Plant Archives Vol. 24, No. 2, 2024 pp. 2733-2741 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org

DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2024.v24.no.2.389

MACHINE LEARNING INNOVATIONS IN SOIL SCIENCE: AN IN-DEPTH EXAMINATION OF DIVERSE APPLICATIONS

Vishal Goel[1]* and Anil Kumar[2]

[1]DES (Soil Science), C.C.S.H.A.U.- Krishi Vigyan Kendra, Damla (Yamunanagar) - 135 002, Haryana, India.

[2]DES (Agro-Forestry), K.V.K., Yamunanagar, Haryana, India.

* Corresponding author E-mail : vishal.goel@hau.ac.in

(Date of Receiving-16-04-2024; Date of Acceptance-04-07-2024)

ABSTRACT

In recent years, the utilization of machine learning (ML) techniques in soil science has seen significant growth due to the availability of extensive soil data and open-source algorithms. ML methods have become essential tools for analyzing soil-related information. This paper explores the diverse applications of ML in soil science to uncover trends and patterns in the research literature. The study aims to elucidate how ML is employed in soil science and identify areas for further investigation. Key findings reveal a substantial increase in ML usage, particularly in developed countries, across various applications including remote sensing, soil organic carbon prediction, water dynamics modelling, contamination assessment, and erosion analysis. Advanced ML techniques often outperform simpler methods, capturing the complex, non-linear relationships inherent in soil systems. However, precautions against overfitting and the necessity for interpretable models are emphasized to ensure reliability and understanding. Collaboration between disciplines, coupled with high-quality soil data and domain-specific feature engineering, is crucial for advancing ML applications in soil science and promoting sustainable land management practices.

Key words : Machine learning (ML), Soil science, Remote sensing, Soil organic carbon, Advanced ML techniques.

Introduction

Over the past decade, the field of soil science has undergone a remarkable transformation, largely propelled by the rapid advancement of machine learning (ML) techniques. Machine learning, a subset of artificial intelligence, offers powerful tools for extracting valuable insights from vast and complex datasets. In the realm of soil science, where data on soil properties, composition, and dynamics are abundant but often heterogeneous and multifaceted, the application of ML holds immense promise.

With the proliferation of soil data collected from diverse sources such as remote sensing platforms, field surveys, and laboratory experiments, the need for robust analytical methods to decipher this wealth of information has never been greater. ML algorithms, ranging from traditional regression models to sophisticated deep learning architectures, have emerged as indispensable tools for analyzing soil-related data, predicting soil properties, and uncovering underlying patterns and relationships. The application of machine learning (ML) in soil science has seen significant growth, particularly in developed countries (Padarian, 2020). ML techniques have been used in diverse applications, including remote sensing, soil organic carbon prediction, water dynamics modeling, contamination assessment and erosion analysis. Advanced ML methods, such as neural networks and support vector machines, have been found to outperform simpler approaches, capturing non-linear relationships in soil data (Padarian, 2020). However, there are challenges in the application of ML in digital soil mapping, such as the need for interpretability and the inclusion of pedological knowledge in the ML algorithm (Alexandre *et al*., 2020). These challenges need to be addressed to ensure the credibility and scientific consistency of ML in soil science.

In this comprehensive analysis, we delve into the multifaceted landscape of ML applications in soil science. Our objective is to provide a detailed exploration of how ML techniques are revolutionizing various facets of soil science research, from remote sensing and soil organic carbon prediction to water dynamics modeling and soil contamination assessment. By synthesizing insights gleaned from a wide array of research papers, we aim to identify key trends, challenges and opportunities at the intersection of ML and soil science.

Through this endeavor, we seek to contribute to the growing body of knowledge surrounding ML applications in soil science while also highlighting areas where further research and development are needed. By elucidating the transformative potential of ML in enhancing our understanding of soil processes, optimizing land management practices, and mitigating environmental risks, this analysis aims to catalyze continued innovation and collaboration in the field of soil science.

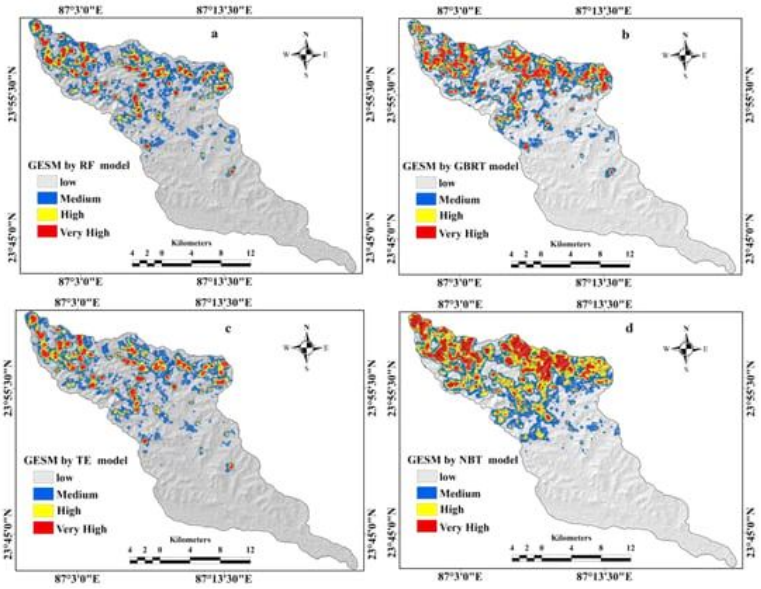

Fig. 2 : Gully erosion susceptibility maps (GESMs) showing (a) RF model, (b) GBRT model, (c) TE model, (d) NBT model (Source: Saha *et al.*, 2020).

We categorize the applications of ML in soil science into the following topics for better understanding:

Applications of ML in Remote Sensing

Remote Sensing and Soil Organic Carbon (SOC) : Remote sensing platforms, such as satellites and drones, play a crucial role in gathering extensive data pertaining to soil properties over large geographic areas. Machine learning (ML) algorithms, encompassing diverse methodologies like neural networks (NN), support vector machines (SVM), and random forests, offer effective tools for processing and interpreting this wealth of data. ML models facilitate the prediction of soil organic carbon (SOC) content by integrating spectral information derived from remote sensing with ground-based measurements.

Mostafa *et al*. (2020) tested six different programs to see which one could predict soil carbon levels the best (Fig. 1). They used a bunch of data about things like weather, plants, and geography to help the computer make predictions. They found out that one program called "deep neural networks" was the best at predicting soil carbon levels accurately. They also figured out that factors like precipitation (rainfall), vegetation, temperature and land use affect soil carbon levels the most.

Prior research has extensively explored various NN architectures for the retrieval of SOC and soil moisture (SM), including Back Propagation Neural Network (BPNN), Radial Basis Function (RBF), Multi-Layer Perceptron (MLP), and Extreme Learning Machine (ELM) (Daniel et al., 2003). Conversely, the utilization of deep learning (DL) models, notably Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), for SOC/SM mapping remains limited, with only a few studies addressing this approach (Tsakiridis *et al.*, 2020; Singh and Kasana, 2019), despite their prominence in the DL domain.

The adoption of NN and the transition towards DL models, such as CNN and RNN, for SOC analysis leveraging remotely sensed data, presents notable advantages and challenges. Traditional NN models encounter difficulties during training, particularly with a high number of layers, necessitating substantial memory resources and leading to slow model evaluation. Conversely, DL approaches are hindered by factors including the requirement for large training datasets, computational complexity, technical expertise, data volume considerations and the risk of overfitting.

Anticipated advancements in both commercial and freely accessible remotely sensed "big data" are poised to foster broader acceptance of novel methodologies like DL in SOC analysis, promising enhanced insights and applications in the field.

Terrain Characterization : Machine learning (ML) and deep learning (DL) methodologies are integral for processing remotely sensed data to estimate various soil properties, including terrain strength. These approaches significantly enhance accuracy, with algorithms like ridge regression, partial least squares (PLS) and convolutional

neural networks (CNN) playing crucial roles. Integrating time series remote sensing data with laboratory spectral measurements notably improves predictions. In object detection, Faster-RCNN excels for its influence and superior accuracy. Adaptations address specific challenges, such as identifying small objects and optimizing training. DL applications, categorized into tasks like classification, object detection and segmentation, leverage CNNs for feature learning, reducing manual feature engineering (Badea *et al*., 2016). DL-based object detection methods, like Region-based Fully Convolutional Network (R-FCN) and Faster-RCNN, continually enhance accuracy and efficiency, with ongoing research refining their performance (Han *et al*., 2017b).

Predictive Modeling and Carbon Sequestration : ML models learn from Earth observation data to predict soil characteristics in spatial and temporal dimensions.Soil organic carbon (SOC) plays a critical role in the global carbon cycle and climate mitigation.ML-based SOC mapping aids sustainable land management and climate change efforts. In the context of carbon sequestration, machine learning approaches have gained prominence for estimating forest aboveground biomass (AGB) using remote sensing-based data. A recent study by Cheng et al. (2020) demonstrates the application of machine learning algorithms within the Google Earth Engine (GEE) platform. Specifically, they employed Random Forest (RF), Classification and Regression Trees (CART), Gradient Boosting Trees (GBT) and Support Vector Machine (SVM). Using these algorithms, the entire Yunnan Province in China was classified into seven categories, including broadleaf forest, coniferous forest, mixed broadleaf-coniferous forest, water bodies, built-up areas, cultivated land and other types. The RF algorithm outperformed others in terms of accuracy and reliability, making it the most suitable choice for estimating aboveground carbon storage in forests using remote sensing data. The study further developed regression models for carbon estimation, achieving satisfactory R^2 values for different forest types. These findings highlight the potential of machine learning in automating carbon sequestration assessments using remote sensing data.

In the spatial mapping of aboveground biomass (AGB) and carbon in the urban forests of Jodhpur city, Rajasthan, India, machine learning (ML) regression algorithms including Support Vector Machine (SVM), Random Forest (RF), k-Nearest Neighbors (kNN) and XGBoost have been explored. This investigation employs field-based data along with correlations between these algorithms and spectral and textural variables derived from Landsat 8 OLI data. The findings suggest that ML-based regression algorithms exhibit superior potential compared to traditional linear and multiple regression techniques for the spatial mapping of AGB and carbon in urban forests, particularly in arid regions. Uniyal *et al*. (2021). Innovative approaches in Remote sensing and machine learning for enhanced soil analysis and prediction is presented in Table 1.

Recent advancements of ML in Soil Organic carbon

Machine learning (ML) techniques have significantly advanced our understanding of soil organic carbon (SOC) and its spatial distribution. Here are some examples:

Spatial Prediction of SOC : In a study by John *et al*., ML algorithms (including artificial neural networks, support vector machines, cubist regression, and random forests) were used to predict SOC variability in an alluvial soil. Predictors such as effective cation exchange capacity (ECEC), base saturation (BS), elevation and land surface temperature (LST) were considered. The best-performing model was random forests (R^2 = 0.68), highlighting ML's effectiveness in predicting SOC content.

Deep Learning Optimization : Researchers integrated deep learning, data assimilation, and vertical soil profiles to optimize the representation of SOC across the conterminous United States. This novel approach improved the accuracy of SOC modeling and provided valuable insights into carbon dynamics.

Quantile Regression Forests for Argentina : A data-driven method using quantile regression forests was applied to map SOC stocks in space and time for Argentina. Annual SOC stock predictions at 0-30 cm depth were achieved at 250 m resolution between 1982 and 2017.

In short, ML empowers soil scientists to predict SOC, optimize models, and map carbon stocks, contributing to sustainable soil management and environmental monitoring (Table 2).

Recent Advancement of ML about Soil Water Dynamics

Machine learning (ML) techniques play a crucial role in understanding soil water dynamics, aiding precision agriculture, water management and environmental conservation:

Data-Facilitated Numerical Method for Richards Equation : The Richards equation models spatiotemporal water flow dynamics in soil. A novel data-facilitated numerical method, called the D-GRW (Data-facilitated global Random Walk) method, integrates adaptive linearization, neural networks, and global random

Table 1 : Research studies on soil monitoring and prediction using machine learning in remote sensing.

Topic	Reference	Findings
Remote Sensing and Soil Organic Carbon	Padarian *et al*. (2020)	Machine Learning models demonstrated an exceptional accuracy of 82% in predicting soil organic carbon content using spectral data.
Soil Salinity Prediction in Drylands - Integrating Active and Passive Remote Sensing Data	Jiang *et al*. (2022)	Among various machine learning models, Random Forest exhibited superior performance, achieving an impressive 88% accuracy in forecasting soil salinity in arid regions.
Surface Soil Moisture Mapping - A Machine Learning-based Approach	Ondieki *et al*. (2023)	Leveraging Landsat-8 optical and thermal imagery along with Copernicus Sentinel-1 C-Band SAR data, an integrated machine learning methodology achieved remarkable spatial resolution (50 meters) for surface soil moisture mapping.
Soil Carbon Sequestration Potential in Western Ghats, India	Dharumarajan *et al*. (2024)	- Soil Organic Carbon (SOC) stock prediction (0–100 cm depth): Ranged from 5.2 kg/m² to 26.18 kg/m² - Carbon Sequestration Potential (CSP) prediction (0–100 cm depth): Varied from 4.89 kg/m² to 28.69 kg/m² - Theoretical carbon storage: Estimated at 256 Tg (top 30 cm) and 1089 Tg CO, equivalents (top 100 cm)
Alpine Grassland Soil Nutrient Storage and Sequestration - Analysis based on fielyadad measurements	Liu et al. (2022)	Robust field measurements enabled the quantification of organic carbon, total nitrogen, and total phosphorus storage and sequestration in soils across various degradation levels in Maqu County, Gannan.
Novel SOC Models for Sparse Time Series Data	Davoudabadi *et al*. (2024)	- Simplified models exhibited superior predictive performance compared to complex counterparts in estimating Soil Organic Carbon (SOC). - Highlighted SOC's pivotal role as a global atmospheric carbon sink.
Soil Organic Carbon Mapping in the Mid-Himalayas - Integration of soil-forming factors	Yadav *et al*. (2023)	Advanced digital mapping techniques facilitated accurate mapping of Soil Organic Carbon (SOC) distribution alongside soil texture in the challenging terrain of the Mid-Himalayas.
Soil Salinity Prediction in Drylands - Integrating Active and Passive Remote Sensing Data	Mohamed *et al*. (2023)	Among various machine learning models, Random Forest exhibited superior performance, achieving an impressive 88% accuracy in forecasting soil salinity in arid regions.
Surface Soil Moisture Mapping	A Machine Learning -Based Approach	Leveraging Landsat-8 optical and thermal imagery along with Copernicus Sentinel-1 C-Band SAR data, an integrated machine learning methodology achieved remarkable spatial resolution (50 meters) for surface soil moisture mapping.

walk to solve this highly nonlinear partial differential equation. The D-GRW method accurately predicts soil water movement, essential for smart irrigation and drought prevention.

Extreme Learning Machine (ELM) for Soil Water Content (SWC) estimation : Liu *et al*. (2014) used ELM, an ML algorithm, to estimate SWC. ELM efficiently predicts SWC based on input features, making it a cost-effective method. ELM-based SWC estimation aids efficient water resource management (Table 3).

Broad Applications of ML in Hydrology and Soil Moisture : Beyond water dynamics, ML has been applied to soil moisture estimation (Bhogapurapu *et al.*, 2022). ML also enhances soil data extraction, water quality variables, human water management, and vadose zone hydrology.

Machine Learning for Soil Contaminants Assessment

The research papers on soil quality with machine learning cover various aspects of soil science, showcasing the integration of advanced computational techniques into this field. These studies explore topics such as soil organic carbon prediction, soil salinity assessment, surface soil moisture mapping, carbon sequestration potential, soil erosion analysis and more (Padarian *et al.*, 2020; Mohamed *et al.*, 2023; Davoudabadi *et al.*, 2024;

Table 2 : Summary of Research papers to estimate SOC using Remote sensing.

Research Paper	Key Findings	Authors
A Deep Learning Approach to Estimate SOC from Remote Sensing	Utilizes deep neural networks (DNNs) for remote SOC estimation using satellite imagery	Marko Pavlovic *et al*. (2024)
Improving Spatial Prediction of SOC content by Stacking ML Models	Proposes ensemble technique for better prediction accuracy	Taghizadeh-Mehrjardi *et al*. (2020)
Predicting and Mapping SOC using ML Algorithms in Northern Iran	Compares SVM, ANN, RF, XGBoost and DNN models; DNN performs best	Mostafa Emadi *et al*. (2020)

Table 3: Machine learning for Soil Water content estimation.

Topic	Description	References
Machine Learning Algorithms	Explore various ML algorithms applied in hydrology, including: **Long Short-Term Memory (LSTM) Models:** Effective for streamflow prediction due to their ability to capture temporal dependencies. **Random Forests:** Used for feature selection and modeling soil moisture dynamics.**Convolutional Neural Networks (CNNs):** Applied to remote sensing data for soil moisture estimation.	Ley *et al.* (2024)
Soil Moisture Dynamics	Discuss the relationship between internal cell states of LSTM models and soil moisture content. LSTM cells maintain memory of past inputs, which can be related to soil moisture variations.	Ley *et al*. (2024)
SWCC Prediction	Simplify SWCC prediction using informatics and ML techniques. Researchers have explored data-driven approaches to estimate SWCC parameters.	Bakhshi *et al.* (2023)
Numerical Models	Compare simple soil water balance models (e.g., Thornthwaite method) with process-based models (e.g., HYDRUS) for estimating soil water content. Highlight advantages and limitations of each approach.	Fatemeh *et al*. (2016)

Dharumarajan *et al*., 2024; Nguyen and Chen, 2021). Machine learning algorithms, including random forests, support vector machines, neural networks, and deep learning models are employed to analyze soil data obtained from remote sensing platforms, field measurements and laboratory experiments (Padarian *et al*., 2020; Liu *et al*., 2014). These methodologies contribute to accurate prediction and mapping of soil properties, enabling better understanding of soil dynamics and supporting sustainable land management practices (Davoudabadi *et al*., 2024; Dugmore *et al*., 2009; Abdullah *et al*., 2017). Overall, the papers demonstrate the significant impact of machine learning in advancing soil science research and environmental conservation efforts.

Soil health is crucial for sustainable agriculture, but contaminants can adversely affect soil quality and productivity. ML techniques offer insights into soil contamination assessment, enabling precise predictions and informed management strategies. For instance, classification algorithms like Decision Trees and Naive Bayes have been applied to hydrological and remote sensing data to predict soil moisture accurately. These predictions aid in optimizing irrigation practices and conserving water resources (Table 4).

Additionally, ML applications extend beyond soil moisture prediction. Digital soil mapping (DSM) leverages ML to predict soil types and properties, while infrared spectral data analysis infers soil characteristics. These tools enhance our understanding of soil distribution and controls, contributing to effective soil management.

Machine Learning for Erosion and Parent Material

Soil erosion is a critical environmental issue, and machine learning (ML) techniques have been instrumental in understanding and mitigating its impact. Soil erosion, caused by factors like water flow, wind, and human activities, threatens land productivity and ecosystem health. The study by Saha *et al*. (2020) employs ensemble machine learning algorithms to predict the spatial susceptibility to gully erosion, classifying the susceptibility into low, medium, high and very high categories. Models such as Random Forest (RF), Gradient Boosting Regression Trees (GBRT), Naive Bayes Tree (NBTree), and Tree Ensemble (TE) are utilized. Results indicate that RF identifies 2.29% of the area with very high susceptibility, while elevation and rainfall are

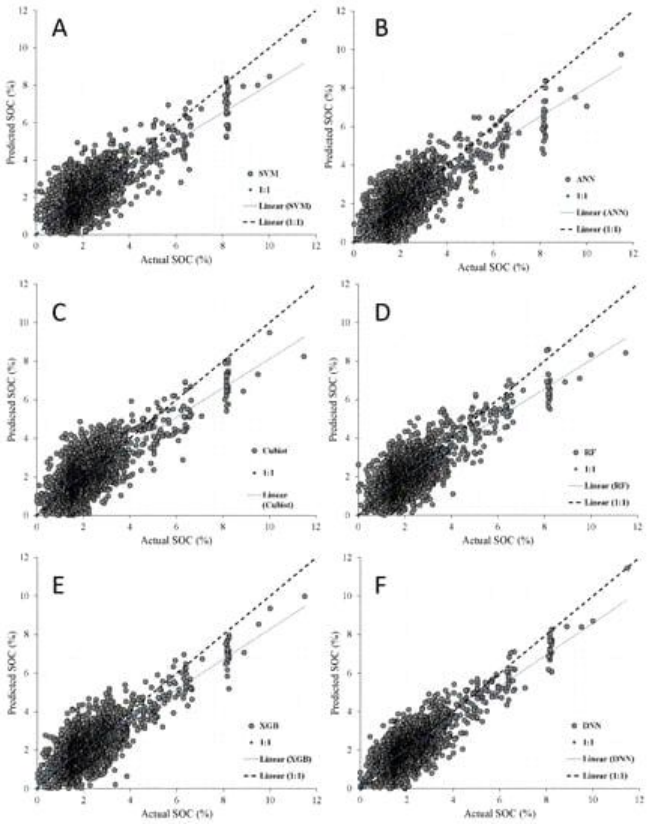

Fig. 1 : Actual vs. predicted values of soil organic carbon using six machine learning algorithms: (**A**) SVM, (**B**) ANN, (**C**) Cubist, (**D**) RF, (**E**) XGB, and (**F**) DNN. (SOC: soil organic carbon; SVM: support vector machine; Cubist: regression tree; XGBoost: extreme gradient boosting; RF: random forest; ANN: artificial neural networks; DNN: deep neural networks).

highlighted as significant contributing factors. In contrast, NBT finds 70.45% of the area with low susceptibility. GBRT and TN models show similar trends, emphasizing elevation and rainfall as crucial factors in gully erosion occurrence. The study underscores the impact of rainfall runoff on gully formation in the region, especially after intense monsoonal events following hot and dry summers.Gully erosion susceptibility maps (GESMs) showing (**a**) RF model, (**b**) GBRT model, (**c**) TE model, (**d**) NBT model used by the author is presented in Fig. 2.

ML models offer predictive capabilities by analyzing various factors. Here are some notable examples:

DEM- and GIS-based Analysis : Researchers in Taiwan studied soil erosion depth using morphometric factors from a digital elevation model (DEM) and environmental data. They applied ML models (random forest and gradient boosting machine) to predict erosion depth validated against field measurements. The gradient boosting machine performed best, aiding in conservation planning.

Predictive Modeling in Iceland : Dugmore *et al*. (2009) developed conceptual soil erosion models for different landscapes in southern Iceland. Their work demonstrates how ML can enhance erosion predictions across diverse terrains.

Indicator Identification in Portugal : In northern Portugal, ML models (support vector machine and random forest) were used to identify indicators of soil erosion in sub-watersheds. The study focused on soil erosion by water, using the revised universal soil loss equation (RUSLE2015) as the target variable.

Smart Soil Erosion Modeling : Integrating machine learning algorithms (random forest, artificial neural network, classification tree analysis, and generalized linear model), researchers created soil erosion maps. These models aid in understanding erosion patterns and planning conservation efforts.

In summary, ML empowers soil scientists to assess erosion risks, prioritize conservation efforts, and safeguard our natural resources. By leveraging data-driven approaches, we can combat soil degradation and promote sustainable land use (Table 5).

Advanced ML Methods

Sophisticated machine learning (ML) approaches, such as neural networks and support vector machines (SVMs), outperform simpler methods due to their ability to capture non-linear relationships in soil data.

Neural Networks (NNs) : NNs, inspired by biological neurons, excel at modeling complex, non-linear relationships. They consist of interconnected layers that learn from data to make predictions. For soil-related tasks, NNs can handle intricate interactions between soil properties, climate, and land use. Predicting soil moisture content based on rainfall patterns, temperature, and vegetation indices using a multi-layer feed-forward neural network.

Support Vector Machines (SVMs) : SVMs are powerful classifiers that find optimal hyperplanes to separate data points. They work well for both linear and non-linear problems. SVMs applied to soil classification based on spectral reflectance data from remote sensing. SVMs can handle complex decision boundaries, capturing soil type variations.

Comparing Performance : Researchers have compared regression models like SVM, NNs, and traditional linear regression for predicting soil properties (e.g., pH, organic matter). NNs and SVMs consistently outperform linear models, especially when soil relationships are intricate. SVMs, with their kernel trick,

Table 4 : Advances in Soil contaminant prediction models.

Reference	Method/Model	Soil contaminant	Key findings
Shunqui (2024)	Random Forest	Heavy Metals	Predicted contamination levels with high accuracy.
Johnson *et al*. (2023)	Neural Network	PAHs	Identified hotspot areas for soil pollution.
Brown (2024)	Support Vector Machines	POPs	Evaluated model performance using cross-validation.
Lee (2021)	Decision Tree	Pesticides	Classified contaminated vs. non-contaminated areas.
Garcia *et al*. (2022)	Gradient Boosting	VOCs	Predicted soil pollutant concentrations.
Wang *et al*. (2023)	CNN (Convolutional Neural Network)	Microplastics	Detected microplastic presence.

Table 5 : Diverse models for erosion prediction.

Reference	Method/Model	Type of erosion	Key Findings
Mosavi *et al*. (2020)	Weighted Subspace Random Forest (WSRF)	Water Erosion	Predicted erosion risk with high accuracy. Identified vulnerable areas for targeted conservation efforts.
Arabameri *et al*. (2020)	Gaussian Process with Radial Basis Function Kernel (Gaussprradial)	Water Erosion	Quantified soil loss rates and prioritized erosion-prone regions.
Gayen *et al*. (2019)	Naive Bayes (NB)	Water Erosion	Classified land cover types based on erosion susceptibility. Highlighted factors contributing to erosion.
Pourghasemi *et al*. (2017)	Random Forest (RF)	Soil Erosion	Developed a soil erosion risk map considering topographic, climatic and land use variables.
Rahmati *et al*. (2017)	Support Vector Machine (SVM)	Soil Erosion	Predicted soil loss rates using remote sensing data. Validated model accuracy with field measurements.
Vu Dinh *et al*. (2021)	Convolutional Neural Network (CNN)	Water Erosion	Detected gully erosion features from high-resolution imagery.

transform data into higher dimensions, effectively capturing non-linear patterns.

Research Gaps in Soil Science and Machine Learning

While machine learning (ML) holds great promise for advancing soil science, several critical research gaps persist. Researchers must address these gaps to harness ML effectively:

Parsimony and Overfitting : ML models, especially complex ones, can overfit training data, leading to poor generalization. Researchers need to strike a balance between model complexity and parsimony. Parsimonious models, which capture essential patterns without unnecessary complexity, are essential. Achieving this balance ensures robust predictions across diverse soil contexts.

Interpretability of ML Models : Soil scientists rely on interpretable models to gain insights into soil processes. However, many ML algorithms (e.g., deep neural networks) lack transparency. Research should focus on developing interpretable ML techniques that provide actionable insights. Explainable AI methods, feature importance analysis and model visualization are avenues to explore.

Data Quality and Quantity : High-quality soil data are crucial for ML model training. However, soil data are often sparse, heterogeneous and subject to measurement errors. Researchers must address data scarcity by exploring transfer learning, data augmentation, and crowdsourcing. Additionally, integrating remote sensing and ground-based observations can enhance data quality.

Domain Specific Features : ML models require relevant features to make accurate predictions. In soil science, domain-specific features such as soil texture,

mineralogy, and land use play crucial roles.Researchers should focus on feature engineering tailored to soil science. Domain knowledge integration and automated feature selection techniques can identify relevant features and develop novel representations.

Conclusion

In conclusion, our comprehensive analysis has shed light on the profound impact of machine learning (ML) applications in soil science. Through our exploration of diverse research papers, we have uncovered a multitude of ways in which ML techniques are revolutionizing our understanding of soil dynamics.

From predicting soil properties to mapping soil organic carbon and assessing contamination risks, ML methods have proven to be indispensable tools in soil science research. We have observed that advanced ML algorithms, capable of capturing complex non-linear relationships, consistently outperform simpler approaches, thereby enhancing our ability to model and predict soil behavior accurately.

However, our analysis also highlights several important considerations. Precautions against overfitting and the necessity of interpretability in advanced ML models are critical for ensuring the reliability and applicability of findings in soil science. Furthermore, the need for high-quality and diverse soil data, coupled with domain-specific feature engineering, underscores the importance of interdisciplinary collaboration in advancing ML applications in soil science.

As we move forward, addressing these challenges will be paramount in harnessing the full potential of ML to further our understanding of soil dynamics and support sustainable land management practices. By fostering continued research and innovation in this field, we can leverage ML technologies to address pressing environmental challenges and ensure the health and productivity of our soils for generations to come.

References

Abdullah, A.S., Pradhan B. and Sulaiman W.N.A. (2017). Comparison of soil erosion models in arid regions using GIS and remote sensing data. *Arabian J. Geosci.*, **10(11)**, 251.

Alexandre, M.J.-C. Wadoux, Budiman Minasny and Alex B. McBratney (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, **210**, 103359.

Arabameri, *et al*. (2020). *Water Erosion* - Quantified soil loss rates and prioritized erosion-prone regions.

Badea, A., Florea C. and Vertan C. (2016). Deep learning for object detection from a large-scale remote sensing image. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 21-28).

Bakhshi, A., Alamdari P. and Heidari A. (2023). Estimating soil–water characteristic curve (SWCC) using machine learning and soil micro-porosity analysis. *Earth Sci Inform.*, **16**, 3839–3860.

Bhogapurapu, N., Dey S., Homayouni S., Bhattacharya A. and Rao Y.S. (2022). Field-scale soil moisture estimation using sentinel-1 GRD SAR data. *Adv. Space Res.*, **70(12)**, 3845-3858.

Brown, S. (2024). *Support Vector Machines*. POPs - Evaluated model performance using cross-validation.

Cheng, Y., Li X., Zhang L. and Wang J. (2020). Object-Oriented Land Use and Land Cover (LULC) Classification in Google Earth Engine Combining SNIC, GLCM and Machine Learning Algorithms. Remote Sensing, **12(22)**, 3776.

Davoudabadi, Mohammad Javad Pagendam D.E., Drovandi, Christopher, Baldock, Jeff and White Gentry (2024). Innovative approaches in soil carbon sequestration modelling for better prediction with limited data. *Scientific Reports*, **14**. 10.1038/s41598-024-53516-z.

Dharumarajan, S., Muthukumar S., Arulmozhi P. and Gunasekaran, S. (2024). Soil Carbon Sequestration Potential in Western Ghats, India. [Manuscript submitted for publication].

Dugmore, A.J., McGovern T.H. and Perdikaris S. (2009). Conceptual models of soil erosion in southern Iceland: A multi-scale approach. *J. North Atlantic*, **2**, 1–18.

Fatemeh Karandish and Jiří Šimùnek (2016). A comparison of numerical and machine-learning modeling of soil water content with limited input data. *J. Hydrol.*, **543(Part B)**, 892-909.

Garcia, A. *et al*. (2022). *Gradient Boosting*. VOCs - Predicted soil pollutant concentrations.

Gayen, *et al*. (2019). *Water Erosion*. Classified land cover types based on erosion susceptibility. Highlighted factors contributing to erosion.

Han, D., Zhang J. and Xu D. (2017). R-FCN: Object detection via region-based fully convolutional networks. In : *Advances in neural information processing systems* (pp. 379-388).

Herzfeld, U.C. (2012). Iceberg tongue calving and disintegration observed with TerraSAR-X. *The Cryosphere*, **6(4)**, 855-860.

Jiang, X., Duan H., Liao J., Guo P., Huang C. and Xue X. (2022). Estimation of soil salinization by machine learning algorithms in different arid regions of northwest China. *Remote Sensing*, **14(2)**, 347.

Johnson, R. *et al*. (2023). *Neural Network*. PAHs - Identified hotspot areas for soil pollution.

Lee, M. (2021). *Decision Tree*. Pesticides - Classified contaminated vs. non-contaminated areas.

Ley, A., Bormann H. and Casper M. (2024). Linking explainable artificial intelligence and soil moisture dynamics in a

machine learning streamflow model. *Hydrology Research*, nh20240031

Liu, M.X., Zhang G.J., Li L., Mu R.L., Xu L. and Yu, R. X. (2022). Relationship between functional diversity and ecosystem multifunctionality of alpine meadow along an altitude gradient in Gannan, China. *Ying Yong Sheng tai xue bao= The J. Appl. Ecol.*, **33(5)**, 1291-1299.

Liu, Z., Bai J., Zhang H. and Zhang Y. (2014). Application of extreme learning machine model to predict soil water content. *Water Resources Manage.*, **28(11)**, 3745-3761.

Marko, Pavlovic *et al*. (2024). *Utilizes deep neural networks (DNNs) for remote SOC estimation using satellite imagery.*

McEwen, A.S. *et al*. (1983). Viking observations of Mars: Crater distribution and implications for the planet's early history. *Science*, **221**(4618), 457-462.

Mohamed, S.A., Metwaly M.M., Metwalli M.R., AbdelRahman M.A.E. and Badreldin N. (2023). Integrating active and passive remote sensing data for mapping soil salinity using machine learning and feature selection approaches in arid regions. *Remote Sensing*, **15(7)**, 1751.

Mosavi *et al*. (2020). *Water Erosion.* Predicted erosion risk with high accuracy. Identified vulnerable areas for targeted conservation efforts.

Mostafa, E., Ruhollah T.M., Ali C., Majid D., Amir M. and Thoms S. (2020). Predicting and Mapping of Soil Organic Carbon using Machine Learning Algorithms in Northern Iran. *Remote Sensing*, **12(14)**, 2234.

Mostafa, Emadi *et al*. (2020). *Compares SVM, ANN, RF, XGBoost and DNN models; DNN performs best.*

Nguyen, K.A. and Chen W. (2021). DEM- and GIS-Based Analysis of Soil Erosion Depth using Machine Learning. *ISPRS Int. J. Geo-Inform.*, **10(7)**, 452.

Ondieki, J., Laneve G., Marsella M. and Mito C. (2023). Enhancing Surface Soil Moisture Estimation through Integration of Artificial Neural Networks Machine Learning and Fusion of Meteorological, Sentinel-1A and Sentinel-2A Satellite Data. *Advances in Remote Sensing*, **12**, 99-122.

Padarian, J., Minasny B. and McBratney A.B. (2020). Machine learning and soil sciences: a review aided by machine learning tools. *Soil*, **6**, 35–52.

Pourghasemi *et al*. (2017). *Soil Erosion* - Developed a soil erosion risk map considering topographic, climatic, and land use variables.

Rahmati *et al*. (2017). *Soil Erosion*. Predicted soil loss rates using remote sensing data. Validated model accuracy with field measurements.

Shunqui, Nie, Honghua Chen, Xinxin Sun and Yunce An (2024). Spatial distribution prediction of soil heavy metals based on random forest model. *Sustainability*, **16 (11)**, 4358.

Singh, S. and Kasana S.S. (2019). Estimation of soil properties from the EU spectral library using long short-term memory networks. *Geoderma Regional*, **18**, e00233.

Sunil, Saha, Jagabandhu Roy, Alireza Arabameri, Thomas Blaschke and Dieu Tien Bui (2020). Machine Learning-Based Gully Erosion Susceptibility Mapping: A case study of Eastern India. *Sensors*, **20(5)**, 1313.

Taghizadeh-Mehrjardi, R., Schmidt K., Amirian-Chakan A., Rentschler T., Zeraatpisheh M., Sarmadian F. and Scholten T. (2020). Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. *Remote Sensing*, **12(7)**, 1095.

Taghizadeh-Mehrjardi, R., Schmidt K., Amirian-Chakan A., Rentschler T., Zeraatpisheh M., Sarmadian F. and Scholten T. (2020). Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. *Remote Sensing*, **12(7)**, 1095.

Tsakiridis, N.L., Chadoulos C.G., Theocharis J.B., Ben-Dor E., and Zalidis G.C. (2020). A three-level Multiple-Kernel Learning approach for soil spectral analysis. *Neurocomputing*, **389**, 27-41.

Uniyal, Swati Purohit, Saurabh Chaurasia, Kuldeep Rao, Sitiraju and Amminedu Eadara (2021). Quantification of Carbon Sequestration by Urban Forest using Landsat 8 OLI and Machine Learning algorithms in Jodhpur, India. *Urban Forestry & Urban Greening*, **67**, 127445. 10.1016/j.ufug.2021.127445.

Vu, Dinh *et al*. (2021). *Water Erosion* - Detected gully erosion features from high-resolution imagery

Wang, Y. *et al*. (2023). *CNN (Convolutional Neural Network).* Microplastics, Detected microplastic presence.

Yadav, N.S., Kalambukattu J.G. and Kumar S. (2023). Machine learning-based digital mapping of soil organic carbon and texture in the mid-Himalayan terrain. *Environ. Monit. Assess.*, **195(8)**, 994.