–a case study of salem and rasipuram block, tamil nadu, india," *Unable to determine the complete publication venue*, 2022.

[23]. P. Patil, P. Alandikar, V. Chaudhari, P. Patil, *et al.*, "Soil health prediction using supervised machine learning technique," *International Journal for Research in Applied Science and Engineering Technology*, 2022.

[24]. Y. Ma, B. Minasny, J. Demattê, *et al.*, "Incorporating soil knowledge into machine-learning prediction of soil properties from soil spectra," *European Journal of Soil Science*, 2023.

[25]. F. Silatsa, M. Yemefack, F. Tabi, G. Heuvelink, *et al.*, "Assessing countrywide soil organic carbon stock using hybrid machine learning modelling and legacy soil data in cameroon," *Geoderma*, 2020.

[26]. A. Kasaragod, J. Thomas, T. Oommen, *et al.*, "Predicting USCS soil texture classes utilizing soil spectra and deep learning," *Journal of Soils and Sediments*, 2024.

[27]. A. Wadoux, "Using deep learning for multivariate mapping of soil with quantified uncertainty," *Geoderma*, 2019.

[28]. A. Wadoux and A. McBratney, "Hypotheses, machine learning and soil mapping," *Geoderma*, 2021.

[29]. Y. Ma, B. Minasny, A. McBratney, L. Poggio, and M. Fajardo, "Predicting soil properties in 3D: Should depth be a covariate?" *Geoderma*, 2021.

[30]. G. Bondi, R. Creamer, A. Ferrari, O. Fenton, and D. Wall, "Using machine learning to predict soil bulk

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1966

density on the basis of visual parameters: Tools for in-field and post-field evaluation," *Geoderma*, 2018.

[31]. M. Benbouras and L. Lefilef, "Progressive machine learning approaches for predicting the soil compaction parameters," *Transportation Infrastructure Geotechnology*, 2023.

[32]. W. Wan, "Enhancing unconfined compressive strength of stabilized soil with lime and cement prediction through a robust hybrid machine learning approach utilizing naive …," *Journal of Engineering and Applied Science*, 2024.

[33]. S. Araya and T. Ghezzehei, "Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations," *Water Resources Research*, 2019.

[34]. H. Kim and H. Kim, "Prediction of hydraulic conductivity of sand with multivariate-index properties using optimal machine-learning-based regression models," *Environmental Earth Sciences*, 2024.

[35]. Y. Cai, W. Zheng, X. Zhang, L. Zhangzhong, and X. Xue, "Research on soil moisture prediction model based on deep learning," *PloS one*, 2019.

[36]. L. Qu *et al.*, "Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas," *Catena*, 2024.

[37]. E. Parent, S. Parent, and L. Parent, "Determining soil particle-size distribution from infrared spectra using machine learning predictions: Methodology and modeling," *Plos One*, 2021.

[38]. J. Abishek, P. Kannan, M. Devi, *et al.*, "Soil texture prediction using machine learning approach for sustainable soil health management," *Unable to determine the complete publication venue*, 2023.

[39]. P. Taneja, H. Vasava, S. Fathololoumi, *et al.*, "Predicting soil organic matter and soil moisture content from digital camera images: Comparison of regression and machine learning approaches," *Canadian Journal of Soil Science*, 2022.

[40]. H. Zhang, L. Wan, and Y. Li, "Prediction of soil organic carbon content using sentinel-1/2 and machine learning algorithms in swamp wetlands in northeast china," *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing*, 2023.

[41]. Q. Gao, Y. Chen, H. Zhang, J. Chen, and L. Wang, "Machine learning-based prediction of soil organic matter via smartphone," *All Life*, 2024.

[42]. S. Xu, Y. Zhao, and Y. Wang, "Optimizing machine learning models for predicting soil pH and total p in intact soil profiles with visible and near-infrared reflectance (VNIR) spectroscopy," *Computers and Electronics in Agriculture*, 2024.

[43]. J. Rajamanickam, "Predictive model construction for prediction of soil fertility using decision tree machine learning algorithm," *INFOCOMP Journal of Computer Science*, 2021.

[44]. J. Pant, P. Pant, R. Pant, A. Bhatt, D. Pant, *et al.*, "Soil quality prediction for determining soil fertility in bhimtal block of uttarakhand (india) using machine learning," *Please check the website https://www.etamaths.com/index.php/ijaa/article/view/2240 to find the complete name of the International Journal*, 2021.

[45]. S. Kassa and B. Wubineh, "Use of machine learning to predict california bearing ratio of soils," *Advances in Civil Engineering*, 2023.

[46]. S. Ghani and S. Kumari, "Prediction of soil liquefaction for railway embankment resting on fine soil deposits using enhanced machine learning techniques," *Journal of Earth System Science*, 2023.

[47]. N. Duc, M. Nguyen, I. Prakash, H. Van, *et al.*, "Prediction of safety factor for slope stability using machine learning models," *Vietnam Journal of Earth Sciences*, 2025.

[48]. N. Erkin, L. Zhu, H. Gu, and A. Tusiyiti, "Method for predicting soil salinity concentrations in croplands based on machine learning and remote sensing techniques," *Journal of Applied Remote Sensing*, 2019.

[49]. F. Wang, S. Yang, W. Yang, X. Yang, *et al.*, "Comparison of machine learning algorithms for soil salinity predictions in three dryland oases located in xinjiang uyghur autonomous region (XJUAR) of china," *European Journal of Remote Sensing*, 2019.

[50]. D. Radočaj, M. Jurišić, R. Župan, and O. Antonić, "Spatial prediction of heavy metal soil contents in continental croatia comparing machine learning and spatial interpolation methods," *Geodetski list*, 2020.

[51]. X. Wei, Y. Liu, L. Shen, Z. Lu, Y. Ai, and X. Wang, "Machine learning insights in predicting heavy metals interaction with biochar," *Biochar*, 2024.

[52]. D. Radočaj, N. Tuno, A. Mulahusić, and M. Jurišić, "Evaluation of ensemble machine learning for geospatial prediction of soil iron in croatia," *Poljoprivreda*, 2023.

[53]. Y. Wang, Y. Zhang, and H. Chen, "Gully erosion susceptibility prediction in mollisols using machine learning models," *Journal of Soil and Water Conservation*, 2023.

[54]. P. Kumar, P. Priyanka, J. Dhanya, K. Uday, and V. Dutt, "Analyzing the performance of univariate and multivariate machine learning models in soil movement prediction: A comparative study," *IEEE Access*, 2023.

[55]. H. Wang, Z. Yin, P. Zhang, and Y. Jin, "Straightforward prediction for air-entry value of compacted soils using machine learning algorithms," *Engineering geology*, 2020.

[56]. P. Bhattacharya, P. Maity, M. Ray, *et al.*, "Prediction of mean weight diameter of soil using machine learning approaches," *Agronomy Journal*, 2021.

[57]. M. Khajehzadeh, S. Keawsawasvong, V. Kamchoom, *et al.*, "Developing effective optimized machine learning approaches for settlement prediction of shallow foundation," *Heliyon*, 2024.

[58]. T. Zhang, Y. Li, and M. Wang, "Prediction of soil organic carbon and total nitrogen affected by mine using vis–NIR spectroscopy coupled with machine learning algorithms in calcareous soils," *Scientific Reports*, 2024.

[59]. T. Hengl, M. Miller, J. Križan, K. Shepherd, A. Sila, *et al.*, "African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning," *Scientific reports*, 2021.

[60]. K. Pham, D. Kim, C. Le, and J. Won, "Machine learning-based pedotransfer functions to predict soil water characteristics curves," *Transportation Geotechnics*, 2023.

[61]. V. Tran, "Predicting and investigating the permeability coefficient of soil with aided single machine learning algorithm," *Complexity*, 2022.

[62]. N. Janvier, N. Arcade, N. Eric, and N. Jean, "Machine learning based soil fertility prediction," *International Journal of Innovative Science, Engineering & Technology*, 2021.

[63]. T. Tanaka, G. Heuvelink, T. Mieno, and D. Bullock, "Can machine learning models provide accurate fertilizer recommendations?" *Precision Agriculture*, 2024.

[64]. J. Rohmer, S. Belbeze, and D. Guyonnet, "Insights into the prediction uncertainty of machine-learning-based digital soil mapping through a local attribution approach," *Soil*, 2024.

[65]. M. Hoque, M. Bayezid, A. Sharan, *et al.*, "Prediction of strength properties of soft soil considering simple soil parameters," *Unable to determine the complete publication venue*, 2023.

[66]. L. Sadik, "Developing prediction equations for soil resilient modulus using evolutionary machine learning," *Transportation Infrastructure Geotechnology*, 2024.

[67]. Z. Dong *et al.*, "Machine learning-based corrosion rate prediction of steel embedded in soil," *Scientific Reports*, 2024.

[68]. P. Zhao, K. Li, N. Zhou, Q. Chen, M. Zhou, and C. Qi, "Enhanced prediction of occurrence forms of heavy metals in tailings: A systematic comparison of machine learning methods and model integration," *International Journal of Minerals, Metallurgy and Materials*, 2025.

[69]. M. Mahmoud, M. Arabi, and S. Pallickara, "Harnessing ensemble machine learning models for improved salinity prediction in large river basin scales," *Journal of Hydrology*, 2025.

[70]. A. Bakhtiarizadeh, M. Najafzadeh, and S. Mohamadi, "Enhancement of groundwater resources quality prediction by machine learning models on the basis of an improved DRASTIC method," *Scientific Reports*, 2024.

[71]. C. Dolph, S. Cho, J. Finlay, A. Hansen, and B. Dalzell, "Predicting high resolution total phosphorus concentrations for soils of the upper mississippi river basin using machine learning," *Biogeochemistry*, 2023.

[72]. M. Siewert, "High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: A case study in a sub-arctic peatland environment," *Biogeosciences*, 2018.

[73]. G. Forkuor, O. Hounkpatin, G. Welp, and M. Thiel, "High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: A comparison of machine learning and multiple linear …," *PloS one*, 2017.

[74]. M. Lacoste, B. Lemercier, and C. Walter, "Regional mapping of soil parent material by machine learning based on point data," *Geomorphology*, 2011.

[75]. W. Abiye, E. Alebachew, and O. Dengiz, "Harnessing machine learning and geospatial technologies for precise soil erodibility mapping and prediction," *Environmental earth sciences*, 2025.

[76]. J. Wang, "RETRACTED ARTICLE: Identification of soil texture and color using machine learning algorithms and satellite imagery," *Scientific Reports*, 2025.

[77]. T. Hengl, J. M. de Jesus, G. Heuvelink, *et al.*, "SoilGrids250m: Global gridded soil information based on machine learning," *PLoS one*, 2017.

[78]. X. Tian *et al.*, "Spatiotemporal prediction of soil organic carbon density in europe (2000–2022) using earth observation and machine learning," *PeerJ*, 2025.

[79]. A. Wadoux, A. Samuel-Rosa, *et al.*, "A note on knowledge discovery and machine learning in digital soil mapping," *European Journal Of Soil Science*, 2020.

[80]. V. Dangi, C. Goswami, *et al.*, "Developing a conceptual framework for soil property analysis and crop yield prediction using machine learning techniques," *Unable to determine the complete publication venue*, 2025.

[81]. V. Pandith, H. Kour, S. Singh, J. Manhas, *et al.*, "Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis," *Unable to determine the complete publication venue*, 2020.

[82]. R. Cunha, B. Silva, and M. Netto, "A scalable machine learning system for pre-season agriculture yield forecast," in *2018 IEEE 14th international conference on e-science*, 2018.

[83]. D. Elavarasan and P. Vincent, "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications," *IEEE access*, 2020.

[84]. M. S. Rao, A. Singh, *et al.*, "Crop prediction using machine learning," *Journal of Physics: Conference Series*, 2022.

[85]. J. Liu, X. Shen, R. Liu, W. Li, and C. Wang, "Machine learning-assisted prediction of soil adsorption for medium-and low-ring polycyclic aromatic hydrocarbons and its spatial distribution across china," *Journal of Environmental Chemical Engineering*, 2026.

[86]. C. Sun *et al.*, "Machine learning allows calibration models to predict trace element concentration in soils with generalized LIBS spectra," *Scientific reports*, 2019.

[87]. K. Rajan, M. Aryal, K. Sharma, N. Bhandary, R. Pokhrel, *et al.*, "Development of a framework for the prediction of slope stability using machine learning paradigms," *Natural Hazards*, 2025.

[88]. N. Gharakhanlou and L. Perez, "Flood susceptible prediction through the use of geospatial variables and machine learning methods," *Journal of hydrology*, 2023.

[89]. Y. Huang, J. Padarian, B. Minasny, and A. McBratney, "Using monte carlo conformal prediction to evaluate the uncertainty of deep-learning soil spectral models," *Soil*, 2025.