land

MDPI

Review

Benefits and Challenges of Artificial Intelligence in Soil Science—A Review

Christos Kikis and Vasileios Antoniadis *

Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, Fytokou Street, 38446 Volos, Greece; chkikis@uth.gr
* Correspondence: antoniadis@uth.gr

Abstract

Artificial intelligence (AI) is rapidly affecting soil science by allowing the analysis of large, complex, and heterogeneous datasets that were previously difficult to exploit. The current review synthesizes the recent advances of AI and highlights how these tools are applied in key soil science domains, such as digital soil mapping, soil fertility management, soil moisture prediction, contamination monitoring, soil carbon assessment, and precision agriculture. This study evaluates the performance of different AI methods, showing that techniques such as random forests, neural networks, and convolutional neural networks often outperform traditional methods in capturing non-linear soil-environment. At the same time, it identifies major limitations such as data scarcity, reproducibility, lack of large datasets, uncertainty, and the "black-box" nature of many models. This review concludes that AI has strong potential to support sustainable soil management, but its real-world impact will depend on better data integration, explainability, standardization, and closer collaboration with scientists, technologists, and end-users.

Keywords: machine learning; soil informatics; data-driven agriculture; remote sensing; smart farming; technology adoption; agroecosystems

1. Introduction

Soil is essential for agricultural production and, therefore, for life, as it helps food crops to grow, stores carbon, and supports a wide range of ecosystem services. However, soil is not a renewable resource and thus is heavily pressured. Around one-third of the world's land is already degraded as a result of erosion, poor nutrient content, contamination, and inappropriate land management [1,2]. Traditionally utilized methods for soil assessment include practices such as manual sampling and lab analyses, which require expensive instrumentation, are time-consuming, and often cannot be scaled to the extent of complexity that modern agriculture requires. Recently, techniques like remote sensing, geospatial technology, and precision agriculture have led to the collection of large datasets [3]. On top of this, artificial intelligence (AI) has emerged as a powerful tool that is utilized to analyze large heterogeneous datasets and support decision-making processes [4].

Recent studies have highlighted the capabilities of AI across a wide range of applications in soil management. For example, AI has achieved remarkable accuracy in predicting various soil properties such as soil carbon, soil moisture patterns, and mapping soil conditions [5]. Data combinations retrieved from drones, satellites, soil sensors, and past databases can lead to AI-based systems capable of producing high-resolution maps,

Check for updates

Academic Editor: Gabriele Broll

Received: 15 January 2026
Revised: 12 February 2026
Accepted: 14 February 2026
Published: 15 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

predicting soil moisture and nutrient content, monitoring contamination in soil, and optimizing agricultural inputs, all of which were not possible before [6]. Such achievements mark a major shift from traditional soil studies based on conventional field measurements to AI continuous monitoring over larger areas. However, adopting AI in soil science raises scientific concerns about data reliability, data quality and quantity, and real-world implementation, which need to be duly addressed.

The application of AI in soil science was first mentioned in the work of Holt [7], who established the field for pattern recognition and rule-based systems. Early studies by McCracken and Cate [8] and Dale et al. [9] demonstrated that soil taxonomy reasoning could be encoded by algorithmic rules. This led scientists to encode pedological knowledge into decision rules via the "if" command for soil classification and land evaluation. From the mid to late 1900s, AI shifted from rule-based systems to artificial neural networks (ANNs) and early machine learning (ML) models, enabling the capture of non-linear relationships. Schaap et al. [10] showed that ANNs outperformed traditional regression in predicting soil hydraulic properties via pedotransfer functions. At the beginning of the 2000s, Breiman [11] introduced random forests (RFs), and in 2003, McBratney et al. [12] proposed the SCORPAN model, based on geographic information system (GIS) layers, ANNs, and environmental covariates, paving the way for digital soil mapping (DSM). Hence, from the mid-2000s, the GlobalSoilMap project was initiated in order to produce high-resolution digital maps [13], while support vector machines (SVMs) were established in DSM creation and prediction [14]. In the 2010s, ensemble methods like RFs and SVMs led to major scientific milestones, such as the creation of high-resolution soil maps [15]. Along with SVMs, deep learning (DL) tools like convolutional neural networks (CNNs) advanced precision agriculture via soil spectroscopy, drones, and sensors, leading to see-and-spray technologies in the late 2010s [16]. During the pandemic in 2020, the IoT explosion enabled real-time data collection from field to cloud-based modeling [17]. In 2021–2022, growing attention to "Net Zero" carbon sequestration monitoring was performed by remote sensing technologies and ML. Hundreds of thousands of soil samples are integrated into AI engines coupled with satellite data to create DSMs of soil carbon (SOC) for carbon credit programs [18]. In 2022, the launch of ChatGPT 3.5 increased interest of soil practitioners in large-language models (LLMs), ending utilizing them as digital advisors with personalized recommendations. However, LLMs still offer little conceptual novelty and largely repeat technological applications that have existed since the 1990s [19]. Nonetheless, generative AI (GenAI) is now transforming soil science by enhancing the accuracy of data analysis and decision-making through advanced models. One of its most important applications is the creation of synthetic soil data to fill gaps in areas with limited sampling, significantly improving the accuracy of SOC and moisture predictions [20,21]. The evolution of AI techniques in soil science is summarized in Figure 1.

Despite the growth in applying AI in soil science, the field still has limitations and gaps that reduce real-world applicability. Many AI models rely on narrow sets of data, ignoring important biological and climatic factors that strongly influence soil behavior. Additionally, most research efforts are rarely involved in long-term experiments, but rather extract outcomes and models based on short-term data. Furthermore, soil, by its nature, provides uneven data, often infrequently measured, which makes datasets hard to reproduce or apply elsewhere. Researchers also tend to underuse advanced AI methods that could better capture complex soil processes, while models lack clear explanations for prediction-making. Ethical and practical issues such as data ownership and privacy are also largely overlooked.

The aim of this review is to (i) provide a comprehensive overview of AI applications in soil science with emphasis on critical analysis of methodologies and findings; (ii) highlight the strengths and limitations of each AI method used in soil science; (iii) discuss the challenges that limit AI's impact in soil science; and (iv) outline future directions for

research and development. To achieve these objectives, this manuscript is structured by first presenting the methodology of approaching the topic, then comparatively reviewing the main AI techniques used in soil science, critically evaluating their strengths, discussing the major limitations and practical barriers to real-world implementation, and finally highlighting emerging trends and future directions.

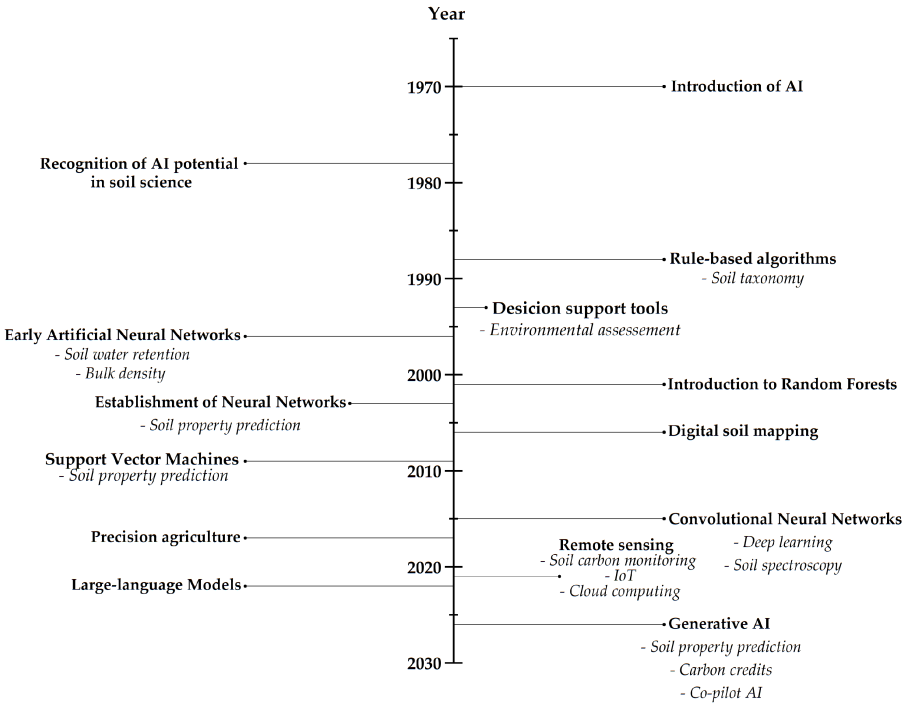


Figure 1. Timeline illustrating the major methodological milestones and evolving applications of AI in soil science. The figure highlights the dominant AI paradigms and their association with soil science.

2. Literature Review and Methodology

The current review was performed after a systematic methodology that led to the creation of network information, based on which the study was conducted. The search was performed on the Web of Science Core Collection database by using the words "artificial intelligence" as a title, followed by the keywords "soil science," "soil properties," "soil mapping," "soil moisture," "soil contamination," and "carbon sequestration." In addition, the search was also modified by using the word "soil" and the following keywords: "modeling," "machine learning," "deep learning," "neural networks," "random forests," "support vector machines," "language models," and "explainable AI." The aim of the search was to cover as much "field" as possible, without missing any relevant topics. The search led to a summary of 5261 studies, which were filtered, keeping only peer-reviewed articles and articles written in English. We also filtered them by publishing date, selecting only articles published in the last decade. Also, the Web of Science Core Collection database provides filters based on the discipline, i.e., Meso Topic, which represents a scientific category within the classification system of the database. Hence, the Meso Topic we used to filter any irrelevant research was the "Soil Science" category, which narrowed the overall number of articles to 523. The remaining articles were further screened for duplicate removal and relevance confirmation.

The total number of the remaining articles was 511, based on which we conducted a bibliometric analysis to identify the most influential and frequently occurring terms, which formed the conceptual foundation of the manuscript. The most frequently used terms were selected based on keyword occurrence and relevance, which led to the creation of a network chart (Figure 2) that helped visualize patterns and aid manuscript construction.

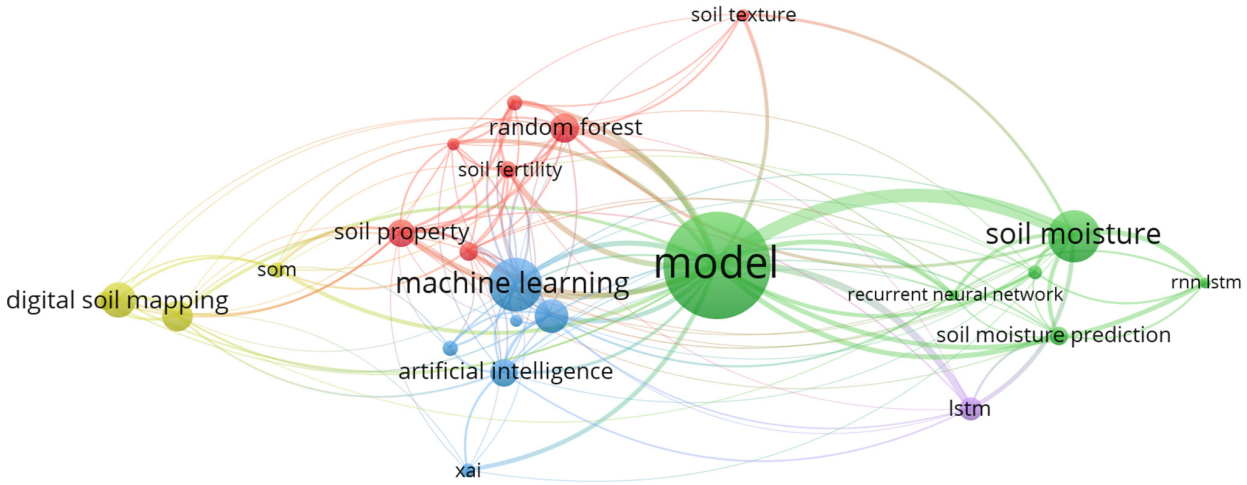


Figure 2. Keyword co-occurrence network showing thematic clusters in soil modeling research. Node size represents keyword frequency, link thickness indicates co-occurrence strength, and colors denote communities detected through clustering analysis.

3. Key AI Techniques in Soil Science

Machine learning methods are widely utilized in soil science. This is due to their ability to identify complex, non-linear relationships among soil properties. The most common methods include ensemble trees like RFs, gradient boosting, SVMs, and others. For example, RFs have emerged as one of the most highly utilized ML tools in creating digital maps for soil fertility prediction [22]. The popularity of RFs derives from their ability to handle versatile decision-making trees, averaging them, reducing overfitting, and delivering accurate and stable results, outperforming traditional linear models. As for SVMs, they are commonly used as well, mainly for classification tasks, performing better in modelling medium-sized non-linear datasets [23]. Furthermore, ANNs are mainly used for modeling complex relationships, such as predicting soil moisture from climate data or estimating soil features from spectral measurements [24]. They often require large amounts of data and more careful calibration of such large data compared to RFs.

Recently, DL has also entered soil science as a result of increased data availability. Deep neural networks contain mainly layers that can automatically learn upcoming representations [25]. In particular, CNNs are increasingly applied for analyzing hyperspectral images and high-resolution satellite or drone images. They are capable of identifying soil patterns, contamination, and spatial variation, even in data-limited regions [26]. Furthermore, CNNs have also been used on soil profile photographs to classify soil types and properties with great accuracy in pilot programs [27]. Recurrent neural networks, like long short-term memory (LSTM) networks, are utilized to model soil moisture dynamics from sensor readings. For example, an LSTM system can learn how soil moisture depends on previous patterns like rainfall events and evaporation incidences, resulting in outperforming simpler models for precision tasks like irrigation scheduling. In agricultural water management, for example, LSTM-based models have been proven to be a sufficient tool to accurately predict soil water needs, helping to reduce costs [28].

Geostatistical models are often combined with AI in soil digital mapping. This means that the difference between traditional spatial methods and AI-based models is less clear. However, traditional approaches like kriging and co-kriging are useful in cases where soil sampling is performed unevenly or when samples are limited [29]. Under this scenario, they can be combined with AI. Machine learning predicts general soil patterns by initially combining them with environmental data (such as terrain or climate) and then kriging is used to adjust to local features by interpolating the remaining errors. Some studies treat geostatistical methods

as a type of ML model (like Gaussian processes), whereas others use two-step predictions followed by interpolation [30]. In either case, the utilization of GIS with AI models enables the creation of high-resolution, accurate, and consistent soil maps [31,32].

Explainable AI (XAI) and GenAI methods are becoming critical as decision-making processes of AI systems that act as "black boxes," indicating that it is hard to understand the causality behind their decision-making processes [33]. To address this, methods like feature importance rankings can show the type of inputs that matter the most in tree-based ensembles, where SHAP values and saliency maps aid in explaining the model's predictions. This explanatory system is imperative when it comes to data reliability, as users need to trust the results. For instance, if a model predicts contamination in an area, end users can trust this result if the model clearly shows it is due to factors like nearby industrial activity or specific soil properties. Indeed, recent studies on soil contamination have utilized XAI to identify heavy metal pollution sources [34–36]. Hence, the models cannot only make accurate calculations but can also provide scientifically aligned information. An overall comparative analysis of the most utilized AI methods applied in soil science is summarized in Table 1.

In soil science, AI applications range from prediction-oriented to process-oriented models, reflecting a trade-off between statistical accuracy and mechanistic understanding. Process-oriented models embed mechanistic knowledge of soil features, offering interpretability and transferability to data-sparse conditions, though often at the cost of lower spatial resolution or calibration complexity [37]. Recent hybrid approaches attempt to combine ML into process-based frameworks and improve both predictive performance and physical plausibility [38]. Conversely, prediction-oriented models, like ML algorithms, have shown high accuracy in soil mapping and forecasting when trained on large datasets [37]. However, their reliance on data correlations without physical constraints limits their explanatory power and can lead to misleading generalizations when applied beyond the conditions they were trained on [39]. One major issue of prediction-oriented models is data leakage, which can lead to optimistic model assessments. If training and test samples are not truly independent in space or depth (e.g., if cross-validation splits nearby areas or samples from the same soil profile in both training and validation sets), the model essentially gets to "see" part of its test data during training. This violates the independence assumption and often inflates performance metrics due to spatial autocorrelation. For instance, John et al. [40] demonstrated that 3D soil mapping using random cross-validation (which allowed soil samples from the soil profile in both train and test) gave accuracy metrics 8–62% higher than a strict model due to vertical autocorrelation. Another limitation in soil predictive modeling is the limited transferability of models across regions. Machine learning models that perform well and degrade when applied elsewhere. For example, Mirzaeitalarposhti et al. [41] found that a soil texture model calibrated in one German region had poor predictive performance when transferred to another distinct region, likely due to differences in environmental and soil-forming factors between the sites.

Finally, overfitting remains a pervasive issue in AI-based soil modelling, especially with high-capacity models like DL. Their flexibility to learn extremely complex data patterns can lead to fitting noise rather than meaningful generalized relationships. In soil science, where datasets are often small or biased toward easily accessible areas, the risk of overfitting is amplified. Moreover, because many studies report cross-validation without spatial separation of training and testing data, this can create a false sense of model reliability and lead to flawed decision-making in applied contexts like precision agriculture. Even RFs have been shown to produce high-accuracy models using irrelevant or synthetic variables [42], highlighting how misleading performance metrics can be when models overfit correlations.

Table 1. Comparative overview of AI applications in soil science.

AI Technique	Application in Soil Science	Key Advantage	Main Limitation	Data Requirement	Interpretability	Computational Cost	Reference
Decision Trees	Predictive soil mapping and classification; decision support when model simplicity is required.	Highly interpretable, non-parametric models with if/then rules;	Prone to overfitting if developed too deeply; less accurate and stable than random forests.	Low	High	Low	[43]
Random Forests (ensemble trees)	Digital soil mapping of soil properties (pH, carbon, texture, contamination, etc.) where high accuracy is needed, and data are limited.	A summary of many decision trees with high predictive accuracy and robust results. Offers future importance outputs, identifying key soil variables.	Complex model with reduced transparency compared to simple decision trees. It can be outperformed by deep learning on large, complex datasets.	Medium	Medium	Medium	[44]
Support Vector Machines (SVRs)	Soil property classification and mapping when sample sizes are moderate. Also used in soil spectroscopy and remote sensing when a non-linear method is needed, and dataset size is moderate.	Effective on complex high-dimensional datasets. Often robust with noisy or limited training data	Low interpretability. Computationally intensive on large datasets and requires careful parameter tuning for optimum performance.	Low	Low	High	[45]
Artificial Neural Networks (ANNs)	Prediction of soil properties and development of pedotransfer functions. Used in soil moisture forecasting, especially when many input variables are involved.	Captures non-linear complex relationships between soil variables, allowing multivariate modelling.	Requires largely trained datasets for optimum generalization; otherwise degrades. Low transparency compared to rule-based methods.	High	Low	Medium	[46]

Table 1. *Cont.*

AI Technique	Application in Soil Science	Key Advantage	Main Limitation	Data Requirement	Interpretability	Computational Cost	Reference
Convolutional Neural Networks (CNNs)	Remote sensing and image-based soil mapping (e.g., hyperspectral for soil texture); analysis of soil profile images to identify structures like pore networks.	Captures complex patterns in soil maps or imagery by extracting spatial features from imaged and spatial data; high accuracy in image-based soil analyses given sufficient training data.	Data-hungry and computationally intensive to train; very low inherent interpretability. In addition, explainable AI can help understand how the model learns.	High	Low	High	[47]
Recurrent Neural Networks	Time series forecasting of soil variables, such as weekly soil moisture; used for modelling soil moisture for irrigation scheduling; applicable whenever soil processes have temporal dynamics.	Designed for sequential data; effectively learns temporal dependencies in soil time series; uses LSTM-based networks, which capture long-term trends and complex relationships better than static models.	Requires sufficient sequential data for training; can overfit if training series are short. Training can be slow for long-time sequences due to recurrent computations.	High	Low	High	[48]

Table 1. *Cont.*

AI Technique	Application in Soil Science	Key Advantage	Main Limitation	Data Requirement	Interpretability	Computational Cost	Reference
Generative AI (GANs)	Soil spectral data augmentation (e.g., soil organic carbon prediction); generating soil profile images for training CNNs; filling spatial gaps in soil property maps.	Can generate synthetic soil data to augment limited datasets; improves performance of predictive models in data-scarce regions; used for super-resolution in soil maps.	Training GANs is unstable due to overfitting; requires validation of the generated data to avoid false patterns; complex to train and tune properly.	High	Low	High	[49]
Large Language Models (LLMs)	Chatbots advisors for farmers on soil practices; text-based interpretation of soil test results; automatic literature summarization (e.g., best practices for conversion tillage).	Integrates large-scale agronomic knowledge from text (research, documents, etc.); enables Q&A systems for soil and crop decision support; useful for literature summarization and recommendation.	It can "hallucinate" (generate false information); requires fine-tuning for an agricultural context; generalized models, unless trained on domain texts.	Medium	Medium	High	[50]

4. Main AI Applications in Soil

4.1. Soil Mapping

Digital soil mapping is considered one of the first and most important uses of AI in soil science. It is a process where ML tools are utilized for soil property prediction and classification, instead of manually drawing soil maps. They work by capturing complex non-linear relationships in the data from soil measurements (texture, pH, conductivity, moisture, nutrient status, etc.) and environmental data (satellite data, climate, and landscape features). Afterwards, ML models are trained to create continuous soil maps across different areas [51]. The review of Adeniyi et al. [52] on the DSM of agricultural areas covering the years 2008–2023 reported that statistical ML models—like RFs, decision trees, and gradient boosting machines—have become popular and widely used for soil property prediction (Figure 3).

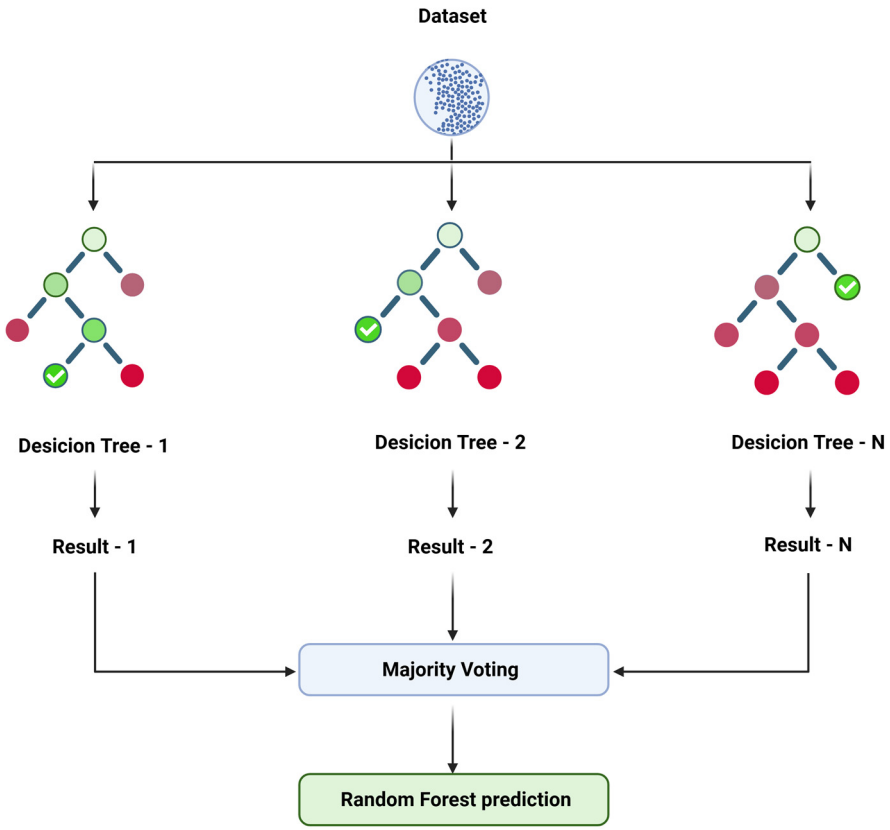


Figure 3. Schematic of the random forest algorithm. A dataset is used to train multiple decision trees independently; each tree produces a prediction, and the final output is obtained by aggregating individual results through majority voting to generate the random forest prediction. Created in BioRender.

When it comes to DSM, RFs are regarded as the most consistent performers among traditional ML tools, due to their robustness to noisy data and their ability to model complex interactions, especially in temperate and semi-arid regions where spatial heterogeneity is high. For instance, RFs utilized to map soil texture in a semi-arid area ended up outperforming multiple linear regression because soil texture is influenced by complex factors like land shape and vegetation [53]. Also, they tend to outperform ANNs and SVMs when data are scarce or non-imagery. For instance, the study of Zhang et al. [54] compared different ML tools for soil texture classification, using 640 soil profiles and multiple environmental covariates, in a semi-arid region in northwest China; RFs achieved the lowest errors—measured as root mean square errors (RMSEs)—and the highest correlation (0.63) for predicting soil texture, outperforming SVMs. However, in terms of computing speed,

its classification time was the slowest, with gradient boosting being the fastest. While SVMs can achieve competitive accuracy in classification tasks, they are sensitive to kernel selection and require more manual parameter tuning. In contrast, CNNs have shown significant improvements when raw spatial data, such as satellite images or soil profile photos, are available. For instance, CNNs trained on hyperspectral data images or terrain tiles can outperform RFs, especially in data-rich environments. For example, in a study conducted in an arid region in Iran with a 1524 soil profile dataset and 164 environmental covariates, CNNs outperformed RFs for modelling soil particle fractions [55]. Similarly, Beucher et al. [56] compared RF and CNN models to map the spatial occurrence of potential acid sulfate soils in the wetlands of Jutland, Denmark, using environmental covariates. While both models performed reasonably well, the CNN achieved the highest overall accuracy (68%), slightly outperforming RF (61–63%), which the authors attribute to its ability to better capture spatial patterns. However, the advantage was limited by the relatively coarse (30.4 m) resolution of the input data and the fact that variable selection had been optimized for RFs rather than for CNNs. Nonetheless, in order for the comparison to be fair, they created RFs trained with spatially smoothed auxiliary data, but even in that case, they found that a multi-layer CNN reduced RMSE for topsoil clay–sand–silt fractions by roughly 10–20% relative to RFs and improved accuracy. This happened due to the ability of CNNs to learn multi-scale landscape patterns automatically. In general, CNNs require tuning and more computational power along with large-labeled datasets, making them less suitable for areas with sparse soil sampling or weak imagery coverage, where RFs are computationally cheap, require less data, and offer interpretability.

4.2. Soil Fertility and Nutrient Management

The importance of AI in soil fertility lies in its capacity to drive precision agriculture. Overuse of fertilizers can result in environmental toxicity and economic waste, while underuse can lead to lower yields and loss of valuable food [57]. The amount of nutrients in soil, like nitrogen, phosphorus, and potassium, is typically measured in a lab using chemical extraction tests, expensive instruments, and intensive labor. In contrast, AI offers a rapid way to determine soil fertility by integrating soil lab measurements, along with satellite data or crop performance data [58]. As a result, AI can predict the nutrient status in soil and aid fertilizer application schemes. Machine learning methods like RFs, SVMs, and ANNs have been applied to predict and classify soil fertility by estimating nutrient availability from easy-to-measure data or categorizing soils into fertility classes [59]. For example, an RF model can utilize inputs like soil color, electrical conductivity, precipitation, and land shape to estimate soil organic carbon and nutrient status in areas where no soil samples were obtained and still achieve high accuracy. To illustrate this, Gunasekaran et al. [60] developed an AI system with real-time sensors measuring parameters like soil pH, moisture, temperature, and nutrient levels, with an RF model achieving 92%. In the same study, the applied DL methods further improved predictions for related tasks, like recommending suitable crops, by capturing small differences between variables, e.g., seasonal changes in fertility. Indirect data can also be used to predict nutrient levels in soil. For example, soil spectroscopy, along with ML, can rapidly and efficiently determine nutrient levels in soil. For example, a recent study utilized VNIR spectroscopy to estimate several soil properties and found that among SVMs, RFs, and CNNs, SVMs performed best for most of the variables. Specifically, SVMs had the highest accuracy for clay (R^2 = 0.79), pH (R^2 = 0.84), total nitrogen (R^2 = 0.800, and cation exchange capacity (R^2 = 0.83) [61]. This is due to the fact that SVMs use a radial basis function kernel, which can model non-linear patterns in spectral data. On the other hand, RFs performed worse due to their random nature, indicating that the performance of AI methods for soil fertility estimation depends

on input type and spatial scale. On the other hand, when rich spectral or image inputs are available, DL methods outperform every other approach. For example, the study of Ma et al. [62] used Sentinel-2A imagery from 800 field soil samples in China, and found that CNNs vastly outperformed RFs and SVMs, with an R^2 equal to 0.89, for SOC prediction. This shows that DL could more effectively capture the complex relationship between multispectral reflectance and organic matter. The same observations were extracted from Deng et al. [63], who also predicted organic matter from 206 forest soil samples in China, via hyperspectral reflectance. Their goal was to predict organic matter from hyperspectral data, and they compared multiple models, including SVMs, CNNs, and an optimized hybrid CNN model, which was found to be the most accurate (R^2 = 0.93, RMSE = 3.04), outperforming standard CNN and SVM methods. Such studies demonstrate DL's high accuracy when given the right type of data. However, SVMs are relatively data-efficient and avoid overfitting in small datasets, while CNNs require more samples and training time but can capture nonlinearity in data.

4.3. Soil Moisture Prediction and Irrigation

Soil moisture is one of the most important soil indicators as it affects plant growth, soil microbial activity, and ultimately soil erosion and runoff. Knowledge of soil moisture temporal dynamics can assist farmers in making decisions on efficient irrigation schedules and drought and flood predictions. Nevertheless, soil moisture measurements are not always possible, as water is usually measured via ground sensors and soil samples at specific points. Traditionally, it is calculated using physical models which account for water movement based on precipitation, evaporation, and related soil properties; all these require high computing effort since they need many inputs (soil hydraulics parameters, fine-scale rainfall data, etc.) [64]. Physics-based models offer high interpretability due to their reliance on established physical laws, where parameters have physical meaning. In contrast, AI-based approaches are typically less interpretable but can be enhanced by XAI frameworks to identify influential environmental factors [65]. While physics models are generally more robust when dealing with out-of-distribution data, their sensitivity to initial conditions and soil heterogeneity can be a limitation [66]. Nevertheless, a critical distinction between the two methods lies in their robustness under extreme events, like drought or heavy rainfall. Physics-based models offer higher reliability because their mechanistic nature remains valid when environmental conditions exceed normality. As a result, they can predict moisture dynamics for rare events that have not occurred in the past, where the accuracy of AI-based models often degrades when they encounter conditions underrepresented in their training sets [67]. In contrast, data-driven approaches achieve accuracy compared to physical models when sufficient data are available and can produce moisture maps at finer spatial resolutions compared to satellite images [68].

Today, soil moisture prediction is largely based on data-driven AI approaches. The selected method depends on the type of data. Common AI approaches include RFs and ANNs. In fact, ANNs have been utilized since the 1990s and can integrate correlations of satellite signals and soil moisture, which are difficult to describe with equations. For example, Satalino et al. [69] studied the ability of satellite data to predict soil moisture by using ANNs. The authors utilized radar data retrieved from the Integral Equation Model (IEM) and found that the model predicted soil moisture with an error of only 6%. Notably, the main source of error was surface roughness, making it hard for the signal to separate moisture from soil texture. Even so, the ANN was able to handle this complexity reasonably well. Similarly, Notarnicola et al. [70] applied ANNs to radar satellite signaling and obtained an average accuracy of R^2 = 0.8, concluding that when compared to commonly used methods, e.g., the Bayesian method, ANNs can offer faster and more stable results.

Deciding between which AI tool to use depends on the variability of the data. To illustrate, a study in Tamil Nadu, India, compared eleven ML models—including linear regression, SVM, RF, ANN, and several hybrid/metaheuristic models (LSTM-based)—to predict monsoon season soil moisture (June–September) from 2001 to 2014, using gridded soil moisture data together with IMD rainfall and high-resolution topographic variables. They found that RFs clearly performed best, substantially outperforming SVMs, basic ANNs, and especially the LSTM-based models [71]. On the other hand, when spatiotemporal data (remote sensing or ground sensor network, satellite images) and time series data are available, DL methods outperform classical ML calculations by far. Han et al. [72] predicted daily soil moisture at four depths (100, 200, 500, and 1000 mm) up to 6 days ahead. The study was performed at the Eagle Lake Observatory in California, USA, using daily data from November 2014 to February 2020 collected by the SCAN monitoring network under natural field conditions. The authors used five input variables (air temperature, precipitation, vapor pressure, soil temperature, and relative humidity), and data were trained on an ANN and an LSTM model. They found that both models worked well, but LSTM consistently outperformed ANN, especially for deeper layers and longer lead times (R^2 = 0.90 for LSTM vs. 0.80–0.97 for ANN, with RMSE < 2.0), while ANN performed slightly better only for very short-term (1-day) predictions at the surface layer.

Satellite observations can provide indirect but highly informative variables related to soil. One such variable is evapotranspiration (ET), which is strongly controlled by soil moisture availability and vegetation water stress. Recent satellite missions, such as ECOSTRESS, aimed at retrieving land surface temperature at very high spatial resolution and enabling the derivation of ET at field and sub-field scales. Figure 4 illustrates ECOSTRESS-derived ET over irrigated agricultural areas, highlighting strong spatial heterogeneity in water use within individual fields.

Although ET is not a direct measurement of soil moisture, it reflects the integrated response of soil-plant-atmosphere interactions and therefore serves as a valuable proxy and input for data-driven soil moisture estimation models. Machine learning tools can be trained to identify the relationship between environmental data (e.g., rainfall, vegetation, etc.) and soil moisture directly by utilizing past data. Studies have shown that when enough data are available, soil moisture prediction can be as accurate as physical models [73]. In an overview, Taheri et al. [74] found that DL tools have the capacity to learn complex soil moisture patterns when given a lot of data, as stated before. However, when data are limited or noisy, simpler tools like SVMs can perform better. Indeed, fine-scaled moisture maps have been developed by ML models in many projects using in situ sensor networks and satellite data as inputs [75]. As a result, farmers can apply smart irrigation systems that irrigate crops only when needed. There are even reports that claim saving on water usage in the order of 20–30% without yield penalty, by utilizing AI-based soil moisture prediction systems [76]. For instance, if the model decides that the amount of moisture in the soil will remain adequate for the next 3 days, given the weather forecast, irrigation can be delayed. Furthermore, on a larger crop scale, AI-based moisture monitoring can significantly improve early drought warning signs and support water management by estimating soil moisture deficits in a region [77].

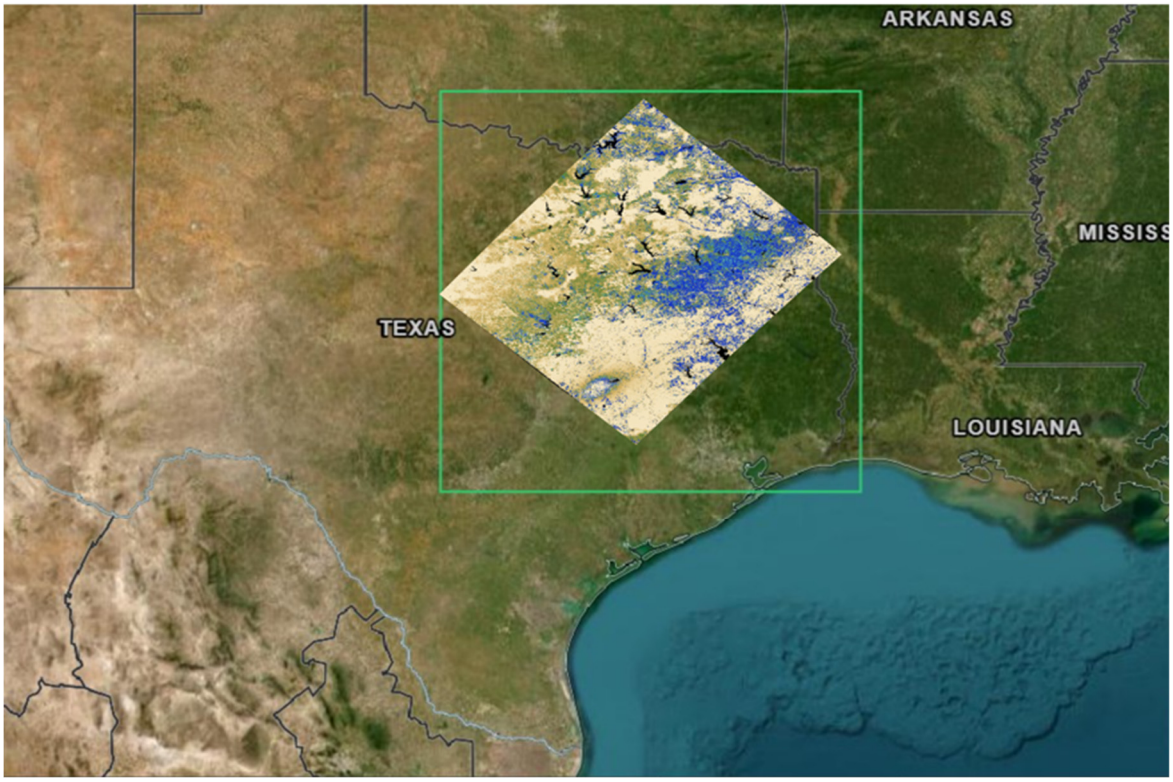


Figure 4. Example of high-resolution evapotranspiration (ET) derived from the ECOSTRESS mission over agricultural regions of the central United States. Blue colors indicate higher ET associated with actively transpiring vegetation, while beige colors indicate lower ET. Image reproduced from NASA/JPL-Caltech Earthdata and shown for illustrative purposes only; the figure was not generated by the authors.

4.4. Soil Contamination Monitoring and Remediation

Soil contamination can rise from both organic and inorganic pollutants, including heavy metals, pesticides, harmful hydrocarbons, PFAS, microplastics, and petroleum byproducts [78,79]. Such substances can be deposited in soil, often due to anthropogenic activity, and be absorbed by plants or leached down towards the groundwater, posing risks for human health [80]. Monitoring soil pollution requires extensive soil sampling, well-trained lab labor, and expensive instruments for analysis. Moreover, soil contamination is usually localized near industrial sites, mines, or urbanized areas, making it difficult to be characterized with a small number of samples [81]. To help improve this, there have been efforts to develop AI maps to identify and predict soil contamination by analyzing indirect data using the technique of complex data recognition.

Most research focuses on heavy metals due to their toxic and persistent nature. As non-degradable substances, they can pose serious risks for the environment, especially for commonly found metals in soils like arsenic, cadmium, and lead [80]. In AI, factors like proximity to highways, past industrial activity, and soil pH are the most crucial for predicting soil contamination [82]. This approach has delivered accurate maps of contamination hotspots and provided information via the model on the root causes of soil pollution. Furthermore, even though heavy metals cannot be observed via satellite images, they often cause secondary effects like toxicity symptoms in plants, unusual soil coloring, and spectral anomalies. As such, high-resolution hyperspectral imagery can capture even minute changes in soil and vegetation as a result of contamination, and AI models like SVMs or CNNs can be trained to compare differences between contaminated and non-contaminated regions. One study used remote sensing analysis combined with

unusual image signals to detect likely pollutant source patterns [83]. This approach was reported to have delivered much faster results compared to manually surveying land for sampling. The model can flag areas that are likely to contain high metal concentrations and help researchers decide where samples should be collected. For example, data from the European LUCAS soil survey, research efforts applied AI methods to show that 5.5% of soil samples contained metal concentrations exceeding safety thresholds (Figure 5) [84].

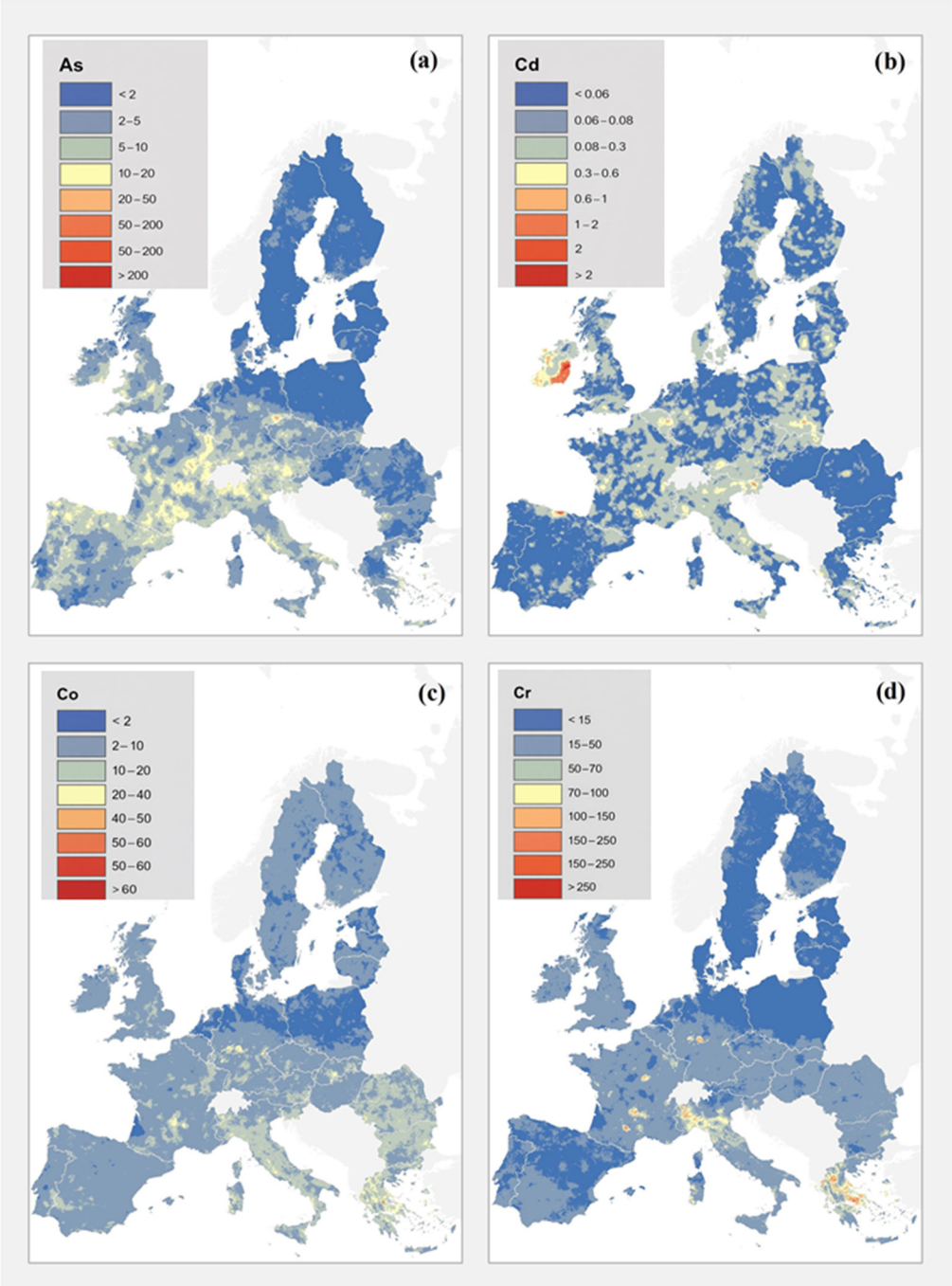


Figure 5. An example of four heavy metal concentrations, i.e., (**a**) As, (**b**) Cd, (**c**) Co, and (**d**) Cr in Europe, provided by the European Soil Data Centre (Sustainable Resources Directorate, JRC Ispra) and retrieved from Tóth et al. [84].

Emerging contaminants, like microplastics and PFAS, are another topic where AI could prove useful. Detecting microplastics in soil is a difficult and time-consuming operation that requires laborious extraction and analysis with microscopy. Nevertheless, researchers have

started utilizing computer vision to automatically detect and count microplastic particles in soil samples. The review of Rosca and Stancu [85] highlighted that plastic pollution in soil is an understudied subject and suggested that AI could be used to help prevent this problem. For instance, AI could link plastic pollution from field data with factors such as proximity to cities or plastic mulch usage in agriculture and create maps showing areas of high risk. Similarly, polycyclic hydrocarbons and pesticides could be monitored via AI by studying their outcomes on spectral signaling or soil biomarkers.

In the context of monitoring and assessing soil contamination, AI outputs should be clearly treated as tools for screening and prioritization, rather than as standalone instruments for regulatory decision making. Artificial intelligence is highly valuable for rapidly identifying potential contamination hotspots, which can significantly improve the efficiency of monitoring programs. Nevertheless, despite these advantages, AI predictions are inherently dependent on the quality, representativeness, and spatial coverage of the input data, and remain subject to uncertainties and limited generalizability across different environmental contexts. Given the legal, environmental, and public health implications of soil contamination, AI output alone cannot replace direct soil measurements, laboratory analyses, and expert evaluation in regulatory frameworks. As a result, AI should be regarded as a screening tool that complements traditional monitoring methods by informing where and how detailed assessments should be conducted, rather than serving as definitive evidence for compliance, remediation decisions, or legal determinations.

4.5. Soil Carbon and Climate Change Mitigation

Soil stores a large amount of Earth's carbon, mainly in the form of soil organic carbon (SOC) [86]. Managing this fraction of carbon is crucial for reducing climate change and keeping soils "healthy" and productive. As such, researchers aim to perform accurate measurements of soil carbon content and its fluctuations. Carbon levels in soil can be mapped via AI, as well as changes predicted as a result of agricultural practices or land use, to confirm carbon credits in sustainability programs [87].

The creation of carbon maps has benefited a lot from the use of AI digital soil mapping. The content of SOC can change even in small areas as a result of climate change, land use, farming practices, and soil type [88]. Hence, older soil maps may have missed this variability, while AI combines covariates from rich datasets (satellites, topography, climate data, and land features) to deliver high-resolution SOC maps. Especially for SOC, AI models can have high R^2 or correlation values due to a strong link between SOC and vegetation or leaf area index from satellites [89]. To further illustrate this, more vegetation means more plant material contributing to SOC and affecting climate change. Utilizing ML models like RFs or gradient boosting can help capture such relations across regions, delivering carbon distribution maps. A great example of such a case is the SoilGrids project (Figure 6), which uses AI to map soil carbon around the world at a fine scale (250 m resolution). Such maps help identify areas of high carbon content, such as forests and peatlands, as well as areas of low carbon content that could be managed.



Figure 6. Soil organic carbon distribution in Europe at 250-m resolution from the SoilGrids 2.0 dataset (ISRIC—World Soil Information). Soil carbon content is predicted using machine learning models based on soil observations and environmental covariates [90]. The image was not generated by the authors and is shown for illustrative purposes only.

Additionally, AI can be utilized to estimate SOC sequestration potential and its dynamics. Long-term experiments and soil monitoring networks can be employed for training on SOC prediction under various scenarios (e.g., the effect of no-till farming on carbon storage in a 20-year period). One novel approach is utilizing ANNs as emulators of process-based carbon models. With this approach, a neural network can be trained to copy the model's output from environmental inputs, providing a fast alternative. This can be later tested under different scenarios or integrated into larger models [91]. Furthermore, when it comes to climate change policies, AI can be proven critical. Global programs to reward soil carbon sequestration, such as funding farmers to increase carbon storage in soil, require strict and reliable measurements, as well as verification of SOC changes. In this case, AI tools (proxies, remote sensing, farmer data, etc.) can be used for SOC estimation and target where sampling is needed, where traditional methods are impractical in larger scale. For instance, an AI-based system might flag a field that has gained significant carbon as a result of changes in farming management. This means that this specific field might become a good candidate for soil sampling to confirm carbon credit. As such, AI can streamline carbon credit verification, making carbon farming more feasible.

In short, AI has become an imperative tool for SOC mapping, assisting decision-making processes on where and how to increase SOC. Research findings demonstrate the power of AI when calibrated with good data. Such advances are crucial, as they support global efforts to track carbon dynamics and develop nature-based solutions. However, as with other domains, attention is needed to model limits and uncertainty in SOC prediction, especially when used for policy guidance.

4.6. Precision Agriculture and Decision Support

Precision agriculture is an integrated approach that utilizes variability in a specific field in order to manage crops and soil in a highly efficient way [92]. Soil plays a crucial role in these variations because its properties can vary over very small distances, which can affect crop yield and the overall input costs. Artificial intelligence can analyze data from satellites, soil maps, drones, sensors, and farm machinery, and transform data into practical decisions, e.g., on the quantity of irrigation water and fertilization in each part of a field.

The key idea in precision agriculture is site-specific management, which means adjusting and optimizing farming practices based on the specific needs of different parts of the field. This is possible via data analysis, including, among others, soil fertility, moisture, and past crop yields, in order to identify which zones of the field require which management. For instance, an AI-driven device support system (DSS) might combine detailed soil nutrient maps with other factors like crop growth models to create a variable-rate fertilizer map. This would precisely indicate the amount of fertilizer that is required in each location. Studies have demonstrated that AI-based DSS used for planning variables such as seeding, fertilization, and smart irrigation can improve input efficiency by approximately 20–25% [5]. Hence, farmers can reduce the costs of fertilization and irrigation with no yield decrease by avoiding overapplication of unnecessarily high inputs.

Machine learning in precision agriculture assists farmers by predicting different production aspects such as yield, input rates, and zones of stress. For instance, when predicting crop yields, ML tools can retrieve soil data as inputs and predict the spatial yield distribution. If the model observes that certain soil data correlate significantly with yield variability, it can help in highlighting management zones [93]. By making such observations, clusters of algorithms are created so that the field can be divided into similar segments rather than treating the entire field equally. Similarly, these tools optimize fertilization levels, water, or other inputs, avoiding extra costs and improving efficiency. While advanced AI control is still in the phase of simulation and basic research, simpler optimization, like algorithms determining the best fertilization rates, is already in commercial use in precision agriculture.

Real-time control is another important topic, where modern farming machinery and tools are equipped with sensors (e.g., nutrient sensors or yield monitors) or machine vision (e.g., cameras to assess soil condition) [94]. Hence, AI models can make quick decisions utilizing this method. For instance, a smart irrigation system is capable of deciding the amount of water every hour based on soil moisture, or a farmer may adjust the seeding rate when it detects optimum soil conditions. The advantage of this method is that AI runs directly on machinery, rather than relying on the internet or cloud computing (the latter would be problematic in rural areas due to poor connectivity) [95]. By inserting AI directly on farm equipment, precision agriculture becomes more responsive and efficient.

5. Challenges and Limitations

5.1. Challenges of AI in Soil Applications

Despite growing use, AI faces limitations and gaps that hinder its practical application. One such case can be the existence of identical or uneven data can decrease model prediction accuracy. If a model is trained mainly, say, on temperate fertile soils, it may not perform well in arid, tropical, or boreal climates. To reduce overfitting, researchers limit input variables. Khaledian and Miller [96], for example, noted that in arable lands, soil differences are less informative compared to hilly regions, which makes AI models create accurate maps; so, common inputs, like soil texture, are not useful. This requires further research into applying new data, such as sensors or geochemical measurements, so that models can account for small differences. Another issue is the uncertainty of DSMs. Uncertainty can arise from

multiple sources, like errors in measurements, sampling bias, limitations in model structure, and spatial uncertainty due to spatial variability [97]. Addressing these issues has led to the development of quantification methods like ensemble modelling techniques (e.g., RFs), which capture model uncertainty by generating a spectrum of predictions from multiple models [16]. Brungard et al. [98] showed that spatially stratified RF ensembles reduced uncertainty compared to a global model. Quantile RFs can directly estimate the distribution of soil property values at each location, retain all measured values at tree nodes (rather than only the mean), and deliver prediction quantiles; for instance, mapping the 5th and 95th percentiles produces uncertainty maps [99]. In DL models, Monte Carlo dropout provides a practical Bayesian approximation of model uncertainty by randomly dropping neurons at inference time over many forward passes; one obtains a distribution of predictions, from which prediction intervals can be computed [16]. This method produces wider, more realistic intervals for novel inputs, avoiding overconfident extrapolations. Another straightforward method is mapping cross-validation residuals, which allows visualization of spatial patterns of model error and identification of high uncertainty areas [100].

The lack of a common standard methodology and data in soil fertility prediction is another challenge of AI. Different studies utilize different data, targeting a variety of aspects, e.g., crop yield instead of nutrient amounts, with different accuracy methods [101]. Selecting which one fits best is difficult, as different accuracy measures make it hard to compare across different soil fertility studies [102]. Creating shared datasets or utilizing standardized methods could help unify this area of research. Another challenge is that most AI models are region-specific and do not easily transfer to other systems, as climate and other environmental factors differ significantly. Soil properties and management practices vary greatly across climates and landscapes, so a model trained on one region's data may perform poorly when applied elsewhere. Global reviews have identified pronounced regional biases in current soil datasets; for example, tropical and arid climates remain underrepresented in the data used to train many predictive models [103]. This lack of diverse training data means the resulting models learn region-specific patterns that do not hold universally. To overcome this, researchers emphasize the need for broader and standardized multi-regional data collection. Expanding soil monitoring to many regions and using consistent protocols would provide the diverse, high-quality datasets needed to train more universal AI models [103]. By pre-training models on large, data-rich regions and then fin-tuning them on target regions, it is possible to apply knowledge to new conditions. For instance, a model trained in one country has been successfully adapted via domain-alignment methods to make reliable predictions in other regions [104]. Such approaches can mitigate the generalizability of region-specific AI models.

Another problem is the dynamic nature of soil fertility. Nutrient availability and soil health parameters can fluctuate within growing seasons due to plant uptake, residue decomposition, leaching, etc., and also shift over years to decades under changing environmental conditions. For instance, long-term studies have documented significant changes in soil fertility indicators such as pH and soil organic carbon over recent decades in response to land use intensification and climate variability [105]. Capturing these temporal dynamics remains a challenge for AI models. Advanced AI techniques (e.g., LSTM-based models) can learn sequential patterns and long-term dependencies in soil data, thereby handling temporal relationships that static models cannot [60]. Likewise, spatiotemporal AI frameworks, such as graph neural networks with recurrent layers, have been proposed to capture temporal dynamics in soil fertility alongside spatial dependencies [106]. Such characteristics are not often accounted for in AI models. Additionally, trust in AI models is one major obstacle that has to be considered. Farmers, agronomists, and landowners are less likely to follow recommendations when causality is missing [107]. A variety of ML

tools deliver predictions without clear explanations, which decelerates AI adaptation on farms. For this to be solved, studies focus on XAI, in which various tools can show which factors, e.g., low organic carbon, poor crop growth, lead to decision making process [108].

Challenges also exist in the application of soil moisture prediction. The relationship of inputs, like satellite signals or weather and soil moisture, is heavily dependent on the site's specific traits, like soil structure, sensors, calibration, and vegetation. This reduces transferring models out of the box to other areas. One solution to this issue is federated learning or training multiple sites, where models are trained with inputs from different regions in order to gain broader applicability. In addition, training data for soil moisture does not always reflect the real world, as direct measurements of soil moisture are mostly performed in a small number of locations, like experimental sites, and even in that case, might not cover the full spectrum of conditions. On the other hand, ML models rely heavily on the quality of the trained data. Insufficient or biased data can lead to overfitting or complete failure of the model. For example, if data originate from a wet season, the model may struggle during the dry period due to mismatches in the input data [109]. However, there are ML strategies like cross-validation, regularization, and data augmentation that are utilized to improve model accuracy [110].

AI use for soil contamination also faces important challenges, primarily due to limited and uneven data [111]. Commonly, when measuring soil contamination, soil data are retrieved from places where pollution is expected, e.g., near industries or mines, which can bias models to consider all sites close to mines as polluted. On-site monitoring in this situation is needed; AI cannot fully replace manual soil sampling. Researchers still need to perform lab analyses, especially when health or legal decisions are involved. Thus, AI-flagging should be used to confirm lab analysis but should not be used as final proof. Also, soil contamination is multidisciplinary, involving different scientific field e.g., environmental chemistry, geology, water chemistry, socio-economic factors, etc. [112]. As a result, an AI model needs to combine all these different aspects in order to extract predictions. For example, for successful predictions on where pesticides could accumulate, data from agricultural practices would be required; such data are usually not included in soil indicator datasets. However, some models have started to include such information. In addition, different types of data, like geological maps, soil tests, pollution records, etc., are also incorporated to ensure reliability [5]. One more challenge is that soil contamination is a dynamic process. AI extracted models only show pollution levels at one point at a time, but soil contaminants may be allocated (e.g., leached down the soil profile) or even decrease if a remediation method is applied [78,79]. AI models will need to include data collected over a period of time (months or years) and explain with ML models their movement. This is similar to the moisture case, where LSTMs can forecast changes, but for contaminants, the process may be slower and the data even scarcer. Even so, in the future, AI models could be used to predict how pollution changes under different scenarios, such as estimating the metal bioavailability decrease when phytoremediation is applied.

Finally, challenges and gaps remain in fully taking advantage of AI for SOC, mainly due to data limitations. Most SOC data are obtained from the top layer of soil, when a large amount of carbon is stored in deeper layers [113], making subsurface predictions less reliable. Another issue is the lack of SOC mapping in several parts of the world. For example, some parts of Africa have very limited data availability [114]. Hence, the model must rely more on indirect information like climate or satellite information; models in data-poor regions must be used with caution and with clear information about uncertainty. Furthermore, SOC transformations, apart from being temporally dynamic, are spatially specific, with differences occurring even on a very small scale. Detecting a meaningful change (e.g., a 5% decline in a 5-year period) is challenging since the "signal" of change might be

swamped by model uncertainty. Thus, models would increase their accuracy if trained for longer periods of time to capture typical ranges of change under different practices. Currently, there is a gap: most of the AI studies do not contain repeated measurements taken over many years, and this leads to predictions of low accuracy. To develop such tools for increased accuracy, hybrid approaches might be needed, such as combining data-driven patterns with already known processes like SOC decomposition rates, kinetics, etc.

5.2. Validation Strategies and Performance Metrics in AI-Based Soil Models

Artificial intelligence and ML studies in soil science often report the accuracy of models using R^2, RMSE, etc., but these metrics can lead to misleading assumptions when not validated. Many authors still rely on random k-fold cross-validation, which assumes independent, identically distributed samples. In spatial data, this is problematic. For example, Piikki et al. [43] showed that random k-fold tends to overestimate map accuracy when samples are clustered. In contrast, spatial cross-validation methods exclude nearby points from the same fold. For instance, Habibi et al. [115] compared random vs. spatially clustered cross-validation and found that random splitting gave optimistic soybean yield predictions, whereas spatially more reliable predicted an independent test field. In practice, researchers have found that spatially structured cross-validation produces a lower R^2 but much better generalization than random splits.

In addition, soil data are often strongly autocorrelated in space and time. If nearby observations are similar, a model can "cheat" by learning local patterns. Spatial autocorrelation, therefore, inflates apparent accuracy unless accounted for. For example, Ploton et al. [116] emphasize that when spatial correlation is ignored, random cross-validation cannot serve as spatially independent validation, meaning test points benefit from nearby training points. Likewise, temporal autocorrelation can mislead time-series models like soil moisture; if training and test data are from the same season or close months, models simply learn seasonal cycles. Tziachris et al. [117] showed this for soil salinity, where a model trained on 2020 data had $R^2 = 0.78$ under random cross-validation, but when tested on 2019 data from the same site, the R^2 fell to 0.70. In short, failing to separate samples will give an overly optimistic R^2.

Furthermore, another problem is the lack of external validation. Most soil AI studies stop at interna cross validation and never test models on truly independent data. Piikki et al. [43] systematically found only a few percent of studies used external test sets, while almost all found that only a few percent of studies used external test sets, while almost all relied on some form of data-splitting cross-validation. They stress that the final map is validated only when an independent sampling is carried out. Without holding out separate sites or dates, one cannot be confident that a model will generalize. Indeed, Bernardini et al. [37] showed that an RF SOC model with high R^2 performed no better than chance under leave-one-site-out testing, revealing its reliance on site-specific effects. Similarly, Habibi et al. [115] emphasize that yield models must be tested on independent fields, showing that random cross-validation models failed to predict an unseen field, whereas spatial cross-validation gave more realistic results. Hence, without independent validation, reported accuracy only reflect interpolation, not true predictive skill.

6. Future Directions

Artificial intelligence is already playing a crucial role, and will continue to do so, in soil science and land management. This is a result of ongoing technological advancements and global needs, such as feeding an increasing population, dealing with climate change, and maintaining sustainable farming. Researchers aim to improve AI tools to solve modern problems more effectively. Hence, we outline key future directions for research

and development that emerge from the state of the art in the field, helping to address existing challenges.

Edge computing and in situ AI are one likely trend for future AI, as discussed before. Putting AI directly on farm equipment, instead of relying on connectivity, is one big trend. Additionally, AI systems could be embedded on tractors, irrigation systems, or hand-held devices to extract information in real-time on site. For example, an edge AI device could monitor soil moisture and adjust irrigation accordingly, or a portable soil scanner could use a smaller neural network to instantly predict soil health indicators [118]. Furthermore, any innovations in hardware (IoT sensors and microcontrollers) combined with AI could lead to the creation of the "Internet of Soils," where soil itself would communicate its needs to farmers or automated systems with the aim of improving soil management and sustainability. Nonetheless, broad development is still at its beginning [119]. Cost and complexity remain hurdles; robust microcomputers or specialized chips are needed at scale, and ensuring they are affordable and user-friendly for farmers and practitioners is a challenge.

Another promising future use is decentralized federated learning, which could find use in battling privacy issues. In this scenario, stakeholders train a shared model on their local data and exchange only model parameter updates [120]. In this way, AI models learn from many farms, improving their predictions for parameters like yield or soil nutrients, while each farm obtains a unified model and keeps its data private. This could solve the problem of poor data regions by benefiting from collective intelligence, without data ownership being compromised. However, data privacy and ownership are paramount concerns. Soil data might not seem personal at first glance; however, it can give clues about a farm's productivity, practices, or even economic situation. In traditional cloud-based systems, when soil data from many sources is aggregated, there is a risk that contributors may lose control over their information. Federated learning can address such concerns, since raw data stays on local devices and only learned parameters are shared [121]. Yet even federated learning requires trust; participants must trust that the system will not leak any information. Hence, techniques like adding noise to updates or secure aggregation protocols are being developed to strengthen security in federated learning. Data ownership is equally tricky. Models trained on farmers might end commercial products; who has the right to those models and predictions? Without clear agreements, farmers may feel exploited if their data are fed to an AI that someone else sells. On the other hand, heavy reliance on AI and automatization on decision making raises questions on accountability; for example, if an AI-driven irrigation system fails, who is responsible? In summary, innovation must be pursued with strong ethical guidelines and clear data governance frameworks.

Another future need in soil science may be the creation of digital twins between soil and farming systems. The term "digital twins" refers to a virtual copy of a real system that is continuously updated with real-time data. For soils, a digital twin might integrate sensor data, AI-based models, and other factors to simulate soil conditions virtually [122]. This creates an opportunity for "what-if" testing scenarios, which helps farmers maximize yields in the most efficient way. For example, a farmer could state a hypothetical question, "If I irrigate 20% less for the next month, how will soil moisture and crop growth be affected?" or "if I switch my practice to cover cropping, how will soil carbon and nitrogen be affected?" The AI can respond by mirroring the twins' predictions against real sensor data. Hence, digital twins can be powerful decision-making tools in virtual reality to guide real-world practices. However, user-accessible digital twins is resource intensive, as it demands extensive data integration, computing power, and expert knowledge on both soil science and modelling. So far, digital twins in agriculture remain largely experimental [123],

and scaling them up to regional or global "twins" would require enormous technical and financial challenges.

To address the problem of trust and reliability, future AI research and development will likely place a strong emphasis on explainability. Expert knowledge could be embedded into models or by explanation tools tailored to soil-crop scenarios. For example, an AI system could automatically generate recommendations in the form of a report. To illustrate this further, a recommendation could be as follows: "The field is expected to produce lower yield due to historically low organic matter content and limited rainfall, so the model recommends the addition of organic fertilizers." Researchers are already in the hunt for interpretable ML frameworks for agriculture, especially if AI outputs are used for policy or justification reasons.

In distinguishing short- and long-term goals, some AI innovations are ready to be applied while others are aspirational. In the short term, those involved in soil science can capitalize on sensors and computing infrastructures; for example, integrating IoT sensors with cloud computing is an immediate practical action. Many soil nutrient and moisture sensors are readily available and can stream data to cloud applications for analysis. Cloud-based services are widely accessible and can handle large datasets from soil surveys and sensors. Similarly, proximal sensing techniques are readily available and, when combined with ML tools, can provide rapid soil property assessment in the field. The advances in satellite technology (higher resolution and LIDAR from space) will lead to enhanced remote sensing, and proximal sensing (mounted scanners on tractors, drones, and sensors) will be widely used [124]. Furthermore, soil spectroscopy is becoming cheaper, leading to multiple soil property predictions by AI models [125]. As these methodologies pass from the lab to on-site measurements, we might be able to perform instantaneous soil testing.

In contrast, long-term goals include more futuristic or large-scale visions such as harnessing quantum computing. Quantum computing, which utilizes qubits to potentially solve complex problems far faster than computers, could revolutionize data-heavy soil modelling. For example, quantum algorithms might one day unravel with highly complex soil biogeochemical interactions. Nevertheless, current quantum hardware is extremely limited and practically not accessible. While researchers are already formulating quantum algorithms for agricultural problems to prepare for future breakthroughs [126], any tangible impact of quantum computing on routine soil science lies well beyond the immediate horizon. Achieving it would require massive investment in sensor infrastructure, standardization, and global cooperation.

7. Conclusions

The future of AI in soil science is very promising. New tools demonstrate the ability to work directly in the field and at a large scale to aid soil management in a sustainable and climate-resilient way. Such technologies can make data-driven decisions a normal part of modern agriculture. Nevertheless, challenges remain. Technologies must be understood in order to be widely adopted; thus, delivering final, easy-to-understand, ethically sound, inexpensive, and friendly to end-users' products is the main challenge for AI systems. In the near future, AI could become an everyday tool in soil management, and within the next decade, it may help enhance soil literacy. By coupling sound knowledge of soil science with advanced AI and encouraging cooperation among farmers, scientists, and policymakers can utilize this technology to support soil management more efficiently.

Author Contributions: Conceptualization, V.A.; methodology, V.A.; validation, C.K. and V.A.; formal analysis, C.K.; investigation, C.K.; data curation, C.K.; writing—original draft preparation, C.K.; writing—review and editing, V.A.; visualization, C.K.; supervision, V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ANN	Artificial neural network
ML	Machine learning
RF	Random forest
GIS	Geographical information system
DSM	Digital soil mapping
SVM	Support vector machine
DL	Deep learning
CNN	Convolutional neural network
SOC	Soil organic carbon
LLM	Large language model
LSTM	Long short-term memory
GenAI	Generative AI
XAI	Explainable artificial intelligence
ET	Evapotranspiration
IoT	Internet of things
DSS	Device support system
RMSE	Root mean square error

References

1. Kopittke, P.M.; Harper, S.M.; Asio, L.G.; Asio, V.B.; Batalon, J.T.; Batuigas, A.M.T.; Gonzaga, A.B.; Gonzaga, N.R.; de Guzman, M.T.L.; Lumanao, D.M.; et al. Soil Degradation: An Integrated Model of the Causes and Drivers. *Int. Soil Water Conserv. Res.* **2025**, *13*, 744–755. [CrossRef]
2. Shokri, N.; Robinson, D.A.; Afshar, M.; Alewell, C.; Aminzadeh, M.; Arthur, E.; Broothaerts, N.; Campbell, G.A.; Eklund, L.; Gupta, S.; et al. Rethinking Global Soil Degradation: Drivers, Impacts, and Solutions. *Rev. Geophys.* **2025**, *63*, e2025RG000883. [CrossRef]
3. Liu, Y.; Cao, J.; Liu, C.; Ding, K.; Jin, L. Datasets for large language models: A comprehensive survey. *Artif. Intell. Rev.* **2025**, *58*, 403. [CrossRef]
4. Balsalobre-Lorente, D.; Pilař, L.; Shah, S.A.R.; Radulescu, M. Is Supply Chain Digitization a Supportive Instrument for Green Energy Resilience? A Heterogeneity Factors Analysis for OECD Economies under Environmental Sustainability. *Energy Convers. Manag.* **2025**, *325*, 119384. [CrossRef]
5. Mishra, H.; Kayusi, F. Artificial intelligence and soil conservation: An overview. *J. Sci. Agric.* **2025**, *9*, 229–249. [CrossRef]
6. Abdelhak, M. Innovative Techniques for Soil and Water Conservation. In *Ecosystem Management: Climate Change and Sustainability*; Wiley: Hoboken, NJ, USA, 2024; pp. 291–326. [CrossRef]
7. Holt, D. Potentials for Artificial Intelligence and Supercomputers in Soil Science. In *Future Developments in Soil Science Research*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 1978; pp. 459–468. [CrossRef]
8. McCracken, R.J.; Cate, R.B. Artificial Intelligence, Cognitive Science, and Measurement Theory Applied in Soil Classification. *Soil Sci. Soc Amer J.* **1986**, *50*, 557–561. [CrossRef]
9. Dale, M.B.; McBRATNEY, A.B.; Russell, J.S. On the Role of Expert Systems and Numerical Taxonomy in Soil Classification. *J. Soil Sci.* **1989**, *40*, 223–234. [CrossRef]
10. Schaap, M.G.; Leij, F.J.; Van Genuchten, M.T. Neural Network Analysis for Hierarchical Prediction of Soil Hydraulic Properties. *Soil Sci. Soc Amer J.* **1998**, *62*, 847–855. [CrossRef]
11. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
12. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]

13. Sanchez, P.A.; Ahamed, S.; Carré, F.; Hartemink, A.E.; Hempel, J.; Huising, J.; Lagacherie, P.; McBratney, A.B.; McKenzie, N.J.; Mendonça-Santos, M.D.L.; et al. Digital Soil Map of the World. *Science* **2009**, *325*, 680–681. [CrossRef]
14. Ballabio, C. Spatial Prediction of Soil Properties in Temperate Mountain Regions Using Support Vector Regression. *Geoderma* **2009**, *151*, 338–350. [CrossRef]
15. Hengl, T.; Mendes De Jesus, J.; Heuvelink, G.B.M.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLoS ONE* **2017**, *12*, e0169748. [CrossRef] [PubMed]
16. Padarian, J.; Minasny, B.; McBratney, A.B. Using Deep Learning for Digital Soil Mapping. *SOIL* **2019**, *5*, 79–89. [CrossRef]
17. Nadporozhskaya, M.; Kovsh, N.; Paolesse, R.; Lvova, L. Recent Advances in Chemical Sensors for Soil Analysis: A Review. *Chemosensors* **2022**, *10*, 35. [CrossRef]
18. Da Silva, A.F.; Nathaniel, J.; Wong, K.C.L.; Watson, C.; Wang, H.; Singh, J.; Chamon, A.A.; Klein, L. NetZeroCO2, an AI Framework for Accelerated Nature-Based Carbon Sequestration. In *Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17 December 2022*; IEEE: Osaka, Japan, 2022; pp. 4881–4887.
19. Minasny, B.; McBratney, A.B. Machine Learning and Artificial Intelligence Applications in Soil Science. *Eur. J. Soil Sci.* **2025**, *76*, e70093. [CrossRef]
20. Singh, R.; De, M.; Banerjee, R.; Nayak, A.; Dasgupta, S.; Das, A.; Dey, S.; Biswas, A.; Weindorf, D.C.; Chakraborty, S. Enhancing Soil Organic Carbon Estimation with Generative AI and Nix Color Sensor. *Sci. Rep.* **2025**, *15*, 40628. [CrossRef]
21. Sai, S.; Kumar, S.; Gaur, A.; Goyal, S.; Chamola, V.; Hussain, A. Unleashing the Power of Generative AI in Agriculture 4.0 for Smart and Sustainable Farming. *Cogn. Comput.* **2025**, *17*, 63. [CrossRef]
22. Keerthan Kumar, T.G.; Shubha, C.A.; Sushma, S.A. Random forest algorithm for soil fertility prediction and grading using machine learning. *Int. J. Innov. Technol. Explor. Eng* **2019**, *9*, 1301–1304. [CrossRef]
23. Zhang, X.-Y.; Zhang, X.-P.; Yu, H.-G.; Liu, Q.-S. A Confident Learning-Based Support Vector Machine for Robust Ground Classification in Noisy Label Environments. *Tunn. Undergr. Space Technol.* **2025**, *155*, 106128. [CrossRef]
24. Soni, P.; Kumar, R.; Mishra, S.; Swain, S.; Mishra, P.; Kumar, S.S. Optimized Neural Network for Soil Moisture Prediction in Precision Agriculture. *Measurement* **2025**, *252*, 117380. [CrossRef]
25. Montúfar, G.; Pascanu, R.; Cho, K.; Bengio, Y. On the Number of Linear Regions of Deep Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2014; Volume 27.
26. Al Masmoudi, Y.; Tallou, A.; Ouaissa, M.; Ouaissa, M. Applications of machine learning and artificial intelligence in soil and environmental sciences: Methods, challenges, and future perspectives. In *Climate Resilience: Impact of Quantum Computing and Artificial Intelligence on Urban Planning*; Springer Nature: Cham, Switzerland, 2025; pp. 231–256.
27. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional Neural Network for Simultaneous Prediction of Several Soil Properties Using Visible/near-Infrared, Mid-Infrared, and Their Combined Spectra. *Geoderma* **2019**, *352*, 251–267. [CrossRef]
28. Datta, P.; Faroughi, S.A. A Multihead LSTM Technique for Prognostic Prediction of Soil Moisture. *Geoderma* **2023**, *433*, 116452. [CrossRef]
29. Wan, H.; Li, J.; Shang, S.; Rahman, K.U. Exploratory Factor Analysis-Based Co-Kriging Method for Spatial Interpolation of Multi-Layered Soil Particle-Size Fractions and Texture. *J. Soils Sediments* **2021**, *21*, 3868–3887. [CrossRef]
30. Awais, M.; Naqvi, S.M.Z.A.; Zhang, H.; Li, L.; Zhang, W.; Awwad, F.A.; Ismail, E.A.A.; Khan, M.I.; Raghavan, V.; Hu, J. AI and machine learning for soil analysis: An assessment of sustainable agricultural practices. *Bioresour. Bioprocess.* **2023**, *10*, 90. [CrossRef]
31. Zhu, A.X.; Hudson, B.; Burt, J.; Lubich, K.; Simonson, D. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. *Soil Sci. Soc. Am. J.* **2001**, *65*, 1463–1472. [CrossRef]
32. Espinel, R.; Herrera-Franco, G.; Rivadeneira García, J.L.; Escandón-Panchana, P. Artificial intelligence in agricultural mapping: A review. *Agriculture* **2024**, *14*, 1071. [CrossRef]
33. Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G.; et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* **2023**, *55*, 194:1–194:33. [CrossRef]
34. Wang, Q.; Li, C.; Hao, D.; Xu, Y.; Shi, X.; Liu, T.; Sun, W.; Zheng, Z.; Liu, J.; Li, W.; et al. A Novel Four-Dimensional Prediction Model of Soil Heavy Metal Pollution: Geographical Explanations beyond Artificial Intelligence "Black Box". *J. Hazard. Mater.* **2023**, *458*, 131900. [CrossRef]
35. Zhao, W.; Ma, J.; Li, X.; Li, T.; Wang, M.; Chen, Y. Explainable Deep Learning Unveils Critical Scenarios Driving Soil Cadmium Pollution in a Coastal Industrial City in China: A Geospatial AI Approach. *Environ. Pollut.* **2025**, *382*, 126769. [CrossRef]
36. Wang, X.; Borjesson, T.; Wetterlind, J.; van der Fels-Klerx, H.J. Prediction of deoxynivalenol contamination in spring oats in Sweden using explainable artificial intelligence. *NPJ Sci. Food* **2024**, *8*, 75. [CrossRef]

37. Bernardini, L.G.; Rosinger, C.; Bodner, G.; Keiblinger, K.M.; Izquierdo-Verdiguier, E.; Spiegel, H.; Retzlaff, C.O.; Holzinger, A. Learning vs. Understanding: When Does Artificial Intelligence Outperform Process-Based Modeling in Soil Organic Carbon Prediction? *New Biotechnol.* **2024**, *81*, 20–31. [CrossRef]
38. Kheir, A.M.S.; Govind, A.; Nangia, V.; El-Maghraby, M.A.; Elnashar, A.; Ahmed, M.; Aboelsoud, H.; Gamal, R.; Feike, T. Hybridization of Process-Based Models, Remote Sensing, and Machine Learning for Enhanced Spatial Predictions of Wheat Yield and Quality. *Comput. Electron. Agric.* **2025**, *234*, 110317. [CrossRef]
39. Gautam, S.; Mishra, U.; Scott, S.N.; Lara, M.J. Machine Learning and Process-Based Modeling of Spatiotemporal Changes in Active Layer Thickness across Alaska. *Sci Rep* **2025**, *15*, 42420. [CrossRef] [PubMed]
40. John, K.; Saurette, D.D.; Heung, B. The Problematic Case of Data Leakage: A Case for Leave-Profile-out Cross-Validation in 3-Dimensional Digital Soil Mapping. *Geoderma* **2025**, *455*, 117223. [CrossRef]
41. Mirzaeitalarposhti, R.; Shafizadeh-Moghadam, H.; Taghizadeh-Mehrjardi, R.; Demyan, M.S. Digital Soil Texture Mapping and Spatial Transferability of Machine Learning Models Using Sentinel-1, Sentinel-2, and Terrain-Derived Covariates. *Remote Sens.* **2022**, *14*, 5909. [CrossRef]
42. Wadoux, A.M.J.-C. Artificial Intelligence in Soil Science. *Eur. J. Soil Sci.* **2025**, *76*, e70080. [CrossRef]
43. Piikki, K.; Wetterlind, J.; Söderström, M.; Stenberg, B. Perspectives on Validation in Digital Soil Mapping of Continuous Attributes—A Review. *Soil Use Manag.* **2021**, *37*, 7–21. [CrossRef]
44. Canero, F.M.; Rodriguez-Galiano, V.; Aragones, D. Machine Learning and Feature Selection for Soil Spectroscopy. An Evaluation of Random Forest Wrappers to Predict Soil Organic Matter, Clay, and Carbonates. *Heliyon* **2024**, *10*, e30228. [CrossRef]
45. Deiss, L.; Margenot, A.J.; Culman, S.W.; Demyan, M.S. Tuning Support Vector Machines Regression Models Improves Prediction Accuracy of Soil Properties in MIR Spectroscopy. *Geoderma* **2020**, *365*, 114227. [CrossRef]
46. Morais, T.G.; Tufik, C.; Rato, A.E.; Rodrigues, N.R.; Gama, I.; Jongen, M.; Serrano, J.; Fangueiro, D.; Domingos, T.; Teixeira, R.F.M. Estimating Soil Organic Carbon of Sown Biodiverse Permanent Pastures in Portugal Using near Infrared Spectral Data and Artificial Neural Networks. *Geoderma* **2021**, *404*, 115387. [CrossRef]
47. Yang, J.; Wang, X.; Wang, R.; Wang, H. Combination of Convolutional Neural Networks and Recurrent Neural Networks for Predicting Soil Properties Using Vis–NIR Spectroscopy. *Geoderma* **2020**, *380*, 114616. [CrossRef]
48. Wang, H.; Zhang, L.; Zhao, J. Application of a Fusion Attention Mechanism-Based Model Combining Bidirectional Gated Recurrent Units and Recurrent Neural Networks in Soil Nutrient Content Estimation. *Agronomy* **2023**, *13*, 2724. [CrossRef]
49. Bharambe, U.; Mahato, M.; Durbha, S.; Dhavale, C. Exploring Opportunities of Generative Artificial Intelligence for Sustainable Soil Analytics in Agriculture. In *Sustainable Development and Geospatial Technology*; Sharma, C., Shukla, A.K., Pathak, S., Singh, V.P., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 23–43.
50. Wu, Y. Large Language Models and the Future of Soil Health: Bridging Knowledge Gaps through Scalable Semantic Intelligence. *Soil Adv.* **2025**, *4*, 100065. [CrossRef]
51. van der Westhuizen, S.; Heuvelink, G.B.; Gardner-Lubbe, S.; Clarke, C.E. Biplots for understanding machine learning predictions in digital soil mapping. *Eco. Inform.* **2024**, *84*, 102892. [CrossRef]
52. Adeniyi, O.D.; Bature, H.; Mearker, M. A systematic review on digital soil mapping approaches in lowland areas. *Land* **2024**, *13*, 379. [CrossRef]
53. Chagas, C.d.S.; de Carvalho Junior, W.; Bhering, S.B.; Calderano Filho, B. Spatial Prediction of Soil Surface Texture in a Semiarid Region Using Random Forest and Multiple Linear Regressions. *Catena* **2016**, *139*, 232–240. [CrossRef]
54. Zhang, M.; Shi, W.; Xu, Z. Systematic Comparison of Five Machine-Learning Models in Classification and Interpolation of Soil Particle Size Fractions Using Different Transformed Data. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 2505–2526. [CrossRef]
55. Taghizadeh-Mehrjardi, R.; Mahdianpari, M.; Mohammadimanesh, F.; Behrens, T.; Toomanian, N.; Scholten, T.; Schmidt, K. Multi-Task Convolutional Neural Networks Outperformed Random Forest for Mapping Soil Particle Size Fractions in Central Iran. *Geoderma* **2020**, *376*, 114552. [CrossRef]
56. Beucher, A.; Rasmussen, C.B.; Moeslund, T.B.; Greve, M.H. Interpretation of Convolutional Neural Networks for Acid Sulfate Soil Classification. *Front. Environ. Sci.* **2022**, *9*, 809995. [CrossRef]
57. Wu, H.; Hao, H.; Lei, H.; Ge, Y.; Shi, H.; Song, Y. Farm Size, Risk Aversion and Overuse of Fertilizer: The Heterogeneity of Large-Scale and Small-Scale Wheat Farmers in Northern China. *Land* **2021**, *10*, 111. [CrossRef]
58. Miran, N.; Rasouli Sadaghiani, M.H.; Feiziasl, V.; Sepehr, E.; Rahmati, M.; Mirzaee, S. Predicting Soil Nutrient Contents Using Landsat OLI Satellite Images in Rain-Fed Agricultural Lands, Northwest of Iran. *Environ Monit Assess* **2021**, *193*, 607. [CrossRef]
59. Jain, S.; Sethia, D. A review on applications of artificial intelligence for identifying soil nutrients. In *International Conference on Agriculture-Centric Computation*; Springer Nature: Cham, Switzerland, 2023; pp. 71–86. [CrossRef]
60. Gunasekaran, K.; A., K.; Sreevardhan, P. Real-Time Soil Fertility Analysis, Crop Prediction, and Insights Using Machine Learning and Deep Learning Algorithms. *Front. Soil Sci.* **2025**, *5*, 1652058. [CrossRef]
61. Jia, X.; Fang, Y.; Hu, B.; Yu, B.; Zhou, Y. Development of Soil Fertility Index Using Machine Learning and Visible-Near-Infrared Spectroscopy. *Land* **2023**, *12*, 2155. [CrossRef]

62. Ma, L.; Zhao, L.; Cao, L.; Li, D.; Chen, G.; Han, Y. Inversion of Soil Organic Matter Content Based on Improved Convolutional Neural Network. *Sensors* **2022**, *22*, 7777. [CrossRef]
63. Deng, Y.; Xiao, L.; Shi, Y. Enhanced Hyperspectral Forest Soil Organic Matter Prediction Using a Black-Winged Kite Algorithm-Optimized Convolutional Neural Network and Support Vector Machine. *Appl. Sci.* **2025**, *15*, 503. [CrossRef]
64. Zhang, D.; Zhou, G. Estimation of soil moisture from optical and thermal remote sensing: A review. *Sensors* **2016**, *16*, 1308. [CrossRef]
65. Wöhling, T.; Delgadillo, A.O.C.; Kraft, M.; Guthke, A. Comparing Physics-Based, Conceptual and Machine-Learning Models to Predict Groundwater Levels by BMA. *Groundwater* **2025**, *63*, 484–505. [CrossRef]
66. Bagheri, A.; Patrignani, A.; Ghanbarian, B.; Pourkargar, D.B. A Hybrid Time Series and Physics-Informed Machine Learning Framework to Predict Soil Water Content. *Eng. Appl. Artif. Intell.* **2025**, *144*, 110105. [CrossRef]
67. Alsumaiei, A.A.; Alrumaidhi, M. Enhancing Soil Water Prediction in Arid Climates Using Multipredictor Machine-Learning Models and SHAP-Based Interpretability. *J. Irrig. Drain Eng.* **2026**, *152*, 04025049. [CrossRef]
68. Leonarduzzi, E.; Tran, H.; Bansal, V.; Hull, R.B.; De La Fuente, L.; Bearup, L.A.; Melchior, P.; Condon, L.E.; Maxwell, R.M. Training Machine Learning with Physics-Based Simulations to Predict 2D Soil Moisture Fields in a Changing Climate. *Front. Water* **2022**, *4*, 927113. [CrossRef]
69. Satalino, G.; Mattia, F.; Davidson, M.W.; Le Toan, T.; Pasquariello, G.; Borgeaud, M. On current limits of soil moisture retrieval from ERS-SAR data. *IEEE Transact. Geosci. Remote Sens.* **2002**, *40*, 2438–2447. [CrossRef]
70. Notarnicola, C.; Angiulli, M.; Posa, F. Soil Moisture Retrieval from Remotely Sensed Data: Neural Network Approach versus Bayesian Method. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 547–557. [CrossRef]
71. Settu, P.; Ramaiah, M. A Data Driven Comparison of Hybrid Machine Learning Techniques for Soil Moisture Modeling Using Remote Sensing Imagery. *Sci. Rep.* **2025**, *15*, 43170. [CrossRef]
72. Han, H.; Choi, C.; Kim, J.; Morrison, R.R.; Jung, J.; Kim, H.S. Multiple-Depth Soil Moisture Estimates Using Artificial Neural Network and Long Short-Term Memory Models. *Water* **2021**, *13*, 2584. [CrossRef]
73. Khanal, S.; Fulton, J.; Shearer, S. An Overview of Current and Potential Applications of Thermal Remote Sensing in Precision Agriculture. *Comput. Electron. Agric.* **2017**, *139*, 22–32. [CrossRef]
74. Taheri, M.; Bigdeli, M.; Imanian, H.; Mohammadian, A. An Overview of Machine-Learning Methods for Soil Moisture Estimation. *Water* **2025**, *17*, 1638. [CrossRef]
75. Huang, J.; Sehgal, V.; Alvarez, L.V.; Brocca, L.; Cai, S.; Cheng, R.; Cheng, X.; Du, J.; El Masri, B.; Endsley, K.A.; et al. Remotely sensed high-resolution soil moisture and evapotranspiration: Bridging the gap between science and society. *Water Resour. Res.* **2025**, *61*, e2024WR037929. [CrossRef]
76. Mortazavizadeh, F.; Bolonio, D.; Mirzaei, M.; Ng, J.L.; Mortazavizadeh, S.V.; Dehghani, A.; Mortezavi, S.; Ghadirzadeh, H. Advances in Machine Learning for Agricultural Water Management: A Review of Techniques and Applications. *J. Hydroinform.* **2025**, *27*, 474–492. [CrossRef]
77. Oyounalsoud, M.S.; Yilmaz, A.G.; Abdallah, M.; Abdeljaber, A. Drought prediction using artificial intelligence models based on climate data and soil moisture. *Sci. Rep.* **2024**, *14*, 19700. [CrossRef]
78. Antoniadis, V.; Levizou, E.; Shaheen, S.M.; Ok, Y.S.; Sebastian, A.; Baum, C.; Prasad, M.N.V.; Wenzel, W.W.; Rinklebe, J. Trace Elements in the Soil-Plant Interface: Phytoavailability, Translocation, and Phytoremediation–A Review. *Earth-Sci. Rev.* **2017**, *171*, 621–645. [CrossRef]
79. Kikis, C.; Giannoulis, K.D.; Thalassinos, G.; Rinklebe, J.; Shaheen, S.M.; Antoniadis, V. From Phytoremediation to Phytomanagement: The Utilization of Industrial Crops for the Restoration of Contaminated Soils—A Review. *J. Environ. Chem. Eng.* **2026**, *14*, 120581. [CrossRef]
80. Grammenou, A.; Thalassinos, G.; Petropoulos, S.A.; Antoniadis, V. Cadmium and zinc sorption and desorption in soil: The impact of humic-fulvic acids, Bacillus sp., insect frass, and soil aging. *Environ. Sci. Pollut. Res.* **2025**, *32*, 17856–17867. [CrossRef]
81. Kikis, C.; Thalassinos, G.; Antoniadis, V. Soil Phytomining: Recent Developments—A Review. *Soil Syst.* **2024**, *8*, 8. [CrossRef]
82. Wang, J.; Deng, Y.; Huang, Z.; Li, D.A.; Zhang, X. Identification of Driving Factors for Heavy Metals and Polycyclic Aromatic Hydrocarbons Pollution in Agricultural Soils Using Interpretable Machine Learning. *Sci. Total Environ.* **2025**, *960*, 178384. [CrossRef]
83. Dean, J.; Ahmed, S.; Cheung, W.; Salaudeen, I.; Reynolds, M.; Bowerbank, S.L.; Nicholson, C.E.; Perry, J.J. Use of Remote Sensing to Assess Vegetative Stress as a Proxy for Soil Contamination. *Environ. Sci. Process. Imp.* **2024**, *26*, 161–176. [CrossRef]
84. Tóth, G.; Hermann, T.; Szatmári, G.; Pásztor, L. Maps of Heavy Metals in the Soils of the European Union and Proposed Priority Areas for Detailed Assessment. *Sci. Total Environ.* **2016**, *565*, 1054–1062. [CrossRef]
85. Rosca, C.-M.; Stancu, A. Emerging Trends in AI-Based Soil Contamination Monitoring and Prevention. *Agriculture* **2025**, *15*, 1280. [CrossRef]

86. Zavarzina, A.; Kulikova, N.; Belov, A.; Demin, V.; Rozanova, M.; Pogozhev, P.; Danilin, I. Soil Carbon Storage in Forest and Grassland Ecosystems Along the Soil-Geographic Transect of the East European Plain: Relation to Soil Biological and Physico-Chemical Properties. *Forests* **2026**, *17*, 69. [CrossRef]
87. Kotsompolis, G.; Cheilas, P.; Konstantakis, K.N.; Sfakianakis, E.; Goutte, S.; Michaelides, P.G. Smart Forecasting of Carbon Prices Using Machine Learning and Neural Networks: When ARIMA Meets XGBoost and LSTM. *J. Forecast.* **2026**, *45*, 47–60. [CrossRef]
88. Song, Y.; Yao, Y.; Kong, W.; Guo, L.; Bao, K.; Qiu, L.; Shao, M.; Wei, X. Effects of Vegetation Loss and Soil Erosion Intensity on Soil Carbon Dynamics across Landscape Position: Evidence from China's Loess Plateau. *Agric. Ecosyst. Environ.* **2026**, *396*, 109992. [CrossRef]
89. Jiang, X.; Wang, H.; Wu, D.; Ren, C. Soil carbon storage and climate change research supported by remote sensing data and AI models: Accurate estimation and dynamic analysis. *Geogr. Res. Bull.* **2024**, *3*, 454–470. [CrossRef]
90. Poggio, L.; De Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil* **2021**, *7*, 217–240. [CrossRef]
91. Guo, L.; Gao, Q.; Zhang, M.; Cheng, P.; He, P.; Li, L.; Ding, D.; Liu, C.; Muga, F.C.; Kamal, M.; et al. Soil Organic Matter Content Prediction Using Multi-Input Convolutional Neural Network Based on Multi-Source Information Fusion. *Agriculture* **2025**, *15*, 1313. [CrossRef]
92. Aarif, K.O.M.; Alam, A.; Hotak, Y. Smart sensor technologies shaping the future of precision agriculture: Recent advances and future outlooks. *J. Sensors* **2025**, *2025*, 2460098. [CrossRef]
93. Gavioli, A.; de Souza, E.G.; Bazzi, C.L.; Schenatto, K.; Betzek, N.M. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosyst. Eng.* **2019**, *181*, 86–102. [CrossRef]
94. Lakhiar, I.A.; Jianmin, G.; Syed, T.N.; Chandio, F.A.; Buttar, N.A.; Qureshi, W.A. Monitoring and control systems in agriculture using intelligent sensor techniques: A review of the aeroponic system. *J. Sens.* **2018**, *2018*, 8672769. [CrossRef]
95. Pierre, N.; Ishimwe Viviane, I.V.; Lambert, U.; Viviane, I.; Shadrack, I.; Erneste, B.; Schadrack, N.; Alexis, N.; Francois, K.; Theogene, H. AI Based Real-Time Weather Condition Prediction with Optimized Agricultural Resources. *EJT* **2023**, *7*, 36–49. [CrossRef]
96. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Modell.* **2020**, *81*, 401–418. [CrossRef]
97. Gavilán-Acuna, G.; Coops, N.C.; Olmedo, G.F.; Tompalski, P.; Roeser, D.; Varhola, A. Assessing Soil Prediction Distributions for Forest Management Using Digital Soil Mapping. *Soil Syst.* **2024**, *8*, 55. [CrossRef]
98. Brungard, C.; Nauman, T.; Duniway, M.; Veblen, K.; Nehring, K.; White, D.; Salley, S.; Anchang, J. Regional Ensemble Modeling Reduces Uncertainty for Digital Soil Mapping. *Geoderma* **2021**, *397*, 114998. [CrossRef]
99. Baltensweiler, A.; Walthert, L.; Hanewinkel, M.; Zimmermann, S.; Nussbaum, M. Machine Learning Based Soil Maps for a Wide Range of Soil Properties for the Forested Area of Switzerland. *Geoderma Reg.* **2021**, *27*, e00437. [CrossRef]
100. Utazi, C.E.; Yankey, O.; Chaudhuri, S.; Olowe, I.D.; Danovaro-Holliday, M.C.; Lazar, A.N.; Tatem, A.J. Geostatistical and Machine Learning Approaches for High-Resolution Mapping of Vaccination Coverage. *Spat. Spatio-Temporal Epidemiol.* **2025**, *54*, 100744. [CrossRef] [PubMed]
101. Granata, F.; Di Nunno, F.; Modoni, G. Hybrid machine learning models for soil saturated conductivity prediction. *Water* **2022**, *14*, 1729. [CrossRef]
102. Jia, Y.; Li, Y.; Biswas, A.; Pang, J.; Song, X.; Yang, G.; Hou, Z.; Luo, H.; Xie, X.; Ishchanov, J.; et al. Evaluation of Cotton Planting Suitability in Xinjiang Based on Climate Change and Soil Fertility Factors Simulated by Coupled Machine Learning Model. *Resour. Environ. Sustain.* **2025**, *20*, 100200. [CrossRef]
103. Schweng, S.; Bernardini, L.; Keiblinger, K.; Kaul, H.-P.; Fister, I., Jr.; Lukač, N.; Del Ser, J.; Holzinger, A. What Can Artificial Intelligence Do for Soil Health in Agriculture? *Comput. Sci. Rev.* **2026**, *59*, 100832. [CrossRef]
104. Priyatikanto, R.; Lu, Y.; Dash, J.; Sheffield, J. Improving Generalisability and Transferability of Machine-Learning-Based Maize Yield Prediction Model through Domain Adaptation. *Agric. For. Meteorol.* **2023**, *341*, 109652. [CrossRef]
105. Zhang, Z.; Ai, S.; Teng, W.; Meng, X.; Li, R.; Yang, F.; Cheng, K. Artificial Carbon Materials' Impact on Soil Fertility and Greenhouse Gas Emission. *J. Soils Sediments* **2024**, *24*, 1623–1638. [CrossRef]
106. Thamil Selvi, C.P.; Manimaraboopathy, M.; Jeyalakshmi, M.; Narmadha, G.; K, S.S. Intelligent Soil Fertility Forecasting Using Enhanced STGNN and Hybrid Swarm-Based Optimization. *Results Eng.* **2025**, *27*, 106866. [CrossRef]
107. De Vries, A.; Bliznyuk, N.; Pinedo, P. Invited Review: Examples and Opportunities for Artificial Intelligence (AI) in Dairy Farms. *Appl. Anim. Sci.* **2023**, *39*, 14–22. [CrossRef]
108. Chandra, H.; Pawar, P.M.; Elakkiya, R.; Tamizharasan, P.S.; Muthalagu, R.; Panthakkan, A. Explainable AI for soil fertility prediction. *IEEE Access.* **2023**, *11*, 97866–97878. [CrossRef]
109. Wang, X.; Lü, H.; Zeng, S.; Corzo Perez, G.A.; Gou, Q.; Yang, L.; Ji, Y.; Yao, Y.; Su, J. A seamless global daily soil moisture dataset (2010–2015) harmonized from SMOS observations and SMAP-era assimilation modeling. *Int. J. Digit. Earth* **2026**, *19*, 2609469. [CrossRef]

110. Colliander, A.; Reichle, R.H.; Crow, W.T.; Cosh, M.H.; Chen, F.; Chan, S.; Das, N.N.; Bindlish, R.; Chaubell, J.; Kim, S.; et al. Validation of Soil Moisture Data Products From the NASA SMAP Mission. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **2022**, *15*, 364–392. [CrossRef]
111. Günal, E.; Budak, M.; Kılıç, M.; Cemek, B.; Sırrı, M. Combining spatial autocorrelation with artificial intelligence models to estimate spatial distribution and risks of heavy metal pollution in agricultural soils. *Environ. Monitor. Assess.* **2023**, *195*, 317. [CrossRef]
112. Chen, H.; Gao, B.; Li, Y. Soil pollution and remediation: Emerging challenges and innovations. *Front. Environ. Sci.* **2025**, *13*, 1606054. [CrossRef]
113. Cao, J.; Zhang, Z.; Ding, J.; Li, L.; Ai, J.; Yang, Y.; Zhu, C.; Ge, X.; Wang, J. Soil organic carbon sequestration potential, storage, and influencing mechanisms in China. *Land Degrad. Dev.* **2025**, *36*, 4304–4319. [CrossRef]
114. Wade, A.M.; Richter, D.D.; Medjibe, V.P.; Bacon, A.R.; Heine, P.R.; White, L.J.T.; Poulsen, J.R. Estimates and Determinants of Stocks of Deep Soil Carbon in Gabon, Central Africa. *Geoderma* **2019**, *341*, 236–248. [CrossRef]
115. Habibi, L.N.; Matsui, T.; Tanaka, T.S.T. Critical Evaluation of the Effects of a Cross-Validation Strategy and Machine Learning Optimization on the Prediction Accuracy and Transferability of a Soybean Yield Prediction Model Using UAV-Based Remote Sensing. *J. Agric. Food Res.* **2024**, *16*, 101096. [CrossRef]
116. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial Validation Reveals Poor Predictive Performance of Large-Scale Ecological Mapping Models. *Nat Commun* **2020**, *11*, 4540. [CrossRef]
117. Tziachris, P.; Nikou, M.; Aschonitis, V.; Kallioras, A.; Sachsamanoglou, K.; Fidelibus, M.D.; Tziritis, E. Spatial or Random Cross-Validation? The Effect of Resampling Methods in Predicting Groundwater Salinity with Machine Learning in Mediterranean Region. *Water* **2023**, *15*, 2278. [CrossRef]
118. Liu, J.; Xiang, J.; Jin, Y.; Liu, R.; Yan, J.; Wang, L. Boost precision agriculture with unmanned aerial vehicle remote sensing and edge intelligence: A survey. *Remote Sens.* **2021**, *13*, 4387. [CrossRef]
119. Akhtar, M.N.; Shaikh, A.J.; Khan, A.; Awais, H.; Bakar, E.A.; Othman, A.R. Smart Sensing with Edge Computing in Precision Agriculture for Soil Assessment and Heavy Metal Monitoring: A Review. *Agriculture* **2021**, *11*, 475. [CrossRef]
120. Gallios, G.; Tsakiridis, N.; Tziolas, N. Federated learning applications in soil spectroscopy. *Geoderma* **2025**, *456*, 117259. [CrossRef]
121. Zakzouk, S.; Said, L.A. Federated Learning for Soil Moisture Prediction: Benchmarking Lightweight CNNs and Robustness in Distributed Agricultural IoT Networks. *MAKE* **2025**, *7*, 132. [CrossRef]
122. Parewai, I.; Köppen, M. A digital twin approach for soil moisture measurement with physically based rendering simulations and machine learning. *Electronics* **2025**, *14*, 395. [CrossRef]
123. Tsakiridis, N.L.; Samarinas, N.; Kalopesa, E.; Zalidis, G.C. Cognitive Soil Digital Twin for Monitoring the Soil Ecosystem: A Conceptual Framework. *Soil Systems* **2023**, *7*, 88. [CrossRef]
124. Wang, Z.; Menenti, M. Challenges and opportunities in Lidar remote sensing. *Front. Remote Sens.* **2025**, *2*, 641723. [CrossRef]
125. Mokere, R.; Ghassan, M.; Barra, I. Soil spectroscopy evolution: A review of homemade sensors, benchtop systems, and mobile instruments coupled with machine learning algorithms in soil diagnosis for precision agriculture. *Crit. Rev. Anal. Chem.* **2025**, *55*, 1304–1323. [CrossRef]
126. Pook, T.; Vandenplas, J.; Boschero, J.C.; Aguilera, E.; Leijnse, K.; Chauhan, A.; Bouzembrak, Y.; Knapen, R.; Aldridge, M. Assessing the Potential of Quantum Computing in Agriculture. *Comput. Electron. Agric.* **2025**, *235*, 110332. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.