big data and cognitive computing

MDPI

Review

Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review

Olusegun Folorunso [1,*], Oluwafolake Ojo [1], Mutiu Busari [2], Muftau Adebayo [3], Adejumobi Joshua [1], Daniel Folorunso [1], Charles Okechukwu Ugwunna [1], Olufemi Olabanjo [1] and Olusola Olabanjo [1]

[1] Department of Computer Science, Federal University of Agriculture, Abeokuta 2240, Nigeria
[2] Department of Soil Science and Land Management, Federal University of Agriculture, Abeokuta 2240, Nigeria
[3] Safefood Africa Agroenterprise, Abeokuta 110123, Nigeria
* Correspondence: folorunsoo@funaab.edu.ng

Abstract: Agriculture is essential to a flourishing economy. Although soil is essential for sustainable food production, its quality can decline as cultivation becomes more intensive and demand increases. The importance of healthy soil cannot be overstated, as a lack of nutrients can significantly lower crop yield. Smart soil prediction and digital soil mapping offer accurate data on soil nutrient distribution needed for precision agriculture. Machine learning techniques are now driving intelligent soil prediction systems. This article provides a comprehensive analysis of the use of machine learning in predicting soil qualities. The components and qualities of soil, the prediction of soil parameters, the existing soil dataset, the soil map, the effect of soil nutrients on crop growth, as well as the soil information system, are the key subjects under inquiry. Smart agriculture, as exemplified by this study, can improve food quality and productivity.

Keywords: machine learning; digital soil mapping; soil properties, smart soil

check for updates

Citation: Folorunso, O.; Ojo, O.; Busari, M.; Adebayo, M.; Joshua, A.; Folorunso, D.; Ugwunna, C.O.; Olabanjo, O.; Olabanjo, O. Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review. *Big Data Cogn. Comput.* **2023**, *7*, 113. https://doi.org/10.3390/bdcc7020113

Academic Editor: Salvador García López

Received: 10 May 2023
Revised: 23 May 2023
Accepted: 26 May 2023
Published: 8 June 2023

1. Introduction

Without a doubt, technological advancements have dramatically improved the efficiency and productivity of numerous industries, including agriculture. Examples of this revolution in technology include the introduction of such terms as "big data", "data analytics", "artificial intelligence", "Internet of Things", "erosion modeling", "smart farming", and "machine learning" [1–4]. To develop and populate spatial soil information systems, digital soil mapping (DSM) applies numerical models to infer the geographical and temporal variations of soil types and attributes based on soil observations, prior knowledge, and pertinent environmental variables [5].

Even though the above ideas have been utilized in many ways, agriculture technology is continually evolving. Fertilizer and weed application, irrigation management, and soil mapping all involve information technology. AI models are becoming increasingly crucial to smart agriculture's long-term success. In agriculture, AI is used in soil and irrigation management, weather forecasting, plant growth, disease prediction, and animal management [6]. Smart farming, in contrast to traditional farming, makes use of state-of-the-art innovations to boost productivity and reduce labour stress in response to the exponential growth and development of data processing, information technology, and artificial intelligence, automating soil and crop management with AI that mimics the way humans learn and solve problems [7].

Artificial intelligence (AI) applied to soil prediction is vital in agriculture since soil composition impacts crop yields in many ways. Soil prediction involves using several methods to evaluate if the soil is suitable for a crop before planting it. Smart soil prediction is a result of new technology. Smart soil prediction is a low-cost way to anticipate a soil's performance across many crops. Digital soil mapping (DSM) creates digital soil type and quality maps using numerical and statistical models that combine soil sensing data with environmental parameters [8]. Recent years have seen a dramatic increase in DSM activities within the field of soil science, which can be attributed to the comingling of several ideal elements, including, but not limited to, massive interest in quantitative and spatial soil information, the accumulation of databases of estimated or construed soil properties combined with thoroughly known environmental variables, and the development of numerical models combined with computer resources to mine these stores of soil data [9]. For supplying the crop model soil input data, DSM can be used instead of choropleth soil maps. For mapping soil parameters at controlled prices, the DSM provides an alternative to traditional soil surveys [10]. Acquiring precise soil nutrient distribution data is a crucial step in the implementation of precision agriculture, and digital soil mapping is a promising innovation [11]. Several artificial intelligence tools, such as fuzzy systems, decision trees, expert knowledge, machine learning algorithms, deep learning methods, and others, can offer more precise forecasts and solutions in DSM. As shown in Figure 1, there are four major processes for evaluating model and map performance in DSM. The first step is to train the model with the dataset (to ensure goodness of fit), the second step is to test the model performance with cross-validation (to ensure robustness), the third step is to test the map validation within a similar geographic degree with an independent dataset, and the fourth step is to test the model's adaptability in an alternate geographic region with a second independent dataset [12].

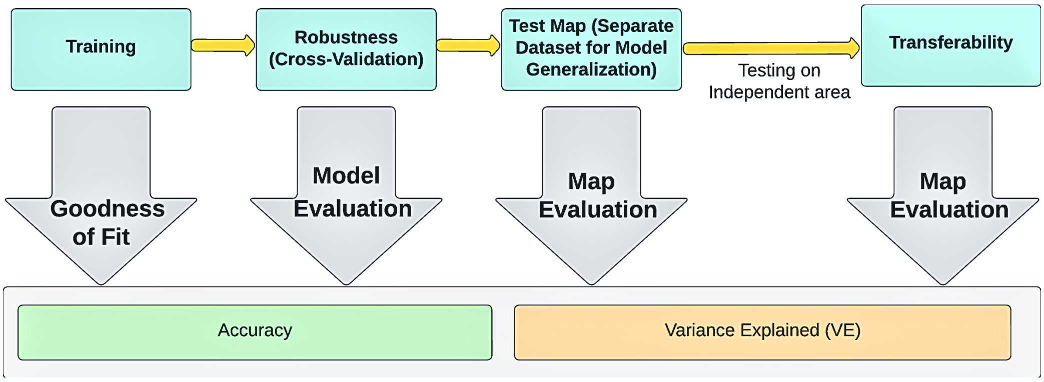

Figure 1. Conceptual View of Assessing Model and Map Performance in DSM [12].

Artificial intelligence models and digital soil mapping have been used in the past to predict soil fertility, providing a decision-making tool capable of predicting the most suited crops to plant based on soil pH, soil nutrients, soil moisture, environmental variables, and other factors [13]. For precision farming, machine learning and deep learning algorithms are the most frequently used types of artificial intelligence [14]. The lack of widespread adoption of digital soil mapping and other digital innovative solutions is a barrier to high productivity in agricultural systems in developing countries, despite the fact that its use has been on the rise internationally. As a result, the primary objective of this research is to investigate the issues that are impeding the deployment of smart soil information systems in developing nations. Furthermore, this study elaborates on numerous examples of digital soil mapping and artificial intelligence-based smart soil systems with emphases on the following contributions:

1. Examining the smart agriculture and digital soil management landscape in developing countries.
2. Existing research literature on soil attributes, classifications, and key components in soil databases for soil fertility prediction.
3. Identify and review the state-of-the-art smart soil system based on artificial intelligence models (machine learning and deep learning models).
4. Overview of the current issues in development and deployment of soil information systems.

5. Establishing a roadmap for future research to improve agricultural productivity with DSM and other digital innovation technologies through the development of a smart soil information system.

The remaining sections of this systematic review are organized as follows. Section 2 examines soil components and qualities, while Section 3 focuses on the use of digital soil mapping and intelligent soil management systems. Section 4 describes the materials and methods employed in this study. Section 5 discusses existing soil information system frameworks, current trends in soil information systems, and problems. Section 6 examines the current state of AI models for soil property management and soil fertility prediction; machine learning and deep learning algorithm applications and accuracy; and existing smart soil mobile applications. Section 7 presents the research findings and discussion. Section 8 provides the conclusion and future research directions.

2. Soil Components and Properties

Sustainable agricultural growth and enhanced crop yields are both feasible consequences of land reclamation and productive resource management. Increased yields can be obtained in intensive cropping by using adequate nutrition sources and application rates [15]. Soil quality fundamentally means “the ability of a soil to function”; this ability can be indicated by the estimated soil’s physical, chemical, and biological qualities, often known as soil quality indicators (SQI) [16]. Several soil investigations may be envisaged to adequately quantify the soil framework, and science-based indices on SQI provide valuable data to farm managers for decision making. These indices incorporate important soil attributes, including supplying suitable amounts of water and nutrients, resisting and recovering from physical degradation, and supporting plant growth with the right management [17]. Sustainable farmland management requires an in-depth familiarity with the relationships between soil physical qualities and many agronomic and environmental factors [18]. The availability of nutrients is influenced by the soil’s chemical and physical properties, such as its parent material and naturally occurring minerals, organic matter, depth to bedrock, sand, or gravel, permeability, water-holding capacity, and drainage. The distribution of nutrients is also determined by plant and atmospheric conditions [19]. The nutrient concentration in the soil solution is influenced by soil water content, depth, pH, cation-exchange capacity, redox potential, soil organic matter, microbial activity, season, and fertilizer application [20]. It is typically time-consuming and costly to estimate and evaluate soil components and qualities. Predictive soil mapping is a common modeling approach used to estimate the spatial distribution of soil components when actual data from samples are unavailable. Many of these approaches rely on predictive maps or the estimation of soil-related variables at unmeasured locations based on field data using mathematical or statistical models of relationships between soil and other environmental elements [21].

2.1. Soil Dataset

To determine the nutrient level, composition, and other properties of a soil sample, scientists conduct a soil test. Soil testing can involve a variety of techniques and fertilizer recommendations to determine the soil’s fertility and pinpoint any deficiencies that need to be addressed. Soil analysis provides information useful to farmers and consumers in deciding when and how much fertilizer and farmyard manure should be administered during a crop’s growth cycle [22]. Soil datasets entail information on land suitability for agricultural production, soil maturity, soil texture, meteorological data, moisture content, soil classes, soil colour, covariate data, soil nutrients, and trace elements. Table 1 lists the most prevalent soil nutrients, trace elements, and their descriptions.

The utilisation of covariate environmental data facilitates the establishment of associations between soil properties and various environmental factors. The process of soil formation and its characteristics are impacted by several factors, including but not limited to climatic conditions, topographical features, vegetation cover, land utilisation, and the nature of the parent material. The integration of covariate data can enhance the efficacy of soil prediction models by enabling a more comprehensive understanding of the intricate interplay between soil and its surrounding ecosystem. The inclusion of covariate environmental data is imperative in soil prediction due to its ability to augment our comprehension of soil-environment associations, capture spatial heterogeneity, offer insights into fundamental mechanisms, enable data amalgamation, and facilitate informed decisions regarding land management. The integration of covariate data into soil prediction models enhances their precision and usefulness in diverse domains, such as agriculture, environmental governance, and land use management [23,24].

Table 1. Description of Soil Nutrients and Trace Elements.

Symbol	Meaning	Units	SPT
N	Nitrogen	%	SN
P	Phosphorus	mg kg^{-1}	SN
K	Potassium	cmol kg^{-1}	SN
Ca	Calcium	cmol kg^{-1}	SN
Mg	Magnesium	cmol kg^{-1}	SN
S	Sulphur	ppm	SN
Fe	Iron	ppm	TE
Mn	Manganese	ppm	TE
Cu	Copper	ppm	TE
Zn	Zinc	ppm	TE
B	Boron	ppm	TE
Mo	Molybdenum	ppm	TE
ESP	Exchangeable sodium percentage	%	SN
CEC	Cation exchange capacity	cmol kg^{-1}	SN

Abbreviations: SN—Soil Nutrients, TE—Trace Elements, SPT—Soil Properties Type.

2.2. Soil Map

Environmental elements pertaining to geology, landforms, or vegetation are identified through the use of aerial photographs, Landsat images, and digital elevation models (DEMs) in traditional digital soil mapping. The method is then checked against real-world data [25]. The final outcome is a map labeled with soil classifications, which can be confusing to users. Furthermore, there are other issues caused by mapping's subjective character [26]. In traditional soil surveys, the soil is mapped according to the surveyor's preconceived notions [27]. Classical mapping's conceptual framework was established using quantitative and statistical methods. The method of developing and updating spatial soil information systems via analytical and experimental observational methods paired with spatial and non-spatial soil inference systems is generally known as digital soil mapping [28]. Digital soil mapping is also known as computer-assisted soil cartography, numerical soil cartography, pedometric mapping, environmental correlation, predictive soil mapping, or geographical extrapolation utilizing models [25,29–32] whose evolution is rapidly rising as depicted in Figure 2. The digital soil map depicted in Figure 3 presents an illustration of the soil nutrient distribution in a specific area located in Ogun State, situated in the south-west region of Nigeria.

In prior studies, a digital soil map was considered a digitized conventional soil map in the form of polygons [33]. However, because the map was not created using statistical inference, it cannot be construed as a digital soil map, but rather a digitized soil map. The initial development of the SCORPAN framework for use in digital soil mapping was accomplished by [34]. SCORPAN is a mnemonic for an empirical quantitative description of relationships between soil and environmental factors with a view to using these as soil spatial prediction functions for the purpose of digital soil mapping where each letter stands for the following:

S = soil classes or attributes
f = function
s = soil, other or previously measured properties of the soil at a point
c = climate, climatic properties of the environment at a point
o = organisms, including land cover and natural vegetation or fauna or human activity
r = relief, topography, landscape attributes
p = parent material, lithology
a = age, the time factor
n = spatial or geographic position.

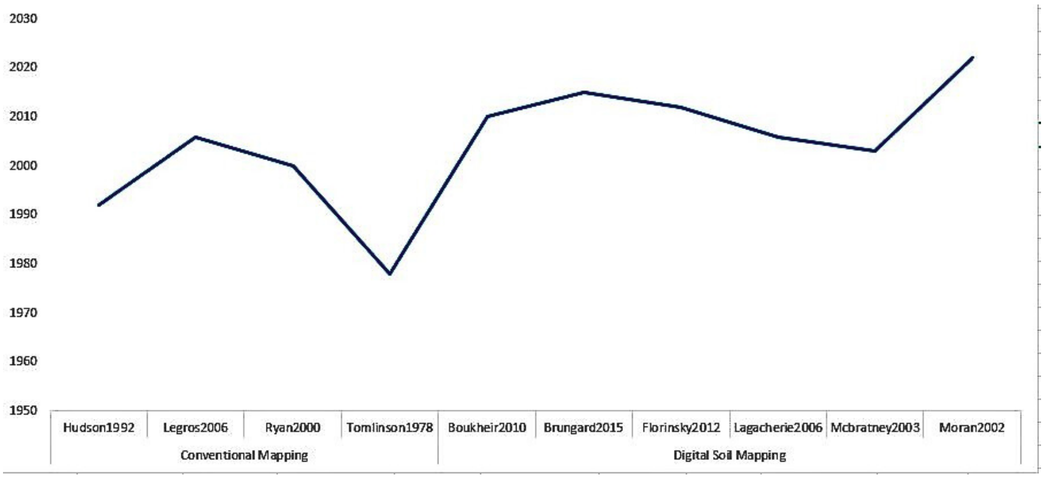

Figure 2. The Evolution of Soil Mapping

Figure 3. Digital soil map depicting the soil's nutrients for a location in South-West Nigeria.

Spatial soil prediction functions with an auto-correlated error are often used to forecast soil class or soil attributes from so-called SCORPAN factors [34].

$$Sc = f(s, c, o, r, p, a, n) + e, \text{or } Sa = f(s, c, o, r, p, a, n) + e$$

'e' stands for spatially correlated residuals, where Sc and Sa are soil classes and soil properties as a function of soil, climate, organisms, relief, parent material, age, and geographical position [35]. For the quantitative prediction of soil groups or dynamic soil properties based on empirical observations, the SCORPAN model is employed.The majority of effort in digital soil mapping is based on developing a mathematical model that connects field soil data and SCORPAN variables [36,37]. Afterwards, the model is used with extensive spatial environmental data. To extrapolate, update, or disaggregate soil maps, digital soil mapping can also employ conventional soil maps as input [38,39]. The underlying principle is to employ machine learning (ML) techniques to find the knowledge inherent in completed surveys or to reverse engineer the surveyor's soil-landscape mental model [40].

2.3. Research Justification

The ability of ML-based methods to accurately forecast soil characteristics, crop growth, and soil fertility has attracted a lot of attention in recent years. Texture, organic matter, pH, nutrient content, soil moisture, and soil structure are just a few of the many soil variables that may be analysed with the ML approach. ML techniques are superior to traditional statistical methods because of their capacity to process massive amounts of complex data and reveal hidden patterns. Several studies have focused on developing ways for applying machine learning to predict soil parameters [41–43], crop growth [44–46], and soil fertility [47,48].

Recently, a systematic literature review that highlights the research gaps in certain applications of deep learning techniques and evaluates the influence of vegetation indicators and environmental factors on agricultural productivity was published in [49]. The authors examined prior studies from 2012 to 2022 from various databases. The article focuses on the benefits of employing deep learning in agricultural yield prediction, the best remote sensing technology depending on data collection requirements, and the numerous factors that influence crop yield prediction. In general, several studies have demonstrated the efficacy of machine learning algorithms in predicting soil properties, soil fertility, and crop yields. It is vital to keep in mind, though, that ML models' accuracy is extremely sensitive to the quantity and quality of data used in training, in addition to the algorithms and parameters with which they are implemented. Further research is needed to investigate how to construct and refine ML models for predicting soil parameters and evaluate how well they function in different environmental and soil circumstances. Farmers, policymakers, plant breeders, and other professionals in the agricultural sector can all benefit from ML recommendations.

3. Materials and Methods

3.1. Database Search Strategy and Eligibility Criteria

In this research, we developed a search strategy and utilized it to scour a variety of databases in search of up-to-date, relevant research publications on the research study of using machine learning models to create digital maps of soil and predict its physical qualities. Google Scholar (https://scholar.google.com) and the ACM Digital Library (https://dl.acm.org/search/) were the primary resources used in the search. Timeframe for the investigation: 2002–2022. These sources were selected because of their extensive indexing of research into the use of machine learning models in DSM and SPP. These can be found with little effort and are easily accessible.

3.2. Review Strategy

The review technique covers research design, search strategy, information sources, study selection, and the method of data collection. Publications that met the predefined inclusion and exclusion criteria were evaluated. Manuscripts that were comments, letters, or editorials were excluded. The search strategy is composed as follows: (a) construct search terms by identifying major keywords, required action, and expected results; (b) determine the synonyms or alternative words for the major keywords; (c) establish exclusion criteria to make exclusions in the course of search; and (d) apply Boolean operators to construct the required search term.

Results for (a): DSM, SPP, ML, deep learning, soil properties, soil nutrients, soil map, soil datasets and crop growth.

Results of (b): smart soil, soil information system and soil fertility.
Results for (c): smart farming, plant disease, crop disease, articles in different languages other than English.
Result (d): a, b, c combined using AND OR.

In this study, publications were chosen from the peer-reviewed literature by conducting a search using the generated search phrase on Science Direct, Scopus, Google Scholar, and MDPI. Conference proceedings, journals, book chapters, and whole books are all examples of vetted resources. The initial number of results returned by Google Scholar was 1328; of those, 480 fulfilled the initial selection criterion and 68 fulfilled the final requirements.The studies were appropriately grouped. Figure 4 shows the preferred reporting items for systematic reviews and meta-analyses (PRISMA) flowchart for study selection.

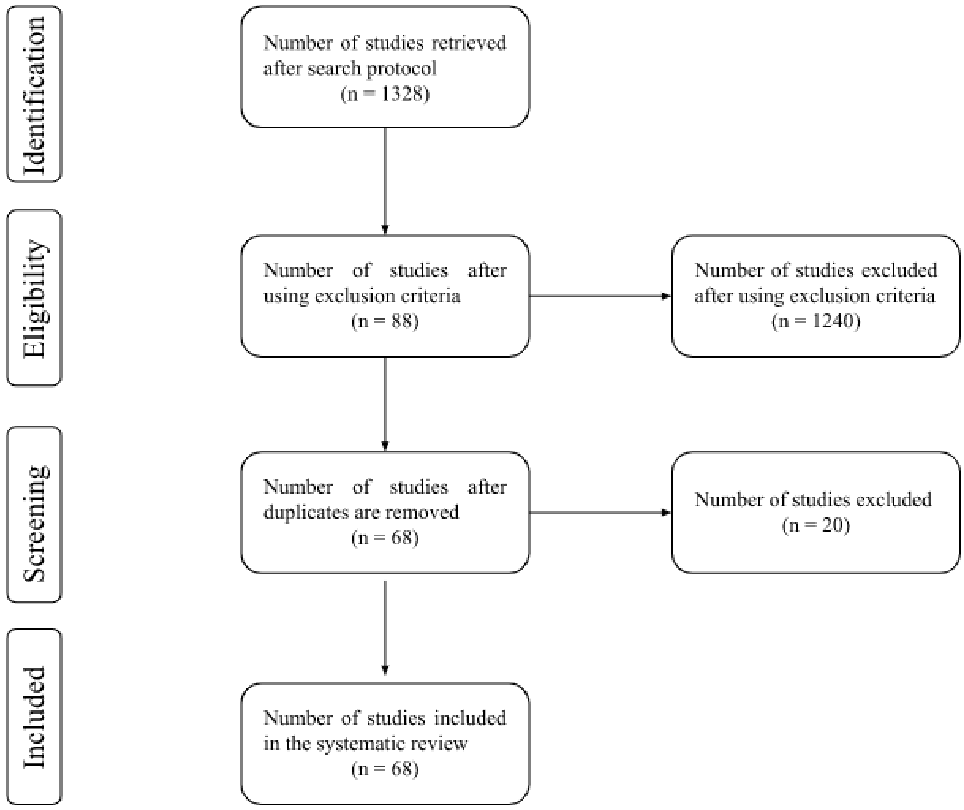

Figure 4. PRISMA Model.

3.3. Characteristics of Studies

The literature search yielded a total of 1328 articles, of which 1308 were retained after duplicates were deleted, 1240 were disregarded as irrelevant based on their article titles and abstracts, and 88 were selected for a detailed review. After a thorough full-text review, we settled on including 68 articles from 1999 to 2022. Only 20 of the 68 articles (as indicated in Table 2) included information on the data type and accuracy achieved.

3.4. Quality Assessment

The vast majority of studies failed to satisfy standards in at least one of the six quality criteria examined. Limited sample size, an inadequate statistical analytical strategy, failure to evaluate for confounders, and failure to disclose results for computational techniques were the most frequently observed issues regarding lack of quality throughout the investigations.

Table 2. Existing Work on AI Models and DSM.

Reference	Problem Addressed	AI Methods	Metric	Accuracy	Dataset Types	Limitations
[50]	DSM to inform gully erosion mitigation measures	MNLR, CM	KC,R^2,RMSE	68%	Covariate and climate data, land type maps	Soil depth map not a good representation of reality (covariate layer map required)
[51]	Assessment of the soil fertility status using DSM and ML	QRF, CM	R^2, CCC, RMSE, MAE	High and average accuracy	Soil dataset (SOC, OM, Kech,, Pass, CEC, SumBas, BS)	Model accuracy was limited for some of the soil properties, such as N and Kech
[52,53]	Improved machine learning models accuracy in DSM	CM, RM, ANFIS, EGB, ERT, ANN, SVR, MARS, KNN, GP	RMSE, MAE, R^2, CCC, F-score	High accuracy	Clay, sand, CaCO3, SOC, SEC, pH, K, Ca + Mg, Na, SAR, EF, MWD	Uncertainty was observed in the predicted values. Small dataset used
[54]	Prediction of soil depth using DSM	QRF, RK	RMSE, R^2, CCC	30%	Covariates dataset	Lower accuracy rate achieved due to the error in locating old coordinates
[55]	Soil maps for a wide range of soil properties using ML	RF, QRF, CM, SVM	Bias, R^2, RMSE	Best accuracy achieved with QRF	Gravel, clay, sand, density, pH,SOC and soil depths (0–200 cm) 0–5, 5–15, 15–30, 30–60, 60–100 and 100–200 cm	Overestimation was observed for some probability values
[56]	Review on DSM algorithms and covariates for SOC mapping	RK, MLR, RF, CM, NN, BRT, SVM, GWR	-	RF performed better than others	Environmental covariates, parent material, climate factor, organic activity, topography	Performance metrics or evaluation methods not reported
[57]	Spatial prediction of soil aggregate using ML algorithms and environmental variables	RF, SVM, kNN, and ANN and ensemble modelling	RMSE, MAE, R^2, and normalized RMSE	Ensemble achieved high accuracy for all soil targets	Soil properties, remote sensing data, legacy soil maps, and DEM derivatives	Lower accuracy achieved for SOC categories
[58–60]	Prediction of SOC and soil total nitrogen using DSM and ML algorithms	RF, BRT, SVM and Bagged -CART	RMSE, MAE and R^2	BRT model performed best in predicting SOC and STN	DEM derivatives, multi-temporal Sentinel data, environmental data	Investigation using other soil properties is required
[61]	Predicting and mapping of SOC using ML algorithms	SVM, ANN, RF, XGBoost, CM, RT, DNN	RMSE, MAE, R^2 and CCC	DNN mapped SOC contents more accurately	Terrain attributes, remote sensing data, climatic data, categorical data	Further investigation on the dataset using hybrid algorithms is required

Table 2. *Cont.*

Reference	Problem Addressed	AI Methods	Metric	Accuracy	Dataset Types	Limitations
[62]	Soil moisture prediction using multi-sensor data and ML algorithm	RFR, XGBoost, SVM, CBR and GA for feature selection	RMSE and R^2	XGBR-GA hybrid model yielded the highest performance (R^2 = 0. 891; RMSE = 0.875%)	DEM derivatives, Sentinel-1 and Sentinel-2 data.	Testing the framework in large-scale areas with various land-use characteristics is required
[63]	Supervised maps for predicting soil moisture	Unsupervised SOM, supervised SOM, semi-Supervised SOM, and RF	R^2, accuracy, and Cohen's KC	Higher accuracy achieved with the SOM methods	Soil moisture and land cover dataset	RMSE and MAE factors are not considered in the performance evaluation
[64]	Predictive mapping using semi-supervised ML	Decision trees, logistic regression (LR), SVMs and graph-based semi-supervised ML (GS-ML)	Mean accuracy (%), accuracy range (%), accuracy standard deviation (%)	GS-ML achieved higher accuracy	Environmental covariate data	Improvement is required for parameter setting, RMSE, R^2 and MAE evaluations are not considered
[37]	ML for predicting soil classes in semi-arid landscapes	Multiple classifications and regression ML	Kappa analysis, Brier scores and confusion index	-	Environmental covariates	Model accuracy was obtained when there are few soil classes, limited dataset to investigate "rare" soil classes
[65]	Mapping of soil water erosion using ML models	Weighted subspace random forest , Gaussian process and naive Bayes (NB) ML methods	Accuracy, Kappa index and probability of detection	-	Soil texture, land and climate dataset	The data collection and sampling of them were not on the same scale. Also, RMSE, R^2 and MAE factors are not considered in the performance evaluation
[66]	Digital mapping of soil carbon fractions using ML	RF, SVM, CaRT, BaRT, BoRT, RK, OK	Mean, standard deviations, prediction error, and R^2	RF achieved the best accuracy	Soil data (0–20 cm), carbon	Further investigation required on the use of more sophisticated predictors

Table 2. *Cont.*

Reference	Problem Addressed	AI Methods	Metric	Accuracy	Dataset Types	Limitations
[67]	Multi-scale DSM with DL	DL-ANN, RF	R^2	DL achieved 4–7 % than RF	Silt, clay, ZC, SFP, DEM resolution	The model is not tested with some environmental data such as climate, lithology, or land cover
[68]	Semi-supervised DNN regression for spatial soil properties prediction	DNN, GA, SVR and regression methods	RMSE, MAE, R^2, Bias, ratio of performance to inter-quartile distance	DNN achieved the highest accuracy	Hyperspectral remote sensing image data	Sensitive to the quality of the initial training dataset and model not tested with a large number of samples
[69]	Assessment of landslide susceptibility using DL with semi-supervised learning	DNN, SVM and LR.	Accuracy, Kappa index, predictive rate curves (AUC), and information gain ratio (IGR)	DNN achieved higher accuracy with AUC of 0.898	Land cover and soil data	The K-means algorithm was tested using fixed value and limitation by the accuracy of layers and sampling process observed

Abbreviations: DSM—Digital soil mapping, ML—Machine learning, DL—Deep learning, MNLR—Multi-nominal logistic regression, CM—Cubist model, QRF—Quantile regression forest, KC—Kappa coefficient, RMSE—Root mean square error, MAE—Mean absolute error (MAE), R^2—Coefficient of determination, CCC—Lin's concordance correlation coefficient, BS—Base saturation, RF—Random forest. SOC—Soil organic carbon, OM—Organic materials, Kech—Exchangeable K, ANFIS—Adaptive-network-based fuzzy inference system, EGB—Extreme gradient boosting, ERT—Extremely randomized trees, ANN—Artificial neural network, SVR—Support vector regression, SFP—Soil formation patterns, DEM—digital elevation models, BRT—Boosted regression tree, GWR—Geographically weighted regression. MARS—Multivariate adaptive regression splines, KNN—k-nearest neighbour, GP—Genetic programming, SAR—Sodium adsorption ratio, SFP-EF—Erodible fraction of the soil, MWD—Mean weight diameter, SEC—Soil electrical conductivity, RK—Regression kriging model, ZC—Zinc concentration, Pass—Assimilable P, CEC—Cation exchange capacity, SumBas–Sum of bases, PLSR—Partial least square regression, OK—Ordinary kriging, CART—Classification and regression trees, CBR—CatBoost gradient boosting regression, GA—Genetic algorithm, SOM—Self-organizing maps.

3.5. Data Sources and Search Strategy

We searched Google Scholar for studies published before October 2022. We considered the top 1328 papers which reported on the application of machine learning for soil properties or soil fertility prediction. Keywords from subject headings or titles or abstracts of the studies were searched for with the help of Boolean operators (and, or) with language restricted to English. In addition, we reviewed the reference lists of primary studies and review articles.

3.6. Inclusion and Exclusion Criteria

All research in which machine learning approaches were applied to predict soil qualities was reported. The included publications had to include the AI technique used or the soil characteristics problem addressed in the article. Articles dealing with DSM's three key datasets and techniques were also included in the study selection. Articles on crop diseases or plant disease prediction, statistical analyses, studies on palm kernel agriculture, and irrigation systems for crop growth monitoring were all excluded. Editorials, narrative review articles, case studies, conference abstracts, and duplicate publications were all discarded from the analysis.

3.7. Data Extraction and Quality Assessments

The full texts of the citations chosen for review were acquired, and the reviewers independently collected all study data, resolving disagreements by consensus. The initial author, year of publication, study setting, ML approach, the data type used or recommended, performance measures used, and accuracy attained were all extracted for every study.

4. The Impact of Soil Nutrients and Fertility on Crop Growth

Nutrients from photosynthesis and soil are two of the most important for any plant's development. This suggests that it may be impossible for any crop to achieve sufficient yields without adequate fertilizer input. Soil nutrients are one of the most crucial types of food for plants. Crops such as corn, cassava, and yam rely heavily on the nutrients in the soil in order to thrive. Three of the most common nutrients in the soil are nitrogen (N), phosphorus (P), and potassium (K). The soil also contains a wide variety of other nutrients, such as calcium, magnesium, sulfur, zinc, boron, copper, iron, manganese, and molybdenum. An available nutrient index is a useful tool for describing soil fertility. Soil fertility is not guaranteed simply by the presence of all these nutrients. Fertile soil is one that contains an abundance of the specific nutrients required by a given crop. The term "soil fertility" refers to the soil's inherent capacity to support plant development. For soil to be considered sustainable, it must meet certain conditions, including but not limited to the following: a suitable soil pH; the presence of suitable microorganisms; adequate internal drainage; and the capacity of the topsoil to contain soil organic materials such as algae, sewage sludge, manure, and many more [70,71]. For this reason, healthy, fertile soil is essential for maximizing harvest production. Soil nutrients and quality have been proven to have a significant impact on the yields of corn, cassava, and yam [72–77].

4.1. Research on Soil Nutrients and Crop Yield in Developing Countries

The authors of [78] analysed the nutrient composition and corresponding crop yield in soil that had been treated with organic manures. The study followed an experimental design, as chosen by the authors. An experiment was conducted by sowing four (4) maize seeds into various earthen containers. To improve the soil's quality, organic manure was spread over it. Poultry manure, composted animal manure, and press mud are the manures used. After six days, the plants were thinned so that each pot would hold two plants. The study discovered that after applying organic manure, soil organic matter, phosphorus, and potassium bioavailability all increased. Both the stature of the maize plants and the total leaf area were boosted by the application of organic manures. These findings demonstrated that soil nutrients can stimulate more robust growth in maize. In Kenya, Ref. [76] examined how maize fared in terms of growth and yield on a specific category of soil. A randomized, complete block nutrition omission trial was used to determine how maize responded to nutrient administration. Ferralsols was the soil type employed. The treatments consisted of applying one of six different inorganic fertilizers: NK, NP, PK, NPK, or NPK + CaMgZnBS. The corn harvest was severely diminished by the use of PK fertilizer. The application of urea resulted in the maximum yield (1800 kg/ha). The author concludes that using fertilizers rich in nitrogen, phosphorus, and potassium will increase crop yields in maize.

In a Northern Zambia study [70], the authors studied the connection between farming methods and soil nutrient levels. Soils in the area are often either orthic arcrisols or feric dirt. The majority of the population in this area is engaged in agriculture, and cassava is their primary crop. Around 40 farmers and 120 fields were chosen from across 10 villages. Fieldwork on the cassava was carried out in the fourth quarter of 2018, thus the plants were between one and three years old. The study found that the potassium content of cassava decreased from the first to the second growing season. Cassava was shown to have nutritional imbalances, which were blamed on its moderate quantities of exchangeable magnesium. The regression analysis also revealed that soil organic carbon and leaf area index were significant predictors of cassava yield.

Research along these paths was also carried out in Southwestern Nigeria. Ref. [79] employed a survey research design to investigate the topic of soil fertility in cassava farms. Soil samples were also collected from each of the 33 farmers' fields in Iwo and Osogbo. The chemical and physical properties of the soil sample were analysed in a laboratory. The research concluded that the soil in roughly 80% of the fields is deficient in organic matter. It was also discovered that the pH of the soil is generally acidic, with readings ranging from 5.4% to 6.4%. Phosphorous and nitrogen levels in the soil were also found to be below the minimum required for cassava cultivation. Soil contains sufficient amounts of essential nutrients such as calcium, potassium, and magnesium. These results suggest that the potential cassava harvest in Osun State is comparable to the national average.

In Ethiopia, Ref. [75] analysed the nutritional levels in the soil of southern smallholder cassava crops. The study's focus was on the town of Wolaita in southern Ethiopia. There were 12 cassava farms in Wolaita, from which data was compiled. Soil samples and information about how local farmers handle their soil were the types of information being collected. The results were interpreted by looking at the physical and chemical characteristics of the soil. Results from the study were mixed in quality. In the soil that was tested, there was an adequate supply of manganese. Soil acidity might be high to mild, and in 83% of farms, the amount of exchangeable calcium (Ca) was below the minimum acceptable level of 5 Cmol (+) kg^{-1}. Boron and copper were both absent from the cassava fields, and iron and zinc levels were low.

Ref. [80] examined the impact of applying inorganic fertilizer and biochar on yam yields in a Ghanaian agroecological zone. The research was a randomized block-design factorial experiment. Three inorganic fertilizers and four biochars made from wood shavings were applied. The research showed that there was no discernible change in soil characteristics in response to the experimental treatment. The amount of nitrogen in the atmosphere decreased. Six months after planting, applying biochar considerably enhanced the number of seed yams per acre, whereas applying fertilizer increased productivity. This means that yam cultivation can benefit from biochar even at high concentrations. Ref. [81] was primarily interested in how soil fertility affected the variations in yam species' growth. The two most common species are D. alata and D. rotundata. The D. alata species was reported to have better growth statistics than the D. rorundata species. The two yam species were found to produce more at the forest location than in the savanna area, which was due to the higher soil fertility there. The deficient nitrogen and potassium nutrients at the savanna location were also responsible for a significant fall in the leaf area index.

4.2. DSM/ML Soil Prediction in Developing Countries: Challenges

In underdeveloped nations, the application of digital soil maps and machine learning for soil prediction is frequently hampered by a number of reasons, including the following:

(a) Data scarcity: In many underdeveloped nations, soil data is scarce or nonexistent, making accurate digital soil maps and training machine learning models problematic. This occurs frequently owing to a scarcity of resources and funds for soil surveys and studies.
(b) Low technical expertise: Poor countries may lack professionals with the technical abilities needed to produce and evaluate digital soil maps as well as developing machine learning models. This can make it challenging to effectively implement these technologies.
(c) Restricted access to technology: Many underdeveloped countries may lack the requisite infrastructure or resources to facilitate the usage of digital soil maps and machine learning. This can involve a lack of internet connectivity, computer equipment, and access to software and data.

(d) Inadequate governmental capacity: Poor countries may lack the institutional ability necessary to properly employ digital soil mapping and machine learning technology. These can include ineffective governance systems, insufficient financing for research and development, and a lack of coordination among various government agencies and stakeholders.

5. Soil Information System

The four main components of soil are minerals, water, air, and soil organic matter (SOM). The ratio and content of these components have a significant impact on the physical properties of soil, including its texture, structure, and porosity (the percentage of pore space). The capacity of the soil to transmit air and water is thus influenced by these features. It is possible to assess the soil's quality using a small collection of data on its properties, such as texture, organic matter, pH, bulk density, and rooting depth. To comprehend soil quality, soil organic matter is very crucial because it can have an impact on a variety of soil properties, including other components of the limited dataset [82]. Soil information systems provide aggregate measurements of soil quality, such as the soil's functional capacity and its performance in relation to a certain application. To "learn" or understand from data how soil components are distributed throughout space and time, statistical models have been employed in soil science research, and more specifically, pedometrics [83]. In order to calibrate, validate, and compare models, Ref. [84] suggests using soil component datasets as standard evaluation datasets, starting values, and system parameters. It is a crucial piece of the puzzle when trying to model the Earth's system.

Given the huge need for quantitative geographic soil data and its current scarcity, it is crucial to create and implement ways of providing this information. Every soil information system needs to be flexible enough to accommodate user needs and requests while also managing datasets that change in space and time [85]. The tremendous growth of computing and digital technology has led to the emergence of enormous quantities of data and tools in every domain. As a result, numerous initiatives have been launched to create data infrastructures for spatial soil information systems [86]. For more efficient land deterioration prevention and control, regional development feasibility studies, disaster risk prediction (such as floods and landslides), environmental quality restoration, and formative strategic planning, accurate and up-to-date information on the environment, extent, spatial distribution, opportunities, and constraints of soil properties are required [87].

Over time, many methods have been developed for collecting soil data. The backbone of most soil information systems consists of databases containing pedotransfer functions, soil profiles and analytical data, and a collection of methodologies. Soil data providers, both public and private, can take advantage of the available technical solutions and apps for data management [88]. It is reported in [89] how a new national soil information system for New Zealand was developed and implemented using a hybrid approach of analogue and digital soil mapping methods. This hybrid approach integrates both traditional soil survey processes and data with modern digital soil mapping techniques and information in order to (eventually) achieve total coverage of New Zealand at a 1:50,000 scale, soil data collection, archiving, and verification by photograph and database.

Several audiences receive customized dynamic fact sheets, maps, and spatial data. The system can conduct pedotransfer functions (PTFs) and other digital soil mapping activities, manage and simulate soil uncertainty, and produce relevant metadata reports. Soil pH, calcium (Ca), and phosphorus levels were predicted using an artificial neural network (ANN) and random forest (RF) machine learning techniques [90]. Farmers can use the Ca, P, and pH readings from a soil sample to determine how much fertilizer to add to the soil. Soil particle-size fractions (PSF) were predicted in Nigeria at six traditional soil depths using GlobalSoilMap criteria (0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm). RF provides reliable predictions of the particle-size fraction composition of Nigeria's soil [91].

Using ESRI software and both main and secondary soil maps based on the geographical subdivision of mapping units found in the dataset source, ref. [92] created a soil information system. Modern soil characteristics are displayed by this system. 250,000 plots were used for sampling, and 100,000 soil mapping units (SMUs) were analysed. Soil characterization units have advanced relational databases and physical and chemical soil categories that facilitate digital descriptions of soil profiles. Soil organic carbon (g kg^1), soil pH, sand, silt, and clay fractions (%), bulk density (kg m^3), cation-exchange capacity (cmol+/kg), coarse fragments (%), soil organic carbon stock (t ha^1), and depth to bedrock are just a few of the local soil properties that [93] takes into account using tree-based models (random forest and gradient tree boosting) at a 250-metre resolution in a 3D soil information system (cm). In order to better assist farmers in managing their crops, Ref. [94] introduced a new IoT and machine learning-based soil information system that would provide them with real-time temperature and soil moisture data for environmental monitoring. Modern technologies allow farmers to instantly report crop, soil variety, and N-P-K levels. The technology is designed to be used by farmers in any location while allowing end users to control their connected farms from afar. There is a rise in climate change adaptation and mitigation efforts.

Ref. [95] built a method for managing soil that makes informed crop suggestions using classifier models. An intuitive web-based content management system is part of the created soil information system, which can be used to make planting predictions. The system is extensible because it can be tested on a wide variety of crops and because it presents the possibility of employing information mining techniques to estimate crop yields based on input parameters for environmental circumstances. However, the soil databases (information systems) currently in use are not extensive or precise enough to incorporate soil data into the global geographic data infrastructure [96]. This is mostly due to the fact that, given their current capacity, they can only store information from sporadic and occasionally available conventional soil surveys. Due to the slow and expensive nature of conventional soil survey methods, there are not many spatial datasets available for soil. The future of conventional soil surveying is also causing some individuals considerable concern due to a general problem in the collection of new field data. The authors of [96] expect technological innovations such as handheld field spectrometers to come to the rescue. To effectively deal with this problem, it was proposed that existing soil information systems be expanded to allow for the generation of new soil maps in addition to the storage and use of digitized (pre-existing) soil maps. One definition of digital soil mapping is the process of creating and populating spatial soil information systems via field and laboratory observational methods in combination with non-spatial and spatial inference systems.

6. Artificial Intelligence Models for Soil Properties Prediction

A quick perusal of related work on artificial intelligence (AI) models and digital soil mapping (DSM), as summarized in Table 2, reveals that AI models are the norm for predicting soil attributes and digital soil mapping. Ref. [50] offered a computerized soil mapping method for preventing gully erosion by advising landowners on preventative steps. Using R-Squared, KC, and RMSE as accuracy metrics, a multiple nonlinear regression model was built with 68% precision. Nonetheless, the low accuracy is understandable given that the soil depth map is not a fair depiction of the sample in reality, making it difficult to conduct research. The use of machine learning algorithms for estimating soil depth has been explored further in [54]. QRF models were utilized, and with RMSE as the measure of evaluation, they were able to reach an accuracy of 30%. It can be inferred from the accuracy percentage that soil depth in digital soil mapping is still a discoverable topic. An evaluation of soil fertility using DSM and machine learning techniques was proposed in [51]. Using the Quality Reference Framework (QRF), great accuracy was attained by utilizing the evaluation metrics RMSE and MAE. However, the model's precision was constrained for some soil characteristics, such as nitrogen (N) and potassium (K). Soil maps for a variety of soil qualities, for which QRF was able to provide the best accuracy, is another issue that was addressed.

Self-organizing maps (SOMs) were also employed as a machine learning model [63]. Supervised maps are used to forecast soil moisture using SOM and random forest (RF) models; when tested on a dataset including both soil moisture and land cover, SOM showed greater model accuracy than RF when evaluated with respect to R2 and KC. Multi-sensor data and ML algorithms, including RF, XGBoost, and SVM (supervised vector machine), were also used to make predictions about soil moisture, with an accuracy of 87.5% [62]. Many deep learning methods, such as deep neural networks (DNN) and artificial neural networks (ANN), have been used to predict soil attributes in space. With an AUC of 89.8%, DNN achieved the highest accuracy. Due to the lack of high-quality artificial intelligence solutions for digital soil mapping, researchers from all over the world are paying close attention to the field.

In addition, a synopsis of the prior research conducted on intelligent soil prediction between 2016 and 2022 is provided in Table 3, along with information regarding the source, solution provided, and dataset type. Finally, some of the existing online and mobile applications pertaining to soil are described in Table 4, along with the documented source, application name, function, and date.

Table 3. Previous Work on Smart Soil Prediction (2016–2022).

Source	Solution	Soil Dataset
[97]	Prediction of clay soil expansion using ML models and meta-heuristic dichotomous ensemble classifier	Soil swelling and soil properties data.
[98]	Predicting crop yield on a particular soil using IoT	Nutritional value of soil data.
[99]	ML approach to simulate soil CO_2 fluxes under cropping systems	Soil classification and temperature data.
[100]	Predicting Africa soil properties using ML techniques	Soil sample measures, soil depth (topsoil or subsoil) and climate data.
[101]	Soil analysis of micro-nutrients using ML and IoT	Soil micro-nutrient and soil pH data.
[102]	Estimation of the moisture of vineyard soils from digital photography using ML.	Soil sample and photographic data.
[103]	Prediction of soil shear strength parameters using ML algorithms	Soil properties and cone penetration test data.
[104]	Analysis of ML methods for agricultural soil health management	Secondary data.
[105]	Crops yield prediction based on mL models in West African countries	Climate, yields, pesticides and chemical data.

Abbreviations: IoT—Internet of Things, ML—Machine learning.

Table 4. Existing Soil Web/Mobile App.

Source	Application Name	Year	Functions
[106]	SQAPP	2015	Sustainability of SM and high productivity
[107]	SoilWeb	1999	Instantaneous soil information
[108]	AgriApp	2014	Crop advisory, soil testing and drone services
[109]	LandPKS	2020	Soil health monitoring and land management

Table 4. *Cont.*

Source	Application Name	Year	Functions
[110]	Crop App Index	2017	Agricultural decision support tool
[111]	MySoil Test Kit	Not Specified	Information to improve soil and plant health
[112]	SIFSS	2017	Provides indication scores for soil types.
[113]	Soil Test Pro	2019	Soil nutrient management system
[114]	SoilScapes	Not Specified	Digital smart information system
[115]	SoilInfo App	2017	Generate open soil data
[116]	SoilCares	2021	Smart application for monitoring soil nutrients and soil fertility

6.1. Data Quality and ML Model Considerations

The efficacy of smart soil systems in predicting accurate outcomes is contingent upon several factors, with the foremost being the quality of the data employed and the identification of a machine-learning model that yields the optimal result. There exist various measures that can be implemented to enhance the accuracy of models employed for predicting soil nutrient levels and to improve the quality of data. A few notions are discussed below:

(a) Data gathering and preprocessing: This entails making sure that the soil types, geographic areas, and environmental conditions represented in the model training data are accurate. In order to understand soil nutrients, data must also be gathered through soil samples, lab testing, remote sensing, and historical records. The final step is data cleaning, which includes handling missing numbers, fixing errors, and removing outliers [117].

(b) Feature engineering: In order to enhance the accuracy of soil nutrient level estimation, it is imperative to identify and extract relevant features from the collected data. The influence of environmental factors, including climate, rainfall, and cultivation of land, as well as the chemical, biological, and physical characteristics of soil, is potentially significant [118].

(c) Integrate domain knowledge: In order to gain further insight into the determinants that impact the levels of nutrients present in the soil, it is recommended to consult with experts in the domain [119], including agricultural scientists or researchers specializing in soil science. Applying this data when constructing the models and determining which attributes to incorporate is essential.

(d) Innovative modelling methods: Conducting research on state-of-the-art machine learning techniques [120] and advanced deep learning architectures is of great significance [121]. Furthermore, it is imperative to consider ensemble methodologies that employ an assemblage of models to enhance the accuracy of predictions.

(e) Model testing and verification: It is imperative to assess the model capacity to extrapolate to new datasets through the application of rigorous evaluation methodologies. Furthermore, assessment criteria are examined and monitored to measure the precision of the models [122].

6.2. Considerations for Choice of ML Technique for Soil Nutrient Properties Prediction

The choice of ML technique to employ for soil nutrient properties prediction and growth response analysis [123], as in any other class of problem, depends on several factors including the nature of the problem under consideration, the available data, and the desired outcome [34,124]. Different machine learning algorithms are designed to address specific types of problems, be it a classification, regression, clustering or recommendation problem [125,126]. The size and quality of available data must also be considered because some algorithms require large amounts of data to generalize well, while others can work effectively with smaller datasets, thereby avoiding fitting problems [127]. Depending on the algorithm, certain types of features may be more suitable, thereby necessitating the need for feature selection and extraction [128]. This is to enhance the predictive power of the features of the dataset.

The interpretability and explainability of a given model [129], when required, may impact the choice of the model over classical models. Some algorithms, such as decision trees or linear regression, provide easily interpretable models, while others, such as neural networks, may be more complex and harder to interpret. The statistical properties of the available dataset also largely determines the choice of ML technique to use in a given instance [130]. Considering the complexity of the relationship between the input variables and the target variable, simple problems may be effectively solved by linear models, while complex problems with non-linear relationships might require more sophisticated algorithms such as random forests or support vector machines.

Domain knowledge is a crucial element in the choice of ML technique used for predicting soil nutrients properties. Incorporating domain knowledge or expert insights into the decision-making process in the preprocessing and model building is essential to the reliability of the outcome of the prediction. Understanding the problem domain, a key component of responsible AI [131], can help guide the selection of appropriate algorithms and feature engineering techniques. Table 5 presents a quick summary of some popular ML techniques with their associated relative strengths and weaknesses which should be considered when determining the technique(s) to employ in predicting soil nutrients.

Table 5. Summary of Some ML Techniques with their Strengths and Weaknesses.

ML Technique	Strengths	Weaknesses
Support Vector Machine [132,133]	Effective in high-dimensional spaces, less prone to overfitting, versatile kernel functions, effective with small to medium datasets, insensitive to irrelevant features	Performs poorly with large or noisy data. Highly sensitive to hyparameter tuning
k-Nearest Neighbours [134,135]	Simple, highly intuitive, non-parametric, flexible decision boundaries, considers the local structure of the data, can be effective with both linear and non-linear relationships, handles outliers relatively more efficiently	There is high computational complexity during prediction phase, distance metric selection may be ambiguous, sensitive to the curse of dimensionality, struggles with imbalance data, has storage issues during prediction
Decision Trees [136,137]	Offers good explainability and interpretability, cognaissant to feature importance, handles non-linear relationships among features relatively well, good for mixed data (categorical + non-categorical), has low computational complexity, handles outliers well	Prone to overfitting, highly unstable, especially to a slight variation in the training set, makes locally optimal decisions without considering the global optimal structure, tends to favour features with a large number of categories or high cardinality, not well-suited for problems where classes are linearly separable, struggle to represent complex relationships that require global knowledge or long-range dependencies in the data

Table 5. *Cont.*

ML Technique	Strengths	Weaknesses
Linear Regression [138,139]	Interpretable, simple, resource efficient, robust feature importance identification, often useful as a baseline model for comparison with more complex algorithms	Often assumes a linear relationship between the input features and the target variable, does not handle outliers efficiently, relatively limited predictive power, naive assumption of homoscedasticity, also sensitive to multicollinearity
Logistic Regression [140–142]	Interpretability, efficiency, probabilistic problems, less prone to overfitting and allows for internal feature selection	Assumes linearity like the linear counterpart, handles limited complexity, cannot handle outliers, limited for binary classification, and can be afftected by imbalance dataset
Artificial Neural Network [143–145]	Ability to learn complex patterns, flexible architecture, automatically learn relevant features, supports parallel processing and has high generalization power thereby reducing fitting problems	Requires large amount of data, has high computational complexity, they lack good interpretability because of their black-box nature, sensitive to hyperparameter tuning
Naive Bayes [146]	Efficient with large datasets, scalable, robust to irrelavant features, effective with limited training sets, interpretable	Sensitive to skewness, does not capture complex relationships between features, highly sensitive to scaling problems
Random Forest [147]	Known for high accuracy, handles outliers and noisy data, handles high cardinality, good with variable importance, resistant to overfitting	Lacks explainability, computationally expensive, requires good hyperparameter tunning for optimal performance, biased towards majority classes
Gradient Boosting [148]	High predictive accuracy, high flexibility in handling mixed data types, provides insights into feature importance, handles outliers internally, handles missing data, can be parrallelized efficiently	Computationally expensive, has a potential problem of overfitting, difficult to interpret, relies heavuly on the order (or sequence) of the training data

7. Findings and Discussion

Figure 5 is a chart depicting the issue that this review seeks to address (as outlined in Table 2). According to the visual analysis, the majority of published works (67.3%) dealt with issues of soil nutrient characteristics; 17.3% handled DSM; 11.1% addressed soil erosion; and 5.5% dealt with soil fertility. Figure 6 also provides a visual representation of the most popular model employed in the research covered in Table 2's meta-analysis, which shows that random forest (RF) is the most popular choice for prediction, followed by support vector machine (SVM) and other ML algorithms as shown in Figure 6.

Our findings show that RF outperformed other ML models in terms of accuracy. Random forest is a popular machine-learning approach that can handle both regression and classification challenges, which makes it an adaptable option for forecasting soil characteristics, nutrients, and soil fertility. At the training phase, the algorithm generates a variety of decision trees and then combines their results to extrapolate. Random forest has a number of advantages that may have led to its excellent success in forecasting soil characteristics, nutrients, and soil fertility:

(a) Resiliency to distortion: When compared to other algorithms, RF is less susceptible to noise and outliers, which might help it deliver precise forecasts even when working with unclear or missing soil data.

(b) Managing massive data: Because RF can accommodate large datasets with many input features, it is well suited for forecasting soil qualities with several factors impacting their values, such as pH, moisture content, organic matter, and nutrient levels.

(c) Features selection: RF automatically chooses the most significant features for making predictions, which can aid in identifying the main soil qualities and nutrients that are most important in determining soil fertility.
(d) Overfitting minimization: Random forest employs numerous decision trees and aggregates their outputs, which can aid in the reduction of overfitting, a typical problem in machine learning in which models perform well on training data but fail to generalize to new data.
(e) Random forest's ensembling feature, in which it integrates many decision trees, aids in bias reduction and prediction accuracy by using the collective wisdom of multiple trees.

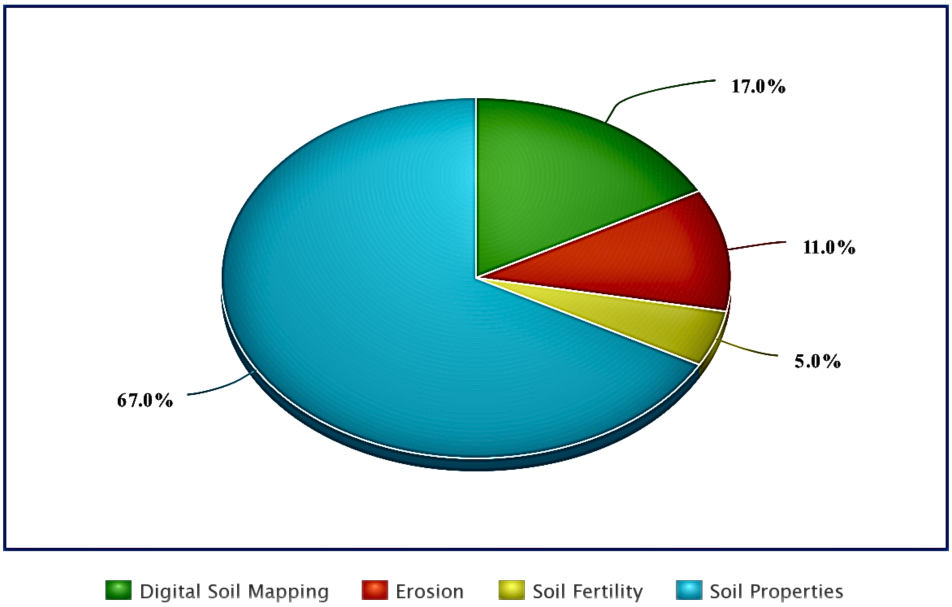

Figure 5. Graphical representation of the problem addressed.

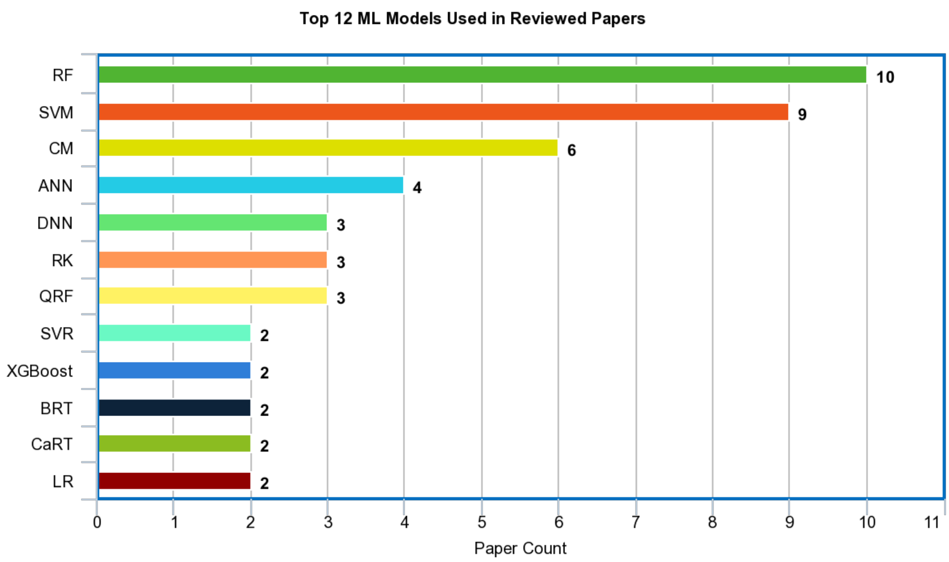

Figure 6. Graphical representation of the top 12 ML models used.

Overall, random forest's superior performance in predicting soil characteristics, nutrients, and soil fertility can be attributed to its capacity to deal with noise, big datasets, feature selection, overfitting reduction, and ensembling, making it a useful tool for soil-related prediction tasks. It should be noted, however, that the performance of any machine learning method is dependent on the quality of the data used for training and testing, as well as suitable parameter tweaking and model evaluation approaches. Furthermore, merging deep learning algorithms with ML can yield an ideal answer. In a nutshell, additional research on intelligent soil prediction and smart agriculture is imperative for broadening the knowledge repository, improving prediction techniques, and addressing the challenges confronting contemporary farming. Through the utilisation of these tools, it is possible to enhance food security, optimise resource utilisation, alleviate the impact of environmental degradation, enable precision farming techniques, and promote sustainable development within the agricultural sector.

8. Conclusions

This study reviews machine learning methods for predicting soil properties, agricultural yield, and soil fertility. This literature evaluation illuminated current research gaps in a specific field of machine learning methodologies and provided useful data on soil attribute prediction. Through this extensive literature study, we learn about the several forms of machine learning used in this subject, the soil characteristics problem that has been addressed, and crop yield prediction criteria. Each study focused on a distinct set of soil qualities, geographical conditions, and other features. For soil prediction, RF and deep learning outperform conventional machine learning methods. The RF machine learning algorithm and deep learning approach can accurately predict soil conditions and inform us if a crop can be grown there given the model's inputs. From the literature evaluation, it is observed that the task of predicting soil or agricultural yield through machine learning poses significant challenges. Inaccurate data has the potential to decrease the precision of forecasting. The process of generalising models is impeded by variations in regional factors, climatic conditions, and farming practices. Additionally, the selection of significant features from multiple influencing factors requires domain expertise and experimentation. In order to employ technology in a responsible manner, it is imperative to address all of these issues. The refinement of machine learning techniques for the purpose of predicting soil characteristics and crop yield is facilitated by expert collaboration, model monitoring, and modification. The application of machine learning techniques to soil information analysis can lead to the optimisation of fertiliser usage, prediction of pest and disease outbreaks, and recommendation of precise irrigation strategies. This can result in enhanced agricultural productivity and efficient management of land resources.

The amalgamation of DSM and ML techniques for soil prediction poses certain challenges in less developed nations. The challenges encountered in the implementation of machine learning and data science initiatives include language and cultural impediments, insufficient financial resources, suboptimal internet connectivity, and restricted availability of reliable and all-encompassing data. In order to address these challenges, it is crucial to allocate resources towards data collection, network enhancements, computing infrastructure, and the promotion of education and training to cultivate local expertise. Partnerships and collaborations with foreign organisations can be advantageous for both information sharing and personnel development. Furthermore, increasing soil investigation, analytical ability, facilities, and public participation would solve these issues. Digital soil mapping and machine learning for soil prediction can improve soil management and agricultural productivity in developing nations.

Funding: This research was funded by European Union Under the Accelerating Inclusive Green Growth through Agri-Based Digital Innovation in West Africa (AGriDI), which is an ACP Innovation Fund (grant ref: FED/2020/420-491) led by the International Centre for Insect Physiology and Ecology (ICIPE), Nairobi Kenya.

Data Availability Statement: Not applicable to this article.

Acknowledgments: We would like to thank the reviewers for their valuable comments and feedback, which have greatly improved the quality and clarity of our manuscript.

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

References

1. Sundari, V.; Anusree, M.; Swetha, U. Crop recommendation and yield prediction using machine learning algorithms. *World J. Adv. Res. Rev.* **2022**, *14*, 452–459. [CrossRef]
2. Muthoni, F.; Thierfelder, C.; Mudereri, B.; Manda, J.; Bekunda, M.; Hoeschle-Zeledon, I. Machine learning model accurately predict maize grain yields in conservation agriculture systems in Southern Africa. In Proceedings of the 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Shenzhen, China, 26–29 July 2021; pp. 1–5.
3. Rao, T.V.N.; Reddy, G.R. Prediction Of Soil Quality Using Machine Learning Techniques. *Int. J. Sci. Technol. Res.* **2019**, *8*, 1309–1313.
4. Phasinam, K.; Kassanuk, T.; Shabaz, M. Applicability of internet of things in smart farming. *J. Food Qual.* **2022**, *2022*, 7692922. [CrossRef]
5. Ma, Y.; Minasny, B.; Malone, B.P.; Mcbratney, A.B. Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* **2019**, *70*, 216–235. [CrossRef]
6. Shaikh, F.K.; Memon, M.A.; Mahoto, N.A.; Zeadally, S.; Nebhen, J. Artificial intelligence best practices in smart agriculture. *IEEE Micro* **2021**, *42*, 17–24. [CrossRef]
7. Chen, Q.; Li, L.; Chong, C.; Wang, X. AI-enhanced soil management and smart farming. *Soil Use Manag.* **2022**, *38*, 7–13. [CrossRef]
8. Dobos, E. *Digital Soil Mapping: As a Support to Production of Functional Maps*; Office for Official Publication of the European Communities: Luxembourg, 2006.
9. Wadoux, A.M.C.; Minasny, B.; McBratney, A.B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. [CrossRef]
10. Lagacherie, P.; Buis, S.; Constantin, J.; Dharumarajan, S.; Ruiz, L.; Sekhar, M. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India). *Geoderma* **2022**, *406*, 115503. [CrossRef]
11. Dong, W.; Wu, T.; Sun, Y.; Luo, J. Digital mapping of soil available phosphorus supported by AI technology for precision agriculture. In Proceedings of the 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), Hangzhou, China, 6–9 August 2018; pp. 1–5.
12. Khaledian, Y.; Miller, B.A. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. [CrossRef]
13. Shahare, Y.; Gautam, V. Soil Nutrient Assessment and Crop Estimation with Machine Learning Method: A Survey. In *Cyber Intelligence and Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 253–266.
14. Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* **2021**, *80*, 2803–2806. [CrossRef]
15. Murmu, K.; Swain, D.K.; Ghosh, B.C. Comparative assessment of conventional and organic nutrient management on crop growth and yield and soil fertility in tomato-sweet corn production system. *Aust. J. Crop. Sci.* **2013**, *7*, 1617–1626.
16. Shukla, M.; Lal, R.; Ebinger, M. Determining soil quality indicators by factor analysis. *Soil Tillage Res.* **2006**, *87*, 194–204. [CrossRef]
17. Rezaei, S.A.; Gilkes, R.J.; Andrews, S.S. A minimum data set for assessing soil quality in rangelands. *Geoderma* **2006**, *136*, 229–234. [CrossRef]
18. Li, Y.; Li, Z.; Cui, S.; Zhang, Q. Trade-off between soil pH, bulk density and other soil physical properties under global no-tillage agriculture. *Geoderma* **2020**, *361*, 114099. [CrossRef]
19. Fernández, F.G.; Hoeft, R.G. Managing soil pH and crop nutrients. *Ill. Agron. Handb.* **2009**, *24*, 91–112.
20. Marschner, P.; Rengel, Z. Nutrient availability in soils. In *Marschners Mineral Nutrition of Higher Plants*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 315–330.
21. Kovačević, M.; Bajat, B.; Gajić, B. Soil type classification and estimation of soil properties using support vector machines. *Geoderma* **2010**, *154*, 340–347. [CrossRef]
22. Baskar, S.; Arockiam, L.; Charles, S. Applying data mining techniques on soil fertility prediction. *Int. J. Comput. Appl. Technol. Res.* **2013**, *2*, 660–662. [CrossRef]
23. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M.; Doukas, I.J.D. Comparing machine learning models and hybrid geostatistical methods using environmental and soil covariates for soil pH prediction. *ISPRS Int. J. -Geo-Inf.* **2020**, *9*, 276. [CrossRef]
24. Zeraatpisheh, M.; Garosi, Y.; Owliaie, H.R.; Ayoubi, S.; Taghizadeh-Mehrjardi, R.; Scholten, T.; Xu, M. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena* **2022**, *208*, 105723. [CrossRef]
25. Legros, J.P. *Mapping of the Soil*; Science Publishers: Singapore, 2006.
26. Ryan, P.; McKenzie, N.; O'Connell, D.; Loughhead, A.; Leppert, P.; Jacquier, D.; Ashton, L. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. *For. Ecol. Manag.* **2000**, *138*, 139–157. [CrossRef]
27. Hudson, B.D. The Soil Survey as Paradigm-based Science. *Soil Sci. Soc. Am. J.* **1992**, *56*, 836–841. [CrossRef]
28. Lagacherie, P.; McBratney, A. Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. In *Digital Soil Mapping*; Lagacherie, P., McBratney, A., Voltz, M., Eds.; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2006; Volume 31, pp. 3–22. [CrossRef]

29. Franklin, J. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr. Earth Environ.* **1995**, *19*, 474–499. [CrossRef]
30. McKenzie, N.J.; Ryan, P.J. Spatial prediction of soil properties using environmental correlation. *Geoderma* **1999**, *89*, 67–94. [CrossRef]
31. Scull, P.; Franklin, J.; Chadwick, O.; McArthur, D. Predictive soil mapping: A review. *Prog. Phys. Geogr.* **2003**, *27*, 171–197. [CrossRef]
32. Kempen, B.; Heuvelink, G.B.M.; Brus, D.J.; Stoorvogel, J.J. Pedometric mapping of soil organic matter using a soil map with quantified uncertainty. *Eur. J. Soil Sci.* **2010**, *61*, 333–347. [CrossRef]
33. Tomlinson, R. Design Considerations for Digital Soil Map Systems. In Proceedings of the 11th Congress of Soil Science, ISSS, Edmonton, AB, Canada, 19–27 June 1978
34. McBratney, A.; Mendonça Santos, M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]
35. Florinsky, I. The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). *Eurasian Soil Sci.* **2012**, *45*, 445–451. [CrossRef]
36. Bou Kheir, R.; Greve, M.H.; Bøcher, P.K.; Greve, M.B.; Larsen, R.; McCloy, K. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. *J. Environ. Manag.* **2010**, *91*, 1150–1160. [CrossRef]
37. Brungard, C.W.; Boettinger, J.L.; Duniway, M.C.; Wills, S.A.; Edwards, T.C. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* **2015**, *239–240*, 68–83. [CrossRef]
38. Subburayalu, S.K.; Slater, B.K. Soil Series Mapping By Knowledge Discovery from an Ohio County Soil Map. *Soil Sci. Soc. Am. J.* **2013**, *77*, 1254–1268. [CrossRef]
39. Odgers, N.P.; Sun, W.; McBratney, A.B.; Minasny, B.; Clifford, D. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* **2014**, *214–215*, 91–100. [CrossRef]
40. Moran, C.J.; Bui, E.N. Spatial data mining for enhanced soil map modelling. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 533–549. [CrossRef]
41. Martinelli, G.; Gasser, M.O. Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. *Soil Sci. Soc. Am. J.* **2022**, *86*, 1509–1522. [CrossRef]
42. Payen, F.T.; Sykes, A.; Aitkenhead, M.; Alexander, P.; Moran, D.; MacLeod, M. Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression. *Clean. Environ. Syst.* **2021**, *2*, 100024. [CrossRef]
43. Liu, D.; Liu, C.; Tang, Y.; Gong, C. A GA-BP neural network regression model for predicting soil moisture in slope ecological protection. *Sustainability* **2022**, *14*, 1386. [CrossRef]
44. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sens.* **2020**, *12*, 236. [CrossRef]
45. Kuwata, K.; Shibasaki, R. Estimating crop yields with deep learning and remotely sensed data. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 858–861.
46. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]
47. Keerthan Kumar, T.; Shubha, C.; Sushma, S. Random forest algorithm for soil fertility prediction and grading using machine learning. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 1301–1304.
48. Benedet, L.; Acuña-Guzman, S.F.; Faria, W.M.; Silva, S.H.G.; Mancini, M.; dos Santos Teixeira, A.F.; Pierangeli, L.M.P.; Júnior, F.W.A.; Gomide, L.R.; Júnior, A.L.P.; et al. Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. *Catena* **2021**, *197*, 105003. [CrossRef]
49. Muruganantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sens.* **2022**, *14*, 1990. [CrossRef]
50. Du Plessis, C.; Van Zijl, G.; Van Tol, J.; Manyevere, A. Machine learning digital soil mapping to inform gully erosion mitigation measures in the Eastern Cape, South Africa. *Geoderma* **2020**, *368*, 114287. [CrossRef]
51. Hounkpatin, K.O.; Bossa, A.Y.; Yira, Y.; Igue, M.A.; Sinsin, B.A. Assessment of the soil fertility status in Benin (West Africa)—Digital soil mapping using machine learning. *Geoderma Reg.* **2022**, *28*, e00444. [CrossRef]
52. Taghizadeh-Mehrjardi, R.; Hamzehpour, N.; Hassanzadeh, M.; Heung, B.; Goydaragh, M.G.; Schmidt, K.; Scholten, T. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. *Geoderma* **2021**, *399*, 115108. [CrossRef]
53. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Eftekhari, K.; Behrens, T.; Jamshidi, M.; Davatgar, N.; Toomanian, N.; Scholten, T. Synthetic resampling strategies and machine learning for digital soil mapping in Iran. *Eur. J. Soil Sci.* **2020**, *71*, 352–368. [CrossRef]
54. Dharumarajan, S.; Vasundhara, R.; Suputhra, A.; Lalitha, M.; Hegde, R. Prediction of soil depth in Karnataka using digital soil mapping approach. *J. Indian Soc. Remote. Sens.* **2020**, *48*, 1593–1600. [CrossRef]
55. 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]
56. Lamichhane, S.; Kumar, L.; Wilson, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* **2019**, *352*, 395–413. [CrossRef]

57. Zeraatpisheh, M.; Ayoubi, S.; Mirbagheri, Z.; Mosaddeghi, M.R.; Xu, M. Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Reg.* **2021**, *27*, e00440. [CrossRef]
58. Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci. Total. Environ.* **2020**, *729*, 138244. [CrossRef] [PubMed]
59. Zhou, T.; Geng, Y.; Ji, C.; Xu, X.; Wang, H.; Pan, J.; Bumberger, J.; Haase, D.; Lausch, A. Prediction of soil organic carbon and the C: N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Sci. Total. Environ.* **2021**, *755*, 142661. [CrossRef] [PubMed]
60. Zhou, T.; Geng, Y.; Chen, J.; Liu, M.; Haase, D.; Lausch, A. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. *Ecol. Indic.* **2020**, *114*, 106288. [CrossRef]
61. Emadi, M.; Taghizadeh-Mehrjardi, R.; Cherati, A.; Danesh, M.; Mosavi, A.; Scholten, T. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote. Sens.* **2020**, *12*, 2234. [CrossRef]
62. Nguyen, T.T.; Ngo, H.H.; Guo, W.; Chang, S.W.; Nguyen, D.D.; Nguyen, C.T.; Zhang, J.; Liang, S.; Bui, X.T.; Hoang, N.B. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Sci. Total. Environ.* **2022**, *833*, 155066. [CrossRef]
63. Riese, F.M.; Keller, S.; Hinz, S. Supervised and semi-supervised self-organizing maps for regression and classification focusing on hyperspectral data. *Remote. Sens.* **2019**, *12*, 7. [CrossRef]
64. Du, F.; Zhu, A.X.; Liu, J.; Yang, L. Predictive mapping with small field sample data using semi-supervised machine learning. *Trans. Gis* **2020**, *24*, 315–331. [CrossRef]
65. Mosavi, A.; Sajedi-Hosseini, F.; Choubin, B.; Taromideh, F.; Rahi, G.; Dineva, A.A. Susceptibility mapping of soil water erosion using machine learning models. *Water* **2020**, *12*, 1995. [CrossRef]
66. Keskin, H.; Grunwald, S.; Harris, W.G. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **2019**, *339*, 40–58. [CrossRef]
67. Behrens, T.; Schmidt, K.; MacMillan, R.A.; Viscarra Rossel, R.A. Multi-scale digital soil mapping with deep learning. *Sci. Rep.* **2018**, *8*, 15244. [CrossRef]
68. Ou, D.; Tan, K.; Lai, J.; Jia, X.; Wang, X.; Chen, Y.; Li, J. Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction. *Geoderma* **2021**, *385*, 114875. [CrossRef]
69. Yao, J.; Qin, S.; Qiao, S.; Che, W.; Chen, Y.; Su, G.; Miao, Q. Assessment of landslide susceptibility combining deep learning with semi-supervised learning in Jiaohe County, Jilin Province, China. *Appl. Sci.* **2020**, *10*, 5640. [CrossRef]
70. Kaluba, P.; Mwamba, S.; Moualeu-Ngangue, D.P.; Chiona, M.; Munyinda, K.; Winter, E.; Stutzel, H.; Chishala, B.H. Cropping Practices and Effects on Soil Nutrient Adequacy Levels and Cassava Yield of Smallholder Farmers in Northern Zambia. *Int. J. Agron.* **2021**, *2021*, 1325964. [CrossRef]
71. Mwamba, S.; Kaluba, P.; Moualeu-Ngangue, D.; Winter, E.; Chiona, M.; Chishala, B.H.; Munyinda, K.; Stutzel, H. Physiological and morphological responses of cassava genotypes to fertilization regimes in chromi-haplic acrisols soils. *Agronomy* **2021**, *11*, 1757. [CrossRef]
72. Agbede, T.; Adekiya, A.; Ogeh, J. Effects of Chromolaena and Tithonia mulches on soil properties, leaf nutrient composition, growth and yam yield. *West Afr. J. Appl. Ecol.* **2013**, *21*, 15–30.
73. Sanchez, D.; Luna, L.; ESPITIA, A.; Cadena, J. Yield response of yam (Dioscorea rotundata Poir.) to inoculation with Azotobacter and nitrogen chemical fertilization in the Caribbean region of Colombia. *RIA Rev. Investig. Agropecu.* **2021**, *47*, 61–70.
74. Byju, G.; Suja, G. Mineral nutrition of cassava. *Adv. Agron.* **2020**, *159*, 169–235.
75. Laekemariam, F. Soil nutrient status of smallholder cassava farms in southern Ethiopia. *J. Biol. Agric. Healthc.* **2016**, *6*, 12–18.
76. Otieno, H.M. Growth and yield response of maize (*Zea mays* L.) to a wide range of nutrients on ferralsols of western Kenya. *World Sci. News* **2019**, *129*, 96–106.
77. Endris, S.; Dawid, J. Yield response of maize to integrated soil fertility management on acidic nitosol of Southwestern Ethiopia. *J. Agron.* **2015**, *14*, 152–157. [CrossRef]
78. Aziz, T.; Ullah, S.; Sattar, A.; Nasim, M.; Farooq, M.; Khan, M.M. Nutrient availability and maize (Zea mays) growth in soil amended with organic manures. *Int. J. Agric. Biol.* **2010**, *12*, 621–624.
79. Salami, B.; Sangoyomi, T. Soil fertility status of cassava fields in South Western Nigeria. *Am. J. Exp. Agric.* **2013**, *3*, 152. [CrossRef]
80. Akom, M.; Oti-Boateng, C.; Otoo, E.; Dawoe, E. Effect of biochar and inorganic fertilizer in yam (Dioscorea rotundata Poir) production in a forest agroecological zone. *J. Agric. Sci.* **2015**, *7*, 211–222. [CrossRef]
81. Mainoo, A.; Banful, B.K. Yam plant growth and tuber yield response to ex-situ mulches of moringa oleifera, chromolaena odorata and panicum maximum under three natural fallow aged systems. *Ann. Ecol. Environ. Sci.* **2018**, *2*, 7–14.
82. McCauley, A.; Jones, C.; Jacobsen, J. Basic soil properties. *Soil Water Manag. Modul.* **2005**, *1*, 1–12.
83. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* **2020**, *6*, 35–52. [CrossRef]
84. Dai, Y.; Shangguan, W.; Wei, N.; Xin, Q.; Yuan, H.; Zhang, S.; Liu, S.; Lu, X.; Wang, D.; Yan, F. A review of the global soil property maps for Earth system models. *Soil* **2019**, *5*, 137–158. [CrossRef]

85. Van Loenen, B.; Kok, B. *Spatial Data Infrastructure and Policy Development in Europe and the United States*; DUP Science: Delft, The Netherlands, **2004**.
86. Masser, I. All shapes and sizes: The first generation of national spatial data infrastructures. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 67–84. [CrossRef]
87. Dwivedi, R.S. *Remote Sensing of Soils*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 497.
88. Eckelmann, W. Soil information for Germany: The 2004 position. *Soil Resour. Eur.* **2005**, *9*, 147.
89. Lilburne, L.; Hewitt, A.; Webb, T. Soil and informatics science combine to develop S-map: A new generation soil information system for New Zealand. *Geoderma* **2012**, *170*, 232–238. [CrossRef]
90. Nshimiyimana, N. Machine Learning based Soil Fertility Prediction. *Int. J. Innov. Sci. Eng. Technol.* **2021**, *8*, 141–146
91. Akpa, S.I.; Odeh, I.O.; Bishop, T.F.; Hartemink, A.E. Digital mapping of soil particle-size fractions for Nigeria. *Soil Sci. Soc. Am. J.* **2014**, *78*, 1953–1966. [CrossRef]
92. Pásztor, L.; Szabó, J.; Bakacsi, Z.; Matus, J.; Laborczi, A. Compilation of 1: 50,000 scale digital soil maps for Hungary based on the digital Kreybig soil information system. *J. Maps* **2012**, *8*, 215–219. [CrossRef]
93. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.; 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]
94. Kumar, T.R.; Aiswarya, B.; Suresh, A.; Jain, D.; Balaji, N. Smart management of crop cultivation using IOT and machine learning. *Int. Res. J. Eng. Technol. (IRJET)* **2018**, *5*, 845–850.
95. Patil, P.; Panpatil, V.; Kokate, S. Crop prediction system using machine learning algorithms. *Int. Res. J. Eng. Technol. (IRJET)* **2020**, *7*, 748–753.
96. Nachtergaele, F.; Van Ranst, E. Qualitative and quantitative aspects of soil databases in tropical countries. *Evolution of Tropical Soil Science: Past and Future*; Koninklijke Academie voor Overzeese Wetenschappen: Brussels, Belgium, 2003; pp. 107–126.
97. Eyo, E.; Abbey, S.; Lawrence, T.; Tetteh, F. Improved prediction of clay soil expansion using machine learning algorithms and meta-heuristic dichotomous ensemble classifiers. *Geosci. Front.* **2022**, *13*, 101296. [CrossRef]
98. Sudha, M.K.; Manorama, M.; Aditi, T. Smart Agricultural Decision Support Systems for Predicting Soil Nutrition Value Using IoT and Ridge Regression. *Agris Line Pap. Econ. Inform.* **2022**, *14*, 95–106. [CrossRef]
99. Adjuik, T.A.; Davis, S.C. Machine Learning Approach to Simulate Soil CO2 Fluxes under Cropping Systems. *Agronomy* **2022**, *12*, 197. [CrossRef]
100. Akinola, I.; Dowd, T. Predicting Africa Soil Properties Using Machine Learning Techniques. *Electr. Eng. Stanf. Univ. Stanford CA* **2016**, *94305*, 50–62.
101. Blesslin Sheeba, T.; Anand, L.; Manohar, G.; Selvan, S.; Wilfred, C.B.; Muthukumar, K.; Padmavathy, S.; Ramesh Kumar, P.; Asfaw, B.T. Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms. *J. Nanomater.* **2022**, *2022*, 5343965. [CrossRef]
102. Hajjar, C.S.; Hajjar, C.; Esta, M.; Chamoun, Y.G. Machine learning methods for soil moisture prediction in vineyards using digital images. In *E3S Web of Conferences*; EDP Sciences: Les Ulis, France, 2020; Volume 167, p. 02004.
103. Zhu, L.; Liao, Q.; Wang, Z.; Chen, J.; Chen, Z.; Bian, Q.; Zhang, Q. Prediction of Soil Shear Strength Parameters Using Combined Data and Different Machine Learning Models. *Appl. Sci.* **2022**, *12*, 5100. [CrossRef]
104. Motia, S.; Reddy, S. Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: A quantitative evaluation. *J. Phys. Conf. Ser.* **2021**, *1950*, 012037. [CrossRef]
105. Cedric, L.S.; Adoni, W.Y.H.; Aworka, R.; Zoueu, J.T.; Mutombo, F.K.; Krichen, M.; Kimpolo, C.L.M. Crops Yield Prediction Based on Machine Learning Models: Case of West African Countries. *Smart Agric. Technol.* **2022**, *2*, 100049. [CrossRef]
106. SQAPP. iSQAPER. 2015. Available online: https://www.isqaper-is.eu/sqapp-the-soil-quality-app (accessed on 15 September 2022).
107. SoilWeb. SOILTEC GmbH. 1999. Available online: https://www.soiltecgeo.com/soilweb-r (accessed on 15 September 2022).
108. AgriApp. AgriApp: Smart Farming App—Apps on Google Play. 2014. Available online: https://play.google.com/store/apps/details (accessed on 18 September 2022).
109. LandPKS. LandPKS. 2020. Available online: https://landpotential.org (accessed on 20 September 2022).
110. CABI. Crop App Index. 2017. Available online: https://cropappindex.org (accessed on 20 September 2022).
111. MySoil. MySoil Test Kit. 2022. Available online: https://www.mysoiltesting.com (accessed on 22 September 2022).
112. SIFSS. Soil Indicators for Scottish Soils (SIFSS) App Update. 2017. Available online: https://soils.environment.gov.scot/news/soil-indicators-for-scottish-soils-sifss-app-update (accessed on 10 October 2022).
113. SoilTestPro. SoilTestPro. 2019. Available online: https://soiltestpro.com (accessed on 7 October 2022).
114. Soil, C.; Institute, A. Soilscapes. Available online: https://www.landis.org.uk/soilscapes (accessed on 10 October 2022).
115. ISRIC. SoilInfo App—Global Soil Data on Your Palm. 2017. Available online: https://www.isric.org/explore/soilinfo (accessed on 17 October 2022).
116. AgroCares. SoilCares-Smart Farming: Nutrient Testing. 2021. Available online: https://www.agrocares.com/soilcares (accessed on 17 October 2022).
117. Rodríguez-Pérez, J.R.; Marcelo, V.; Pereira-Obaya, D.; García-Fernández, M.; Sanz-Ablanedo, E. Estimating soil properties and nutrients by visible and infrared diffuse reflectance spectroscopy to characterize vineyards. *Agronomy* **2021**, *11*, 1895. [CrossRef]

118. Bocca, F.F.; Rodrigues, L.H.A. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Comput. Electron. Agric.* **2016**, *128*, 67–76. [CrossRef]
119. Kansou, K.; Laurier, W.; Charalambides, M.N.; Della-Valle, G.; Djekic, I.; Feyissa, A.H.; Marra, F.; Thomopoulos, R.; Bredeweg, B. Food modelling strategies and approaches for knowledge transfer. *Trends Food Sci. Technol.* **2022**, *120*, 363–373. [CrossRef]
120. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **2020**, *104*, 101822. [CrossRef]
121. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 104440. [CrossRef]
122. Christin, S.; Hervet, É.; Lecomte, N. Going further with model verification and deep learning. *Methods Ecol. Evol.* **2021**, *12*, 130–134. [CrossRef]
123. Akin, M.; Eyduran, E.; Reed, B.M. Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. *Plant Cell Tissue Organ Cult. (PCTOC)* **2017**, *128*, 303–316. [CrossRef]
124. Elavarasan, D.; Vincent, D.R.; Sharma, V.; Zomaya, A.Y.; Srinivasan, K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.* **2018**, *155*, 257–282. [CrossRef]
125. Coble, K.H.; Mishra, A.K.; Ferrell, S.; Griffin, T. Big data in agriculture: A challenge for the future. *Appl. Econ. Perspect. Policy* **2018**, *40*, 79–96. [CrossRef]
126. Akin, M.; Eyduran, S.P.; Eyduran, E.; Reed, B.M. Analysis of macro nutrient related growth responses using multivariate adaptive regression splines. *Plant Cell Tissue Organ Cult. (PCTOC)* **2020**, *140*, 661–670. [CrossRef]
127. Jabbar, H.; Khan, R.Z. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Comput. Sci. Commun. Instrum. Devices* **2015**, *70*, 163–172.
128. Dong, G.; Liu, H. *Feature Engineering for Machine Learning and Data Analytics*; CRC Press: Boca Raton, FL, USA, 2018.
129. Khare, S.K.; Acharya, U.R. An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals. *Comput. Biol. Med.* **2023**, *155*, 106676. [CrossRef] [PubMed]
130. Williams, N.; Zander, S.; Armitage, G. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *ACM Sigcomm. Comput. Commun. Rev.* **2006**, *36*, 5–16. [CrossRef]
131. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
132. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]
133. Suthaharan, S.; Suthaharan, S. Support vector machine. Machine learning models and algorithms for big data classification: Thinking with examples for effective learning. *Integr. Ser. Inf. Syst.* **2016**, *36*, 207–235.
134. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [CrossRef]
135. Cunningham, P.; Delany, S.J. k-Nearest neighbour classifiers—A Tutorial. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–25. [CrossRef]
136. Kingsford, C.; Salzberg, S.L. What are decision trees? *Nat. Biotechnol.* **2008**, *26*, 1011–1013. [CrossRef] [PubMed]
137. Kotsiantis, S.B. Decision trees: A recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [CrossRef]
138. Su, X.; Yan, X.; Tsai, C.L. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 275–294. [CrossRef]
139. Gross, J.; Groß, J. *Linear Regression*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003; Volume 175.
140. LaValley, M.P. Logistic regression. *Circulation* **2008**, *117*, 2395–2399. [CrossRef] [PubMed]
141. Nick, T.G.; Campbell, K.M. Logistic regression. In *Topics in Biostatistics*; Humana Press: New York, NY, USA, 2007; pp. 273–301.
142. Kleinbaum, D.G.; Dietz, K.; Gail, M.; Klein, M.; Klein, M. *Logistic Regression*; Springer: Berlin/Heidelberg, Germany, 2002.
143. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [CrossRef]
144. Aribisala, B.; Odusanya, O.; Olabanjo, O.; Wahab, E.; Atilola, O.; Saheed, A. Development of an Artificial Neural Network Model for Detection of COVID-19. *Int. J. Sci. Adv.* **2022**, *3*, 377–385. [CrossRef]
145. Olabanjo, O.A.; Wusu, A.S.; Manuel, M. A Machine Learning Prediction of Academic Performance of Secondary School Students Using Radial Basis Function Neural Network. *Trends Neurosci. Educ.* **2022**, *29*, 100190. [CrossRef]
146. Leung, K.M. Naive bayesian classifier. *Polytech. Univ. Dep. Comput. Sci. Risk Eng.* **2007**, *2007*, 123–156.
147. Macaulay, B.O.; Aribisala, B.S.; Akande, S.A.; Akinnuwesi, B.A.; Olabanjo, O.A. Breast cancer risk prediction in African women using random forest classifier. *Cancer Treat. Res. Commun.* **2021**, *28*, 100396. [CrossRef]
148. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobotics* **2013**, *7*, 21. [CrossRef]