land

MDPI

Review

Distributed Deep Learning and Intelligent Soil–Water Analytics in Precision Agriculture: A Comprehensive Review

Polina Lemenkova [1,2,3,4]

[1] Bureau de Recherches Géologiques et Minières (BRGM), 3 Avenue Claude-Guillemin, 45060 Orléans, France; polina.lemenkova@tech-orleans.fr; Tel.: +33-0633148601
[2] Le Studium, Institut d'Etudes Avancées du Val de Loire, 1 rue Dupanloup, 45000 Orléans, France
[3] Le Laboratoire PRISME (Laboratoire Pluridisciplinaire de Recherche en Ingénierie des Systèmes, Mécanique, Énergétique, UR 4229), Institut National des Sciences Appliquées Centre-Val de Loire (INSA CVL), Université d'Orléans, Château de la Source, 45100 Orléans, France
[4] La Technopole d'Orléans, 1 Avenue du Champ de Mars, 45100 Orléans, France

Abstract

Efficient management of soil–water resources is critical for global food security under intensifying climatic and demographic pressures. This review provides a comprehensive synthesis of artificial intelligence (AI) and distributed deep learning methodologies applied to soil–water interactions in precision agriculture. The physical and hydraulic foundations of soil–water systems—including water retention, unsaturated flow governed by the Richards equation, and soil degradation processes—are examined and situated within a unified framework of AI-based modeling and decision support. Classical machine learning (ML) algorithms (Random Forests, Support Vector Machines, gradient boosting) and deep learning architectures (convolutional neural networks, long short-term memory networks, transformers) are evaluated with respect to their capacity to predict soil moisture dynamics, estimate hydraulic properties, support smart irrigation scheduling, and generate digital soil maps at field-to-regional scales. Distributed training paradigms, federated learning for privacy-preserving multi-farm analytics, and edge AI deployment on low-power IoT hardware are assessed as enabling infrastructures for scalable agricultural intelligence. This review further addresses explainability, uncertainty quantification, and ethical dimensions inherent to AI-driven agricultural systems. Key challenges—including training data scarcity in data-poor regions, model interpretability, integration with physics-based hydrological models, and real-time deployment constraints—are critically discussed. Prospective research directions encompass physics-informed neural networks, foundation models for earth observation, autonomous digital twins of soil–water systems, and federated learning architectures aligned with data sovereignty frameworks. The synthesis underscores AI's transformative potential for sustainable agricultural water management while delineating the technical and sociotechnical barriers that must be resolved to realize this potential at a global scale.

Keywords: soil–water interactions; machine learning; deep learning; distributed learning; precision agriculture; smart irrigation; soil hydrology; artificial intelligence; remote sensing; digital agriculture

Academic Editors: Anastasia Angelaki and Parveen Sihag

Received: 16 May 2026
Revised: 15 June 2026
Accepted: 20 June 2026
Published: 24 June 2026

1. Introduction

1.1. Global Water Crisis and Agricultural Sustainability

Water scarcity has emerged as one of the most pressing and multidimensional challenges confronting global agricultural systems in the twenty-first century. Agriculture accounts for approximately 70% of total global freshwater withdrawals, making it by far the largest consumer of freshwater resources worldwide [1]. Concurrently, a confluence of accelerating forces—rapid population growth, urbanization, changing dietary preferences, and the intensifying effects of anthropogenic climate change—is placing unprecedented pressure on the hydrological systems that sustain food production. Projections by the Intergovernmental Panel on Climate Change (IPCC) indicate that altered precipitation regimes, increased frequency of extreme drought events, and accelerated glacial retreat will substantially reduce water availability across key agricultural regions in Asia, Sub-Saharan Africa, the Mediterranean basin, and parts of the Americas by mid-century [2]. In this context, the development of intelligent, data-driven frameworks for efficient soil–water management is no longer merely desirable—it is a scientific and societal imperative.

The challenge of sustainable water management in agriculture is compounded by the heterogeneity of soil systems, the spatial and temporal variability in precipitation, and the complex feedback between land use, vegetation, and the hydrological cycle; see Figure 1.

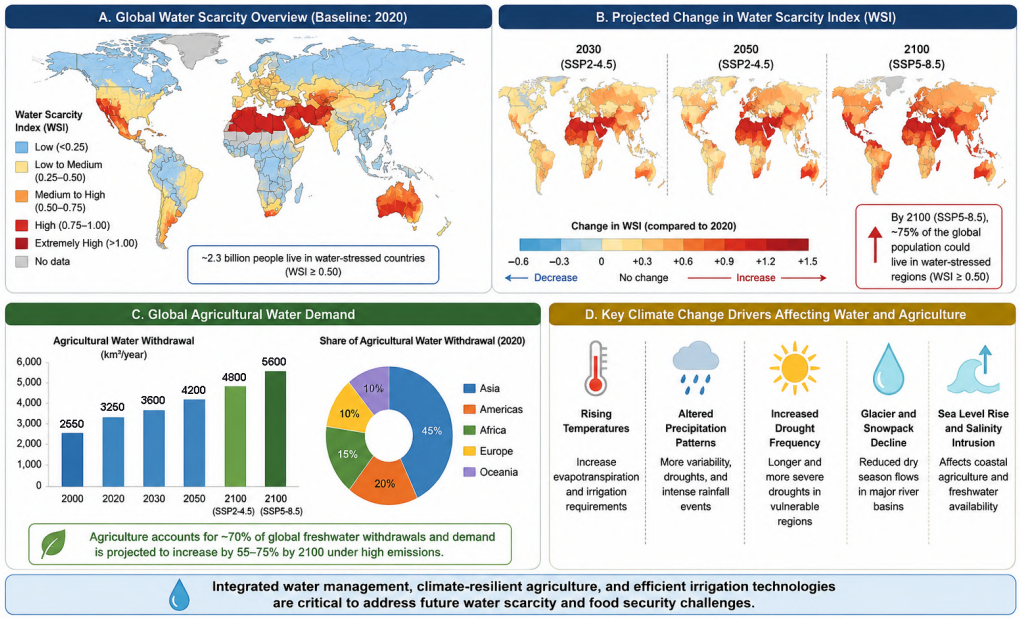

Figure 1. Global overview of water scarcity and agricultural water demand under climate change scenarios. Data sources: Water Scarcity Index (WSI) and projected changes derived from the IPCC [2]; agricultural water withdrawal statistics from the Food and Agriculture Organization of the United Nations [1]. Software used: QGIS v. 3.44 for mapping; R v. 4.6.1 and ggplot2 v. 4.0.3 for bar charts and pie charts; and Inkscape v. 1.4.4 for dashboard assembly. All map data and projections are the author's original compilation from the cited open-access sources. Dashboard source: author.

Traditional irrigation practices, which frequently rely on fixed schedules and empirical rules, are ill suited to the dynamic and location-specific character of soil–water interactions. Globally, it is estimated that between 40% and 60% of water applied through irrigation is lost to inefficient distribution, evaporation, deep percolation, and surface runoff [3]. Such inefficiencies not only deplete scarce freshwater reserves but also accelerate soil salinization, waterlogging, and nutrient leaching—degradation processes that undermine long-term agricultural productivity.

1.2. Importance of Soil–Water Interactions

At the heart of agricultural water management lies the intimate and dynamic relationship between soil physical properties and water movement. Soil–water interactions govern a broad suite of processes that are fundamental to plant growth and ecosystem function: infiltration, redistribution, evapotranspiration, capillary rise, groundwater recharge, surface runoff generation, and nutrient transport. The capacity of a soil to retain and transmit water is determined by its texture, structure, organic matter content, bulk density, and pore-size distribution—properties that vary considerably across landscapes and that are sensitive to management practices and climatic forcing.

The hydraulic behavior of soils is inherently nonlinear and spatially heterogeneous. In the unsaturated (vadose) zone, water retention and conductivity functions exhibit complex, history-dependent relationships that make deterministic modeling extremely demanding. The van Genuchten–Mualem model and the Brooks–Corey parametrization have been widely adopted to describe water retention curves, yet their parameterization requires intensive field and laboratory measurement campaigns that are impractical at regional or continental scales [4]. Similarly, processes such as preferential flow through macropores, crack flow in shrink–swell soils, and lateral redistribution in layered profiles introduce further complexity that classical Richards equation-based simulators struggle to capture adequately without very fine spatial discretization.

Understanding these dynamics is essential not only for optimizing irrigation efficiency but also for predicting the hydrological response of catchments to rainfall events, assessing the risk of contaminant transport to groundwater, and designing effective soil conservation measures. Soil degradation—through salinization, compaction, erosion, and loss of organic carbon—directly diminishes the water-holding capacity of agricultural soils, amplifying both drought stress during dry periods and runoff-related flooding during intense rainfall events. Climate-driven shifts in precipitation patterns are expected to exacerbate these impacts, particularly in already water-stressed regions, underscoring the urgent need for adaptive soil and water management strategies informed by robust predictive tools [5].

1.3. Emergence of AI and Data-Driven Agriculture

The past decade has witnessed a technological transformation in the methods available for monitoring, modeling, and managing soil–water systems; see Table 1. The proliferation of low-cost Internet of Things (IoT) sensor networks, unmanned aerial vehicles (UAVs), and high-resolution satellite remote sensing platforms—including Sentinel-1/2, Landsat-8/9, MODIS, and SMAP—has resulted in an unprecedented abundance of spatially and temporally dense environmental data [6]. Simultaneously, advances in cloud computing infrastructure, edge computing, and high-performance computing have democratized access to the computational resources required to process and analyze these large, heterogeneous datasets. These twin developments have created fertile conditions for the rapid adoption of artificial intelligence (AI) and machine learning (ML) methods across the agricultural and environmental sciences.

Table 1. Evolution of modeling approaches in soil–water systems from empirical and physics-based models to modern artificial intelligence and distributed deep learning frameworks.

Era	Modeling Approach	Representative Models	Strengths	Limitations	Key Sources
1950s–1980s	Empirical and Statistical Models	Linear regression, empirical infiltration equations, water balance models	Simple implementation, low computational requirements, easy interpretation, suitable for small datasets	Limited capability for nonlinear processes, poor generalization under heterogeneous environmental conditions	[7–9]

Table 1. *Cont.*

Era	Modeling Approach	Representative Models	Strengths	Limitations	Key Sources
1980s–2000s	Physics-Based Hydrological Models	SWAT, HYDRUS, STANMOD, MODFLOW, Richards equation, Darcy-based models	Physically interpretable, process-based understanding, strong theoretical foundation, effective for hydrological simulations	Require extensive calibration, sensitive to parameter uncertainty, computationally expensive for large-scale applications	[10–13]
2000s–2015	ML Models	Support Vector Machines (SVMs), Random Forest (RF), Decision Trees, *k*-Nearest Neighbors (kNNs), Artificial Neural Networks (ANNs)	Capable of modeling nonlinear relationships, robust prediction accuracy, reduced dependence on physical assumptions	Require high-quality datasets, feature engineering is often necessary, limited interpretability in some models	[14–16]
2015–2020	Deep Learning Approaches	Convolutional neural networks (CNNs), Recurrent Neural Networks (RNNs), long short-term memory (LSTM)	Excellent spatiotemporal learning capability, automatic feature extraction, strong performance with remote sensing and time-series data	Large training datasets required, high computational cost, risk of overfitting, black-box behavior	[17–19]
2020–Present	Advanced AI and Distributed Intelligence	Transformers, federated learning, edge AI, physics-informed neural networks (PINNs), distributed deep learning	Real-time analytics, scalability, decentralized learning, privacy preservation, integration with IoT and smart farming systems	Infrastructure complexity, high energy consumption, limited explainability, challenges in standardization and deployment	[20,21]
Future Trends	Hybrid Intelligent Systems	Digital twins, explainable AI (XAI), autonomous irrigation systems, climate-integrated AI frameworks	Adaptive decision-making, improved sustainability, interpretable predictions, integration with climate resilience strategies	Requires interdisciplinary integration, lack of benchmark datasets, regulatory and ethical concerns	[22–25]

AI-driven frameworks offer several capabilities that are difficult or impossible to achieve with conventional physically-based models. Deep learning architectures—including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer-based attention models—are capable of learning highly complex, nonlinear mappings between input variables (soil properties, meteorological forcing, remote sensing signals) and target outputs (soil moisture content, evapotranspiration rates, crop yield) directly from observational data [26]. This is achieved without requiring the explicit specification of process equations, making such models particularly attractive for systems where mechanistic understanding remains incomplete or computationally intractable [27]. These models can be trained on historical datasets of an arbitrary size, updated in real time as new observations become available, and deployed across diverse agroecological contexts through transfer learning and domain adaptation techniques.

The concept of precision agriculture—tailoring agronomic inputs to the specific needs of individual field zones in space and time—has been substantially advanced by AI-based decision-support systems. Intelligent irrigation controllers that integrate real-time sensor data, numerical weather prediction outputs, and ML-based soil moisture forecasts are now being evaluated in research settings and, increasingly, in commercial deployments [28]. Such systems hold the potential to reduce agricultural water consumption substantially while maintaining or improving crop yields, contributing to the twin sustainability goals of food security and water conservation.

1.4. Distributed and Federated Deep Learning

The scale of modern agricultural data introduces both opportunities and challenges for AI-based modeling. Individual farms or monitoring stations may generate millions of sensor readings per year, while satellite platforms produce petabytes of earth observation data globally. Training deep learning models on such volumes of data at centralized facilities is increasingly impractical, both because of the cost of data transmission and storage and because of emerging regulatory constraints on the cross-border transfer of commercially sensitive agricultural data; see Table 2. Distributed deep learning—in which model training is parallelized across multiple compute nodes or edge devices—has emerged as a key technological response to these challenges [29].

Table 2. Distributed AI frameworks and agricultural applications.

Framework	Architecture Type	Key Characteristics	Agricultural Applications	Major Challenges	Key Sources
Cloud-Based Deep Learning	Centralized Distributed Computing	High computational power, scalable storage, parallel processing, large-scale analytics	Large-scale soil moisture prediction, satellite image analysis, crop monitoring	Latency, data transfer costs, dependence on internet connectivity	[30,31]
Edge AI Systems	Edge Computing	Real-time processing, low latency, localized intelligence, reduced bandwidth usage	Smart irrigation control, autonomous sensors, real-time field monitoring	Limited computational resources and energy constraints	[32,33]
Federated Learning	Decentralized Collaborative Learning	Privacy-preserving distributed training, local data retention, collaborative model optimization	Multi-farm analytics, privacy-sensitive agricultural data sharing, distributed crop modeling	Communication overhead and model synchronization complexity	[34,35]
Internet of Things (IoT)-Integrated AI	Sensor-Based Distributed Systems	Continuous environmental monitoring, real-time data acquisition, automated analytics	Precision irrigation, soil moisture sensing, climate monitoring	Sensor reliability, cybersecurity risks, heterogeneous data integration	[36–38]
Distributed Deep Learning Clusters	Parallel GPU/ Cloud Clusters	High-performance distributed training, scalable neural network optimization	Remote sensing analytics, climate prediction, large-scale hydrological simulations	High infrastructure cost and energy consumption	[39,40]
Fog Computing Frameworks	Hybrid Cloud–Edge Systems	Intermediate processing layer between cloud and edge devices, reduced latency	Real-time agricultural analytics, adaptive irrigation systems	Complex deployment architecture and maintenance challenges	[41–43]
Blockchain-Integrated AI Systems	Secure Distributed Intelligence	Secure decentralized data sharing, transparency, traceability	Agricultural supply chains, water-use monitoring, smart contracts for irrigation management	Computational overhead and scalability limitations	[44–47]
Digital Twin Platforms	Virtual Simulation Frameworks	Real-time virtual representation of agricultural systems, predictive optimization	Soil–water simulation, crop management optimization, climate adaptation	High data requirements and integration complexity	[22,23,48]
Multi-Agent AI Systems	Autonomous Distributed Agents	Collaborative intelligent agents for adaptive decision-making	Autonomous irrigation scheduling, robotic agriculture, precision resource allocation	Coordination complexity and communication overhead	[49–51]

1.5. Research Gap and Objectives

Despite the rapid growth of the literature on AI applications in agriculture, a comprehensive and integrative synthesis of the field remains lacking. Most existing reviews address either ML methods for soil property prediction, AI-based irrigation management,

or remote sensing applications in isolation, without situating them within a unified conceptual framework that encompasses distributed computation, explainable AI, uncertainty quantification, and the physical principles governing soil–water systems. Furthermore, the practical implications of privacy-preserving AI architectures and edge-deployed deep learning models for field-scale water management have received limited systematic attention.

This review addresses these gaps by providing a structured synthesis of recent developments across four interrelated domains:

(i) The physical and hydraulic fundamentals of soil–water systems;
(ii) AI and deep learning methodologies applicable to soil science and agricultural water management;
(iii) Distributed and federated learning frameworks for large-scale agricultural analytics;
(iv) Explainability, uncertainty, and ethical dimensions of AI deployment in precision agriculture.

The overarching aim is to provide researchers, practitioners, and policy makers with an integrated perspective on the current state of the art and the most promising directions for future research and development.

This manuscript is organized as follows:

(i) Section 1 is the introductory part of this manuscript.
(ii) Section 2 establishes the physical and hydraulic fundamentals of soil–water systems, covering water retention theory, unsaturated flow governed by the Richards equation, and the principal soil degradation processes relevant to agricultural water management.
(iii) Section 3 surveys the AI and deep learning methodologies applicable to soil science and agricultural water management, including classical ML algorithms, deep learning architectures, and physics-informed hybrid models.
(iv) Section 4 reviews the data sources underpinning AI-based soil–water modeling, encompassing ground-based sensor networks, satellite and UAV remote sensing platforms, and data fusion challenges.
(v) Section 5 presents applications of AI in soil–water interactions, covering soil moisture prediction, smart irrigation, digital soil mapping, pedotransfer functions, and crop yield forecasting.
(vi) Section 6 provides a comparative analysis of model performance and discusses generalization and transferability across environments.
(vii) Section 7 addresses explainability, uncertainty quantification, and the ethical and socioeconomic dimensions of AI deployment in precision agriculture.
(viii) Section 8 critically examines current challenges and limitations.
(ix) Section 9 delineates prospective research directions and technical horizons.
(x) Section 10 presents the conclusions.

1.6. Review Methodology and PRISMA Compliance

This article is a comprehensive narrative review of the application of artificial intelligence and distributed deep learning to soil–water systems in precision agriculture. It constitutes a systematic meta-analysis where the review methodology is aligned with the PRISMA 2020 guidelines [52] to the extent applicable to comprehensive narrative reviews, in order to ensure transparency, reproducibility of the literature search, and rigorous reporting of the selection process. A PRISMA-adapted flow diagram is provided in Figure 2.

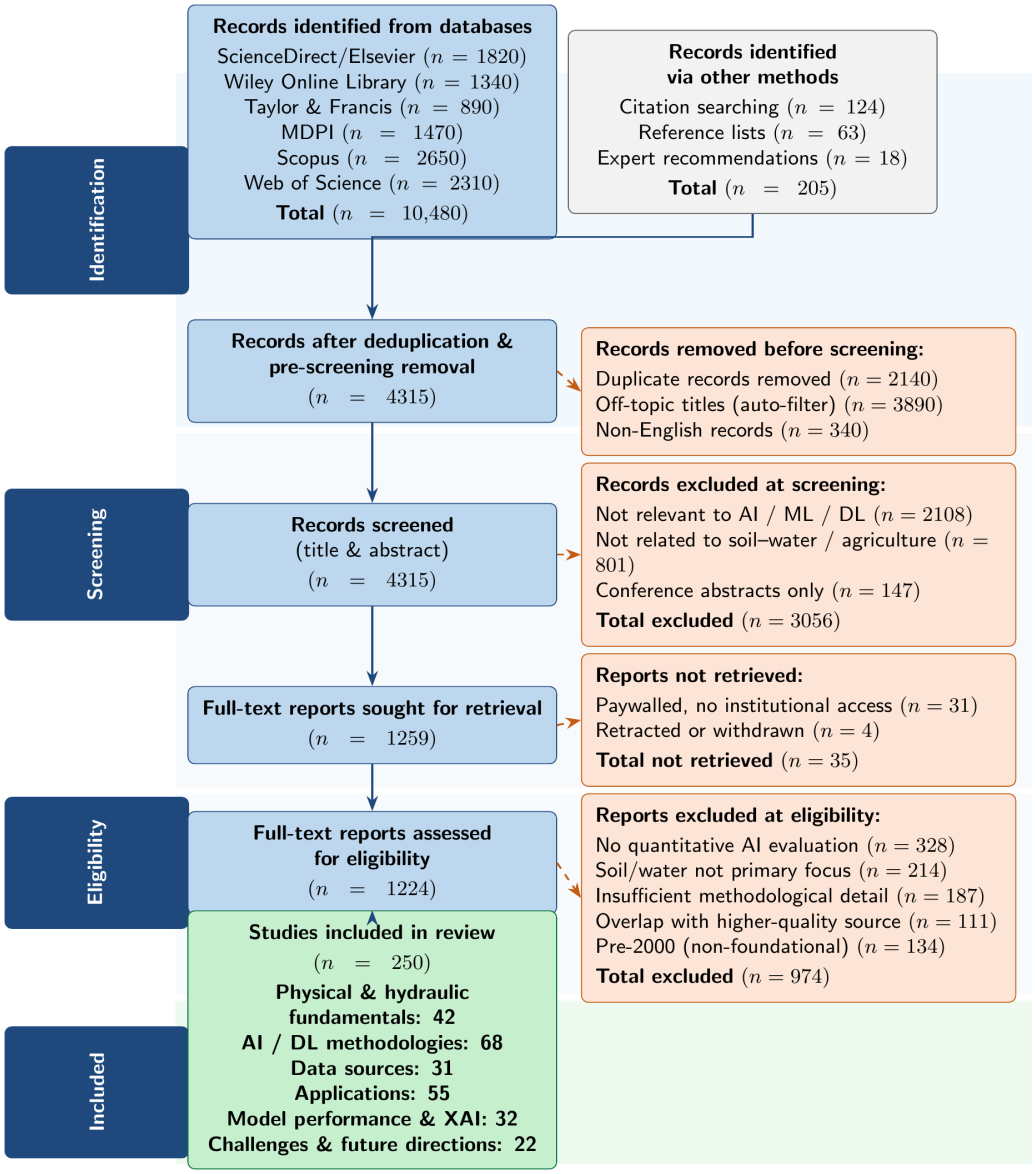

Figure 2. PRISMA 2020 flow diagram (adapted for a comprehensive narrative review) documenting the literature identification, screening, eligibility assessment, and inclusion process. Adapted from: ref. [53]. Licensed under CC BY 4.0. Software used for plotting diagram: LaTeX tikz package. Diagram source: author.

Databases and sources. The primary literature search was conducted across seven major peer-reviewed academic publishers: ScienceDirect (Elsevier), Wiley Online Library, Springer Nature, MDPI Open Access Journals, Taylor & Francis Online, PLOS (Public Library of Science), and Frontiers Media SA. These sources were further cross-referenced through the Scopus and Web of Science (WoS) citation indexing databases to ensure comprehensive coverage and systematic deduplication of retrieved records. These platforms were selected to provide comprehensive coverage of the soil science, hydrology, precision agriculture, and computer science literature. Supplementary records were identified through citation searching of key review papers, examination of reference lists of retrieved articles, and expert recommendations within the author's research network.

Search terms. Search queries were constructed by combining domain-specific thematic terms—"soil moisture", "soil hydraulic properties", "vadose zone", "pedotransfer functions", "irrigation scheduling", "precision agriculture", "digital soil mapping", "soil water content", and "evapotranspiration"—with methodological and computational

terms—"machine learning", "deep learning", "convolutional neural network", "long short-term memory", "transformer", "federated learning", "physics-informed neural network", "explainable AI", "uncertainty quantification", "distributed learning", and "edge computing". Boolean AND/OR operators were used to construct structured queries within each database interface.

Temporal scope. The search covered publications from January 2000 to May 2026. Priority was assigned to publications from 2015 onwards, reflecting the emergence of modern deep learning methods and their adoption in soil and agricultural sciences. Foundational pre-2015 works establishing theoretical frameworks (Richards equation, van Genuchten model, HYDRUS, MODFLOW, classical ML algorithms) were included, irrespective of the publication date, given their continued relevance to the field.

Inclusion criteria. Records were included if they: (i) were published in peer-reviewed journals or peer-reviewed conference proceedings; (ii) addressed at least one of the four thematic domains defined in the review objectives; (iii) were available in English; and (iv) reported sufficient methodological detail to allow assessment of the AI approach and its application context. The gray literature, preprints, and non-peer-reviewed reports were excluded unless they represented canonical technical documentation with no peer-reviewed equivalent (e.g., USGS technical manuals, FAO technical reports, IPCC assessment reports).

Screening and selection process. An initial pool of 10,480 database records was identified, supplemented by 205 records from other sources. After deduplication and automated pre-screening for relevance (removing 6370 records), 4315 records were retained for title and abstract screening. Full-text retrieval was sought for 1259 records that passed the screening stage; 35 were not retrievable. Of the 1224 full-text reports assessed for eligibility, 974 were excluded for the reasons detailed in Figure 2. The final reference corpus comprises 250 sources, grouped thematically into the section structure of this review. The complete selection process is documented in the PRISMA-adapted flow diagram below.

2. Fundamentals of Soil–Water Systems in Agriculture

2.1. Soil Physical and Hydraulic Properties

A thorough understanding of soil physical and hydraulic properties is foundational to any quantitative treatment of soil–water dynamics and forms an essential prerequisite for the effective application of AI-based predictive models. The key physical descriptors of a soil include its particle-size distribution (texture), aggregate structure, bulk density (ρ_b), total porosity (ϕ), and organic matter content; see Table 3. Together, these properties determine the geometry of the pore network through which water and solutes move and within which microbial communities and plant roots develop [54–57].

Table 3. Major soil hydraulic properties and their agricultural significance.

Hydraulic Property	Definition	Agricultural Significance	Factors Affecting Soil Property	Key Sources
Porosity	Fraction of soil volume occupied by pores or voids	Controls water storage capacity, aeration, root penetration, and microbial activity	Soil texture, organic matter, compaction, aggregation, tillage practices	[58–60]
Bulk Density	Mass of dry soil per unit bulk volume	Influences root growth, infiltration, and water retention; high bulk density restricts crop productivity	Soil compaction, machinery traffic, soil texture, moisture content	[61–63]

Table 3. *Cont.*

Hydraulic Property	Definition	Agricultural Significance	Factors Affecting Soil Property	Key Sources
Field Capacity	Amount of water retained in soil after gravitational drainage	Represents the upper limit of plant-available water and determines irrigation scheduling	Texture, structure, organic matter, pore-size distribution	[64–66]
Permanent Wilting Point	Soil moisture level below which plants cannot extract water	Defines the lower threshold of plant-available water and drought stress conditions	Clay content, salinity, organic matter, soil mineralogy	[67–69]
Available Water Capacity	Difference between field capacity and wilting point	Determines the quantity of water accessible for crop uptake and irrigation efficiency	Texture, rooting depth, soil structure, organic matter	[64,70–72]
Hydraulic Conductivity	Ability of soil to transmit water under saturated or unsaturated conditions	Controls infiltration, drainage, groundwater recharge, and irrigation performance	Texture, pore connectivity, compaction, soil moisture conditions	[4,73,74]
Infiltration Rate	Rate at which water enters the soil surface	Affects runoff generation, irrigation uniformity, erosion risk, and water conservation	Surface crusting, vegetation cover, soil texture, tillage practices	[7,8,61,75]
Soil Water Retention	Capacity of soil to retain water against gravitational forces	Critical for drought resistance, irrigation planning, and crop water availability	Clay minerals, organic matter, pore-size distribution	[4,67,76,77]
Matric Potential	Energy status of water held within soil pores due to capillary and adsorptive forces	Influences plant water uptake, unsaturated flow, and root-zone dynamics	Soil texture, salinity, moisture content, pore geometry	[13,67,78]
Capillary Rise	Upward movement of water through soil pores due to capillary forces	Supports crop water availability in shallow groundwater regions and arid environments	Pore-size distribution, groundwater depth, soil texture	[79–82]
Water-Holding Capacity	Maximum amount of water retained by soil after saturation and drainage	Determines irrigation frequency, crop resilience, and water-use efficiency	Organic matter, texture, aggregation, compaction	[58,83–85]
Tortuosity	Ratio of the actual path length of solute movement through soil pores to the straight-line distance; describes the sinuosity of the pore network	Controls effective diffusion and dispersion of solutes, nutrients, and contaminants in the vadose zone; critical for predicting leaching risk	Pore geometry, soil structure, water content, clay content	[86]
Pore-Water Velocity	Average linear velocity of water (and dissolved solutes) through the soil pore space; equals Darcy flux divided by volumetric water content	Governs the rate of solute transport, nutrient leaching, and pesticide movement toward groundwater	Hydraulic conductivity, matric potential gradient, water content, pore-size distribution	[78]
Dispersion Coefficient	Parameter quantifying the spreading of a solute plume due to mechanical dispersion and molecular diffusion during transport through the soil	Determines the spatial extent of contaminant plumes, nutrient front migration, and mixing in the vadose zone; essential for solute transport modeling	Pore-water velocity, dispersivity, diffusion coefficient, soil texture and structure	[86]

Soil texture—the relative proportions of sand, silt, and clay fractions—exerts a primary control on hydraulic behavior; see Figure 3. Sandy soils are characterized by large interparticle pores, rapid drainage, low water-holding capacity, and high saturated hydraulic conductivity (K_s). Clay-rich soils, by contrast, possess predominantly micropores and mesopores that confer high water retention and low K_s but are also prone to shrinkage-induced cracking upon desiccation and to surface sealing under rainfall impact. Loamy soils, occupying an intermediate textural position, generally offer the most favorable combination of water retention and aeration for crop production. The United States Department of Agriculture (USDA) and the Food and Agriculture Organization (FAO) textural classification schemes provide internationally standardized frameworks for communicating soil texture, and texture-based pedotransfer functions (PTFs) are routinely employed to estimate hydraulic parameters from routinely measured soil properties [87–89].

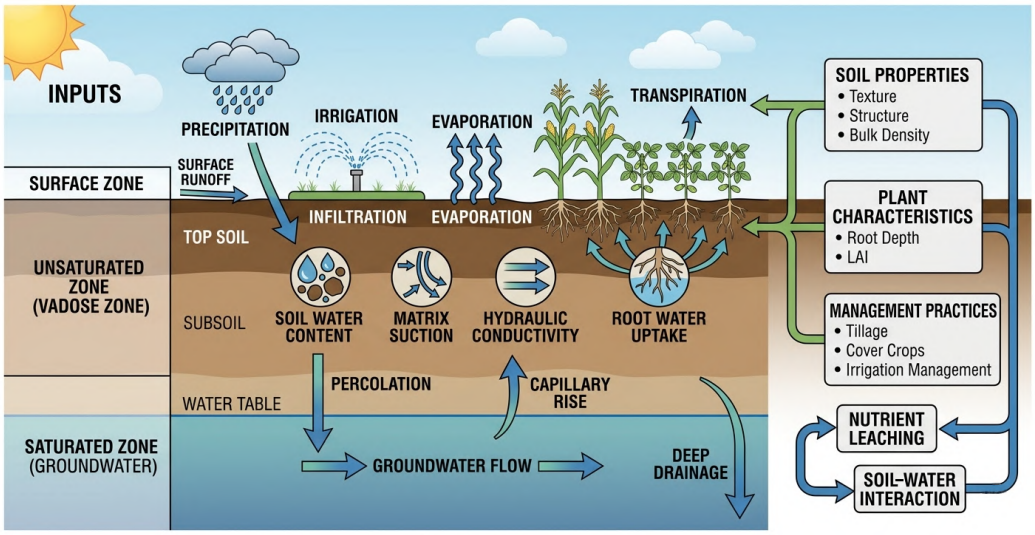

Figure 3. Conceptual diagram scheme of soil–water interaction processes in agricultural systems. Abbreviations: LAI = Leaf Area Index. Software: Inkscape v. 1.4.4 for vector elements; GIMP v. 3.2.4 for raster elements; OpenClipart (openclipart.org) for free SVG clipart of plants, clouds, and the sun. Infographic source: author.

Several hydraulic properties are of particular practical importance for irrigation management. The field capacity (θ_{FC}) is the volumetric water content retained by a well-drained soil approximately 24–48 h after saturation, typically corresponding to a matric potential of -33 kPa for most medium-textured soils. The permanent wilting point (θ_{WP}) is the water content at which the matric potential falls to approximately -1500 kPa, beyond which most crops can no longer extract water from the soil against the combined resistance of soil and root hydraulic pathways. The plant-available water capacity (PAWC), defined as $\theta_{FC} - \theta_{WP}$, quantifies the reservoir of soil water accessible for crop growth between irrigation events and is a critical parameter in irrigation scheduling models.

Hydraulic conductivity describes the ease with which water moves through the soil under a given hydraulic gradient. Its dependence on water content is described by the unsaturated hydraulic conductivity function $K(\theta)$ or, equivalently, $K(h)$, where h is the soil matric potential. The van Genuchten–Mualem model [4] provides a widely used analytical description. The water retention function is first defined as:

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + |\alpha h|^n]^m}, \quad m = 1 - \frac{1}{n}, \tag{1}$$

where $\theta(h)$ is the volumetric water content at matric potential h, θ_r and θ_s are the residual and saturated water contents, α (m^{-1}) is the inverse of the air-entry suction, n is a pore-size distribution index, and $m = 1 - 1/n$, following the Mualem constraint [4]. The unsaturated hydraulic conductivity function is then expressed as:

$$K(h) = K_s S_e^l \left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2, \quad (2)$$

where $S_e = (\theta - \theta_r)/(\theta_s - \theta_r)$ is the effective saturation and l is a pore-connectivity parameter commonly set to 0.5. This formulation is consistent with Equation (13) in Šimůnek et al. [86]. Estimating these parameters for field conditions is a significant practical challenge that AI-based PTFs increasingly seek to address.

2.2. Water Movement in Saturated and Unsaturated Soils

Water movement through saturated soils is governed by Darcy's law, which relates the volumetric flux density q ($m\ s^{-1}$) to the hydraulic gradient:

$$q = -K_s \frac{dH}{dl}, \quad (3)$$

where H is the total hydraulic head (sum of matric and gravitational heads) and l is the flow path length. Darcy's law provides an adequate description of slow, laminar flow through the saturated zone and forms the basis of widely used groundwater models, such as MODFLOW. In the unsaturated zone, which is of primary relevance for crop water uptake and irrigation management, flow is described by the Richards equation, obtained by combining Darcy's law with the continuity equation:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z} + 1\right)\right] - S, \quad (4)$$

where θ is the volumetric water content, t is time, z is the vertical coordinate (positive upward), h is the matric potential head, $K(h)$ is the unsaturated hydraulic conductivity, and S is a sink term representing root water uptake [13,90]. The Richards equation is highly nonlinear—because both θ and K are strongly nonlinear functions of h—and its numerical solution is computationally demanding, particularly when the spatial domain is large, and the soil profile is heterogeneous. This computational burden has been a significant driver of interest in AI-based surrogate models that can approximate Richards equation solutions at a fraction of the computational cost.

Soil water redistribution following rainfall or irrigation events involves a succession of processes: infiltration into the surface layer, downward redistribution by gravity and capillary gradients, lateral preferential flow through macropores and biological channels, redistribution driven by osmotic gradients in saline soils, and capillary rise from shallow water tables. Each of these processes operates at characteristic time and length scales, and their interaction creates spatially and temporally complex patterns of soil moisture distribution that challenge both field measurement and numerical simulation.

2.3. Soil Degradation and Water Stress

Soil degradation constitutes one of the most significant threats to the long-term sustainability of agricultural water use [91]. Globally, it is estimated that more than one-third of agricultural land is moderately to severely degraded, with salinization, erosion, compaction, and loss of soil organic carbon (SOC) representing the dominant degradation pathways [92]. Each of these processes exerts distinct effects on soil–water dynamics and, consequently, on crop water availability and irrigation requirements.

Soil salinization—arising from irrigation with saline water, seawater intrusion into coastal aquifers, or evaporative concentration of dissolved salts in poorly drained soils—reduces plant-available water by increasing osmotic stress and suppressing root water uptake. It has been estimated that salinity-affected soils account for approximately 20% of irrigated agricultural land worldwide and that this proportion is increasing [93]. Soil compaction, caused by heavy machinery traffic and tillage, reduces macroporosity and increases soil strength, thereby impeding root penetration, restricting drainage, and increasing surface runoff [94]. Erosion by water removes fertile topsoil and reduces the depth of the rooting zone, diminishing the soil's water storage capacity.

Water stress exerts a profound influence on crop physiology, suppressing leaf expansion, stomatal conductance, photosynthesis, and ultimately yield formation. The relationship between soil water deficit and crop response is captured by crop-specific stress functions implemented in widely used crop growth models, such as AquaCrop, DSSAT, and APSIM. AI-based models have increasingly been used to learn these stress response functions empirically from observed yield data, bypassing the need for manually specified functional forms [95].

2.4. Conventional Soil–Water Modeling Approaches

Physically-based hydrological and soil–water models have underpinned agricultural water management research for several decades. Widely used frameworks include: the Soil and Water Assessment Tool (SWAT) for watershed-scale water balance and non-point source pollution modeling [10,96]; HYDRUS-1D/2D/3D for variably saturated flow and solute transport in the vadose zone [86]; MODFLOW for regional groundwater flow simulation; and DSSAT and AquaCrop for coupled crop–water productivity modeling. These tools have contributed enormously to the understanding of soil–water processes and to the design of irrigation systems and water conservation strategies.

Nevertheless, physically-based models are subject to several important limitations that motivate the complementary use of AI approaches; see Figure 4.

Empirical and statistical modeling approaches—including multiple linear regression, generalized additive models, and geostatistical methods, such as kriging—offer lower computational cost and require fewer parameters, but their predictive accuracy is limited in strongly nonlinear systems and their extrapolation behavior outside the range of calibration data is unreliable. These limitations have collectively catalyzed the rapid growth of ML and deep learning as complementary or alternative tools for soil–water modeling. The limitations of physically-based hydrological models include extensive parameterization requirements, restricted spatial resolution, potential inaccuracies arising from structural assumptions under complex field conditions, and high computational demands that hinder real-time operational applications:

- First, physically-based models require extensive parameterization. SWAT, for example, may involve hundreds of spatially distributed parameters describing soil properties, land use, and channel hydraulics, many of which cannot be directly measured and must be inferred through computationally intensive calibration procedures.
- Second, the spatial resolution of these models is constrained by the availability of input data and computational resources. Representing sub-field heterogeneity in regional or global applications is generally infeasible.
- Third, the model structural assumptions (e.g., the Richards equation and the Green–Ampt infiltration model) may not hold under all conditions encountered in practice, particularly in degraded soils with complex pore geometries or in the presence of non-equilibrium flow phenomena.

- Fourth, the computational cost of running physically-based models at scales relevant to real-time operational management is prohibitive without significant simplification.

Hybrid model architectures, which integrate the process knowledge embedded in physically-based models with the pattern-recognition capabilities of ML, are increasingly recognized as a promising path forward. Physics-informed neural networks (PINNs), for example, incorporate differential equation constraints directly into the neural network loss function, ensuring that the learned solution remains consistent with governing physical laws even in data-sparse regions [97]. Similarly, ML models serve to replace computationally expensive process model components (e.g., the solution of the Richards equation) with fast, accurate emulators that can be integrated into real-time decision systems.

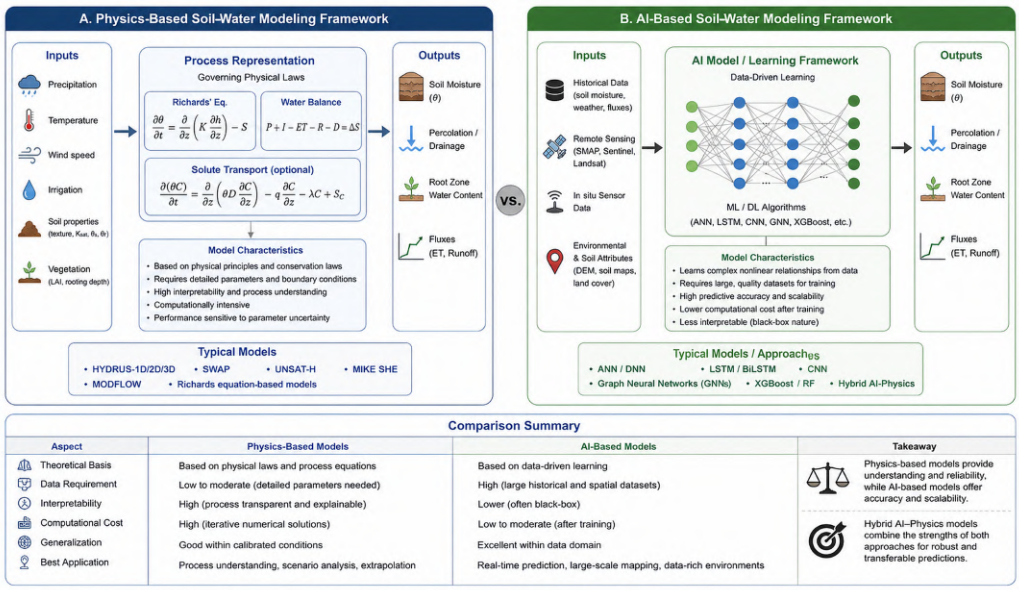

Figure 4. Comparison between physics-based and AI-based soil–water modeling frameworks. Abbreviations in panel equations: $\partial\theta/\partial t$ = rate of change in volumetric water content; K = unsaturated hydraulic conductivity; h = matric potential head; z = vertical coordinate; S = root water uptake sink term; C = solute concentration; D = hydrodynamic dispersion coefficient; q = pore-water flux density (Darcy velocity); λ = first-order degradation rate; S_C = solute source/sink term; and P, I, ET, R, D, ΔS = precipitation, irrigation, evapotranspiration, runoff, drainage, and storage change in the water balance. Software used: LaTeX version 2e 2025-11-01 for equations and TexText plugin for rendering LaTeX equations; Inkscape v. 1.4.4 for vector elements; TikZ package version 3.1.11a for the neural network diagram (panel B); Material Design Icons for symbols. Flowchart source: author.

3. Artificial Intelligence in Soil and Water Sciences

3.1. Overview of AI Techniques and Conceptual Framework

Artificial intelligence encompasses a broad spectrum of computational approaches that enable machines to perform tasks that would conventionally require human expertise: pattern recognition, prediction, classification, optimization, and adaptive decision-making [98]. Within the context of soil and water sciences, AI methods can be broadly categorized into classical ML algorithms (which rely primarily on handcrafted feature engineering) [99], deep learning architectures (which learn hierarchical feature representations directly from raw data) [100], and hybrid or physics-guided models that integrate AI with domain knowledge [101]. Each category offers distinct advantages and is suited to different problem types, data modalities, and application scales.

A unifying characteristic of all AI approaches is their reliance on historical data for model training: the model parameters (weights, split thresholds, kernel coefficients, etc.)

are adjusted to minimize a loss function measuring the discrepancy between model predictions and observed target variables. This data-driven paradigm contrasts with the a priori specification of process equations in physically-based models, and its predictive accuracy depends critically on the volume, quality, and representativeness of the training data for classification [102]. In domains such as soil science, where training data may be sparse, geographically biased, or collected with heterogeneous measurement protocols, this dependence introduces significant challenges that have motivated the development of transfer learning, data augmentation, and physics-informed regularization strategies.

The taxonomy of AI applications in soil and water sciences is increasingly sophisticated; see Figure 5. At the input end, AI models ingest data from a diverse range of sources: ground-based sensors measuring soil moisture, electrical conductivity, temperature, and matric potential; meteorological stations providing precipitation, air temperature, humidity, wind speed, and solar radiation records; multispectral, hyperspectral, and synthetic aperture radar (SAR) satellite imagery; UAV-based RGB, multispectral, and thermal imagery; and agronomic records of crop variety, sowing date, fertilizer application, and yield.

The heterogeneity of these data sources—in spatial and temporal resolution, measurement precision, and data format—poses significant pre-processing and data fusion challenges that must be addressed before AI models can be effectively trained and deployed.

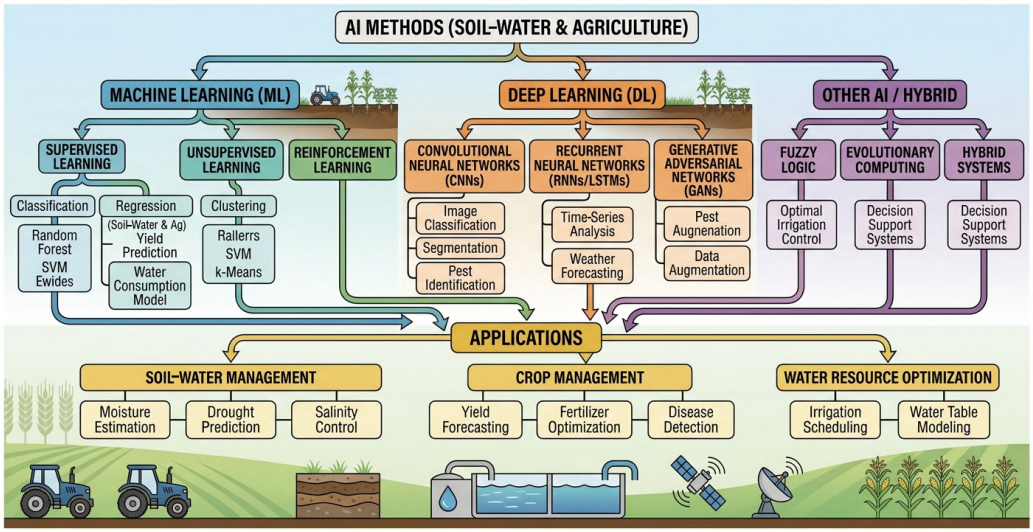

Figure 5. Taxonomy of AI methods applied in soil–water and agricultural systems. Software used: Inkscape v. 1.4.4 for vector elements; GIMP v. 3.2.4 for raster elements and gradient fill; and Vecteezy free tier for schematic icons. Flowchart source: author.

3.2. Classical ML Algorithms

Table 4 provides a comprehensive comparative analysis of various ML algorithms utilized in soil–water modeling, evaluating their core architectures, data requirements, and predictive performance across diverse agroclimatic conditions.

Table 4. Comparison of ML algorithms in soil–water modeling.

Algorithm	Category	Strengths	Limitations	Typical Applications	Key Citations
Linear Regression	Supervised Learning	Simple implementation, computational efficiency, high interpretability	Limited ability to capture nonlinear relationships and complex interactions	Water balance estimation, evapotranspiration prediction, trend analysis	[103–107]
Decision Trees	Supervised Learning	Easy interpretation, handles nonlinear relationships, low pre-processing requirements	Prone to overfitting and instability with noisy datasets	Soil classification, irrigation decision support, land suitability assessment	[108–110]

Table 4. *Cont.*

Algorithm	Category	Strengths	Limitations	Typical Applications	Key Citations
Random Forest (RF)	Ensemble Learning	High predictive accuracy, robust against overfitting, effective for high-dimensional datasets	Reduced interpretability and increased computational cost for large datasets	Digital soil mapping, soil moisture prediction, crop yield forecasting	[14,111–114]
Support Vector Machines (SVMs)	Supervised Learning	Strong generalization capability, effective with limited datasets, suitable for nonlinear modeling	Sensitive to parameter selection and kernel choice, computationally intensive for large datasets	Salinity prediction, irrigation scheduling, soil property estimation	[115–118]
k-Nearest Neighbors (kNNs)	Instance-Based Learning	Simple implementation, nonparametric nature, adaptable to local data patterns	Sensitive to noise and data scaling, computationally expensive for large datasets	Soil texture classification, local hydrological prediction	[16,119–122]
Artificial Neural Networks (ANNs)	Deep Learning	Excellent nonlinear modeling capability, adaptive learning, robust predictive performance	Requires large datasets and extensive training, limited interpretability	Hydraulic conductivity estimation, rainfall-runoff modeling, evapotranspiration forecasting	[15,123–126]
Convolutional Neural Networks (CNNs)	Deep Learning	Strong spatial feature extraction capability, highly effective for image analysis and remote sensing	High computational requirements and large data dependency	Remote sensing analysis, digital soil mapping, crop water stress detection	[27,127–131]
Recurrent Neural Networks (RNNs)	Deep Learning	Efficient sequence and temporal dependency modeling	Gradient vanishing problems and limited long-term memory retention	Time-series soil moisture prediction, hydrological forecasting	[132–136]
Long Short-Term Memory (LSTM)	Deep Learning	Excellent long-term temporal learning capability, effective for sequential environmental datasets	Computationally intensive and complex hyperparameter tuning	Soil moisture forecasting, rainfall prediction, irrigation demand estimation	[17,137–140]
Gradient Boosting Machines (GBMs)	Ensemble Learning	High predictive accuracy, efficient handling of heterogeneous datasets	Sensitive to parameter tuning and potential overfitting	Crop yield prediction, groundwater quality assessment	[141–144]
Extreme Gradient Boosting (XGBoost)	Ensemble Learning	Fast training, high scalability, superior predictive performance	Complex optimization and reduced interpretability	Precision agriculture, smart irrigation systems, hydrological modeling	[145–148]
Transformer Models	Deep Learning	Superior spatiotemporal learning capability, parallel processing efficiency, scalable architectures	Very high computational demand and large training data requirements	Climate-smart agriculture, large-scale hydrological forecasting, environmental analytics	[149–151]

3.2.1. Support Vector Machines

Support Vector Machines (SVMs) represent a class of kernel-based supervised learning algorithms originally developed by Cortes and Vapnik [152] for binary classification and subsequently extended to regression tasks (support vector regression, SVR). The core principle of SVMs is to identify the hyperplane in a high-dimensional feature space that maximizes the margin between classes (for classification) or that contains the largest proportion of training points within a specified ε-tube around the regression surface (for SVR).

For binary classification, the optimal separating hyperplane is obtained by solving:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \tag{5}$$

where $\mathbf{w}$ is the weight vector, b is the bias term, and y_i denotes the class labels.

The decision function of a kernel-based SVM is expressed as:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b, \tag{6}$$

where α_i are Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function.

The ability of SVMs to operate effectively in high-dimensional feature spaces—through the kernel trick, which implicitly maps inputs to high-dimensional representations without explicit computation—makes them particularly well suited to problems involving spectral remote sensing data, where the number of spectral bands may greatly exceed the number of training samples.

In soil science, SVMs have been successfully applied to soil moisture prediction from meteorological and remote sensing inputs, classification of soil types from multispectral imagery, estimation of soil organic carbon content from hyperspectral reflectance spectra, and prediction of soil salinity distribution from electrical conductivity survey data [153,154]. SVMs are attractive for these applications because their regularization framework mitigates overfitting even when training sample sizes are limited—a chronic problem in soil survey databases—and because their kernel hyperparameters can be efficiently optimized using cross-validation. However, SVMs scale poorly with the training set size (their computational complexity is at least $\mathcal{O}(n^2)$ in the number of training samples), limiting their application to large national or global soil databases without approximation methods.

3.2.2. Random Forests

Random Forest (RF) is an ensemble learning algorithm based on the aggregation of multiple Decision Trees trained on bootstrap samples of the training data and evaluated on randomly selected subsets of predictor variables at each node split [14,155]. By averaging the predictions of a large number of individually weak learners, RF achieves substantially lower variance and higher generalization accuracy than individual Decision Trees, while retaining their ability to model complex nonlinear interactions and to handle mixed-type predictor variables without pre-processing.

The RF prediction is computed as the average of predictions from all trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x}), \tag{7}$$

where T is the number of trees and $h_t(\mathbf{x})$ is the prediction of the tth tree.

Feature importance in RF can be estimated using permutation importance:

$$FI_j = \frac{1}{T} \sum_{t=1}^{T} \left(\mathrm{Err}_{t,j}^{\mathrm{perm}} - \mathrm{Err}_t \right), \tag{8}$$

where $\mathrm{Err}_{t,j}^{\mathrm{perm}}$ denotes the prediction error after permuting feature j.

RF has become one of the most widely used algorithms in digital soil mapping (DSM), driven by its strong empirical performance, its built-in feature importance metric (based on the mean decrease in impurity or permutation accuracy), and its robustness to irrelevant predictor variables and moderate amounts of missing data. In the global DSM project SoilGrids, RF models were trained on a global compilation of soil profile observations (over 150,000 profiles) combined with a comprehensive stack of environmental covariates (climate, terrain, vegetation indices, parent material, land use) derived from remote sensing and global datasets, producing predictions of soil properties at 250 m resolution for all soil layers down to 2 m depth [111,156]. Similar approaches have been applied at national and

regional scales to map soil organic carbon, pH, clay content, bulk density, and available water capacity with high spatial resolution [157,158].

Beyond DSM, RF has been applied to: soil moisture prediction from combinations of in situ sensor data, satellite-derived land surface temperature and vegetation indices, and meteorological variables; the estimation of soil hydraulic conductivity from texture, bulk density, and organic matter measurements; and the identification of soil contamination hotspots from geochemical survey data. The interpretability of RF through variable importance analysis is an important advantage in scientific applications where understanding the drivers of model predictions is as important as the predictions themselves.

3.2.3. Gradient Boosting Methods and XGBoost

Gradient boosting is an ensemble technique in which learners are added sequentially, with each new learner fitting the residuals of the ensemble assembled so far [142,159]. Unlike RF, which builds trees independently in parallel, gradient boosting constructs trees in a stage-wise fashion, explicitly minimizing a differentiable loss function.

The additive boosting model is expressed as:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \tag{9}$$

where $F_m(x)$ is the boosted model at iteration m, $h_m(x)$ is the newly added weak learner, and γ_m is the learning rate.

In XGBoost, the objective function combines the training loss and regularization term:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{10}$$

where $l(y_i, \hat{y}_i)$ is the loss function and $\Omega(f_k)$ controls model complexity.

The extreme gradient boosting algorithm (XGBoost) [145,160] and its successors LightGBM and CatBoost have emerged as the state of the art among tabular ML algorithms, consistently achieving top performance in data science competitions across a wide range of prediction tasks.

In soil and water science applications, XGBoost and gradient boosting have demonstrated superior performance compared with RF and SVMs in several benchmarking studies on soil hydraulic property prediction, irrigation demand forecasting, and soil nutrient mapping [111,161,162]. Key advantages include the ability to handle missing predictor values natively, efficient implementation with parallel computation, and built-in regularization terms that control model complexity. Gradient boosting methods are increasingly combined with SHAP (SHapley Additive exPlanations) values [163] to provide feature attribution scores that quantify the contribution of each input variable to individual predictions, substantially improving model interpretability without sacrificing predictive accuracy [162,164].

A critical comparative assessment of classical ML algorithms reveals that their relative performance is highly conditioned on the nature of the prediction task, the volume and quality of training data, and the soil–climatic context of deployment. Random Forest consistently demonstrates the strongest generalization performance for digital soil mapping and soil property prediction tasks, particularly when training datasets are moderately sized (hundreds to low thousands of observations) and predictor sets include strongly correlated remote sensing covariates, owing to its inherent feature bagging mechanism that mitigates multicollinearity effects [111].

XGBoost and gradient boosting variants outperform RF in dynamic process prediction tasks—such as irrigation demand forecasting and evapotranspiration estimation—where

the sequential residual-fitting mechanism provides superior capture of nonlinear temporal dependencies in tabular meteorological records. SVMs offer a statistically principled framework for small-sample problems (fewer than approximately 500 labeled observations) through structural risk minimization, but their $\mathcal{O}(n^2) - \mathcal{O}(n^3)$ training complexity renders them computationally prohibitive for continental-scale soil databases. A key limitation shared across all classical ML approaches—and often understated in the literature—is their inability to provide physically consistent predictions outside the training distribution: they do not encode the conservation laws or differential equation constraints that physically-based models enforce, making their extrapolative reliability under novel climate-forcing conditions fundamentally uncertain without physics-informed regularization [165].

3.3. Deep Learning Architectures

Representative deep learning architectures for spatiotemporal soil moisture prediction are compared in terms of their approaches to spatial feature learning and temporal dependency modeling. Differences among CNN–LSTM, ConvLSTM, and graph-based temporal frameworks in data representation, network connectivity, and prediction mechanisms are illustrated; see Figure 6.

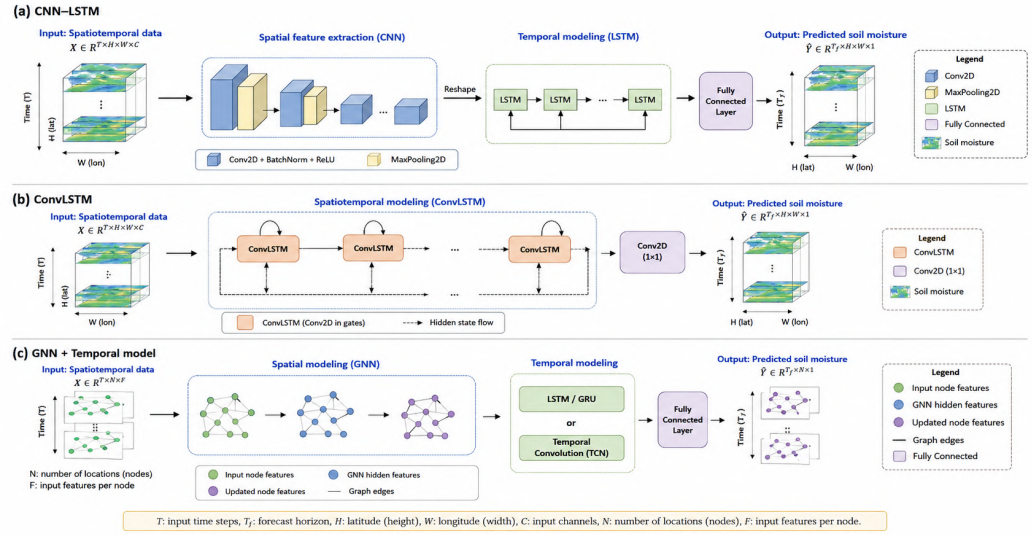

Figure 6. Deep learning architectures for spatiotemporal soil moisture prediction. Software used: Python's Matplotlib (version 3.8.3) library (*mpl_toolkits.mplot3d*) toolkit for 3D plotting; draw.io (diagrams.net) for plotting LSTM chain diagram. Node–edge graphs with colored nodes were plotted using Python version 3.14, NetworkX version 3.6.1 and Matplotlib (version 3.8.3) . Mathematical notations were rendered in LaTeX verison 2e 2025-11-01. Final assembly of three panels in layout: Inkscape v. 1.4.4. Flowchart source: author.

3.3.1. Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are the dominant deep learning architecture for tasks involving spatially structured data, particularly images [166]. CNNs exploit the spatial locality of features in images through the use of local receptive fields, weight sharing across spatial positions, and hierarchical feature extraction through successive convolutional, normalization, and pooling operations. These design principles confer powerful translation invariance and computational efficiency relative to fully connected networks, enabling CNNs to learn rich spatial representations from large image datasets.

In the soil and water sciences, CNNs have found widespread application in the analysis of satellite and UAV imagery for soil classification, soil moisture estimation, crop stress detection, and irrigation mapping. SAR-based soil moisture retrieval has been substantially

improved by deep CNN architectures that exploit the spatial context of radar backscatter signals and can account for topographic, vegetation, and surface roughness effects that confound classical retrieval algorithms [167,168]. Hyperspectral CNNs—designed to process image cubes with hundreds of spectral channels—have achieved state-of-the-art accuracy in soil organic carbon mapping from airborne and proximal sensing data by simultaneously exploiting spectral and spatial information. Transfer learning from large-scale pretrained CNN models (e.g., those trained on ImageNet) to domain-specific soil science tasks has been shown to dramatically reduce the amount of labeled training data required for fine-tuning, facilitating the development of high-accuracy models even in data-scarce agricultural regions.

3.3.2. Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNNs)

Many critical soil–water processes are inherently temporal in character: soil moisture evolves in response to sequences of rainfall, evapotranspiration, and irrigation events; streamflow is generated by the cumulative effect of antecedent soil moisture conditions and precipitation over preceding hours to days; crop growth and water demand follow phenological trajectories that unfold over the growing season. Modeling these temporal dependencies requires architectures capable of representing dynamic system state over extended time horizons.

Recurrent Neural Networks (RNNs) address this requirement by maintaining a hidden state vector that is updated at each time step as a function of the current input and the previous hidden state. However, standard RNNs are notoriously difficult to train on long sequences due to the vanishing and exploding gradient problem—the tendency for backpropagated gradients to diminish or amplify exponentially through many time steps [169]. Long short-term memory (LSTM) networks [137,170] overcome this limitation through the introduction of a gated memory cell architecture comprising input, forget, and output gates that selectively regulate the flow of information into, within, and out of the cell. This architecture enables LSTMs to learn long-range dependencies spanning hundreds to thousands of time steps, making them exceptionally well suited to hydrological and soil moisture time-series modeling.

LSTM networks have achieved state-of-the-art performance in rainfall-runoff modeling, outperforming conceptual hydrological models calibrated on thousands of catchments in large-sample benchmarking studies [17,171]. In soil moisture prediction, LSTMs trained on combinations of meteorological forcing and in situ measurements have demonstrated superior temporal generalization relative to shallow ML models, capturing soil moisture dynamics across wet and dry periods, freeze–thaw cycles, and inter-annual variability. Bidirectional LSTM (BiLSTM) architectures, which process sequences in both forward and reverse directions, and encoder–decoder LSTMs for multi-step ahead forecasting, represent further refinements that extend the applicability of recurrent architectures to complex temporal prediction tasks in precision agriculture.

3.3.3. Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational systems inspired by the organization of the biological neural network, comprising interconnected processing units (neurons) arranged in layers and trained by backpropagating gradients of a loss function through the network [172]. In their shallow multi-layer perceptron (MLP) form—with one or two hidden layers—ANNs have been applied to soil science problems since the early 1990s, initially for the estimation of soil hydraulic properties from texture and bulk density [173] and subsequently for a wide range of prediction tasks, including soil moisture content, evapotranspiration, crop water demand, and water quality indices.

The output of a neuron in an ANN is computed as:

$$a_j = \phi\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right), \tag{11}$$

where x_i are the input features, w_{ij} are the connection weights, b_j is the bias term, and $\phi(\cdot)$ is the activation function (e.g., sigmoid, ReLU, or Leaky ReLU).

The training objective is typically formulated through the minimization of a loss function using backpropagation and gradient descent:

$$\theta^{(t+1)} = \theta^{(t)} - \eta\frac{\partial \mathcal{L}}{\partial \theta}, \tag{12}$$

where θ represents the network parameters (weights and biases), η is the learning rate, and $\mathcal{L}$ is the loss function.

Although shallow ANNs demonstrated useful predictive skill in many of these applications, their universal approximation properties are only fully realized with sufficient depth and width, and their performance tends to degrade relative to ensemble tree methods when training data are limited. Deep feedforward networks, with many hidden layers and nonlinear activation functions (ReLU, Leaky ReLU, sigmoid), can learn highly complex mappings but are prone to overfitting without appropriate regularization (dropout, weight decay, batch normalization) and require larger training datasets.

A commonly used regularized loss function for deep ANNs is:

$$\mathcal{L}_{\text{reg}} = \mathcal{L} + \lambda\sum_{k} w_k^2, \tag{13}$$

where λ is the regularization coefficient controlling weight decay.

Recent work has demonstrated the utility of deep ANNs combined with ensemble techniques for estimating soil water retention curves and hydraulic conductivity functions from basic soil descriptors across diverse global soil datasets [174,175].

3.3.4. Transformer-Based Architectures and Attention Mechanisms

Transformer architectures, introduced by Vaswani et al. [176] for natural language processing, have rapidly emerged as the dominant paradigm for sequence modeling across a broad range of domains, including geoscience and remote sensing. The key innovation of the transformer is the self-attention mechanism, which computes a weighted average of all positions in the input sequence for each output position, with attention weights determined by the compatibility between query and key vectors derived from the input.

The scaled dot-product attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{14}$$

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

The self-attention weights for each token are computed as:

$$\alpha_{ij} = \frac{\exp\left(\frac{\mathbf{q}_i^\top \mathbf{k}_j}{\sqrt{d_k}}\right)}{\sum_{m=1}^{n}\exp\left(\frac{\mathbf{q}_i^\top \mathbf{k}_m}{\sqrt{d_k}}\right)}, \tag{15}$$

where α_{ij} represents the attention weight assigned by position i to position j.

Unlike RNNs and LSTMs, transformers process all elements of the input sequence simultaneously (enabling parallel computation during training) and can capture long-range dependencies without the information bottleneck of a fixed-size hidden state.

Multi-head attention further extends this mechanism by allowing the model to jointly attend to information from multiple representation subspaces:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O, \tag{16}$$

where each attention head is computed independently as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V). \tag{17}$$

In soil and water science, transformer-based models have been applied to time-series prediction of soil moisture, streamflow, groundwater levels, and crop yields, consistently achieving competitive or superior performance relative to LSTM baselines, particularly on long sequences and in multi-variable forecasting tasks. Spatial transformers and vision transformers (ViT) [177] have been adapted for soil classification and moisture mapping from satellite imagery, leveraging the attention mechanism to selectively weight spatially distant pixels with high relevance to the prediction target. Hybrid spatiotemporal transformer architectures—combining spatial attention across the image dimension with temporal attention across the time-series dimension—represent the current frontier for AI-based monitoring of dynamic soil–water states from dense time series of satellite observations.

3.4. Distributed and Federated Deep Learning in Agriculture

3.4.1. Distributed Training Frameworks

Training large deep learning models on agricultural datasets of realistic scale—encompassing multiple seasons, multiple farms, and multiple sensor modalities—demands computational resources that exceed the capacity of a single machine. Distributed deep learning frameworks distribute both the data and the computational workload across multiple processors or nodes, enabling training to complete in hours rather than weeks [29]. Two principal paradigms are employed: data parallelism, in which each worker processes a distinct shard of the training data and the computed gradients are aggregated (by summation or averaging) before the model weights are updated, and model parallelism, in which the model itself is partitioned across workers, with each worker responsible for the forward and backward passes through a subset of layers; see Figure 7.

Large-scale distributed training has enabled the development of foundation models for earth observation—deep learning models pretrained on vast corpora of satellite imagery spanning diverse land cover types, seasons, and geographic regions—that can be fine-tuned with relatively small amounts of task-specific data for downstream applications, including soil property mapping, irrigation classification, and crop-type delineation. Projects such as the NASA-IBM Prithvi model and ESA's EuroSAT benchmark have demonstrated the value of this approach for accelerating the development of high-accuracy agricultural monitoring systems.

3.4.2. Federated Learning for Privacy-Preserving Agricultural AI

Federated learning, a specific paradigm of distributed AI, allows ML models to be trained collaboratively across a network of decentralized participants—individual farms, regional data hubs, or national agricultural agencies—without requiring the raw data to leave the local node. Only model gradients or parameter updates are communicated to a central aggregator, which combines them to produce a globally improved model. This architecture preserves data privacy and sovereignty while enabling the collective intelligence of a distributed sensor network to be leveraged for model training. In the context of soil–water

management, federated learning is particularly promising for aggregating heterogeneous field data from geographically diverse agricultural regions, enabling the development of robust models that generalize across a wide range of soil types and climatic conditions.

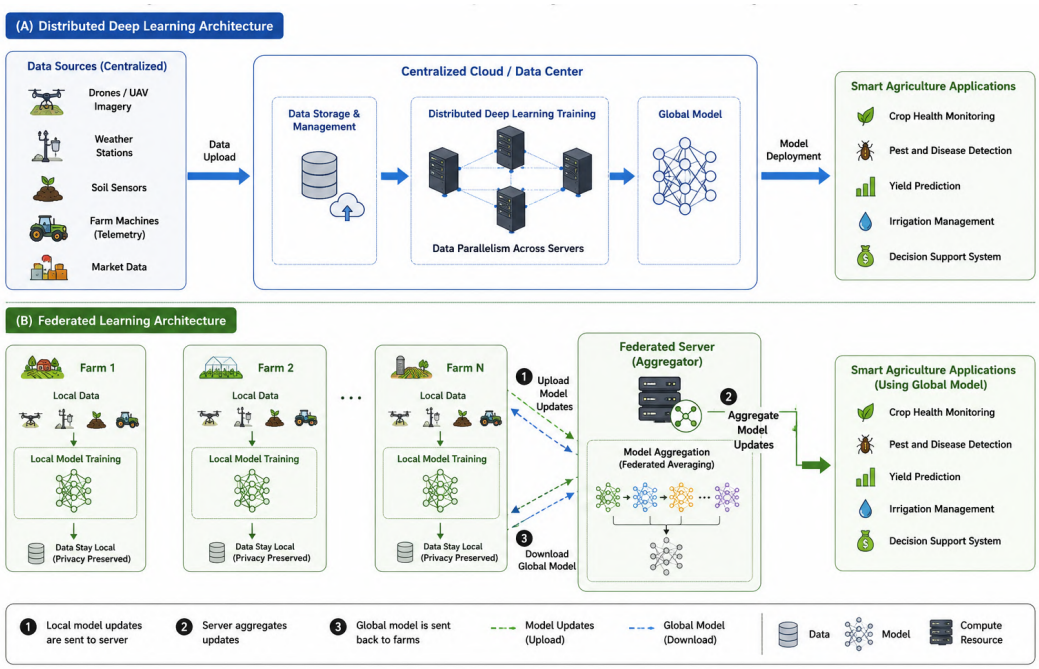

Figure 7. Architecture of distributed deep learning and federated learning in smart agriculture. Software used: draw.io built-in network library for plotting server racks (data center pics); OpenClipart for icons (UAV/drone, weather station, soil sensor, tractor, market); and Noun Project free tier for plotting icons of Smart Agriculture Application panel. Assembly: Inkscape v. 1.4.4. Flowchart source: author.

Federated learning (FL) was introduced by McMahan et al. [178] as a framework for training ML models collaboratively across a network of distributed clients while keeping each client's data local. In the canonical FL algorithm (FedAvg), each communication round proceeds as follows: the central server distributes the current global model parameters to a selected subset of clients; each client performs a fixed number of gradient descent steps on its local data; the updated local parameters are transmitted back to the server; and the server aggregates the local updates (typically by weighted averaging proportional to local dataset size) to produce an improved global model.

The local model update on client k is performed using gradient descent:

$$\theta_k^{(t+1)} = \theta_k^{(t)} - \eta \nabla \mathcal{L}_k(\theta_k^{(t)}), \tag{18}$$

where θ_k denotes the local model parameters for client k, $t \in \{0, 1, \ldots, T-1\}$ is the local iteration index with T denoting the total number of local gradient steps per communication round, η is the learning rate, and $\mathcal{L}_k$ is the local loss function computed on the client dataset.

The global aggregation step in the FedAvg algorithm is defined as:

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_k^{(t+1)}, \tag{19}$$

where K is the number of participating clients, n_k is the number of samples held by client k, and

$$n = \sum_{k=1}^{K} n_k \tag{20}$$

is the total number of training samples across all clients.

This process is iterated until the global model converges. The federated optimization framework substantially reduces the need for centralized data collection while preserving data privacy and enabling collaborative learning across distributed agricultural, environmental, and geospatial datasets.

The application of FL to agricultural AI is motivated by several practical considerations. Agricultural sensor data from individual farms may constitute commercially sensitive information—soil fertility maps, yield records, and irrigation logs can reveal proprietary agronomic practices that farm operators are reluctant to share with competitors or data aggregators. Regulatory frameworks, such as the European Union General Data Protection Regulation (GDPR) and national agricultural data governance policies, impose legal constraints on the transfer and centralized storage of personal and commercially sensitive data [52]. FL provides a technically and legally robust mechanism for harnessing the predictive power of large distributed agricultural datasets while respecting these constraints.

Challenges specific to FL in agricultural contexts include: extreme heterogeneity in the statistical distribution of local data (non-IID data), driven by variation in soil types, climate zones, crop species, and management practices across participating farms; communication bandwidth constraints in rural areas with limited connectivity; and the potential for malicious or malfunctioning clients to degrade the global model through byzantine gradient attacks or hardware failures. Differential privacy mechanisms [179], secure aggregation protocols, and robust aggregation algorithms that down-weight anomalous gradient contributions have been proposed to address these challenges.

3.4.3. Edge AI and Internet of Things (IoT) Integration

The deployment of AI inference at the network edge—on IoT devices, field-deployed microcontrollers, or gateway servers located at the farm level—enables real-time, autonomous decision-making for irrigation control, pest detection, and soil condition monitoring without dependence on reliable cloud connectivity [28]. Edge AI is particularly important in smallholder agricultural contexts across developing regions where internet infrastructure is unreliable and the cost of cloud computing is prohibitive.

The algorithmic families described in the preceding subsections—classical ML, deep learning, and physics-informed hybrid architectures—underpin a broad spectrum of soil–water applications that are treated in detail in Section 5. These include point-to-regional-scale soil moisture prediction (Section 5.1), AI-driven smart irrigation scheduling and water-use optimization (Section 5.2), estimation of soil hydraulic properties through deep learning pedotransfer functions (Section 5.3), salinity, pollution, and soil remediation (Section 5.4), digital soil mapping and spatial land-use analytics (Section 5.5), crop yield prediction and precision agriculture (Section 5.6), and climate change adaptation (Section 5.7). An illustrative overview of the AI workflow for smart irrigation and water management is provided in Figure 8.

Compact deep learning models, optimized for deployment on resource-constrained hardware through techniques such as weight quantization (reducing the numerical precision of model parameters from a 32-bit to an 8-bit or 4-bit floating point), structured pruning (removing low-importance connections or neurons), and knowledge distillation (training a small student model to mimic the predictions of a large teacher model), can achieve accuracy approaching that of large cloud-deployed models while running on microcontrollers with kilobytes of RAM and milliwatts of power.

Integrated soil–water IoT systems typically combine capacitance or time-domain reflectometry (TDR) soil moisture sensors, temperature and rainfall probes, and plant water status sensors (dendrometers, stem psychrometers) with edge AI processors running lightweight LSTM or transformer models fine-tuned for the specific farm and soil

context. These systems close the feedback loop between sensing and actuation, enabling precision variable-rate irrigation that responds to real-time soil moisture dynamics at sub-field resolution.

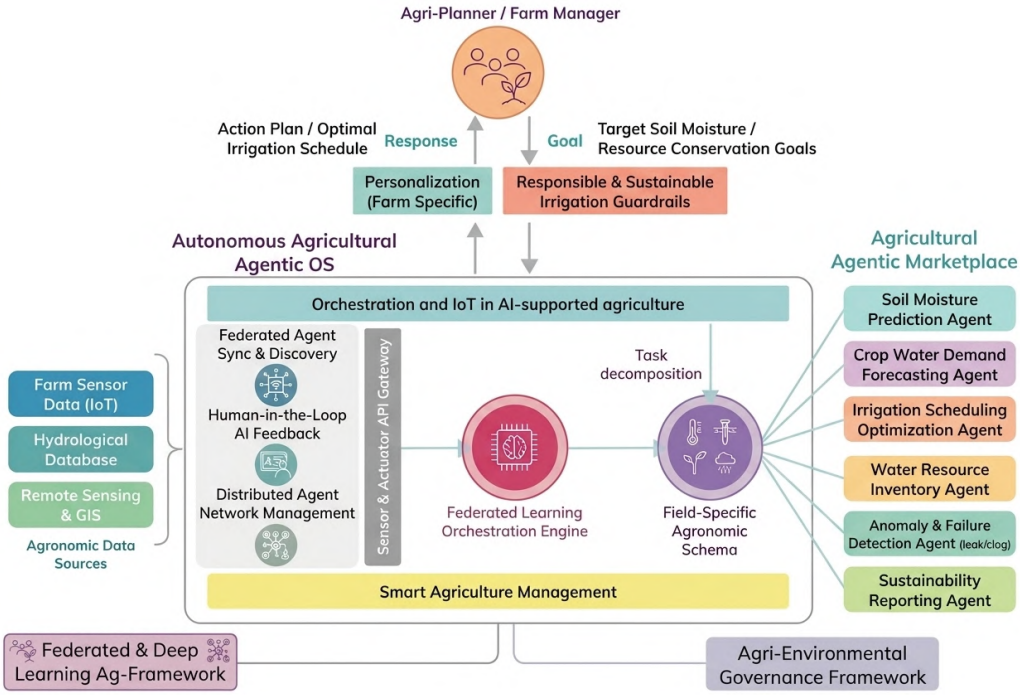

Figure 8. Conceptual system architecture diagram of AI workflow for smart irrigation and water management systems. Software and tools used: GIMP v. 3.2.4; Phosphor Icons for the circular icon at the top (Agri-Planner/Farm Manager); draw.io container shape for plotting panels; and Inkscape v. 1.4.4 for circular engine icons and assembling the layout. Diagram source: author.

4. Data Sources for AI-Based Soil–Water Modeling

The performance of any AI or deep learning model is fundamentally bounded by the quality, volume, and spatial and temporal representativeness of its training data. In soil–water science, this dependency is particularly consequential because target variables—soil moisture content, hydraulic conductivity, solute concentration, and evapotranspiration—are inherently continuous in space and time and yet are sampled at discrete, often sparse locations and intervals. Bridging this observational gap requires the integration of data streams from a diverse ensemble of sensing technologies, each with a characteristic spatial and temporal resolution, measurement uncertainty, and cost profile. The principal data sources reviewed in this section were identified through the literature search described in Section 1.3, with inclusion prioritizing studies reporting quantitative validation of AI models against independently collected field or satellite observations.

4.1. Ground-Based Sensor Networks

In situ sensors remain the primary source of direct, high-frequency soil–water measurements and provide the ground-truth observations against which remote sensing retrievals and model predictions are validated. The principal measurement technologies include: time-domain reflectometry (TDR), frequency-domain reflectometry (FDR), and capacitance probes for volumetric soil moisture content; tension-cup tensiometers and dielectric matric potential sensors for soil water potential; electrical resistivity tomography (ERT) for spatially continuous sub-surface moisture imaging; and electromagnetic induction (EMI) instruments for bulk electrical conductivity surveys of soil salinity. Thermometric sensors, weighing lysimeters, and eddy-covariance flux towers complement these measurements

with continuous records of soil temperature, actual evapotranspiration, and surface energy balance components.

The emergence of low-cost IoT-enabled sensor platforms has dramatically expanded the density and geographic reach of soil moisture networks over the past decade. National and continental networks—including the USDA Soil Climate Analysis Network (SCAN), the International Soil Moisture Network (ISMN), the OzNet network in Australia, and the Cosmic-Ray Neutron Sensing (CRNS) network—now maintain archives of multi-decadal, multi-depth soil moisture records that are publicly accessible and increasingly exploited for ML model training [180]. The ISMN alone aggregates data from over 2600 stations across 71 networks worldwide, providing a globally distributed training corpus for soil moisture prediction models. Wireless sensor networks (WSNs) deployed at the farm scale can resolve sub-field spatial variability at meter-scale resolution, generating high-frequency data streams that feed directly into real-time edge AI inference pipelines [28]. However, sensor drift, fouling, cable damage, and power outages introduce systematic gaps and outliers that must be detected and corrected through automated quality-control algorithms before the data can be ingested by ML models [181].

4.2. Remote Sensing and Satellite Data

Satellite remote sensing has transformed the spatial coverage and temporal frequency of soil–water observations, providing synoptic, repeatable measurements over large areas that are inaccessible to ground-based sensor networks. The electromagnetic spectrum from visible and near-infrared to passive microwave and active radar wavelengths encodes information about soil moisture, surface temperature, vegetation water content, and land surface properties, which can be retrieved through empirical algorithms, physically-based forward models, or, increasingly, deep learning-based inversion schemes [182].

Passive microwave missions—particularly the NASA Soil Moisture Active Passive (SMAP) satellite launched in 2015 and the ESA Soil Moisture and Ocean Salinity (SMOS) mission—provide daily global surface soil moisture retrievals at coarse spatial resolution (36 km for a SMAP L-band radiometer). SMAP data have been extensively used as training targets for deep learning downscaling models that leverage high-resolution optical and SAR auxiliary data to sharpen retrievals to 1–10 km resolution [138,183]. The Copernicus Sentinel constellation has substantially expanded the repertoire of operationally available remote sensing datasets for agricultural monitoring: Sentinel-1 C-band SAR enables all-weather soil moisture retrieval and crop mapping at 10 m resolution with a 6-day revisit time [167], while Sentinel-2 provides 10–20 m multispectral imagery with a 5-day revisit that is widely used for vegetation index time-series extraction, crop-type classification, and surface energy balance estimation [184].

Thermal infrared remote sensing from Landsat-8/9 TIRS, MODIS MOD11 land surface temperature products and the ECOSTRESS instrument aboard the International Space Station provides estimates of land surface temperature (LST) that, in conjunction with vegetation indices, enable the retrieval of evapotranspiration through the surface energy balance equation and data-driven models, such as the Surface Energy Balance Algorithm for Land (SEBAL) and ML-based variants [104,185,186]. Hyperspectral missions—including the PRISMA satellite launched by the Italian Space Agency in 2019 and the upcoming NASA Surface Biology and Geology (SBG) mission—extend spectral resolution to hundreds of narrow bands, enabling the detection of soil organic carbon, clay mineral composition, iron oxide content, and other geochemical properties through spectral mixture analysis and convolutional neural network regression [187].

The integration of observations across satellite platforms at different spatial and temporal scales—a process known as spatiotemporal fusion—is an active research area.

Algorithms based on dictionary learning, sparse representation, and encoder–decoder deep learning architectures have been developed to fuse coarse-resolution, high-temporal-frequency data (e.g., MODIS daily reflectance) with fine-resolution, low-temporal-frequency data (e.g., Landsat bi-weekly imagery) to produce synthetic time series at Landsat resolution and MODIS revisit frequency [188]. Such fused products substantially increase the density of high-resolution observations available for training and validating AI-based soil–water models in cloudy tropical and temperate regions.

4.3. UAV and Drone-Based Monitoring

Unmanned aerial vehicles (UAVs) have emerged as a powerful and flexible complement to satellite remote sensing for high-resolution, on-demand monitoring of soil–water dynamics and crop water status at the field scale [189]. Equipped with RGB, multispectral, hyperspectral, thermal infrared, and LiDAR sensors, UAVs can acquire centimeter-resolution imagery at user-defined acquisition times and revisit frequencies, unconstrained by satellite overpass schedules or cloud cover limitations. This flexibility makes them particularly valuable for capturing transient soil moisture patterns following rainfall or irrigation events, detecting early-stage crop water stress before it manifests in visible symptoms, and generating high-density training datasets for ML models that cannot be obtained from coarser satellite observations alone.

Thermal UAV surveys, in which canopy temperature is retrieved from shortwave infrared cameras, enable the estimation of crop water stress through the Crop Water Stress Index (CWSI) [190] and serve as inputs to precision variable-rate irrigation controllers. Multispectral UAV data enable estimation of the Leaf Area Index (LAI), fractional vegetation cover, and normalized difference vegetation index (NDVI) at sub-meter resolution, which are key covariates in evapotranspiration models and soil moisture downscaling algorithms. Recent advances in structure-from-motion (SfM) photogrammetry applied to UAV imagery enable the generation of high-resolution digital elevation models (DEMs) and canopy height models that support topographic wetness index computation and precision drainage management [191]. The primary constraints on UAV deployment are regulatory restrictions on flight altitude and beyond-visual-line-of-sight operation, battery endurance limiting individual flight coverage, and the substantial processing burden associated with handling very large point clouds and image mosaics.

4.4. Big Data Integration and Data Fusion

Modern AI-based soil–water modeling draws on an increasingly broad and heterogeneous portfolio of data streams that span multiple observational platforms, spatial scales, thematic domains, and temporal resolutions. Effective exploitation of this data richness requires robust data fusion pipelines that harmonize inputs with different coordinate systems, temporal sampling frequencies, missing data patterns, and measurement uncertainties into a coherent, model-ready feature space [18,192]. Cloud computing platforms—including the Google Earth Engine, the Microsoft Planetary Computer, and the AWS Open Data Registry—have dramatically lowered the barrier to large-scale geospatial data processing by providing petabyte-scale archives of satellite imagery and derived products alongside a scalable compute infrastructure, enabling researchers to train and evaluate AI models on global-scale datasets without maintaining the local HPC infrastructure [193].

Data-driven approaches to data fusion, including multimodal deep learning architectures that process inputs from heterogeneous sensor modalities in parallel feature extraction streams before fusing representations in a joint embedding space, have demonstrated superior performance to traditional pixel-based or layer-stacking approaches in agricultural mapping and soil property prediction tasks [187]. Graph neural networks (GNNs) offer

a natural framework for representing the spatial relationships between sensor locations, field parcels, and catchments as graphs, enabling the propagation of information across the observation network in a spatially aware manner [194]. The integration of agronomic management records—crop variety, sowing and harvest dates, tillage operations, fertilizer and pesticide applications, and irrigation logs—with geophysical and remote sensing datasets within unified ML frameworks represents a frontier of data fusion that promises substantial improvements in the precision and interpretability of soil–water models [31,195].

4.5. Data Challenges

Despite the expanding volume and diversity of available data, several persistent challenges limit the effectiveness of AI-based soil–water modeling in practice. The most fundamental is class imbalance and spatial sampling bias: existing soil profile databases are disproportionately concentrated in North America, Western Europe, and Australia, leaving vast areas of Africa, Central Asia, and South America severely underrepresented [111]. ML models trained on spatially biased datasets exhibit degraded prediction accuracy in underrepresented regions and may propagate systematic errors into global soil property maps and water balance assessments.

Temporal data gaps, arising from sensor failures, cloud cover, satellite overpass gaps, and seasonal inaccessibility of field sites, introduce missingness patterns that are frequently non-random and that pose challenges for time-series ML architectures that require complete input sequences [181]. Imputation methods based on temporal interpolation, matrix completion, or generative adversarial networks (GANs) have been explored, but their effectiveness depends critically on the nature and extent of the missingness. Label noise—arising from measurement errors in reference soil samples, inconsistent sampling protocols between laboratories and surveys, and the inherent spatial support mismatch between point observations and gridded model predictions—is a further source of degraded training signal that requires explicit treatment through robust loss functions or uncertainty-aware training procedures.

Data sharing and standardization barriers also constrain the development of large, globally representative training corpora. Harmonizing soil profile data collected under different national standards, with different analytical methods and variable nomenclature, requires substantial data curation effort. International initiatives, such as ISRIC World Soil Information, the Global Soil Biodiversity Initiative, and the FAO Global Soil Partnership, are working to address these challenges through the development of open data standards and federated data infrastructure, but significant interoperability gaps remain.

5. Applications of AI in Soil–Water Interactions

5.1. Soil Moisture Prediction

Soil moisture prediction constitutes one of the most extensively studied applications of AI in soil–water science, driven by the central role of root-zone water content in irrigation scheduling, drought monitoring, and hydrological forecasting [196]. AI-based models have been developed across a hierarchy of spatial and temporal scales, from point-scale time-series prediction at individual sensor stations to regional-scale spatiotemporal mapping using satellite-derived inputs [138].

At the point scale, LSTM networks have demonstrated exceptional capacity to learn the complex nonlinear dynamics of soil moisture evolution in response to precipitation, evapotranspiration, and irrigation inputs [197]. In temporally blocked cross-validated evaluations conducted on multi-year records from point-scale stations distributed across temperate and semi-arid agricultural regions of North America and Europe—drawn from the International Soil Moisture Network (ISMN) and the USDA Soil Climate Analysis

Network (SCAN)—LSTM models have achieved root mean square errors (RMSEs) in the range of 0.02–0.04 m^3 m^{-3} for surface and root-zone volumetric water content at depths of 5–50 cm. These accuracies are competitive with, and, in several cases, superior to, those of calibrated HYDRUS-1D simulations driven by identical meteorological forcing at the same stations, with the critical distinction that the LSTM models require no prior estimation of van Genuchten–Mualem hydraulic parameters from laboratory or pedotransfer function data [17]. It should be noted, however, that these results apply specifically to the point scale under the temperate and semi-arid climatic conditions represented in the ISMN and SCAN networks; generalization to tropical soils, organic soils, or freeze–thaw-affected profiles has not yet been systematically established. Bidirectional LSTM (BiLSTM) and encoder–decoder architectures extend the predictive horizon to multi-day and multi-week soil moisture forecasts, enabling proactive rather than reactive irrigation management.

At the field scale, hybrid CNN–LSTM models that simultaneously exploit the spatial structure of remotely sensed inputs and their temporal evolution have substantially improved prediction accuracy. These models integrate surface reflectance, land surface temperature, and vegetation indices from Sentinel-2 and MODIS with in situ sensor records to generate spatially distributed soil moisture estimates at sub-field resolution. Transformer architectures incorporating self-attention across both the spatial and temporal dimensions are increasingly competitive with LSTM-based approaches, particularly for long-sequence prediction tasks where the attention mechanism can selectively weight distant historical states.

At the regional scale, deep learning downscaling algorithms have been developed to sharpen coarse-resolution SMAP passive microwave retrievals (36 km footprint) to field-scale resolutions of 1–10 km by incorporating high-resolution SAR backscatter, optical vegetation indices, terrain attributes, and soil texture maps as auxiliary covariates in an encoder–decoder framework [138]. Physics-informed neural networks (PINNs) that embed the Richards equation as a soft constraint in the loss function provide a principled mechanism for ensuring that soil moisture predictions remain consistent with mass conservation and hydraulic conductivity relationships, even in data-sparse regions where purely data-driven models tend to extrapolate unreliably [20,198].

Data assimilation frameworks that combine Kalman filter-based state estimation with AI-based observation operators represent a further frontier, enabling the systematic incorporation of heterogeneous satellite and sensor observations into soil moisture prediction systems with rigorously quantified uncertainty bounds.

5.2. Smart Irrigation and Water Management

Intelligent irrigation systems use ML algorithms and sensor feedback to optimize irrigation timing and water allocation, improving water-use efficiency and crop productivity.

Operational smart irrigation systems integrate AI-based soil moisture prediction with real-time IoT sensor telemetry, numerical weather prediction outputs, and crop growth stage information to compute irrigation recommendations that minimize water application while avoiding root-zone moisture deficits that induce yield-limiting crop water stress [28]. Reinforcement learning (RL) provides a particularly natural framework for the sequential decision-making structure of irrigation scheduling: an agent observes the current state of the soil–crop–atmosphere system and selects irrigation actions to maximize a cumulative reward encoding crop water productivity and input cost efficiency. Deep Q-networks (DQNs) and actor-critic algorithms (Proximal Policy Optimization, Soft Actor-Critic) trained within calibrated AquaCrop and DSSAT crop model environments—used as physics-based digital surrogates for agent–environment interaction—have demonstrated irrigation water savings of 20–35% relative to fixed-schedule and soil-moisture-threshold-based control strategies that are representative of current commercial practice [199].

These savings were reported across simulation experiments covering maize, soybean, and wheat under humid temperate and semi-arid agroclimatic conditions in the United States and the Mediterranean basin, evaluated at the field scale (1–10 ha management units) over growing seasons of 90–180 days. The absence of a statistically significant yield penalty under these conditions was established by comparing the simulated end-of-season biomass and harvestable yield between RL-controlled and baseline scenarios within the respective crop model environments; it should be emphasized that these outcomes are contingent on the fidelity of the simulation surrogate to real field conditions, and that transferability to heterogeneous field soils, imperfect sensors, and uncertain weather forecasts has not yet been fully validated under operational conditions.

Model predictive control (MPC) frameworks informed by LSTM-based soil moisture forecasts constitute an alternative AI-driven approach: the controller solves a constrained optimization problem at each decision step, selecting irrigation amounts that minimize the predicted moisture deficit over a rolling forecast horizon subject to water availability and application rate constraints, Table 5. The integration of probabilistic forecast uncertainty into MPC cost functions—penalizing decisions that carry high risk of soil moisture falling below the wilting point—substantially improves the robustness of irrigation recommendations under uncertain weather conditions.

Table 5. AI techniques for smart irrigation and water conservation.

AI and Algorithmic Category	Specific AI/ ML Model	Data Source/Inputs	Primary Role in Water Conservation	Water Saving Metric
Supervised Learning (Regression and Prediction)	Random Forest (RF), Support Vector Regression (SVR)	IoT telemetry (soil moisture, ambient temp, air humidity, UV index)	Predicts the precise localized surface soil moisture (SSM) to prevent over-watering or plant stress.	Reduces baseline manual irrigation water use by 30–40% [200].
Deep Learning and Sequence Modeling	Long Short-Term Memory (LSTM), Transformers	Historical agrometeorological time series, regional weather forecasts	Forecasts daily reference evapotranspiration (ET_0) and crop water depletion rates days in advance [201].	Minimizes evaporative runoff, allowing highly adaptive preemptive irrigation scheduling.
Computer Vision and Image Processing	Convolutional Neural Networks (CNNs), U-Net	Multispectral satellite imagery (Sentinel/Landsat), UAV/drone aerial imagery [202]	Identifies vegetative indices (NDVI, NDWI) to segment crop water stress zones across large geographic fields.	Prevents blanket-field flooding by enabling zone-specific drip irrigation.
Meta-Heuristic Optimization	Genetic Algorithms (GAs), Particle Swarm Optimization (PSO)	Multi-objective constraints (available water reserves, electricity costs, crop growth stage) [36]	Runs algorithmic optimization loops to discover the mathematically most efficient water distribution scheduling schedule.	Increases overall resource efficiency by 30–50% across diverse agroclimatic zones [203].
Edge and Intelligent Control	Fuzzy Logic Controllers (FLCs)	Real-time sensor micro-data (pH levels, salinity, soil matric potential) [204]	Acts as the automated rules engine that triggers, throttles, or shuts off motorized irrigation valves dynamically without human intervention.	Cuts water wastage from system lags; optimizes nutrient–water (“fertigation”) ratios [200].
Anomaly Detection and Diagnostics	Autoencoders, Isolation Forests	Flow rate meters, acoustic sensors, distributed pressure gauges [205]	Monitors water distribution networks to detect sub-surface pipeline leaks or anomalous pressure drops instantly.	Prevents structural water loss and localized over-saturation from damaged plumbing infrastructure.

5.3. Estimation of Soil Hydraulic Properties

Deep learning approaches are increasingly applied to estimate hydraulic conductivity, infiltration rates, and water retention curves from soil texture and environmental datasets.

Pedotransfer functions (PTFs) are the primary tool for estimating soil hydraulic properties—saturated hydraulic conductivity (K_s), water retention curve parameters (α, n, θ_r, θ_s), and unsaturated conductivity function parameters—from routinely measured basic soil descriptors, such as texture fractions, bulk density, and organic matter content [87]. Classical regression-based PTFs, fitted to relatively small and geographically constrained calibration datasets, exhibit limited transferability across soil types and geographic regions. Deep learning PTFs trained on large globally compiled databases—including UNSODA, HYDRUSbase, and ISRIC World Soil Information archives—have substantially outperformed classical regression PTFs in independent validation studies, particularly for soils with atypical textural compositions or unusual structural properties [161].

Convolutional neural networks operating on proximal soil sensing data—including visible–near-infrared (Vis-NIR) and mid-infrared (MIR) spectra, X-ray computed tomography (CT) images of soil pore architecture, and electrical resistivity tomography profiles—have achieved state-of-the-art accuracy in predicting K_s and water retention parameters at the sample scale. Graph neural networks (GNNs) that represent the soil pore network as a graph structure and propagate hydraulic information across connected pore bodies and throats offer a physically grounded approach to learning conductivity from pore-scale CT imaging data, bypassing the need for manual extraction of geometric pore parameters. Physics-informed neural network PTFs that constrain predictions to satisfy the van Genuchten–Mualem model structure—ensuring that estimated water retention and conductivity functions are thermodynamically consistent and monotonic—provide improved extrapolation reliability relative to unconstrained black-box deep learning models [20]. Transfer learning from globally pretrained PTF models to locally fine-tuned variants using small regional calibration datasets represents a practically important strategy for deploying high-accuracy hydraulic property estimation in data-sparse agricultural regions.

5.4. Salinity, Pollution, and Soil Remediation

AI techniques enable the prediction of salinity distribution, pollutant transport, and remediation efficiency under varying climatic and soil conditions.

Soil salinization constitutes a globally significant and accelerating form of soil degradation, affecting approximately 20% of irrigated agricultural land and directly reducing plant-available water through osmotic stress [93]. AI-based models have demonstrated strong predictive capability for mapping and monitoring soil salinity distribution from electromagnetic induction (EMI) survey data, satellite-derived spectral indices (e.g., the Salinity Index derived from SWIR reflectance), and combinations of environmental covariates, including groundwater depth, evaporation rate, and irrigation history. Random Forest and gradient boosting models trained on multi-depth EMI transect data and surface reflectance imagery have achieved spatial prediction accuracy ($R^2 > 0.80$) for soil electrical conductivity at field scales, providing the high-resolution salinity maps required for targeted leaching irrigation scheduling and salt-tolerant crop variety selection [206].

Pollutant transport and groundwater contamination risk prediction represent a further domain in which AI models provide significant analytical value. Recurrent Neural Network models trained on groundwater quality time series—including nitrate, pesticide, and heavy metal concentration records—enable the projection of future contamination levels under different irrigation and fertilization scenarios, enabling proactive management of non-point source agricultural pollution [207–209]. Convolutional neural network models applied to hyperspectral imagery acquired over contaminated agricultural sites have demonstrated

the capacity to map spatial distributions of cadmium, lead, and arsenic concentrations in topsoil with accuracy approaching laboratory analysis, at a fraction of the cost and time of conventional sampling campaigns [168,187].

Physics-informed ML models that couple advection–dispersion equations with neural network surrogate functions for spatially variable hydraulic conductivity and retardation coefficients offer a computationally efficient alternative to finite-element solute transport simulators for scenario analysis of remediation strategies [20,97,210]. AI-assisted optimization of pump-and-treat and phytoremediation interventions—using genetic algorithms or reinforcement learning to identify remediation designs that minimize cost and duration for a target contaminant concentration threshold—is an emerging application with strong practical relevance for soil and groundwater restoration in agriculturally impacted landscapes.

5.5. Digital Soil Mapping and Land-Use Planning

GIS-integrated AI frameworks support precision agriculture, land suitability analysis, and sustainable land-use planning. Digital soil mapping (DSM) uses spatially referenced soil profile observations as training data for AI models that predict the continuous spatial distribution of soil properties across unsurveyed areas from environmental covariate layers derived from remote sensing, terrain analysis, climate datasets, and geological maps [211]. The SCORPAN framework—postulating that soil properties are functions of parent material, organisms, relief, climate, time, and spatial position—guides covariate selection in DSM, with terrain derivatives (slope, curvature, topographic wetness index, valley depth), multitemporal vegetation index composites, and lithological maps constituting the most consistently informative predictor groups across global benchmarking studies.

Random Forest and gradient boosting models have dominated DSM benchmarking exercises, demonstrating particularly strong performance for mapping soil organic carbon, clay content, pH, bulk density, and available water capacity at 100–250 m resolution across national and continental domains [111]. The global SoilGrids250m product (ISRIC—World Soil Information (Wageningen, The Netherlands)), produced using RF models trained on over 150,000 geo-referenced soil profiles and a comprehensive covariate stack, has established an internationally recognized baseline for high-resolution global soil property mapping [111,211]. Deep learning architectures incorporating spatial context through CNN operations on image-structured covariate stacks, or through graph neural network propagation across spatial adjacency graphs of mapping units, are increasingly competitive with ensemble tree methods, particularly for structurally complex properties, such as soil water retention parameters and field-scale hydraulic conductivity [128,187,212].

Multi-task learning frameworks that simultaneously predict multiple correlated soil properties in a shared representation space—exploiting the known pedological relationships between texture, organic carbon, pH, and hydraulic properties—achieve substantially lower prediction errors than independently trained single-property models, particularly in data-sparse regions where joint training provides implicit regularization [161,174]. GIS-integrated AI-based land suitability analysis frameworks that combine DSM outputs with hydrological, climatic, and agronomic constraints in a spatially explicit decision-support framework are increasingly deployed by national agricultural agencies to guide irrigation scheme planning, land-use zoning, and soil conservation investment prioritization.

5.6. Crop Yield Prediction and Precision Agriculture

AI-driven yield forecasting systems integrate soil moisture, weather, and crop growth data to improve productivity and resource management. AI-based crop yield prediction models combine soil moisture status, meteorological forcing, satellite-derived vegetation indices (NDVI, EVI, LAI), and agronomic management records to generate spatially explicit,

in-season yield estimates [213]. CNN–LSTM hybrid architectures that simultaneously capture the spatial pattern of remotely sensed crop reflectance and its temporal evolution through phenological development have achieved county-level maize and soybean yield prediction accuracies comparable to official agronomic survey estimates, several months before harvest [17,213]. Transformer models equipped with cross-attention between the spectral and temporal dimensions of multi-date satellite image time series represent the current state of the art for in-season yield forecasting, demonstrating robust performance across diverse crop types, climatic zones, and management systems.

The integration of root-zone soil moisture dynamics—derived from LSTM-based prediction models informed by in situ sensors and satellite retrievals—as a key predictive covariate substantially improves yield forecast accuracy, particularly in rainfed systems and during drought years when soil water availability is the primary limiting factor on productivity [138]. Precision variable-rate application of irrigation, guided by within-field soil moisture variability maps generated by ensemble ML models, has been demonstrated to reduce total irrigation water inputs by 15–30%, relative to uniform application, while maintaining or increasing field-average yield, by directing water to the highest-yield-potential zones during critical crop growth stages [28]. Spatially explicit AI-based yield gap analysis—comparing predicted attainable yield under optimized soil moisture management with actual yield under observed conditions—provides quantitative diagnostics of where and why water-limited productivity shortfalls occur, enabling targeted interventions in precision agriculture programs.

5.7. Climate Change Adaptation

ML models support drought forecasting, climate-risk assessment, and resilient agricultural planning under changing climatic conditions. Climate change is projected to substantially alter the spatial and temporal distribution of precipitation, intensify drought frequency and severity, increase evapotranspiration demand, and shift agroclimatic zones across major agricultural regions [2]. AI models trained on historical soil–water observations and climate reanalysis data provide computationally efficient surrogates for the complex and nonlinear interactions between changing atmospheric forcing and soil hydrological response, enabling the rapid screening of climate change impacts and adaptation strategies across large ensembles of climate model projections [214].

ML emulators of physically-based crop–water models—trained to reproduce the input–output behavior of DSSAT, AquaCrop, or SWAT simulations across a wide range of climate, soil, and management scenarios—can evaluate thousands of adaptation scenario combinations (drought-tolerant varieties, deficit irrigation strategies, cover cropping, soil carbon sequestration) within minutes, compared with weeks required for explicit physically-based simulation [215]. Ensemble neural network models that incorporate Coupled Model Intercomparison Project Phase 6 (CMIP6) climate model outputs as forcing inputs enable projection of mid-century changes in root-zone soil moisture deficits, growing season irrigation water requirements, and drought-induced yield losses for major staple crops under Representative Concentration Pathway scenarios, providing quantitative inputs for national adaptation planning and investment prioritization [214].

Transfer learning frameworks that adapt AI models pretrained on historical climate–soil–yield relationships to future climate conditions—retraining on synthetic datasets generated from climate model projections to correct systematic biases in future forcing distributions—represent a promising approach for extending the temporal validity of data-driven agricultural models to climate states outside the historical training range. Physics-guided deep learning models that embed thermodynamic constraints on evapotranspiration and water balance closure as hard constraints provide an additional safeguard against

physically implausible extrapolation under novel climatic conditions, ensuring that AI-generated climate adaptation assessments remain grounded in fundamental hydrological principles [165].

6. Comparative Performance of AI Models

6.1. Evaluation Metrics

Rigorous comparative evaluation of AI models in soil–water applications requires a coherent set of quantitative performance metrics that capture different dimensions of predictive quality: overall error magnitude, bias, variance, and the capacity to reproduce the dynamic range and temporal structure of target variables. The selection of appropriate metrics is not merely a technical formality but fundamentally conditions which model properties are rewarded during training and how inter-model comparisons should be interpreted [216].

The root mean square error (RMSE) is the most widely reported accuracy metric in soil moisture prediction, evapotranspiration estimation, and streamflow forecasting applications. It measures the standard deviation of residuals between predicted ($\hat{y}_i$) and observed (y_i) values over n samples:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \quad (21)$$

Because RMSE weights large errors quadratically, it is particularly sensitive to outlier predictions and to errors during extreme events (heavy rainfall, severe drought), making it informative for applications where the tails of the error distribution are consequential [216]. The Mean Absolute Error (MAE), defined as $MAE = n^{-1}\sum|y_i - \hat{y}_i|$, provides a complementary measure that treats all errors equally and is more robust to outliers and is preferred when characterizing typical prediction performance across a broad range of conditions.

The Nash–Sutcliffe Efficiency (NSE), originally proposed for hydrological model evaluation [217], normalizes the mean squared error of the model against the variance in the observations:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (22)$$

where $\bar{y}$ is the observed mean. NSE values range from $-\infty$ to 1, with NSE $= 1$ indicating perfect prediction and NSE $= 0$ indicating that the model performs no better than the mean of the observations as a predictor. NSE is widely adopted in hydrological and soil–water modeling because it penalizes both systematic bias and dynamic range errors, providing a single scalar summary of overall model fidelity. The Kling–Gupta Efficiency (KGE), a decomposition of NSE into correlation, bias ratio, and variability ratio components [218], is increasingly preferred in recent benchmarking studies because it provides a more diagnostically informative characterization of model error structure and avoids certain known deficiencies of NSE in bias-dominated error regimes.

The coefficient of determination (R^2) measures the proportion of variance in the observations explained by the model and is ubiquitous in soil property prediction and digital soil mapping benchmarks. However, R^2 is insensitive to systematic multiplicative or additive bias—a model that consistently over- or under-predicts by a fixed factor can achieve $R^2 = 1.0$ while exhibiting a substantial absolute error—and should therefore always be reported alongside the RMSE or MAE. For probabilistic prediction outputs, the Continuous Ranked Probability Score (CRPS) and coverage probability of prediction intervals at specified nominal confidence levels provide complementary measures of uncertainty calibration

that are essential for evaluating Bayesian and ensemble AI models applied to soil moisture forecasting [219].

6.2. Comparative Analysis of Model Performance

Recent studies indicate that deep learning architectures generally outperform traditional statistical and empirical models in highly nonlinear soil–water systems; see Table 6. The mathematical variations and empirical performance boundaries of these predictive frameworks are systematized below. Of particular importance for soil moisture modeling is the class of physically constrained machine learning models, which embed physical laws, such as the Richards equation, water balance constraints, and soil hydraulic relationships, directly into the model training process. Recent studies have demonstrated that such physics-constrained approaches consistently outperform purely data-driven counterparts in predictive accuracy, physical consistency, and generalization to sparse or extreme-value regimes [210,220,221].

Table 6. Comparative performance of AI models in soil–water applications.

AI Model	Key Advantages	Primary Limitations	Typical Accuracy	Key Citations
Extreme Gradient Boosting (XGBoost)	Exceptional handling of heterogeneous multi-source datasets (soil organic carbon, terrain indices) across multi-depth soil layers.	Prone to overfitting if hyperparameters (learning rate, max depth) are poorly optimized; high memory consumption.	R^2: 0.66–0.74 *NSE*: High	[222]
Temporal Convolutional Networks (TCNs)	Outperforms sequential models like LSTM by retaining longer effective history sizes using causal, dilated convolutions; highly parallelizable.	Demands a strictly regular, gap-free temporal sampling interval; can exhibit high inference latency during recursive prediction loops.	R^2: 0.96–0.99 *MAE*: ≈ 0.05	[223]
Ensemble Learning (RF + XGBoost Hybrid)	Captures complex, highly nonlinear interactions between the crop Leaf Area Index (LAI) and climate; improves *NSE* metrics by up to 25%.	Increased structural complexity leads to limited model interpretability and higher deployment friction at the IoT edge.	R^2: 0.81–0.84 *NSE*: ≥ 0.75	
Multi-Layer Perceptron (MLP)	Demonstrates superior out-of-sample generalization capabilities for daily reference evapotranspiration (ET_0) in humid zones.	Extremely sensitive to input data scarcity; easily trapped in local minima during gradient descent backpropagation.	R^2: 0.95–0.98 *RMSE*: ≈ 0.21	[224,225]
Support Vector Machines (SVMs)	Highly accurate for localized digital soil mapping of saturated soil hydraulic conductivity (K_{sat}) in complex headwater terrains.	Computationally expensive when scaled to vast, grid-based continental geographic data frameworks; poor text/ symbolic processing.	R^2: 0.70–0.81 *RMSE*: Variable	[226]

Synthesizing quantitative performance indicators reported across the reviewed literature permits a more objective cross-model comparison than qualitative description alone. For soil moisture prediction at the point scale, LSTM and BiLSTM networks consistently achieve RMSE values of 0.02–0.05 m^3 m^{-3} and R^2 of 0.85–0.97 in temperate and semi-arid climates, outperforming calibrated HYDRUS-1D simulations by 15–30% in the RMSE on the same station records [17]. CNN–LSTM hybrid models applied to field-scale spatiotemporal soil moisture mapping from Sentinel-1/2 inputs report R^2 of 0.82–0.93 and a RMSE of 0.03–0.06 m^3 m^{-3} under cross-validated evaluation. For evapotranspiration estimation, MLP and RF models achieve MAE values of 0.15–0.35 mm day^{-1} and a NSE of 0.88–0.96 on FAO-56 reference ET benchmarks, while transformer-based architectures have been reported to attain a NSE > 0.95 on multi-site streamflow and ET prediction tasks. For digital soil mapping, RF and XGBoost models predict soil organic carbon with R^2 of 0.60–0.80 and a RMSE of 3–8 g kg^{-1} at 250 m resolution in continental benchmarks [111], with deep learning CNN architectures achieving comparable or superior accuracy when spatially structured covariates are available. These quantitative benchmarks should be

interpreted with caution, as they reflect heterogeneous experimental designs, geographic regions, soil types, and validation strategies; direct cross-study comparison is possible only when identical benchmark datasets and evaluation protocols are applied.

6.3. Generalization and Transferability

Transfer learning and domain adaptation techniques improve the scalability and robustness of AI models across diverse agricultural environments.

The generalizability of AI models beyond the specific geographic region, soil type, climate zone, and crop system represented in their training data is one of the most practically consequential challenges in agricultural AI [18]. A model that achieves high accuracy in cross-validated evaluation within its training domain but fails when applied to unseen conditions provides limited scientific and operational value, yet this failure mode is commonplace in the soil–water modeling literature, where evaluation is frequently confined to the data-rich regions and environmental conditions that dominate the training corpus.

Transfer learning provides a principled approach to improving cross-domain model performance by leveraging knowledge acquired from a source domain (where abundant labeled training data are available) to improve learning in a target domain (where data are sparse). In the context of soil moisture prediction, models pretrained on dense networks of well-instrumented agricultural sites in North America or Central Europe can be fine-tuned on small amounts of locally collected data from target sites in data-sparse regions (sub-Saharan Africa, Central Asia), achieving substantially lower prediction errors than models trained from scratch on the limited target-domain data alone. The degree of transferability depends critically on the distributional similarity between source and target domains: when soil types, climate regimes, and vegetation systems differ markedly, domain adversarial training—which explicitly minimizes feature distribution discrepancy between source and target domains—is necessary to prevent negative transfer, in which pretraining on a dissimilar source domain degrades target-domain performance [227].

Large pretrained foundation models for earth observation—trained on petabyte-scale archives of multitemporal, multi-sensor satellite imagery spanning diverse global land cover conditions—exhibit strong zero-shot and few-shot generalization capacity for downstream soil and water property prediction tasks, owing to the breadth of spatiotemporal patterns encoded in their learned representations. Models such as the NASA-IBM Prithvi geospatial foundation model have demonstrated competitive fine-tuning performance on soil moisture downscaling, flood extent mapping, and crop-type classification with substantially less task-specific training data than conventional architectures, indicating that large-scale pretraining effectively compresses transferable environmental knowledge that supports rapid adaptation to new domains [228]. Developing rigorous evaluation protocols for transferability assessment—including spatially blocked cross-validation, leave-region-out evaluation, and prospective temporal holdout—is essential for honest reporting of model generalizability and for identifying failure modes before operational deployment.

6.4. Decision Framework for Model Selection

A persistent gap in the AI soil–water modeling literature is the absence of structured guidance for practitioners on which modeling approach is most appropriate under a given combination of agricultural context, data availability, computational resources, and operational requirements. The following synthesis addresses this gap by mapping key selection criteria across the model families reviewed in this paper.

When the primary constraint is training data scarcity (fewer than approximately 500 labeled observations, common in field-scale soil profile surveys), SVMs and shallow RF models should be preferred over deep learning architectures, as their regularization

mechanisms provide robust performance under small-sample conditions and avoid overfitting of high-capacity neural networks. When datasets are moderately sized (thousands of observations) and involve spatially structured covariates (remote sensing bands, terrain derivatives), RF and gradient boosting represent the most reliable choice: they handle mixed-type predictors and missing values natively, provide interpretable SHAP-based feature importance, and generalize well across diverse soil–climatic conditions. When the task involves a long time series with pronounced temporal dependencies—soil moisture forecasting, streamflow prediction, and seasonal irrigation demand—LSTM and BiLSTM architectures are most appropriate, as their gated memory mechanisms model sequential dependencies over hundreds of time steps. For spatiotemporal field-scale tasks with satellite image inputs, CNN–LSTM hybrids or ConvLSTM architectures that simultaneously exploit spatial and temporal patterns are preferred.

For multi-farm or privacy-sensitive deployments where raw data cannot be centralized, federated learning frameworks with FedAvg or personalized FL aggregation are the technically and legally appropriate architecture regardless of the base model family. When physical consistency of predictions is required—for regulatory compliance, hydrological balance closure, or deployment under extrapolative climate conditions—physics-informed neural networks or hybrid PINN–data-driven architectures should be selected even at the cost of additional implementation complexity. Interpretability requirements further condition model selection: if practitioners require human-interpretable explanations at the individual prediction level, intrinsically interpretable models (RF with SHAP, GAMs, Decision Trees) or post hoc XAI wrappers applied to deep learning outputs should be incorporated from the design stage rather than as an afterthought.

7. Explainable AI, Uncertainty, and Ethical Considerations

The deployment of AI and deep learning systems in agricultural water management raises important questions that extend beyond predictive accuracy to encompass interpretability, trustworthiness, and the equitable distribution of benefits and risks. Farmers, water resource managers, and policy makers who rely on AI-based decision-support tools require not only accurate predictions but also understandable explanations of why a model produced a particular output, honest quantification of the uncertainty surrounding that output, and assurance that the underlying system was designed, trained, and deployed in a manner consistent with principles of fairness, transparency, and accountability [229]. This section reviews the current state of explainable AI (XAI), uncertainty quantification, and ethical frameworks as they apply to AI-based soil–water systems.

7.1. Explainable Artificial Intelligence

The opacity of complex ML models—particularly deep neural networks with millions of parameters—has been a persistent concern since their widespread adoption. Although these models achieve impressive predictive accuracy, their internal representations are not directly interpretable by human experts, limiting their acceptance in high-stakes decision contexts, such as irrigation management, drought early warning, and land-use planning, where the consequences of erroneous decisions can be severe and difficult to reverse. The field of explainable AI (XAI) has developed a rich toolkit of post hoc and intrinsic explanation methods designed to make model behavior interpretable without sacrificing predictive performance; see Figure 9.

Among post hoc explanation methods, SHapley Additive exPlanations (SHAP) [163] have emerged as the most widely adopted approach in soil and water science applications, owing to their rigorous game-theoretic foundation (based on Shapley values from cooperative game theory), their consistency across model classes, and their availability in efficient

tree-based and kernel-based implementations. SHAP decomposes the prediction for each individual sample into additive contributions from each input feature, providing both local (sample-level) and global (dataset-level) attribution scores.

In digital soil mapping applications, SHAP analysis has revealed that environmental covariates such as the normalized difference vegetation index (NDVI), terrain curvature, and land surface temperature consistently rank among the most important predictors of soil moisture and organic carbon content across diverse geographic regions, providing physically interpretable validation of model behavior [128].

In irrigation scheduling applications, SHAP-based attribution enables the identification of which meteorological variables—evapotranspiration demand, antecedent rainfall, and vapor pressure deficit—most strongly drive irrigation timing recommendations, enabling agronomists to audit model decisions against domain knowledge [230].

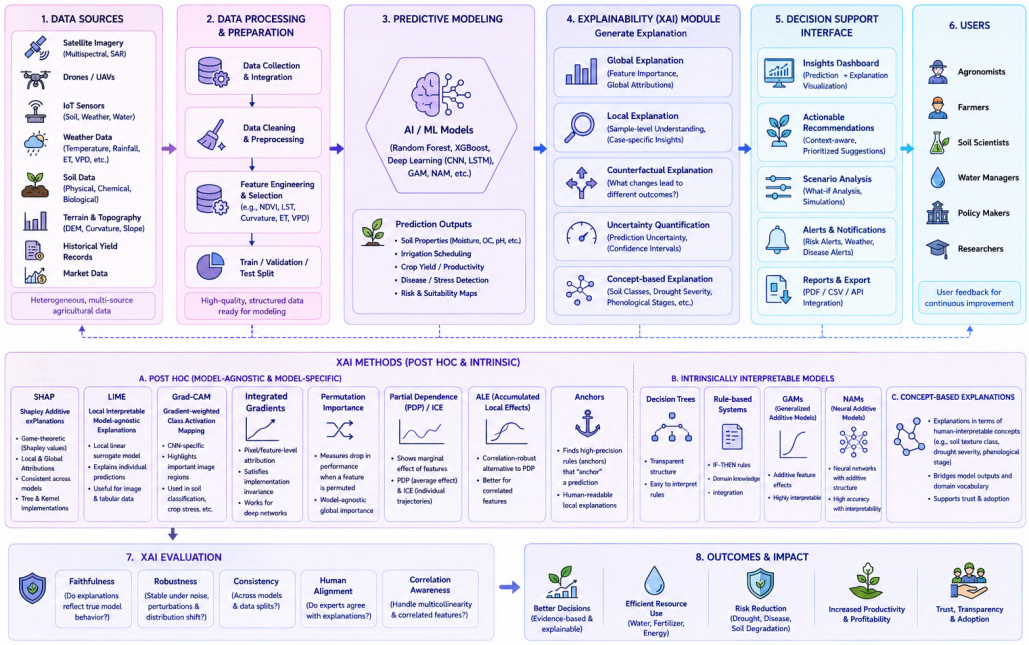

Figure 9. Explainable AI framework for agricultural decision support. Software used: Inkscape v. 1.4.4 for plotting data dashboard and clipart; draw.io for plotting taxonomy flowchart; GIMP v. 3.2.4 for plotting farm scene; and PlotNeuralNet, NetworkX and Matplotlib for plotting DL architecture diagram. Noun Project for plotting conceptual OS architecture. Diagram source: author.

Local Interpretable Model-Agnostic Explanations (LIME) [231] provide an alternative post hoc approach in which a locally faithful linear surrogate model is fitted to the predictions of a complex model in the neighborhood of each input sample. LIME is particularly useful for explaining predictions in image-based applications—for example, identifying which regions of a multispectral or SAR image most strongly influenced a soil classification or moisture retrieval prediction—and has been applied to CNN-based crop stress detection and soil type mapping from UAV and satellite imagery. Gradient-based saliency methods, including Gradient-weighted Class Activation Mapping (Grad-CAM) [232] and integrated gradients, provide pixel-level attribution maps for CNN models, visually highlighting the image regions that most influenced a prediction.

Intrinsically interpretable model architectures—including Decision Trees, rule-based systems, and generalized additive models (GAMs)—remain valuable in soil–water applications where domain experts require full transparency of the decision logic. The Neural Additive Model (NAM) [233] extends GAMs to neural network feature functions while preserving additive interpretability and has been applied to soil hydraulic property estima-

tion with competitive accuracy relative to black-box deep learning models. Concept-based explanations, in which model predictions are expressed in terms of human-interpretable domain concepts (soil textural classes, drought severity indices, phenological stages), represent an emerging frontier that promises to bridge the gap between statistical model outputs and the conceptual vocabulary of agricultural practitioners.

Recent work has emphasized the importance of evaluating the faithfulness and robustness of XAI explanations: it has been shown that some popular attribution methods can produce misleading or unstable explanations when applied to deep networks with correlated input features—a common situation in soil–water datasets where meteorological and remote sensing variables exhibit strong multicollinearity [234]. Developing XAI methods that are provably faithful to model behavior under input correlation and that degrade gracefully under distribution shift remains an active and important research agenda for AI-based agricultural decision support.

7.2. Uncertainty Quantification

Quantifying the uncertainty of AI model predictions is as important as maximizing their expected accuracy, particularly in risk-sensitive agricultural decision contexts where the tails of the predictive distribution determine the probability of crop failure, over-irrigation, or groundwater depletion. Predictive uncertainty in ML models arises from two distinct sources: aleatoric uncertainty , which reflects irreducible noise in the measurement process (sensor error, natural variability) and cannot be reduced by collecting more training data, and epistemic uncertainty, which reflects incomplete knowledge of the model parameters due to limited training data and can in principle be reduced with additional observations [235].

Bayesian deep learning provides a principled framework for quantifying epistemic uncertainty by treating model weights as random variables with prior distributions that are updated to posterior distributions upon observing training data. In practice, exact Bayesian inference is computationally intractable for large neural networks, and approximate methods—including Monte Carlo Dropout [236], variational inference with the local reparametrization trick, and Stochastic Weight Averaging Gaussian (SWAG)—generate samples from an approximate weight posterior. When applied to soil moisture prediction and streamflow forecasting, Bayesian LSTM models with MC Dropout have been shown to produce well-calibrated predictive intervals that accurately capture the frequency of observed values falling within stated confidence bounds across diverse climatic conditions.

Deep ensemble methods [237]—in which a fixed number (typically 5–20) of independently initialized and trained models are combined by averaging their predictions—have demonstrated strong empirical performance as a practical alternative to Bayesian approximations, consistently producing sharper and better-calibrated uncertainty estimates than single models across multiple soil–water benchmarking studies. Conformal prediction [238] offers a distribution-free, statistically rigorous framework for constructing prediction intervals with guaranteed marginal coverage under exchangeability assumptions and is increasingly applied to spatiotemporal agricultural forecasting tasks, where the assumption of independent and identically distributed (IID) test data may not hold. Communicating uncertainty to end users in agriculturally meaningful terms—for example, expressing soil moisture forecast uncertainty as the probability of crossing the wilting point threshold—remains an important and underexplored research challenge at the interface of AI, decision science, and agricultural extension [216].

7.3. Ethical and Socioeconomic Considerations

The deployment of AI-based soil–water management systems at scale raises profound ethical and socioeconomic questions concerning data ownership and sovereignty, algorithmic fairness, digital access inequality, and the distribution of risks and benefits across different categories of agricultural stakeholders [239,240]. These concerns are not peripheral to the technical enterprise of developing AI for precision agriculture but are central to determining whether the technology ultimately contributes to or undermines sustainable and equitable food systems.

Data ownership and privacy are among the most contentious issues in agricultural AI. Farm-level sensor data, yield maps, and soil survey results encode commercially sensitive information about land productivity and management practices. When this data is aggregated by commercial AI platform providers, farmers may inadvertently transfer proprietary agronomic intelligence to entities that could use it to inform commodity trading, land pricing, or the development of competing agricultural services [241]. Federated learning and differential privacy mechanisms [179] offer partial technical remedies by enabling AI model training without centralized data collection, but their adoption requires standardized data governance frameworks and legally enforceable data sharing agreements that remain underdeveloped in most jurisdictions. The European Union's proposed Data Act and the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [242] provide important policy foundations, but their practical implementation in the agricultural sector is nascent.

Algorithmic fairness is a critical concern in AI systems that inform resource allocation decisions with distributional consequences. Soil–water AI models trained predominantly on data from technologically advanced, well-instrumented agricultural regions—as most current models are—systematically underperform in data-sparse smallholder farming contexts in sub-Saharan Africa, South Asia, and Latin America, where the consequences of irrigation mismanagement are most severe [243]. This performance gap constitutes a form of algorithmic inequity that reinforces existing disparities in agricultural productivity and climate resilience. Addressing it requires deliberate investment in the data collection, model adaptation, and co-design of AI tools with smallholder farming communities, respecting principles of participatory and responsible innovation [239].

Digital infrastructure inequality—in connectivity, electrification, device availability, and data literacy—poses structural barriers to the adoption of AI-based precision irrigation in rural and low-income agricultural contexts. Designing AI systems that are robust to intermittent connectivity, operable on low-cost hardware, and explainable without specialist technical training is essential for ensuring that the benefits of agricultural AI are broadly accessible rather than accruing exclusively to large-scale, technologically endowed farm operators [240]. The integration of local and indigenous soil and water knowledge into the design and validation of AI systems, alongside formal scientific datasets, represents both an ethical imperative and a scientifically productive strategy for developing models that are contextually appropriate and trusted by the communities they are intended to serve.

8. Challenges and Limitations

8.1. Data Scarcity and Quality

Many agricultural regions lack high-quality datasets for robust AI model training and validation. The performance of AI and deep learning models is fundamentally bound by the quality, volume, and spatial representativeness of their training data. In soil–water science, this dependency is particularly consequential because target variables—soil moisture, hydraulic conductivity, solute concentration, and evapotranspiration—are continuous in space and time yet sampled at discrete, geographically biased locations. Existing soil profile

databases are disproportionately concentrated in North America, Western Europe, and Australia; vast agricultural areas of sub-Saharan Africa, South Asia, and Central Asia are severely underrepresented [111]. ML models trained on spatially biased datasets exhibit systematically degraded prediction accuracy in underrepresented regions, propagating inequities in agricultural productivity data into global soil and water assessments.

Data quality limitations compound the scarcity problem. Sensor drift, cable damage, and power outages in field-deployed IoT networks introduce systematic gaps and outliers that corrupt training signals if not rigorously screened by automated quality-control algorithms [181]. Label noise arising from measurement errors in soil reference samples, inconsistent analytical protocols between laboratories, and spatial support mismatch between point observations and gridded model targets imposes a practical accuracy ceiling that cannot be overcome by model architecture improvements alone. Temporal gaps caused by cloud cover in optical satellite time series are frequently non-random—cloud cover covaries with rainfall events that are of primary hydrological interest—introducing systematic biases in models trained on cloud-filtered archives. Addressing these data infrastructure limitations through investment in globally distributed sensor networks, standardized measurement protocols, and open data sharing platforms is a prerequisite for realizing the full predictive potential of AI in agricultural water management.

8.2. Computational Complexity

Large-scale distributed deep learning models require substantial computational resources and infrastructure. Training state-of-the-art deep learning architectures for soil–water applications—particularly transformer-based foundation models and spatiotemporal ensemble systems—demands computational resources that are unavailable in most agricultural research institutions and entirely inaccessible in developing-country contexts. The energy consumption of large-scale Graphics Processing Unit (GPU) cluster training carries a significant carbon footprint that is increasingly scrutinized from a sustainability perspective: estimates suggest that training a large transformer language model can emit hundreds of tonnes of CO_2 equivalent, a concern that extends to geoscience foundation models trained on comparable data volumes [244]. Memory requirements for processing high-resolution spatiotemporal image sequences—as required for field-scale soil moisture downscaling or regional DSM—often exceed the capacity of individual GPU devices, necessitating model and data parallelism strategies with associated communication overhead and engineering complexity.

Inference latency introduces a further practical constraint for real-time irrigation control applications: edge-deployed AI processors on battery-powered IoT nodes must produce soil moisture estimates and irrigation decisions within milliseconds while consuming milliwatts of power, imposing severe model size and computational budget constraints that conflict with the accuracy requirements of precision management. Bridging the gap between the computational demands of high-accuracy deep learning models and the resource constraints of edge deployment is a fundamental systems engineering challenge that requires advances in hardware-aware neural architecture design, model compression, and energy-efficient inference acceleration.

8.3. Model Interpretability

Black-box AI systems often limit stakeholder trust and operational transparency. The opacity of deep neural network models constitutes a structural barrier to their adoption in agricultural decision-support contexts, where farmers, irrigation managers, and water regulatory agencies require not only accurate predictions but also intelligible explanations of the reasoning underlying model outputs [24]. When a LSTM-based irrigation controller

recommends withholding irrigation during a phenologically sensitive crop growth stage, or a gradient boosting model predicts a substantially elevated salinity risk in a specific field zone, practitioners need to understand which input factors drove these recommendations, whether the model is functioning within its domain of competence, and what the uncertainty bounds on its predictions are. Without this understanding, the adoption of AI tools remains limited to technically sophisticated early adopters and fails to penetrate the broad agricultural practitioner community.

Current XAI methods—including SHAP, LIME, and gradient-based attribution—provide useful but imperfect approximations to model explanation: they can be computationally expensive, sensitive to correlated input features, and may produce misleading attributions under distribution shift [234]. Developing XAI approaches that are provably faithful to model behavior, efficient enough for routine use in operational agricultural AI systems, and expressible in terms of domain concepts (soil classes, phenological stages, drought severity indices) meaningful to agronomists and farmers remains an active and insufficiently resourced research frontier.

8.4. Integration with Physical Models

Hybrid frameworks integrating physics-based hydrological models with AI approaches remain an important research frontier. Despite the empirical success of purely data-driven AI models in soil–water prediction tasks, their inability to guarantee physically consistent outputs under distributional shift—predicting soil moisture values outside the physically plausible range, violating water balance closure, or producing conductivity estimates inconsistent with the van Genuchten–Mualem framework—limits their reliability in operational hydrological applications and under novel climate-forcing conditions [165]. Physics-informed neural networks (PINNs) that embed governing partial differential equations as loss function constraints represent a technically promising but computationally challenging approach to addressing this limitation: optimizing PINN loss functions that combine data fit and partial differential equation (PDE) residual terms is numerically ill conditioned for stiff PDEs, such as the Richards equation, and practical PINNs for vadose zone flow remain an active research area. Modular hybrid frameworks—in which AI surrogate models replace computationally expensive components of established physically-based simulators while the overall water balance and process structure are maintained—offer a pragmatic near-term pathway for combining the physical consistency of process models with the data assimilation flexibility and pattern-recognition capacity of deep learning.

8.5. Real-Time Deployment Constraints

Latency, connectivity limitations, and energy requirements constrain real-time deployment in rural agricultural environments. The translation of AI-based soil–water models from research prototypes to operational field deployments is frequently impeded by infrastructure and connectivity constraints that are especially acute in smallholder and developing-country agricultural contexts. Reliable internet connectivity—required for cloud-based AI inference and remote model updating—is unavailable across large fractions of the world's agricultural land, particularly in sub-Saharan Africa and South Asia, where water management improvements would generate the greatest food security benefits [240]. Edge AI deployment on locally powered IoT nodes mitigates this connectivity dependency but introduces its own constraints: model inference must complete within strict latency budgets; hardware must withstand field temperature extremes, humidity, and mechanical vibration; and power must be supplied by solar panels or batteries with limited capacity. Firmware updates for edge-deployed models across spatially distributed sensor networks require secure over-the-air update protocols and robust version management systems that

represent a significant operational engineering burden. Harmonizing these deployment realities with the accuracy and latency requirements of precision irrigation control constitutes a systems-level challenge that demands closer collaboration between AI researchers and agricultural engineering practitioners.

8.6. Model Validation, Uncertainty Analysis, and Data Quality Effects

The practical adoption of AI-based soil–water systems depends critically on rigorous model validation procedures and transparent uncertainty quantification, yet these topics receive insufficient coverage in most published studies. A persistent methodological weakness in the reviewed literature is the use of random cross-validation splits that do not account for spatial autocorrelation in soil property datasets: when training and test samples are spatially proximate, cross-validated accuracy estimates are optimistically biased because the model can exploit spatial structure rather than learning generalizable pedological relationships. Spatially blocked cross-validation—partitioning observations into spatially contiguous folds that enforce minimum separation distances between training and test sets—is the methodologically correct evaluation strategy for soil mapping and field-scale prediction tasks and should be adopted as the standard reporting requirement. Temporal holdout validation, in which models are evaluated on complete seasons or years withheld from training, is the appropriate strategy for time-series soil moisture and irrigation forecasting tasks, as it tests the model's ability to generalize forward in time under potentially novel forcing conditions.

Sensitivity analysis of AI model predictions with respect to input data quality—quantifying how much the prediction error increases as a function of the sensor noise level, temporal gap frequency, spatial sampling density, or label measurement error—provides essential guidance for field deployment design and data collection investment decisions and yet is rarely reported alongside accuracy benchmarks. Uncertainty quantification frameworks—whether Bayesian (Monte Carlo Dropout, variational inference), ensemble-based (deep ensembles, Random Forest prediction intervals), or conformal prediction—should be incorporated into AI soil–water models from the design stage rather than added post hoc, as the predictive uncertainty is as operationally consequential as the point estimate for irrigation decisions near the wilting point threshold or drought early-warning triggers.

The effect of training data quality on AI model prediction reliability is further compounded by label noise: soil moisture measurements from field sensors carry calibration uncertainties of 0.02–0.05 m^3 m^{-3}, which sets a practical accuracy floor that cannot be overcome by model architecture improvements, and should be explicitly acknowledged when reporting RMSE values approaching this range. Developing standardized reporting guidelines for validation design, uncertainty communication, and sensitivity analysis in AI soil–water modeling studies would substantially improve the comparability and practical credibility of published results.

9. Future Research Directions and Technical Horizons

The systemic bottlenecks hindering widespread field implementation of agricultural AI frameworks, alongside corresponding next-generation research opportunities, are detailed below. Despite rapid progress, the field faces several unresolved challenges that define specific and actionable future research priorities (Table 7):

1. The absence of standardized benchmark datasets for soil moisture prediction, pedotransfer function development, and digital soil mapping impedes rigorous cross-study comparison and slows the cumulative scientific progress achievable through reproducible benchmarking; establishing openly accessible, globally distributed benchmark

corpora analogous to ImageNet for computer vision represents a critical infrastructure investment.

2. Improving the transferability of AI models across agroecological regions requires systematic investigation of domain adaptation architectures—including meta-learning, domain adversarial training, and few-shot fine-tuning strategies—benchmarked against consistent leave-region-out evaluation protocols that honestly assess cross-regional generalization failure modes.
3. The integration of physics-informed AI models with operational precision irrigation systems requires the development of computationally efficient PINN architectures capable of training at field-to-catchment scales, with particular attention to the numerical stiffness of the Richards equation and the spatial heterogeneity of vadose zone hydraulic properties.
4. Advancing privacy-preserving federated learning for agricultural decision support requires the development of communication-efficient aggregation algorithms robust to the extreme non-IID data distributions arising from soil and climate heterogeneity across participating farm nodes, as well as legally enforceable data governance frameworks aligned with national agricultural data sovereignty policies.

Table 7. Future research gaps and opportunities in AI-driven agriculture.

Research Domain	Identified Knowledge and Technical Gap	Emerging Opportunity and Solution	Key Sources
Model Interpretability	Standard deep architectures function as "black boxes," causing low practitioner trust when models output highly irregular irrigation schedules.	Development of explainable AI (XAI) frameworks (e.g., SHAP, integrated gradients) to map physical features to neuron activations.	[245]
Physical Consistency	Purely data-driven models frequently output predictions that violate fundamental laws of thermodynamics and mass conservation.	Integration of physics-informed neural networks (PINNs) that bake Darcy's law and the Richards equation directly into the loss function.	
Data Privacy and Sovereignty	Centralized cloud storage of granular farm telemetry creates corporate surveillance vulnerabilities and data ownership disputes.	Implementing federated learning (FL) to train localized models on edge gateways without moving raw agricultural data off-site.	[246]
Cross-Regional Generalization	Models optimized for a specific soil type or micro-climate exhibit drastic performance degradation when deployed in unseen regions.	Development of multimodal Agricultural Foundation Models pretrained on continental remote sensing and fine-tuned locally.	
Edge Compute Constraints	High-fidelity transformer networks are computationally too heavy for low-power, solar-powered microcontroller valves in the field.	Application of hardware-aware knowledge distillation and structural network pruning to compress multi-million parameter models.	

9.1. Physics-Informed AI

Physics-informed neural networks (PINNs) offer opportunities to integrate physical laws with data-driven learning for improved generalization. The integration of domain physical knowledge with deep learning through physics-guided loss functions represents one of the most promising frontiers for improving the reliability and generalizability of AI-based soil–water models [20].

By embedding differential equation constraints—such as the Richards equation governing unsaturated flow, Darcy's law relating flux to the hydraulic gradient, or the water balance equation linking precipitation, evapotranspiration, runoff, and soil moisture storage change—directly into the neural network loss function, PINNs constrain predictions

to remain consistent with known physical laws, even in data-sparse regions where purely data-driven models are prone to physically implausible extrapolation. The combination of PINNs with ensemble uncertainty quantification and Bayesian parameter inference frameworks would provide a principled mechanism for simultaneously learning soil hydraulic parameters, predicting soil moisture dynamics, and quantifying prediction uncertainty within a unified physics-consistent model. Hybrid architectures that couple PINN-based PDE solvers with data-driven parameterization of spatially variable hydraulic properties—learning the spatially distributed van Genuchten parameters as neural network outputs rather than prescribing them from PTFs—represent a particularly compelling direction for reducing the parameter estimation burden of physically-based soil–water models without sacrificing their physical interpretability.

9.2. Federated Learning in Agriculture

Federated learning can facilitate collaborative agricultural intelligence while preserving privacy and decentralization. Federated learning (FL) offers a technically and legally robust mechanism for aggregating AI model training across geographically distributed farms, research stations, and national agricultural networks without requiring sensitive farm-level data to be centralized or disclosed [247]. The agricultural domain presents distinctive challenges for FL that motivate important future research directions. Non-IID data heterogeneity—arising from the pronounced variation in soil types, climate regimes, crop species, and management practices across participating nodes—causes significant performance degradation under standard FedAvg aggregation, necessitating personalized FL approaches that maintain local model components adapted to site-specific conditions while sharing globally beneficial representations.

Communication-efficient FL algorithms that compress gradient updates through quantization, sparsification, or local Stochastic Gradient Descent (local SGD, in which each client performs multiple gradient steps between communication rounds) with infrequent synchronization are needed to accommodate the limited and intermittent bandwidth of rural agricultural networks. The combination of FL with differential privacy mechanisms—adding calibrated noise to local model updates before transmission to prevent inference of individual farm records from the aggregate model—requires careful calibration to balance privacy protection against the accuracy degradation imposed by noise injection, particularly for small-farm nodes with limited local training data. Future FL systems for agricultural AI should integrate standardized data harmonization protocols, interoperable communication interfaces, and transparent audit mechanisms that enable participating farms to verify the integrity of the aggregation process and retain meaningful control over their data contributions, in alignment with emerging agricultural data governance frameworks.

9.3. Edge AI and Smart Farming

Edge computing and low-power AI devices will enable autonomous irrigation and real-time field-scale analytics. The continuing miniaturization and energy efficiency improvements of neuromorphic and application-specific integrated circuit (ASIC) hardware, combined with advances in model compression through knowledge distillation, quantization-aware training, and neural architecture search, are progressively closing the gap between the computational requirements of state-of-the-art deep learning models and the resource budgets of field-deployable edge processors.

Future smart farming systems will likely comprise hierarchical edge–fog–cloud AI architectures in which ultra-low-power microcontrollers at the sensor node level run highly compressed models for real-time anomaly detection and actuation control; see Figure 10.

Among others, intermediate fog gateway processors aggregate sensor streams and run more complex soil moisture and evapotranspiration prediction models; cloud or high-performance computing (HPC) resources periodically retrain global models and push updated lightweight models to the edge tier. Autonomous variable-rate irrigation systems—comprising distributed wireless sensor arrays, AI-driven soil moisture prediction at sub-field resolution, weather-forecast-integrated optimization, and automated valve actuation—represent a realistically near-term technology convergence that could substantially reduce agricultural water consumption in irrigated systems. Standardizing sensor communication protocols (LoRaWAN, NB-IoT, MQTT), data formats, and AI model interchange standards across hardware vendors and platform providers is essential to enable the interoperability required for scaling these systems from experimental deployments to commercial precision agriculture ecosystems.

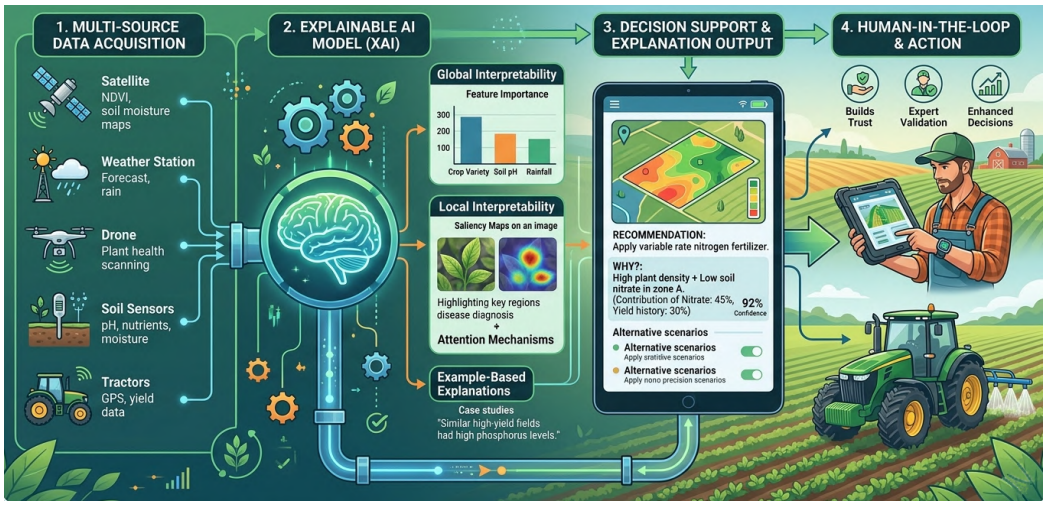

Figure 10. Future ecosystem of AI-enabled sustainable agriculture. Software: Midjourney v. 6.1 for infographic illustration of farm scenes; Inkscape v. 1.4.4 for figure composition and layout. Diagram source: author.

9.4. Digital Twins for Soil–Water Systems

Digital twins (DTs) can simulate soil–water interactions and optimize agricultural management under dynamic environmental conditions. DT technology—creating real-time, continuously updated virtual replicas of physical agricultural systems that synchronize observations from sensor networks and satellite platforms with AI and physically-based simulation models—offers transformative potential for soil–water management and climate adaptation planning [22,23,248,249]. An agricultural digital twin integrates data streams from field sensors, satellite observations, weather forecasts, and agronomic management records into a dynamically updated state representation of the soil–water–plant–atmosphere system, which serves as both a predictive engine for forward simulation under different irrigation and climate scenarios and a decision-support interface for farm managers and water resource planners.

ML surrogate models trained to emulate high-fidelity HYDRUS or SWAT simulations at low computational cost are the enabling technology for real-time DT updating: they provide sub-second response times necessary for interactive what-if scenario exploration while maintaining quantitative consistency with the physical process representations embedded in the reference simulators. Future research should address the development of DT architectures that quantify and propagate model and observational uncertainties throughout the virtual system state, provide rigorous validation frameworks for assessing DT predictive fidelity under novel forcing conditions, and support multi-stakeholder decision-making

across the spatial scales relevant to both individual farm management and watershed-level water allocation governance.

The integration of agricultural DTs with climate service platforms—providing dynamically downscaled seasonal climate forecasts as boundary conditions for soil–water system simulation—would enable adaptive water management that anticipates seasonal rainfall variability months in advance rather than merely reacting to observed soil moisture deficits.

10. Conclusions

Artificial intelligence, ML, and distributed deep learning are transforming soil science and sustainable agricultural water management. AI-driven frameworks provide powerful tools for soil moisture prediction, irrigation optimization, digital soil mapping, and climate resilience. Despite significant progress, challenges related to data quality, interpretability, computational requirements, and real-time deployment remain unresolved. Future research should emphasize explainable AI, physics-informed learning, federated intelligence, and integrated digital agricultural ecosystems to support resilient and sustainable food production systems. Future research should prioritize the development of computationally efficient PINN architectures tailored to the numerical stiffness and spatial heterogeneity characteristic of vadose zone flow, including adaptive collocation point sampling strategies and domain decomposition methods that enable PINN training at field-to-catchment scales.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the reviewers for reading and reviewing this manuscript. AI Use Statement: Artificial intelligence tools were used during the preparation of this manuscript. Specifically, Grammarly (https://www.grammarly.com/) was used for English language polishing, including grammar checking, sentence restructuring, and improving the clarity and readability of the text. The AI-generated suggestions were critically reviewed and edited by the author, who takes full responsibility for the final content. In addition, Midjourney v. 6.1 was used for the generation of the infographic illustration of farm scenes included in Figure 10. The figure was subsequently composed and assembled using Inkscape v. 1.4.4 by the author.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
DL	Deep Learning
DSM	Digital Soil Mapping
DQN	Deep Q-Network
FL	Federated Learning
GNN	Graph Neural Network
IoT	Internet of Things
IPCC	Intergovernmental Panel on Climate Change
LSTM	Long Short-Term Memory

MAE	Mean Absolute Error
MDPI	Multidisciplinary Digital Publishing Institute
ML	Machine Learning
MLP	Multi-Layer Perceptron
MODFLOW	Modular Groundwater Flow Model
NSE	Nash–Sutcliffe Efficiency
PAWC	Plant-Available Water Capacity
PINN	Physics-Informed Neural Network
PTF	Pedotransfer Function
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SAR	Synthetic Aperture Radar
SHAP	SHapley Additive exPlanations
SMAP	Soil Moisture Active Passive
SOC	Soil Organic Carbon
SVM	Support Vector Machine
SVR	Support Vector Regression
SWAT	Soil and Water Assessment Tool
TDR	Time-Domain Reflectometry
UAV	Unmanned Aerial Vehicle
ViT	Vision Transformer
XGBoost	Extreme Gradient Boosting

References

1. Food and Agriculture Organization of the United Nations. *The State of the World's Land and Water Resources for Food and Agriculture: Systems at Breaking Point*; FAO: Rome, Italy, 2020.
2. IPCC. Climate Change 2022: Impacts, Adaptation and Vulnerability. In *Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2022. [CrossRef]
3. Galland, F. Climate Change and Water Resources: Excess, Shortage, and Pollution. *Field Actions Sci. Rep.* **2025**, *27*, 80–84.
4. van Genuchten, M.T. A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [CrossRef]
5. Lehmann, J.; Bossio, D.A.; Kögel-Knabner, I.; Rillig, M.C. The Concept and Future Prospects of Soil Health. *Nat. Rev. Earth Environ.* **2020**, *1*, 544–553. [CrossRef] [PubMed]
6. Ran, Y.; Li, X.; Lu, L. Evaluation of four remote sensing based land cover products over China. *Int. J. Remote Sens.* **2010**, *31*, 391–401. [CrossRef]
7. Philip, J.R. The theory of infiltration: 1. The infiltration equation and its solution. *Soil Sci.* **1957**, *83*, 345–357. [CrossRef]
8. Green, W.H.; Ampt, C.A. Studies on Soil Physics: I. The Flow of Air and Water through Soils. *J. Agric. Sci.* **1911**, *4*, 1–24. [CrossRef]
9. Thornthwaite, C.W.; Mather, J.R. The water balance. *Climatology* **1955**, *8*, 1–104.
10. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: Model development. *J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89. [CrossRef]
11. Šimůnek, J.; Brunetti, G.; Jacques, D.; van Genuchten, M.T.; Šejna, M. Developments and Applications of the HYDRUS Computer Software Packages since 2016. *Vadose Zone J.* **2024**, *23*, e20310. [CrossRef]
12. Harbaugh, A.W. *MODFLOW-2005, The U.S. Geological Survey Modular Ground-Water Model—The Ground-Water Flow Process*; Techniques and Methods 6-A16; US Department of the Interior, US Geological Survey: Reston, VA, USA, 2005.
13. Richards, L.A. Capillary conduction of liquids through porous mediums. *Physics* **1931**, *1*, 318–333. [CrossRef]
14. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
15. Maier, H.R.; Dandy, G.C. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Model. Softw.* **2000**, *15*, 101–124. [CrossRef]
16. Elshorbagy, A.; Corzo, G.; Srinivasulu, S.; Solomatine, D.P. Experimental investigation of the predictive capabilities of data-driven modeling techniques in hydrology—Part 1: Concepts and methodology. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1931–1941. [CrossRef]
17. Kratzert, F.; Klotz, D.; Brenner, B.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [CrossRef]

18. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
19. Shen, C. A transdisciplinary review of deep learning the hydrologic sciences. *Water Resour. Res.* **2018**, *54*, 5965–5981. [CrossRef]
20. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
21. Lai, N.; Dewi, D.A.; Maidin, S.S.; Xiao, W.; Zhao, S.; Hu, Q. A comprehensive review of lightweight deep learning models for edge computing with future directions. *Discov. Comput.* **2026**, *29*, 110. [CrossRef]
22. Pylianidis, C.; Osinga, S.; Athanasiadis, I.N. Introducing digital twins to agriculture. *Comput. Electron. Agric.* **2021**, *184*, 105942. [CrossRef]
23. Verdouw, C.; Tekinerdogan, B.; Beulens, A.; Wolfert, S. Digital twins in smart farming. *Agric. Syst.* **2021**, *189*, 103046. [CrossRef]
24. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, 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]
25. Sit, M.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A Comprehensive Review of Deep Learning Applications in Hydrology and Water Resources. *Water Sci. Technol.* **2020**, *82*, 2635–2670. [CrossRef] [PubMed]
26. Tripathy, K.P.; Mishra, A.K. Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. *J. Hydrol.* **2024**, *628*, 130458. [CrossRef]
27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
28. Vij, A.; Vijendra, S.; Jain, A.; Bajaj, S.; Bassi, A.; Sharma, A. IoT and Machine Learning Approaches for Automation of Farm Irrigation System. *Procedia Comput. Sci.* **2020**, *167*, 1250–1257. [CrossRef]
29. Ben-Nun, T.; Hoefler, T. Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis. *ACM Comput. Surv.* **2019**, *52*, 1–43. [CrossRef]
30. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
31. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
32. Tariq, M.U.; Saqib, S.M.; Mazhar, T.; Khan, M.A.; Shahzad, T.; Hamam, H. Edge-enabled smart agriculture framework: Integrating IoT, lightweight deep learning, and agentic AI for context-aware farming. *Results Eng.* **2025**, *28*, 107342. [CrossRef]
33. Gong, R.; Zhang, H.; Li, G.; He, J. Edge Computing-Enabled Smart Agriculture: Technical Architectures, Practical Evolution, and Bottleneck Breakthroughs. *Sensors* **2025**, *25*, 5302. [CrossRef] [PubMed]
34. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [CrossRef]
35. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [CrossRef]
36. Paria, A.; Giri, A.; Dutta, S.; Neogy, S. Smart Irrigation for Coriander Plant: Saving Water with AI and IoT. *Water Resour. Manag.* **2025**, *39*, 3379–3395. [CrossRef]
37. Khanna, A.; Kaur, S. Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Comput. Electron. Agric.* **2019**, *157*, 218–231. [CrossRef]
38. Ayaz, M.; Ammad-Uddin, M.; Sharif, Z.; Mansour, A.; Aggoune, E.H.M. Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE Access* **2019**, *7*, 129551–129583. [CrossRef]
39. Dean, J.; Corrado, G.; Monga, R.; Chen, K.; Devin, M.; Mao, M.; Ranzato, M.A.; Senior, A.; Tucker, P.; Yang, K.; et al. Large Scale Distributed Deep Networks. In *Advances in Neural Information Processing Systems: Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 1*; Pereira, F., Burges, C., Bottou, L., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; NIPS'12; Volume 25, pp. 1223–1231.
40. Verbraeken, J.; Wolting, M.; Katzy, J.; Kloppenburg, J.; Verbelen, T.; Rellermeyer, J.S. A survey on distributed machine learning. *ACM Comput. Surv.* **2020**, *53*, 1–33. [CrossRef]
41. Bonomi, F.; Milito, R.; Zhu, J.; Addepalli, S. Fog computing and its role in the internet of things. In Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, Helsinki, Finland, 17 August 2012 ; pp. 13–16. [CrossRef]
42. Kalyani, Y.; Collier, R. A Systematic Survey on the Role of Cloud, Fog, and Edge Computing Combination in Smart Agriculture. *Sensors* **2021**, *21*, 5922. [CrossRef] [PubMed]
43. Guardo, E.; Stefano, A.D.; Corte, A.L.; Sapienza, M.; Scatà, M. A Fog Computing-based IoT Framework for Precision Agriculture. *J. Syst. Archit.* **2018**, *19*, 1401–1411. [CrossRef]
44. Ferrag, M.A.; Shu, L.; Yang, X.; Derhab, A.; Maglaras, L. Security and Privacy for Green IoT-Based Agriculture: Review, Blockchain Solutions, and Challenges. *IEEE Access* **2020**, *8*, 32031–32053. [CrossRef]
45. Nayak, A.; Raghatate, K.S.; Negi, G.S. Blockchain-Enabled Decentralized Water Management System (BD-WMS) for Sustainable Irrigation. *Int. J. Basic Appl. Sci.* **2025**, *14*, 344–351. [CrossRef]

46. Casino, F.; Dasaklis, T.K.; Patsakis, C. A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telemat. Inform.* **2019**, *36*, 55–81. [CrossRef]
47. Reyna, A.; Martín, C.; Chen, J.; Soler, E.; Díaz, M. On blockchain and its integration with IoT. Challenges and opportunities. *Future Gener. Comput. Syst.* **2018**, *88*, 173–190. [CrossRef]
48. Alves, R.G.; Maia, R.F.; Lima, F. Development of a Digital Twin for smart farming: Irrigation management system for water saving. *J. Clean. Prod.* **2023**, *388*, 135920. [CrossRef]
49. Wooldridge, M. *An Introduction to MultiAgent Systems*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
50. Dragomir, O.E.; Dragomir, F. A Decentralized Hierarchical Multi-Agent Framework for Smart Grid Sustainable Energy Management. *Sustainability* **2025**, *17*, 5423. [CrossRef]
51. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [CrossRef]
52. European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). *Off. J. Eur. Union* **2016**, *OJ L 119*, 1–88.
53. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
54. Hillel, D. 19—Plant Uptake of Soil Moisture. In *Introduction to Environmental Soil Physics*; Academic Press: Burlington, VT, USA, 2003; pp. 365–384.
55. Lindh, P.; Lemenkova, P. Effects of Water—Binder Ratio on Strength and Seismic Behavior of Stabilized Soil from Kongshavn, Port of Oslo. *Sustainability* **2023**, *15*, 12016. [CrossRef]
56. Hillel, D. 21—Irrigation and Water-Use Efficiency. In *Introduction to Environmental Soil Physics*; Academic Press: Burlington, VT, USA, 2003; pp. 407–425.
57. Lindh, P.; Lemenkova, P. Dynamics of Strength Gain in Sandy Soil Stabilised with Mixed Binders Evaluated by Elastic P-Waves during Compressive Loading. *Materials* **2022**, *15*, 7798. [CrossRef] [PubMed]
58. Çal, S.; Barik, K. Hydraulic Conductivity Values of Soils in Different Soil Processing Conditions. *Alinteri J. Agric. Sci.* **2020**, *35*, 132–138. [CrossRef]
59. de Oliveira, J.A.; Cássaro, F.A.; Pires, L.F. Estimating soil porosity and pore size distribution changes due to wetting-drying cycles by morphometric image analysis. *Soil Tillage Res.* **2021**, *205*, 104814. [CrossRef]
60. Rabot, E.; Wiesmeier, M.; Schlüter, S.; Vogel, H.J. Soil structure as an indicator of soil functions: A review. *Geoderma* **2018**, *314*, 122–137. [CrossRef]
61. Assouline, S. Infiltration into soils: Conceptual approaches and solutions. *Water Resour. Res.* **2013**, *49*, 1755–1772. [CrossRef]
62. Chaudhari, P.R.; Ahire, D.V.; Ahire, V.D.; Chkravarty, M.; Maity, S. Soil Bulk Density as related to Soil Texture, Organic Matter and Available Water Capacity of Soils of Central India. *Int. J. Sci. Res. Publ.* **2013**, *3*, 1–8.
63. Keller, T.; Lamandé, M.; Schjønning, P.; Dexter, A.R. Analysis of soil compression curves from uniaxial confined compression tests. *Geoderma* **2011**, *163*, 13–23. [CrossRef]
64. Amsili, J.P.; van Es, H.M.; Schindelbeck, R.R. Pedotransfer Functions for Field Capacity, Permanent Wilting Point, and Available Water Capacity Based on Random Forest Models for Routine Soil Health Analysis. *Commun. Soil Sci. Plant Anal.* **2024**, *55*, 1967–1984. [CrossRef]
65. Twarakavi, N.K.C.; Sakai, M.; Šimůnek, J. An objective analysis of the dynamic nature of field capacity. *Water Resour. Res.* **2009**, *45*, W10410. [CrossRef]
66. Nachabe, M.H. Refining the definition of field capacity in the literature. *J. Irrig. Drain. Eng.* **1998**, *124*, 230–232. [CrossRef]
67. Gamagedara, Y.; Wijewardane, N.K.; Feng, G. Soil Hydraulic Properties Database: Volumetric Water Content and Matric Potential Parameters Using Centrifugation Workflows. Mendeley Data, V1, 2025. Available online: https://data.mendeley.com/datasets/2x57v8wbbv/1 (accessed on 1 June 2026).
68. Kirkham, M. 7—Water Movement in Saturated Soil. In *Principles of Soil and Plant Water Relations*; Academic Press: Burlington, VT, USA, 2005; pp. 85–100. [CrossRef]
69. Pidgeon, J.D. The Measurement and Prediction of Available Water Capacity of Ferrallitic Soils in Uganda. *J. Soil Sci.* **1972**, *23*, 431–441. [CrossRef]
70. Altobelli, F.; Vargas, R.; Corti, G.; Dazzi, C.; Montanarella, L.; Monteleone, A.; Caon, L.; Piazza, M.G.; Calzolari, C.; Munafò, M.; et al. Improving soil and water conservation and ecosystem services by sustainable soil management practices: From a global to an Italian soil partnership. *Ital. J. Agron.* **2020**, *15*, 1765. [CrossRef]
71. Lal, R.; Shukla, M.K. *Principles of Soil Physics*; CRC Press: Boca Raton, FL, USA, 2004; p. 736. [CrossRef]
72. Veihmeyer, F.J.; Hendrickson, A.H. The moisture equivalent as a measure of the field capacity of soils. *Soil Sci.* **1931**, *32*, 181–194. [CrossRef]

73. Mualem, Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* **1976**, *12*, 513–522. [CrossRef]
74. Famiglietti, J.S.; Ryu, D.; Berg, A.A.; Rodell, M.; Jackson, T.J. Field observations of soil moisture variability across scales. *Water Resour. Res.* **2008**, *44*, W01423. [CrossRef]
75. Alkassem Alosman, M.; Ruy, S.; Buis, S.; Lecharpentier, P.; Bader, J.C.; Charron, F.; Olioso, A. An Improved Method to Estimate Soil Hydrodynamic and Hydraulic Roughness Parameters by Using Easily Measurable Data During Flood Irrigation Experiments and Inverse Modelling. *Water* **2018**, *10*, 1581. [CrossRef]
76. Hodnett, M.G.; Tomasella, J. Marked differences between van Genuchten soil water retention parameters for temperate and tropical soils: A new set of pedotransfer functions. *Geoderma* **2002**, *108*, 155–180. [CrossRef]
77. Vereecken, H.; Maes, J.; Feyen, J.; Darius, P. Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. *Soil Sci.* **1989**, *148*, 389–403. [CrossRef]
78. Dane, J.H.; Topp, G.C. *Methods of Soil Analysis, Part 4: Physical Methods*; Soil Science Society of America: Washington, DC, USA, 2002. [CrossRef]
79. Waleed, M.; Inam, M.A.; Albano, R.; Samad, A.; Farid, H.U.; Shoaib, M.; Ali, M.U. Statistical Model Development for Estimating Soil Hydraulic Conductivity Through On-Site Investigations. *Hydrology* **2025**, *12*, 55. [CrossRef]
80. Gardner, W.R. Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. *Soil Sci.* **1958**, *85*, 228–232. [CrossRef]
81. Ali, U.; Ali, A.; Aziz, M.; Umar, M.; Shah, S.K.H.; Malik, A.A. Capillary force models for unsaturated granular soils using discrete element method: A review. *J. Rock Mech. Geotech. Eng.* 2026, *in press*. [CrossRef]
82. Zhao, Z.; Luo, Z.; Sun, H.; Li, H.; Liu, Q.; Liu, H. Capillary Rise in Layered Soils. *Appl. Sci.* **2023**, *13*, 3374. [CrossRef]
83. Flint, L.E.; Flint, A.L.; Stern, M.A.; Mayer, A.; Silver, W.L.; Casey, C.; Franco, F.; Byrd, K.B.; Sleeter, B.M.; Alvarez, P.; et al. *Increasing Soil Organic Carbon to Mitigate Greenhouse Gases and Increase Climate Resiliency for California*; Technical Report CCCA4-CNRA-2018-006; California Natural Resources Agency: Sacramento, CA, USA, 2018.
84. Hudson, B.D. Soil organic matter and available water capacity. *J. Soil Water Conserv.* **1994**, *49*, 189–194. [CrossRef]
85. Rawls, W.J.; Pachepsky, Y.A.; Ritchie, J.C.; Sobecki, T.M.; Bloodworth, H. Effect of soil organic carbon on soil water retention. *Geoderma* **2003**, *116*, 61–76. [CrossRef]
86. Šimůnek, J.; Sejna, M.; Saito, H.; Sakai, M.; van Genuchten, M.T. The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media. *Univ. Calif. Riverside Res. Rep.* **2008**, *3*, 1–240.
87. Wösten, J.H.M.; Pachepsky, Y.A.; Rawls, W.J. Pedotransfer Functions: Bridging the Gap between Available Basic Soil Data and Missing Soil Hydraulic Characteristics. *J. Hydrol.* **2001**, *251*, 123–150. [CrossRef]
88. Lindh, P.; Lemenkova, P. Laboratory Experiments on Soil Stabilization to Enhance Strength Parameters for Road Pavement. *Transp. Telecommun. J.* **2023**, *24*, 73–82. [CrossRef]
89. Lindh, P.; Lemenkova, P. Behaviour of Moraine Soils Stabilised with OPC, GGBFS and Hydrated Lime. *Arch. Min. Sci.* **2023**, *68*, 319–334. [CrossRef]
90. Vereecken, H.; Huisman, J.A.; Bogena, H.; Vanderborght, J.; Vrugt, J.A.; Hopmans, J.W. On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resour. Res.* **2008**, *44*, W00D06. [CrossRef]
91. Safriel, U.N. The Assessment of Global Trends in Land Degradation. In *Climate and Land Degradation*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1–38. [CrossRef]
92. Oldeman, L.R. Global Extent of Soil Degradation. In *ISRIC Bi-Annual Report 1991–1992*; International Soil Reference and Information Centre (ISRIC): Wageningen, The Netherlands, 1992; pp. 19–36.
93. Nachshon, U. Cropland Soil Salinization and Associated Hydrology: Trends, Processes and Examples. *Water* **2018**, *10*, 1030. [CrossRef]
94. Lindh, P.; Lemenkova, P. Permeability, compressive strength and Proctor parameters of silts stabilised by Portland cement and ground granulated blast furnace slag (GGBFS). *Arch. Mech. Eng.* **2022**, *69*, 667–692. [CrossRef]
95. van Klompenburg, T.; Kassahun, A.; Peng, C. Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. *Comput. Electron. Agric.* **2019**, *177*, 105709. [CrossRef]
96. Fatichi, S.; Vivoni, E.R.; Ogden, F.L.; Ivanov, V.Y.; Mirus, B.; Gochis, D.; Downer, C.W.; Camporese, M.; Davison, J.H.; Ebel, B.; et al. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J. Hydrol.* **2016**, *537*, 45–60. [CrossRef]
97. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
98. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [CrossRef]

99. Amani, S.; Shafizadeh-Moghadam, H. A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. *Agric. Water Manag.* **2023**, *284*, 108324. [CrossRef]
100. Attri, I.; Awasthi, L.K.; Sharma, T.P.; Rathee, P. A review of deep learning techniques used in agriculture. *Ecol. Inform.* **2023**, *77*, 102217. [CrossRef]
101. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
102. Kavzoglu, T.; Tso, B.; Mather, P.M. *Classification Methods for Remotely Sensed Data*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2024. [CrossRef]
103. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2021; p. 704.
104. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; Number 56 in FAO Irrigation and Drainage Paper; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
105. Gao, F.; Feng, G.; Ouyang, Y.; Wang, H.; Fisher, D.; Adeli, A.; Jenkins, J. Evaluation of Reference Evapotranspiration Methods in Arid, Semiarid, and Humid Regions. *JAWRA J. Am. Water Resour. Assoc.* **2017**, *53*, 791–808. [CrossRef]
106. Konukcu, F. Modification of the Penman method for computing bare soil evaporation. *Hydrol. Processes* **2007**, *21*, 3627–3634. [CrossRef]
107. Scott, R.L.; Knowles, J.F.; Nelson, J.A.; Gentine, P.; Li, X.; Barron-Gafford, G.; Bryant, R.; Biederman, J.A. Water Availability Impacts on Evapotranspiration Partitioning. *Agric. For. Meteorol.* **2021**, *297*, 108251. [CrossRef]
108. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
109. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]
110. Padarian, J.; McBratney, A.B.; Minasny, B. Game theory interpretation of digital soil mapping convolutional neural networks. *SOIL* **2020**, *6*, 389–397. [CrossRef]
111. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.M.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [CrossRef] [PubMed]
112. Carranza, C.; Nolet, C.; Pezij, M.; van der Ploeg, M. Root zone soil moisture estimation with Random Forest. *J. Hydrol.* **2021**, *593*, 125840. [CrossRef]
113. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
114. Biau, G.; Scornet, E. A random forest tutorial. *Test* **2016**, *25*, 197–227. [CrossRef]
115. Gill, M.K.; Asefa, T.; Kemblowski, M.W.; McKee, M. Soil moisture forecasting using support vector machines. *J. Am. Water Resour. Assoc.* **2006**, *42*, 1035–1046. [CrossRef]
116. Asefa, T.; Kemblowski, M.W.; Urroz, G.; McKee, M.; Khalil, A. Support vectors–based groundwater head observation networks design. *Water Resour. Res.* **2004**, *40*, W11509. [CrossRef]
117. Behzad, M.; Asghari, K.; Coppola, E.A. Comparative Study of SVMs and ANNs in Aquifer Water Level Prediction. *J. Comput. Civ. Eng.* **2010**, *24*, 408–413. [CrossRef]
118. Karandish, F.; Šimůnek, J. A comparison of numerical and machine-learning modeling of soil water content with limited input data. *J. Hydrol.* **2016**, *543*, 892–909. [CrossRef]
119. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
120. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [CrossRef]
121. Veloso, M.F.; Rodrigues, L.N.; Filho, E.I.F. Evaluation of machine learning algorithms in the prediction of hydraulic conductivity and soil moisture at the Brazilian Savannah. *Geoderma Reg.* **2022**, *30*, e00569. [CrossRef]
122. Fix, E.; Hodges, J.L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **1989**, *57*, 238–247. [CrossRef]
123. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
124. Govindaraju, R.S. Artificial neural networks in hydrology. II: Applications. *J. Hydrol. Eng.* **2000**, *5*, 124–137. [CrossRef]
125. ASCE Task Committee. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123. Available online: https://www.ars.usda.gov/ARSUserFiles/50701000/cswq-0332-hjelmfelt.pdf (accessed on 1 June 2026). [CrossRef]
126. Dawson, C.W.; Wilby, R.L. Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr.* **2001**, *25*, 80–108. [CrossRef]
127. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
128. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning for digital soil mapping. *Soil* **2019**, *5*, 79–89. [CrossRef]
129. Lehouel, K.; Saber, C.; Bouziani, M.; Yaagoubi, R. Remote Sensing Crop Water Stress Determination Using CNN-ViT Architecture. *AI* **2024**, *5*, 618–634. [CrossRef]

130. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
131. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared spectroscopy. *Geoderma* **2019**, *352*, 251–266. [CrossRef]
132. Jordan, M.I. Serial order: A parallel distributed processing approach. *Inst. Cogn. Sci. Rep.* **1986**, *8604*, 471–495.
133. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]
134. Coulibaly, P.; Anctil, F.; Bobée, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* **2000**, *230*, 244–257. [CrossRef]
135. Rugină, A.M. Alternative Hydraulic Modeling Method Based on Recurrent Neural Networks: From HEC-RAS to AI. *Hydrology* **2025**, *12*, 207. [CrossRef]
136. Medsker, L.R.; Jain, L.C. *Recurrent Neural Networks: Design and Applications*; CRC Press: Boca Raton, FL, USA, 2001; p. 416. [CrossRef]
137. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
138. Fang, K.; Shen, C.; Kifer, D.; Yang, X. Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network. *Geophys. Res. Lett.* **2017**, *44*, 11030–11039. [CrossRef]
139. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2019**, *561*, 918–929. [CrossRef]
140. A, Y.; Wang, G.; Hu, P.; Lai, X.; Xue, B.; Fang, Q. Root-zone soil moisture estimation based on remote sensing data and deep learning. *Environ. Res.* **2022**, *212*, 113278. [CrossRef] [PubMed]
141. Subudhi, S.; Pati, A.K.; Bose, S.; Sahoo, S.; Pattanaik, A.; Acharya, B.M.; Thakur, R.R. Prediction of groundwater quality assessment by integrating boosted learning with DE optimizer. *Sci. Rep.* **2025**, *16*, 618. [CrossRef] [PubMed]
142. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
143. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobotics* **2013**, *7*, 21. [CrossRef] [PubMed]
144. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [CrossRef] [PubMed]
145. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
146. Tola, D.; Bustillos, L.; Arragan, F.; Chipana, R.; Hostache, R.; Resongles, E.; Espinoza-Villar, R.; Zolá, R.P.; Uscamayta, E.; Perez-Flores, M.; et al. High Spatial Resolution Soil Moisture Mapping over Agricultural Field Integrating SMAP, IMERG, and Sentinel-1 Data in Machine Learning Models. *Remote Sens.* **2025**, *17*, 2129. [CrossRef]
147. Ghiat, I.; Govindan, R.; Al-Ansari, T. Data-driven model predictive control for irrigation management in agricultural greenhouses under CO_2 enrichment. *Smart Agric. Technol.* **2025**, *12*, 101071. [CrossRef]
148. Benassi, A.; Kardous, F.; Moussa, M.; Hachicha, M.; Grayaa, K. Online MPC irrigation control using ML-based soil moisture prediction. *Control Eng. Pract.* **2026**, *173*, 106975. [CrossRef]
149. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805. [CrossRef]
150. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430. [CrossRef]
151. Nguyen, D.; de Voil, P.; Potgieter, A.; Dang, Y.P.; Orton, T.G.; Nguyen, D.T.; Nguyen, T.T.; Chapman, S.C. Multimodal sequential cross-modal transformer for predicting plant available water capacity (PAWC) from time series of weather and crop biological data. *Agric. Water Manag.* **2025**, *307*, 109124. [CrossRef]
152. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
153. Viscarra Rossel, R.A.; Webster, R. Predicting Soil Properties from the Australian Soil Visible–Near Infrared Spectroscopic Database. *Eur. J. Soil Sci.* **2012**, *63*, 848–860. [CrossRef]
154. Lemenkova, P. Support Vector Machine Algorithm for Mapping Land Cover Dynamics in Senegal, West Africa, Using Earth Observation Data. *Earth* **2024**, *5*, 420–462. [CrossRef]
155. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
156. Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* **2021**, *7*, 217–240. [CrossRef]
157. Lemenkova, P. Automation of image processing through ML algorithms of GRASS GIS using embedded Scikit-Learn library of Python. *Examples Counterexamples* **2025**, *7*, 100180. [CrossRef]
158. Lemenkova, P. Improving Bimonthly Landscape Monitoring in Morocco, North Africa, by Integrating Machine Learning with GRASS GIS. *Geomatics* **2025**, *5*, 5. [CrossRef]

159. Tyralis, H.; Papacharalampous, G.; Langousis, A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **2019**, *11*, 910. [CrossRef]
160. Probst, P.; Wright, M.N.; Boulesteix, A.-L. Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
161. Zhang, Y.; Schaap, M.G. Weighted Recalibration of the Rosetta Pedotransfer Model with Improved Estimates of Hydraulic Parameter Distributions and Summary Statistics (Rosetta3). *J. Hydrol.* **2017**, *547*, 39–53. [CrossRef]
162. Lemenkova, P. Reclassification Scheme for Image Analysis in GRASS GIS Using Gradient Boosting Algorithm: A Case of Djibouti, East Africa. *J. Imaging* **2025**, *11*, 249. [CrossRef] [PubMed]
163. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 4765–4774. [CrossRef]
164. Lemenkova, P. Machine Learning Algorithms of Remote Sensing Data Processing for Mapping Changes in Land Cover Types over Central Apennines, Italy. *J. Imaging* **2025**, *11*, 153. [CrossRef] [PubMed]
165. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *Acm Comput. Surv.* **2022**, *55*, 1–37. [CrossRef]
166. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
167. El Hajj, M.; Baghdadi, N.; Zribi, M.; Bazzi, H. Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. *Remote Sens.* **2017**, *9*, 1292. [CrossRef]
168. Lemenkova, P. Deep Learning Methods of Satellite Image Processing for Monitoring of Flood Dynamics in the Ganges Delta, Bangladesh. *Water* **2024**, *16*, 1141. [CrossRef]
169. Bengio, Y.; Simard, P.; Frasconi, P. Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef] [PubMed]
170. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef] [PubMed]
171. Kratzert, F.; Klotz, D.; Shalev, G.; Klambauer, G.; Hochreiter, S.; Nearing, G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 5089–5110. [CrossRef]
172. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
173. Pachepsky, Y.A.; Timlin, D.; Varallyay, G. Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. *Soil Sci. Soc. Am. J.* **1996**, *60*, 727–733. [CrossRef]
174. Lemenkova, P. Artificial Neural Networks for Mapping Coastal Lagoon of Chilika Lake, India, Using Earth Observation Data. *J. Mar. Sci. Eng.* **2024**, *12*, 709. [CrossRef]
175. Lemenkova, P. Artificial Intelligence for Computational Remote Sensing: Quantifying Patterns of Land Cover Types Around Cheetham Wetlands, Port Phillip Bay, Australia. *J. Mar. Sci. Eng.* **2024**, *12*, 1279. [CrossRef]
176. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. [CrossRef]
177. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *Int. Conf. Learn. Represent.* **2021**. [CrossRef]
178. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A.Y. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*; Singh, A., Zhu, J., Eds.; Proceedings of Machine Learning Research; PMLR: New York, NY, USA, 2017; Volume 54, pp. 1273–1282. [CrossRef]
179. Dwork, C.; Roth, A. *The Algorithmic Foundations of Differential Privacy*; Emerald Insight: Leeds, UK, 2014; Volume 9, pp. 211–407. [CrossRef]
180. Dorigo, W.; Himmelbauer, I.; Aberer, D.; Schremmer, L.; Petrakovic, I.; Zappa, L.; Preimesberger, W.; Xaver, A.; Annor, F.; Arh, J.; et al. The International Soil Moisture Network: Serving Earth System Science for Over a Decade. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5749–5804. [CrossRef]
181. Gruber, A.; De Lannoy, G.; Albergel, C.; Al-Yaari, A.; Brocca, L.; Calvet, J.C.; Colliander, A.; Cosh, M.; Crow, W.; Dorigo, W.; et al. Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sens. Environ.* **2020**, *244*, 111806. [CrossRef]
182. Petropoulos, G.P.; Ireland, G.; Barrett, B. Surface Soil Moisture Retrievals from Remote Sensing: Current Status, Products & Future Trends. *Phys. Chem. Earth* **2015**, *83–84*, 36–56. [CrossRef]
183. Fang, K.; Shen, C. Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel. *J. Hydrometeorol.* **2020**, *21*, 399–413. [CrossRef]
184. Lemenkova, P. Exploitation d'images satellitaires Landsat de la région du Cap (Afrique du Sud) pour le calcul et la cartographie d'indices de végétation à l'aide du logiciel GRASS GIS. *Physio-Géo* **2024**, *20*, 113–129. [CrossRef]

185. Song, L.; Liu, S.; Kustas, W.P.; Nieto, H.; Sun, L.; Xu, Z.; Todd, H.; Ma, Y.; Li, M.; Xu, T. Applications of a Thermal-Based Two-Source Energy Balance Model Using Priestley–Taylor Approach for Surface Temperature Partitioning under Advective Conditions. *J. Hydrol.* **2016**, *540*, 574–587. [CrossRef]
186. Lemenkova, P. Evapotranspiration, vapour pressure and climatic water deficit in Ethiopia mapped using GMT and TerraClimate dataset. *J. Water Land Dev.* **2022**, *54*, 201–209. [CrossRef]
187. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better: Multimodal Deep Learning Meets Remote Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4340–4354. [CrossRef]
188. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the Blending of the Landsat and MODIS Surface Reflectance: Predicting Daily Landsat Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218. [CrossRef]
189. Mulla, D.J. Twenty Five Years of Remote Sensing in Precision Agriculture: Key Advances and Remaining Knowledge Gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]
190. Jones, H.G. Irrigation Scheduling: Advantages and Pitfalls of Plant-Based Methods. *J. Exp. Bot.* **2004**, *55*, 2427–2436. [CrossRef] [PubMed]
191. Yao, H.; Qin, R.; Chen, X. Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. *Remote Sens.* **2019**, *11*, 1443. [CrossRef]
192. Bergen, K.J.; Johnson, P.A.; de Hoop, M.V.; Beroza, G.C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **2019**, *363*, eaau0323. [CrossRef] [PubMed]
193. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
194. Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How Powerful Are Graph Neural Networks? *Int. Conf. Learn. Represent.* **2019**. [CrossRef]
195. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine learning in agriculture: A comprehensive updated review. *Sensors* **2021**, *21*, 3758. [CrossRef] [PubMed]
196. Wang, Y.; Shi, L.; Hu, Y.; Hu, X.; Song, W.; Wang, L. A comprehensive study of deep learning for soil moisture prediction. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 917–943. [CrossRef]
197. Datta, P.; Faroughi, S.A. A multihead LSTM technique for prognostic prediction of soil moisture. *Geoderma* **2023**, *433*, 116452. [CrossRef]
198. Cuomo, S.; Di Cola, V.S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *J. Sci. Comput.* **2022**, *92*, 88. [CrossRef]
199. Saikai, Y.; Peake, A.; Chenu, K. Deep reinforcement learning for irrigation scheduling using high-dimensional sensor feedback. *PLoS Water* **2023**, *2*, e0000169. [CrossRef]
200. Ercan Oğuztürk, G.; Murat, C.; Yurtseven, M.; Oğuztürk, T. The Effects of AI-Supported Autonomous Irrigation Systems on Water Efficiency and Plant Quality: A Case Study of Geranium psilostemon Ledeb. *Plants* **2025**, *14*, 770. [CrossRef] [PubMed]
201. Zhang, J.; Ding, Y.; Zhu, L.; Wan, Y.; Chai, M.; Ding, P. Estimating and forecasting daily reference crop evapotranspiration in China with temperature-driven deep learning models. *Agric. Water Manag.* **2025**, *307*, 109268. [CrossRef]
202. Rodríguez-Yparraguirre, C.D.; Rodríguez-Yparraguirre, A.J.; Moreno-Rojo, C.; Castañeda-Rodríguez, W.A.; Saavedra-Vera, J.V.; Lopez-Carranza, A.R.; Olivares-Espino, I.M.; Epifania-Huerta, A.D.; Guarniz-Vásquez, E.; Maco-Vasquez, W.A. Advances in Emerging Digital Technologies for Sustainable Agriculture: Applications and Future Perspectives. *Earth* **2026**, *7*, 63. [CrossRef]
203. Hamdouni, A. Artificial Intelligence-Driven Integrated Water Management and Agricultural Sustainability: Evidence from Saudi Arabia. *Resources* **2026**, *15*, 38. [CrossRef]
204. Ali, A.; Hussain, T.; Zahid, A. Smart Irrigation Technologies and Prospects for Enhancing Water Use Efficiency for Sustainable Agriculture. *AgriEngineering* **2025**, *7*, 106. [CrossRef]
205. Rahman, A.; Chung, G.C.; Ng, Y.H. Applications, Challenges, and Future Trends of Artificial Intelligence of Things (AIoT)-Enabled Water Quality and Resource Management. *Water* **2026**, *18*, 919. [CrossRef]
206. Lemenkova, P. Machine Learning Methods of Remote Sensing Data Processing for Mapping Salt Pan Crust Dynamics in Sebkha de Ndrhamcha, Mauritania. *Artif. Satell.* **2025**, *60*, 37–69. [CrossRef]
207. Rahmati, O.; Panahi, M.; Kalantari, Z.; Soltani, E.; Falah, F.; Dayal, K.S.; Mohammadi, F.; Deo, R.C.; Tiefenbacher, J.; Tien Bui, D. Capability and robustness of novel hybridized models used for drought hazard modeling in southeast Queensland, Australia. *Sci. Total Environ.* **2020**, *718*, 134656. [CrossRef] [PubMed]
208. Lindh, P.; Lemenkova, P. Soil contamination from heavy metals and persistent organic pollutants (PAH, PCB and HCB) in the coastal area of Västernorrland, Sweden. *Gospod. Surowcami Miner. Miner. Resour. Manag.* **2022**, *38*, 147–168. [CrossRef]
209. Lindh, P.; Lemenkova, P. Leaching of Heavy Metals from Contaminated Soil Stabilised by Portland Cement and Slag Bremen. *Ecol. Chem. Eng. S* **2022**, *29*, 537–552. [CrossRef]
210. Ghorbani, A.; Sadeghi, M.; Jones, S.B. Towards New Soil Water Flow Equations Using Physics-Constrained Machine Learning. *Vadose Zone J.* **2021**, *20*, e20136. [CrossRef]
211. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]

212. Wadoux, A.M.J.-C.; Minasny, B.; McBratney, A.B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. [CrossRef]
213. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4559–4565. [CrossRef]
214. Trnka, M.; Brázdil, R.; Možný, M.; Štěpánek, P.; Dobrovolný, P.; Zahradníček, P.; Balek, J.; Semerádová, D.; Dubrovský, M.; Hlavinka, P.; et al. Soil moisture trends in the Czech Republic between 1961 and 2012. *Int. J. Climatol.* **2015**, *35*, 3733–3747. [CrossRef]
215. Meng, X.; Lu, S.; Gao, Y.; Guo, J. Simulated effects of soil moisture on oasis self-maintenance in a surrounding desert environment in Northwest China. *Int. J. Climatol.* **2015**, *35*, 4116–4125. [CrossRef]
216. Moges, E.; Demissie, Y.; Larsen, L.; Yassin, F. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water* **2021**, *13*, 28. [CrossRef]
217. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
218. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
219. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [CrossRef]
220. Wang, Y.; Wang, W.; Ma, Z.; Zhao, M.; Li, W.; Hou, X.; Li, J.; Ye, F.; Ma, W. A Deep Learning Approach Based on Physical Constraints for Predicting Soil Moisture in Unsaturated Zones. *Water Resour. Res.* **2023**, *59*, e2023WR035194. [CrossRef]
221. Alsumaiei, A.A. Water Balance Regularised Modelling of Multi-Depth Soil Moisture Dynamics in Arid Hydrological Systems. *Hydrol. Processes* **2026**, *40*, e70551. [CrossRef]
222. Demir, M.S. Multi-Depth Soil Moisture Prediction Using Machine Learning Across Türkiye's Diverse Environments. *J. Agric. Sci.* **2026**, *32*, 423–438. [CrossRef]
223. Sarkar, S.S.; Bedi, J.; Jain, S. A deep learning based framework for enhanced reference evapotranspiration estimation: Evaluating accuracy and forecasting strategies. *Sci. Rep.* **2025**, *15*, 15136. [CrossRef] [PubMed]
224. Acharki, S.; El-Haddad, A.; El-Fadel, M. Comparative assessment of empirical and hybrid machine learning models for estimating daily reference evapotranspiration in sub-humid and semi-arid climates. *Sci. Rep.* **2025**, *15*, 2542. [CrossRef] [PubMed]
225. Bouregaa, T.; Benbrahim, M. Machine Learning Models (MLP, Random Forest, LightGBM) for Daily ET_0 Estimation with Limited Data in Humid Mediterranean Region (JIJEL) Algeria. *Veredas Direito* **2026**, *23*, 112–129. [CrossRef]
226. Lamichhane, M.; Mehan, S.; Mankin, K.R. Soil Moisture Prediction Using Remote Sensing and Machine Learning Algorithms: A Review on Progress, Challenges, and Opportunities. *Remote Sens.* **2025**, *17*, 2397. [CrossRef]
227. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
228. Jakubik, J.; Roy, S.; Phillips, C.E.; Fraccaro, P.; Godwin, D.; Zadrozny, B.; Szwarcman, D.; Gomes, C.; Nyirjesy, G.; Edwards, B.; et al. Foundation Models for Generalist Geospatial Artificial Intelligence. *arXiv* **2023**, arXiv:2310.18660. http://arxiv.org/abs/2310.18660.
229. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
230. Zhou, J.; Li, E.; Yang, S.; Wang, M.; Shi, X.; Yao, S.; Mitri, H.S. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Saf. Sci.* **2019**, *118*, 505–518. [CrossRef]
231. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. [CrossRef]
232. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 17–29 October 2017; pp. 618–626. [CrossRef]
233. Agarwal, R.; Melnick, L.; Frosst, N.; Zhang, X.; Lengerich, B.; Caruana, R.; Hinton, G.E. Neural Additive Models: Interpretable Machine Learning with Neural Nets. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 4699–4711. [CrossRef]
234. Adebayo, J.; Gilmer, J.; Muelly, M.; Goodfellow, I.; Hardt, M.; Kim, B. Sanity Checks for Saliency Maps. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2018; Volume 31. [CrossRef]
235. Hüllermeier, E.; Waegeman, W. Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods. *Mach. Learn.* **2021**, *110*, 457–506. [CrossRef]
236. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *Proceedings of the 33rd International Conference on Machine Learning*; PMLR: New York, NY, USA, 2016; Volume 48, pp. 1050–1059.

237. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30. [CrossRef]
238. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [CrossRef]
239. Janssen, M.; Brous, P.; Estevez, E.; Barbosa, L.S.; Janowski, T. Data Governance: Organizing Data for Trustworthy Artificial Intelligence. *Gov. Inf. Q.* **2020**, *37*, 101493. [CrossRef]
240. Shepherd, M.; Turner, J.A.; Small, B.; Wheeler, D. Priorities for Science to Overcome Hurdles Thwarting the Full Promise of the Precision Agriculture Revolution. *J. Sci. Food Agric.* **2020**, *100*, 5083–5092. [CrossRef] [PubMed]
241. Wolf, S.A.; Buttel, F.H. The Political Economy of Precision Farming. *Am. J. Agric. Econ.* **1996**, *78*, 1269–1274. [CrossRef] [PubMed]
242. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
243. Xu, B.; Fan, J.; Chao, J.; Arsenijevic, N.; Werle, R.; Zhang, Z. Instance segmentation method for weed detection using UAV imagery in soybean fields. *Comput. Electron. Agric.* **2023**, *211*, 107994. [CrossRef]
244. Schwartz, R.; Dodge, J.; Smith, N.A.; Etzioni, O. Green AI. *Commun. ACM* **2020**, *63*, 54–63. [CrossRef]
245. Pai, D.G.; Balachandra, M.; Kamath, R. Explainable AI in agriculture: Review of applications, methodologies, and future directions. *Eng. Res. Express* **2025**, *7*, 032202. [CrossRef]
246. Zhan, S.; Huang, L.; Luo, G.; Zheng, S.; Gao, Z.; Chao, H.C. A Review on Federated Learning Architectures for Privacy-Preserving AI: Lightweight and Secure Cloud–Edge–End Collaboration. *Electronics* **2025**, *14*, 2512. [CrossRef]
247. Žalik, K.R.; Žalik, M. A review of federated learning in agriculture. *Sensors* **2023**, *23*, 9566. [CrossRef] [PubMed]
248. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J. Big data in smart farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [CrossRef]
249. Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F.X. A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **2017**, *143*, 23–37. [CrossRef]