Intelligent Agriculture
https://ojs.ukscip.com/index.php/ia/index

Review

Crop Yield Prediction Using Precision Agriculture and Smart Farming Technologies: A Systematic Review and Future Research Trends

Mariam Shaaban Sayed Hassan [1], Abdel Rahman Shaaban [2] and Mohammed Safy [1,*]

[1] College of Computing and Information Technology, Egypt Arab Academy for Science, Technology, and Maritime Transport, Cairo 11511, Egypt

[2] Biomedical Engineering Department, Higher Institute of Engineering, El-Shorouk Academy, Cairo 11837, Egypt

[*] Correspondence: msafy@aas.edu

Received: 16 December 2025; **Revised:** 18 February 2026; **Accepted:** 20 February 2026; **Published:** 12 March 2026

Abstract: The use of crop yield predictions is essential in today's agricultural systems as the global food system is facing increased pressures from climate change, soil degradation, and water scarcity. Precision agriculture is based on the development of data-driven solutions for yield predictions using remote sensors, IoT (Internet of Things) devices, machine learning, and crop growth models. This systematic literature review synthesizes recent research (2014–2026) concerning crop yield prediction by reviewing what types of data are available, factors affecting yield, what predictive models have been developed, and how those models are used in field, regional, and large-scale agricultural situations. This review also demonstrates how climatic and hydroclimatic variability interact with soil properties and management practices to create yield outcomes, while evaluating the accuracy and performance of statistical, machine learning, deep learning, and hybrid modelling approaches. The study concludes that multimodal data fusion, real-time data assimilation, and hybrid physics/AI modelling approaches improve prediction accuracy and robustness. Ongoing challenges include (but are not limited to) model transferability, the lack of standardized benchmark datasets for model development and evaluation, insufficient real-time integration of multiple types of information to assist in yield prediction, and difficulty in explaining how advanced AI models develop yield predictions. The authors propose a future research agenda based on the findings of this study that will provide a standard multimodal dataset for the development of explainable AI, a scalable edge/cloud architecture for model deployment, and models that can be transferred across crops, climates, and production systems.

Keywords: Crop Yield Prediction; Precision Agriculture; Smart Farming; Machine Learning; Remote Sensing; IoT Sensors; Climate Change; Soil Degradation

1. Introduction and Background

Issues around food security have always been a top priority for both national authorities and global organizations. In September 2015, the United Nations approved a strategic document titled "Transforming our World: The 2030 Agenda for Sustainable Development". The agenda included multiple goals known as Sustainable Development Goals (SDGs), which aimed to address global hunger levels and provide food security while promoting sustainable agriculture throughout the world [1]. At the outset, this goal seemed possible; yet, after almost 8 years since the goal was launched, the global situation continues to worsen [2,3].

The stability of global food supplies is increasingly threatened by the rising frequency and intensity of climate shocks, which drive significant crop yield instability [4]. This volatility acts as a primary stressor on food-system

https://doi.org/10.54963/ia.v2i1.2377

resilience, defined as the capacity of a system to maintain nutritional security and functional integrity in the face of such disturbances [5,6]. When yields fluctuate unpredictably, the cascading effects can destabilize local and global supply chains, leading to price spikes and localized shortages [7]. To mitigate these risks, the development of robust early yield forecasting models has become a critical necessity [8]. These predictive tools provide the lead time required for governments to implement social safety nets, for markets to adjust trade flows, for insurance providers to calibrate risk assessments, and for farmers to make informed tactical decisions regarding harvest and storage. Consequently, integrating high-frequency forecasting into agricultural policy is essential for transforming reactive crisis management into proactive resilience planning [9,10].

The global food system is undergoing major change as it faces growing environmental, population, and resource constraints. In addition to environmental stressors, global agriculture is under pressure from rising food demand driven by population growth, urbanization, and changing dietary patterns, making accurate yield forecasting essential for planning, resource allocation, and food-supply stabilization [11,12]. Climate change is one of the greatest threats to crop productivity. Increased temperatures, changes in precipitation patterns, and more frequent extreme weather events are leading many crops around the world to have measurable decreases in their yields due to climate change [13,14]. Current levels of high temperature stress, drought, flooding, and variability in atmospheric CO_2 are all contributing to wide fluctuations in yield from year to year in many important crops located in vulnerable regions of the world [4,15,16].

Anthropogenic climate change has introduced unprecedented levels of climate variability, fundamentally altering the biophysical conditions necessary for stable agricultural production [4,17]. The intensification of seasonal anomalies, manifesting as extreme heat stress, prolonged drought, and catastrophic flooding, directly impairs crop physiology and soil health, leading to unpredictable harvest outcomes. Traditional yield models, often based on historical averages, are increasingly insufficient in the face of such non-linear climatic shifts [18]. To bridge this gap, integrating climate-informed crop forecasting has become essential. These advanced frameworks leverage real-time meteorological data and satellite imagery to simulate how specific stressors affect plant phenology [19,20]. By embedding these variables into early-warning systems, researchers can provide stakeholders with high-resolution predictions of potential shortfalls months in advance [21]. This direct linkage between atmospheric dynamics and agricultural modeling transforms raw climate data into actionable intelligence, allowing for a more robust response to the volatile environmental conditions of the 21st century [22].

Soil degradation is another significant barrier to agriculture and is an enormous global problem with negative consequences for agricultural productivity and food system resilience. In the FAO report, The State of Food and Agriculture 2025, it is estimated that nearly 1.7 billion people live in a location where the yields of their crops are declining as a result of human-induced land degradation and other contributing factors, such as soil erosion, salinization, nutrient depletion, and unsustainable agricultural practices [23,24].

Water scarcity is one of the top challenges to food production, which makes up about 70% of all freshwater usage globally, and nearly 50% of the global population suffers from significant water scarcity at least once a year. Increasing demand for freshwater used in agricultural irrigation, along with climate-related changes to the hydrologic cycle, is putting additional stress on many places already under stress from water shortages [25]. Drought and flooding events have affected over 3 billion people between 2002 and 2021, resulting in billion-dollar losses of crops and instability of food systems. Water scarcity has led to an increase in the usage of precision systems for irrigation, soil moisture sensing technology, and computer modeling tools for predicting crop water use and maintaining stable annual yield predictions [26,27].

Therefore, precision agriculture and smart farming technology are seen as viable avenues for improving the prediction of crop yields, maximizing the efficiency of resource use, and supporting data-driven decision-making. In precision agriculture, real-time sensors, remote sensing, geographical analysis, and artificial intelligence work together to provide data for areas of variability at the field level relating to soil, weather around the field, and the plants in the field. Recent systematic reviews on the use of machine learning and deep learning methods have found these approaches to improve predictive accuracy, reducing error rates by 15% to 30% compared with traditional statistical methods [28,29].

Modern agricultural research increasingly leverages remote sensing as a cornerstone for large-scale yield estimation, moving beyond traditional point-based surveys to high-resolution, continuous observation [30,31]. By capturing diverse biophysical conditions, satellite and aerial platforms provide objective data on crop health, soil moisture, and phenological development across field, regional, and national scales [32,33]. The integration of multispectral and hyperspectral data allows for the identification of narrow spectral signatures associated with nutrient deficiencies or disease [34], while Radar/SAR data offers a distinct advantage by penetrating cloud cover and providing information on crop structure and biomass even in adverse weather conditions [35,36].

Modern agriculture has been transformed by the application of precision agriculture (PA) and the application of technology, such as new sensing devices, advanced data analytics, and automated decision-making support systems. Recent systematic reviews of PA show that PA can greatly improve the prediction of crop yields by using machine learning and remote sensing [37]. In addition to other digital technologies, such as using real-time data from the field, PA can use sophisticated modelling techniques to better understand how various elements interact, including the interaction among pests, the quality of the soil, and the balance of nutrients, as well as the climate in which the crops are grown. Artificial intelligence (AI)-based approaches have demonstrated substantial improvements over traditional models in predicting yields based on soil, climate, and management variables. IoT-integrated frameworks further enhance predictive accuracy by enabling continuous monitoring of temperature, soil moisture, humidity, and nutrient levels, thus supporting dynamic, real-time yield prediction systems [37].

The integration of the Internet of Things (IoT) into precision agriculture has revolutionized yield prediction by shifting from periodic observations to continuous, real-time monitoring of the farm environment [38,39]. By deploying an interconnected network of hardware, farmers can capture granular data that satellite imagery might miss, particularly regarding the subterranean and micro-climatic factors that drive crop success [39,40]. Soil moisture sensors provide high-frequency data on water availability at the root zone, while nutrient sensors track nitrogen, phosphorus, and potassium levels to ensure optimal fertilization. Simultaneously, on-site weather stations and temperature and humidity sensors capture the specific micro-climate of a field, allowing models to account for localized heat pockets or frost risks [41,42].

The convergence of Precision Agriculture (PA) with IoT, AI, and Remote Sensing facilitates a paradigm shift from reactive to proactive agricultural management, optimizing every stage of the food value chain [43]. At the field level, the synthesis of soil moisture sensors and satellite-derived indices enables precise irrigation scheduling and targeted fertilizer management, ensuring resource efficiency and environmental sustainability [32]. Beyond the farm gate, AI-driven yield forecasting informs harvest planning and streamlines supply-chain management, while providing governments and international agencies with an early warning for food crises [44]. These technological insights are equally transformative for the financial sector, allowing for objective crop insurance payouts based on verified satellite data and aiding government policy and subsidy planning through high-resolution land-use monitoring [45].

Numerous research gaps still exist despite the substantial advancement in precision agriculture and smart farming technologies. Many studies use fragmented methods with isolated datasets, differing ways to pre-process the data, and inconsistent metrics to evaluate their respective performance. As such, there is little generalization of findings across the agricultural context [46]. Further, there are limited grounds for the comparison of crop yield prediction algorithms. Many of the referenced reviews note that there are no accepted benchmarks against which to make comparative assessments, and little analysis of the performance of various models with respect to each other results in a lack of ability to provide insight into which algorithm performs better across various environments or under different agronomic conditions [29].

The paper provides the following innovative contributions: (i) a comprehensive synthesis using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA guidelines) involving a systematic review of 82 total unique studies that utilize an agricultural yield predictions prediction model; (ii) systematic review of methodological trends and gaps across these studies; (iii) proposes an architectural taxonomy for yield predictions systems across multiple scales; (iv) integrated analysis of both methods, data sources, and factors influencing the yield that was forecasted; and (v) proposes a data-driven agenda for future scalable, explainable, and transferable precision agriculture systems.

The remainder of this paper is organized as follows. Section 2 presents the systematic literature review methodology, including the PRISMA-based search strategy, study selection process, inclusion and exclusion criteria, and procedures used to identify relevant publications. Section 3 provides a statistical analysis of the selected studies, highlighting publication trends, research distributions, methodological approaches, and publication sources.

Section 4 examines the key factors influencing crop development and yield, while Section 5 focuses specifically on climatic variables affecting crop productivity and their impact on agricultural outcomes. Section 6 reviews the role of precision agriculture and artificial intelligence technologies in crop yield prediction, emphasizing recent advancements in data-driven farming practices. Section 7 discusses the architectures of prediction models employed in crop yield forecasting, including machine learning and deep learning frameworks. Section 8 analyzes the various data sources utilized for crop yield prediction, such as climatic, soil, remote sensing, and management datasets.

Section 9 provides a comprehensive discussion of the reviewed literature and synthesizes the major findings. Section 10 presents a comparative analysis of existing studies, highlighting differences in methodologies, datasets, and predictive performance. Section 11 discusses the challenges and limitations encountered in current crop yield prediction research. Section 12 outlines future research directions and emerging opportunities for advancing precision agriculture and intelligent farming systems. Finally, Section 13 concludes the paper by summarizing the key findings and their implications for future research and practical applications in crop yield prediction and precision agriculture.

2. Systematic Literature Review Methodology

To obtain a solid understanding of the academic discipline of predicting crop yields, researchers followed the steps defined in the Structured Systematic Literature Review (SSLR) methodology to systematically compile, appraise, and synthesize peer-reviewed publications. This included defining the review's scope, developing general research questions, applying a systematic search strategy across multiple data sources using an array of relevant crop yield prediction-related keyword categories including (but not limited to): crop yield predictions and precision agriculture; remote sensing, satellite images, UAV-based monitoring; Internet of Things (IoT) and field sensors; soil variables; hydro-climatic variables; machine learning and deep learning; explainable AI; and crop growth models/data assimilation techniques for real-time forecasting using spatial/temporal modeling techniques. A comprehensive review of relevant peer-reviewed publications was conducted using a wide variety of digital libraries (IEEE, Springer, Elsevier/ScienceDirect, MDPI, Taylor and Francis, Wiley) to maximize dissemination across all fields of study related to the agricultural domain.

The study selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as illustrated in **Figure 1**. A total of 841 records were initially identified through major digital libraries, including IEEE Xplore, Elsevier (ScienceDirect), SpringerLink, ACM Digital Library, and MDPI. After removing 113 duplicate records, 728 studies remained for title and abstract screening. Subsequently, 516 full-text articles were assessed for eligibility, of which 405 studies were excluded based on predefined inclusion and exclusion criteria. This resulted in 111 eligible studies, which were further subjected to quality assessment. Following this step, 31 studies were excluded due to insufficient methodological rigor or lack of validation, leading to a final set of 82 studies included in the systematic review.

To provide a comprehensive and objective synthesis of progress made towards improving yield predictions with data-driven, process-based, and hybrid modelling techniques, clear inclusion and exclusion criteria have been defined to ensure only high-quality, methodologically sound, and applicable to predicting Yield were included. Studies were required to provide specific information regarding Yield Prediction or Forecasting, plus provide an evaluative or validating methodology (e.g., Algorithm, (in this instance via an empirical Comparative Criteria, (e.g., through quantitative performance metrics) methodology) in order to be included. Studies must meet additional criteria such as being published in peer-reviewed journals and/or in reputable Associations' Conferences (with a minimum editorial requirement for such publications) and be recorded (via the aforementioned) in the same database; and (of course, only publications that were written in English have been included). To establish an accepted time period for all included works, only articles published between 2014 and 2026 were eligible for inclusion. The following studies were excluded: those that only addressed crop classification without Yield Prediction; studies that do not employ a rigorous (and therefore reproducible/valid) methodology; works that addressed solely conceptual or review-based work without original analysis; those that replicate previously published works; or works that describe agricultural technologies that are not directly relevant to Yield Prediction.

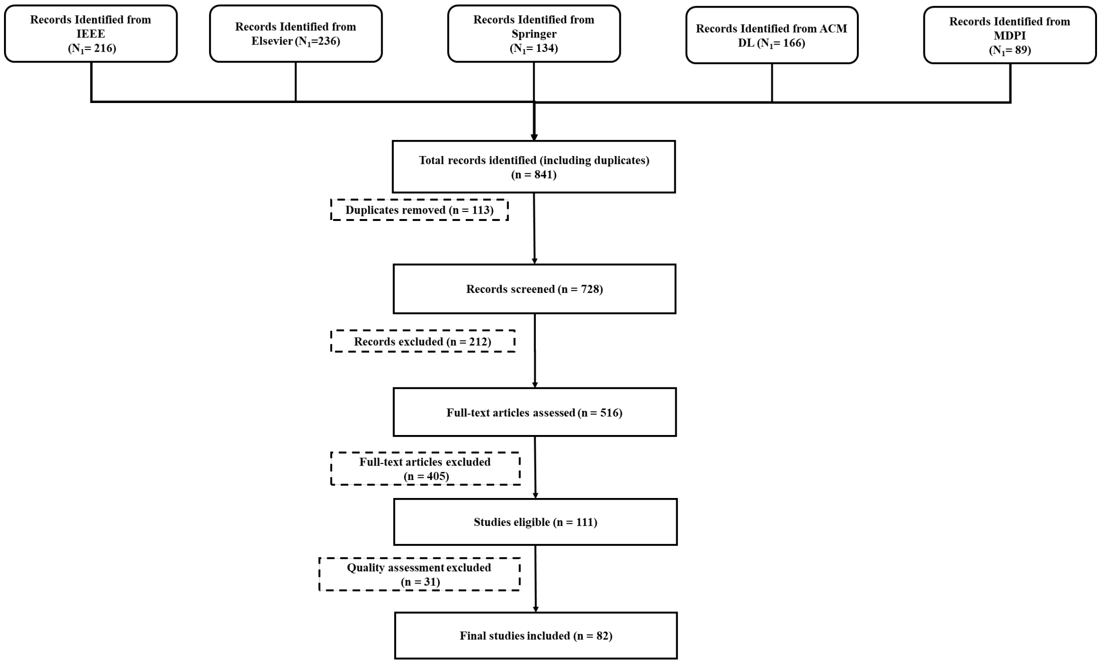

Figure 1. PRISMA flow diagram illustrates the study selection process for the systematic literature review.

3. Statistical Analysis of Selected Studies

The statistical evaluation of the literature indicates that crop yield prediction research has seen a clear increase in publishing over time; total numbers and numbers of journal contributors are presently at all-time highs. It was not until about 2020 that there were significant increases in both publication levels and publication numbers, with the highest levels of publications occurring around 2022, and again in 2024. This growth coincides with the rapid adoption of precision agricultural technologies, the expanded use and availability of satellite and IoT data, and the increased concern related to the climate's impact on production stability. The distribution of authors and publishers shows that Elsevier and Springer account for almost half of the included studies, suggesting that many of the high-impact journals for agricultural and environmental sciences are concentrated in this area of research. Other contributing publishers are MDPI and IEEE Xplore, suggesting that crop yield prediction research is interdisciplinary and covers agricultural scientists, remote sensing scientists and AI researchers. The bibliometric analysis was refined to focus on scientifically meaningful dimensions rather than publisher-based distributions. Specifically, the reviewed studies were categorized according to methodological approaches, spatial scales, and application domains of yield prediction systems. Furthermore, the reviewed studies were analyzed based on their spatial scale, distinguishing between field-level, regional, and large-scale prediction systems. This classification reveals a clear trend toward multi-scale modeling approaches that balance accuracy and scalability.

With respect to methodology, machine learning-based yield predictions (22.0%) and IoT (18.3%) represent the two primary methodologies in the literature, with remote sensing (14.6%) and deep learning (13.4%) at the next level. Hybrid approaches are less frequent (7.3%), although several studies have documented that hybrid methodologies combining crop models using AI offer more robust and interpretable results. Optimization-based approaches and studies related to policies represent the smallest percentage of studies in literature when compared to other methods and suggest a research gap in the use of operational optimization methods and the use of decision-level integration. Overall, the statistical trend data confirms

Following PRISMA reporting principles, the included evidence base comprises N = 82 studies, which were then quantitatively characterized by time, source outlets, and methodological focus.

The temporal distribution indicates a rapid increase since 2020 in reporting, as shown in **Figure 2**. The peak is in 2022 (23/82, or approximately 28% of all studies), with subsequent distributions of studies occurring in 2024 (19/82, or 23.2%), 2025 (15/82, or 18.3%), and 2023 (13/82, or approximately 16%). Publisher/library distribution shows that the literature is anchored in high-impact venues, as shown in **Figure 3**, led by Elsevier (24/82;

29.3%), then Springer (15/82; 18.3%), MDPI (12/82; 14.6%), and IEEE Xplore (11/82; 13.4%), with additional representation from Taylor & Francis (8/82; 9.8%), Wiley (6/82; 7.3%), and FAO/UN reports (6/82; 7.3%). Methodological synthesis indicates dominance of data-driven pipelines, as shown in **Figure 4**, ML-based yield prediction (18/82; 22.0%) and IoT-enabled approaches (15/82; 18.3%), followed by remote sensing (12/82; 14.6%), deep learning (11/82; 13.4%), and climate-focused studies (10/82; 12.2%), while hybrid modeling remains comparatively underrepresented (6/82; 7.3%), alongside policy (6/82; 7.3%) and optimization (4/82; 4.9%).

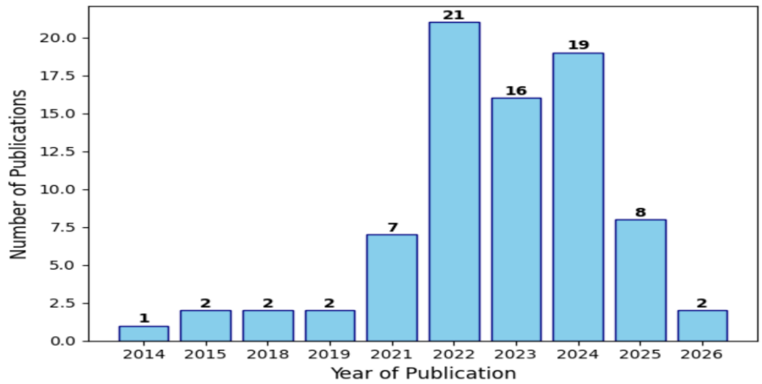

Figure 2. Growth trend of publications on crop yield prediction using precision agriculture and smart farming technologies (2014–2026).

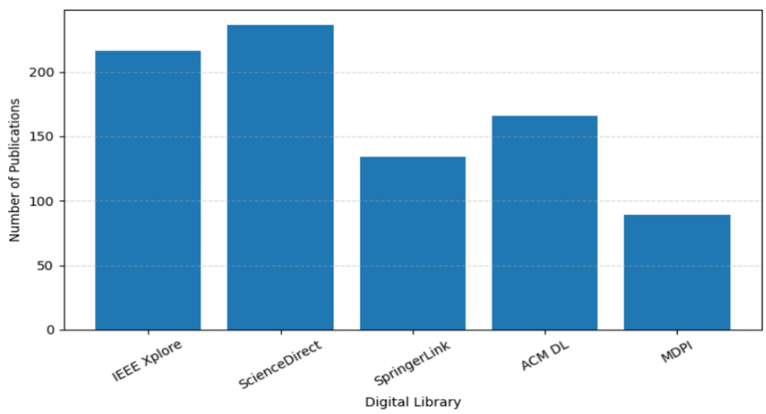

Figure 3. Distribution of selected publications across major scientific publishers and digital libraries.

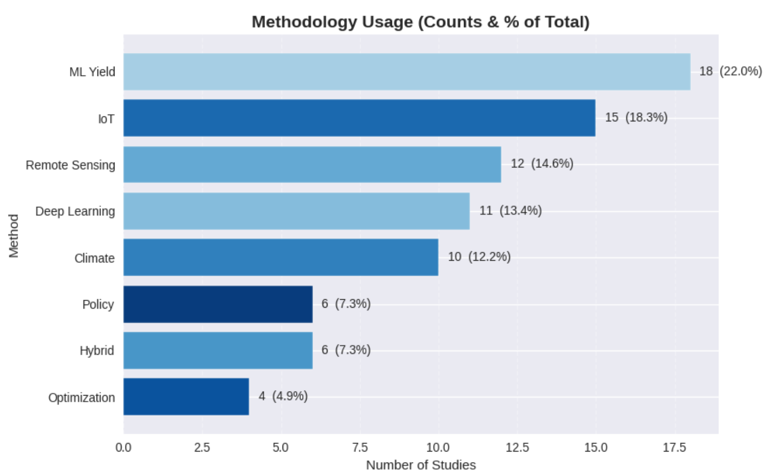

Figure 4. Distribution of selected studies by methodological approach in crop yield prediction.

The proposed conceptual framework, **Figure 5**, produces the findings of this review by integrating the key components of crop yield prediction systems, including multi-source data inputs, data processing techniques, modeling approaches, and application domains. The framework reflects the PRISMA-based analysis and highlights the central role of data fusion and advanced machine learning techniques in improving predictive performance.

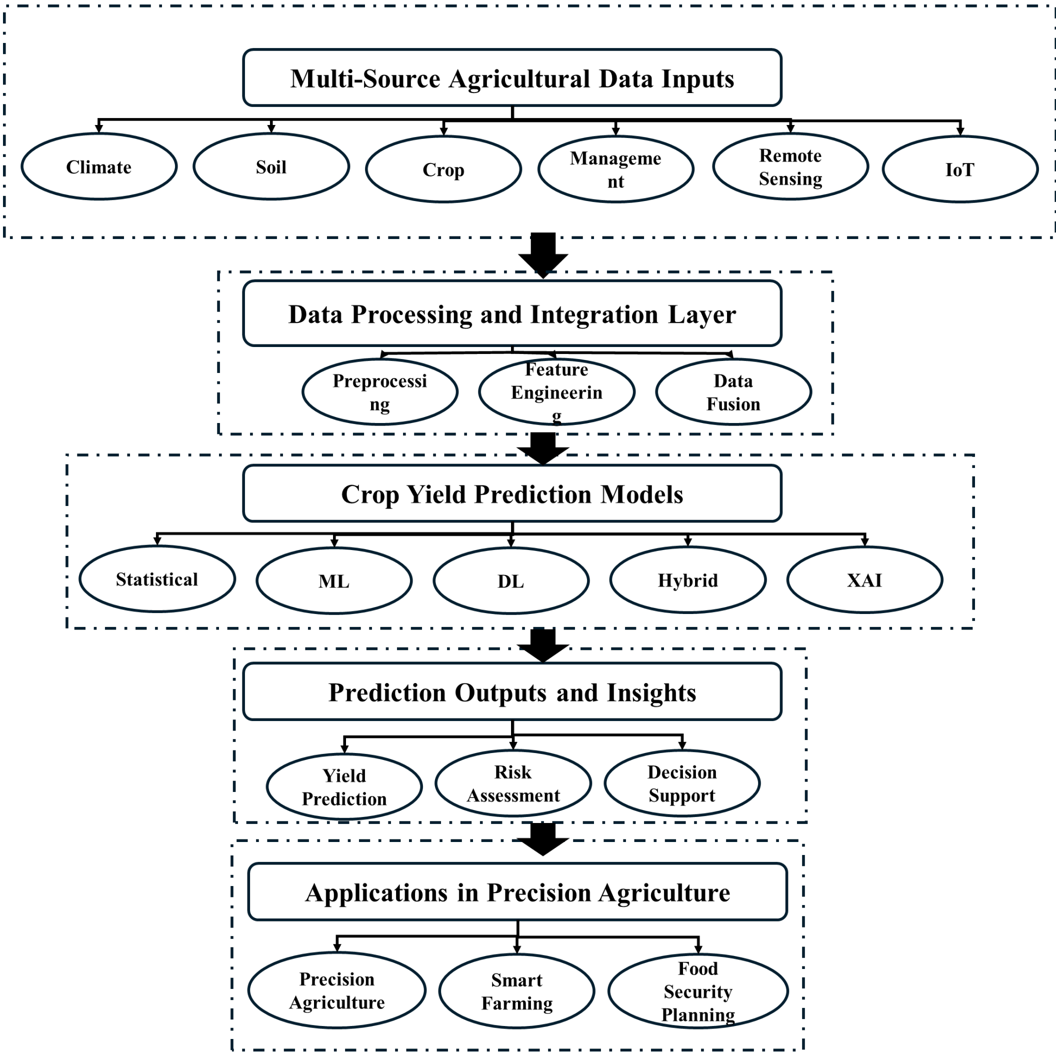

Figure 5. Conceptual framework for crop yield prediction integrating layers.

4. Factors Influencing Crop Development and Yield

Crop yield prediction is fundamentally driven by the interaction of multiple factors, including climate conditions, soil properties, crop characteristics, and agronomic practices [47], as shown in **Figure 6.** However, in modern yield prediction systems, these factors are not only studied descriptively but are operationalized as input variables within data-driven and process-based models. Understanding how these factors are captured, represented, and integrated into predictive frameworks is essential for designing accurate and scalable crop yield prediction systems [48–50]. In combination with these environmental constraints, agronomic decisions related to the management of crops have a large impact on how a crop will respond to the environmental constraints. Thus, understanding and modeling the interactions between these four types of factors (environmental, soil, crop, and agronomic) is necessary to accurately predict crop yield and provide the foundation for developing resilient precision agriculture and smart farming systems [51].

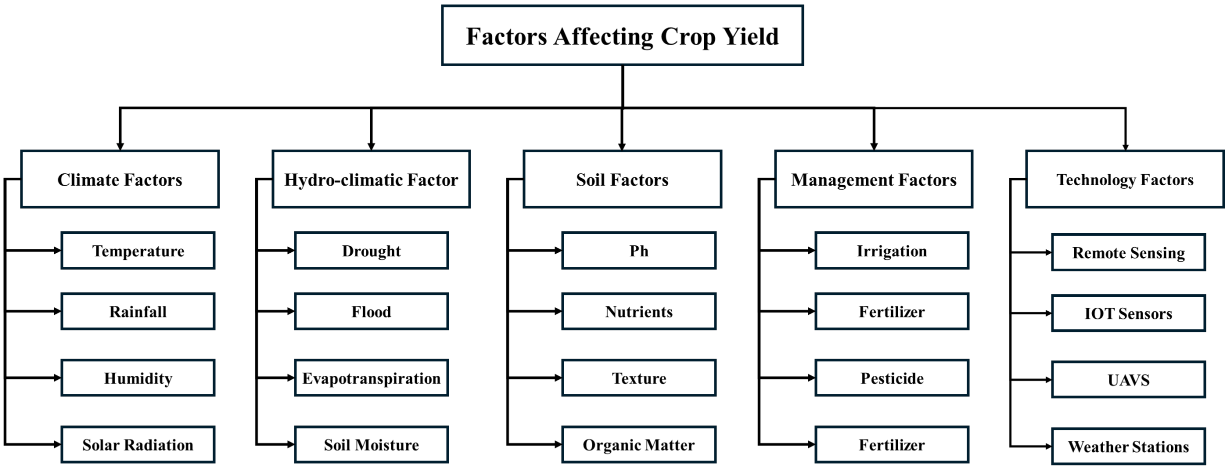

Figure 6. Conceptual framework showing the major climatic, hydro-climatic, soil, crop-management, and technological factors influencing crop yield prediction.

5. Climatic Variables Affecting Crop Productivity

5.1. Temperature-Induced Constraints

Agricultural production depends primarily on climate variables, which establish the maximum yield of a crop (biological metric) via their effect on physiological/metabolic aspects like energy from the sun driving photosynthesis, temperature controlling metabolic rates of crop growth, and making the length of the growing season longer or shorter. Excess or insufficient amounts of rain and relative humidity can affect how well water and nutrients are transferred; when there is too much or too little of an element within this triad, the crops can be adversely affected by either drought or increased disease. Ultimately, this triad will determine if a crop grows to achieve its full potential yield or falls prey to environmental restrictions [47,52].

The use of temperature to predict crop yield is hampered by the following limiting factors: how plants respond to temperature changes nonlinearly; extreme heat waves can occur; there is variability in both space and time of temperatures; there are interactive effects with soil and water; and there are some limitations with respect to the use of sensors and data. Therefore, temperature-related variability in the environment poses numerous constraints to the accuracy, reliability, and generalizability of crop production forecasting models. As climate changes continue to accelerate, these hindrances become more apparent, hence the requirement for the continued development of hybrid (physics and artificial intelligence (AI)) models that merge the understanding of crop physiology with the ability of data to improve the reliability of crop yield predictions under extreme temperatures [52].

From an econometric and statistical perspective, Chandio et al. (2022) developed a quantitative model to examine the effects of climatic and non-climatic factors on cereal production in India using historical econometric data. Their findings show that increasing mean temperature negatively affects cereal yields, especially when heat stress occurs during sensitive phenological stages such as flowering and grain filling. The study further indicates that temperature-induced increases in evapotranspiration and reductions in photosynthetic efficiency can offset the benefits gained from technological improvement, economic growth, or improved agricultural practices [52]. Similarly, Gul et al. (2022) investigated the impact of climate change on major crop yields in Pakistan and found that rising temperatures significantly reduce crop productivity in both the short and long term [53]. Short-term temperature shocks directly reduce yields through heat stress, whereas long-term warming alters cropping calendars, increases irrigation requirements, and affects soil moisture balance. Both studies provide strong regional evidence from South Asia, where cereal production is highly sensitive to thermal stress and water limitations.

In contrast to econometric models, AI-based and explainable artificial intelligence approaches provide a more flexible framework for capturing nonlinear relationships between climatic variables and yield responses. Mohan et al. (2025) developed an AI-based crop yield prediction model supported by Explainable Artificial Intelligence (XAI). Their results show that temperature-related variables, including daily average temperature, daily maximum temperature, cumulative heat units, and heat stress indices, consistently rank among the most influential predictors [18]. Unlike traditional statistical models, the AI-based approach was able to detect nonlinear thresholds, yield plateaus, and sharp yield declines beyond crop-specific temperature limits. This is particularly important because crop responses to temperature are rarely linear; moderate warming may have limited effects in some conditions, while extreme heat during critical growth stages can cause sudden and severe yield reductions.

Compared with the econometric approaches of Chandio et al. (2022) and Gul et al. (2022), the XAI-based model of Mohan et al. (2025) offers two main advantages. First, it can model complex and nonlinear climate–yield relationships more effectively. Second, it improves interpretability by identifying which temperature-related variables contribute most strongly to prediction outcomes. Nevertheless, AI-based models generally require larger and more diverse datasets, and their transferability across crops, regions, and management systems remains a key challenge. Therefore, while econometric models are useful for macro-level climate impact assessment, XAI-based models are more suitable for high-resolution prediction and decision-support applications where interpretability and variable importance are required.

Compared with studies that examine temperature as an isolated variable, Chen et al. (2022) provide stronger evidence that yield prediction models should include interaction effects among climatic variables [54]. This is important because models that consider only seasonal average temperature may underestimate yield losses caused by short-duration heatwaves, drought periods, or simultaneous heat–moisture stress. Therefore, compound-event studies add an important modeling insight: yield prediction systems should incorporate not only mean climatic conditions but also climatic variability, extreme thresholds, and interactions among temperature, precipitation, soil moisture, and evapotranspiration.

The climatic factors that have a major impact on crop yield are temperature, which negatively affects yields through numerous interconnected pathways. For example, econometric studies (Chandio et al., Gul et al.) provide evidence at the regional/national level (South Asia) that yields are reduced with increasing temperature, as there are thermal thresholds and limits that crops can reach. Mechanistic/physiological analyses (Chen et al., Qaderi et al.) explain reduced yields due to heat stress on plants—namely, reproductive development, and resource allocation. Furthermore, global-level analyses (Chen et al.) have shown that the effects of temperature increases would also be magnified in conjunction with moisture stress, meaning compounded extreme climatic events would have more impact than just an increase in temperature by itself. From a modeling perspective, these findings highlight the need for advanced machine learning and hybrid models capable of capturing nonlinear relationships, extreme events, and interactions between climatic variables, which are critical for improving yield prediction accuracy.

5.2. Hydro-Climatic Factors

The growth of crops, the dynamics of soil moisture, and overall agricultural productivity all rely heavily on hydro-climatic conditions. Specifically, hydro-climatic factors such as the timing, intensity, and amount of precipitation received, rates of evapotranspiration, relative humidity of the atmosphere, and the quantity of available water significantly impact the above. Variability in rainfall amounts or timing can have direct effects on the timing of planting, how much irrigation will be required, and if the crops will experience stress. Human-induced variability in humidity and Evapotranspiration will impact natural processes that result in plants losing their available water (water loss) and will have an effect throughout the crop-growth cycle on how resilient each plant will be when faced with adverse conditions. Climate Change and extreme hydro-climatic events (droughts, floods, and long-term dry spells) create additional challenges in the accuracy of Yield Predictions due to the significant uncertainty associated with both time and region. The variability of precipitation amounts and the long-term potential of hydrologic processes under Climate Change will create significant variability in the ability of Crop Modelling to provide accurate predictions and develop Risk Forecasting Models, as well as adapt their approach to Farm Management Strategies [55,56].

The reviewed hydro-climatic studies can be classified into four main methodological groups: long-term vegetation–hydroclimate trend analysis, hydrological hazard and flood-risk mapping, watershed-scale climate and water-resource assessment, and remote-sensing/time-frequency analysis of vegetation response. This classification helps clarify how different solution strategies contribute to crop yield prediction. While some studies provide direct indicators for yield forecasting, such as vegetation indices and evapotranspiration, others provide indirect but important information related to drought risk, flood exposure, irrigation planning, and climate adaptation.

From a long-term vegetation–hydroclimate trend perspective, Xu et al. (2026) to conduct an analysis of multi-decadal vegetation greenness dynamics and the role of hydrological and climate factors affecting vegetation greenness across China. Based on their findings, they concluded that "greening" of vegetation may lead to greater evaporation demand and may also alter the amount of available water for produce by increasing monthly yield rates in locations where there is enough water to meet that demand and by decreasing monthly yield rate potential when evaporation exceeds available water supply. In addition, their extensive temporal and spatial coverage has revealed the heterogeneous nature of vegetation greenness across China and provided insight into the role of vegetation greenness and climate in the risk of droughts and in sustaining yield predictions [57].

A second methodological group includes hydrological hazard and flood-risk mapping studies, Kakavand et al. (2025) established methods for conducting trend analysis in order to create space-time maps that can help end users better understand the nature of flood hazard risk, including an estimate of where hydrologic risk will increase or decrease due to climate change. Furthermore, this research can help establish risks of flooding, as well as identify where better flood management practices are already being implemented or are not necessary given the changing climate and weather patterns. This research provides a method for developing early warning thresholds and infrastructure planning for drainage systems, as well as targeting and promoting climate-resilient agricultural practices. There are two key limitations with this analysis: flood hazard is not an agricultural yield prediction; therefore, direct impacts of flooding on agricultural productivity cannot be determined from this study. Additionally, land use change, including urbanization, deforestation, and river channel modifications, will confound hydroclimate signals produced by this work, making it essential for the future stage to couple the data obtained from this study with agronomic data and necessary controls for non-hydroclimatic factors [58].

A third group of studies adopts a watershed-scale hydro-climatic assessment. Alehu et al. (2022) combined long-term climate observations, trend analysis, and regional climate model projections to evaluate changes in water storage, precipitation distribution, temperature, and snowmelt timing. This watershed-based perspective is highly relevant to agriculture because water availability is often determined not only by local rainfall but also by upstream hydrological processes, river flow, reservoir storage, and seasonal snowmelt. Compared with flood-risk mapping, watershed-scale assessment provides broader information for irrigation scheduling, water allocation, and drought preparedness. Nevertheless, its limitation is that hydro-climatic indicators are not always directly linked to observed crop yield outcomes. Therefore, integrating watershed-level hydrological variables with crop-growth models and yield databases would strengthen its relevance for yield forecasting [59].

Similarly, Sigalla et al. (2023) examined hydro-climatic trends in Tanzania using a combination of gridded climate data and station-based observational records. Their study considered precipitation, maximum and minimum temperature, potential evapotranspiration, and river discharge. Unlike studies focusing on a single hydro-climatic variable, this work provides a multi-variable and spatially explicit assessment of hydro-climatic zones. The main advantage of this approach is that it supports climate-smart agricultural planning by identifying zones with different water availability, evapotranspiration patterns, and discharge behavior. However, as with other hydro-climatic trend studies, the analysis would be more useful for yield prediction if linked with crop-specific yield records, irrigation practices, soil characteristics, and management variables [60].

A fourth group applies remote sensing and time-frequency analysis to evaluate vegetation response to hydro-climatic variability. Batungwanayo et al. (2024) employed wavelet-based analysis to measure how vegetation dynamics (measured through NDVI) have been affected by hydro-climatic drivers (precipitation, moisture stress, etc.) across various agro-climatic zones. This approach is particularly useful for crop yield prediction because NDVI anomalies can act as early indicators of crop stress before harvest, especially in rain-fed systems. Unlike conventional trend analysis, wavelet analysis captures both the timing and frequency of hydro-climatic influences on vegetation response, making it valuable for early-warning systems. However, the relationship between NDVI and yield varies by crop type, growth stage, canopy structure, and management practice. In addition, wavelet coherence indicates association rather than causation, so NDVI-based hydro-climatic signals should be validated against observed yield data [61].

Mojid et al. (2024) analyze multi-decadal climate variables and reference evapotranspiration (ETo) from 1990 to 2019 to demonstrate how climate information at appropriate temporal scales. The study specifically identifies ETo as an indicator of atmospheric water demand; thus, ETo data may assist farmers when scheduling irrigation, optimizing planting windows, and assessing drought stress, all of which have direct implications for agricultural yield. A primary strength of this research is the decision-centered approach taken in linking hydro-climatic signals to on-farm actions. However, due to its focus on characterizing climate and ETo data rather than developing a model directly related to yield, the outcome of this research has a limited application to sub-regions and types of crops because of differences in the management of crops and limitations of data [62].

Dorjsuren et al. (2023) studied how hydro-climatic properties and the amount, timing and location of vegetation have changed over time as a function of climate change impact across the arid and semi-arid Great Lakes Depression. The results indicate that the growth of vegetation is concentrated around the peak growth months of May through August, during which time rainfall and temperature play a critical role in determining plant conditions. In areas with limited water, variability in precipitation and water demand resulting from temperature are major determinants of the state and vigor of vegetation, which is particularly useful for predicting yield variability for both dryland cropping and pastoral systems. The openness and relevance to climate-responsive agriculture of this proposal make it a valuable contribution; but, since basin-level vegetation indices do not directly represent crop yield, and additional grazing or land use pressures may bias hydro-climate/vegetation correlations, further crop-specific and management-specific analysis would be necessary to accurately interpret agricultural yield impacts of these overall findings [63].

From a modeling perspective, these findings indicate that crop yield prediction systems should not rely on rainfall or temperature alone. Instead, robust models should incorporate multiple hydro-climatic indicators, including precipitation timing, rainfall intensity, evapotranspiration, soil moisture, drought indices, river discharge, snowmelt timing, and flood-risk variables. Machine learning and hybrid models are particularly suitable for integrating these heterogeneous variables because they can capture nonlinear relationships and interaction effects between water availability, crop growth stage, and climatic stress. However, model performance will depend on the temporal resolution of the data, spatial scale of analysis, crop specificity, and availability of reliable ground-truth yield records. Therefore, future yield prediction frameworks should combine hydro-climatic observations, remote-sensing vegetation indicators, watershed-level water availability, and crop-management information to improve prediction accuracy, interpretability, and operational usefulness. As shown in **Table 1**, a comparative analysis of recent studies of Hydro-Climatic Factors in Crop Yield Prediction.

Table 1. Comparative Analysis of Hydro-Climatic Factors in Crop Yield Prediction.

Methodological Category	Hydro-Climatic Focus	Application Scenario	Main Contribution to Yield Prediction	Key Limitation
Long-term vegetation–hydroclimate trend analysis, Xu et al. (2026)	Vegetation greening, evaporation demand, water availability	Regional vegetation monitoring and drought-risk assessment	Shows how long-term greening and hydro-climatic change can alter water demand and productivity potential	Vegetation greenness does not directly equal crop yield; requires crop-specific yield and management data
Hydrological hazard and flood-risk mapping, Kakavand et al. (2025)	Flood hazard, hydrologic risk, drainage vulnerability	Flood-risk planning, drainage infrastructure, early-warning thresholds	Identifies areas where flood exposure may affect agricultural operations and crop-loss risk	Does not directly model yield; land-use change and river modification may confound hydro-climatic signals
hydro-climatic assessment, Alehu et al. (2022)	Precipitation distribution, water storage, temperature, snowmelt timing	Irrigation planning, water allocation, drought preparedness	Provides watershed-level understanding of agricultural water availability	Hydro-climatic variables are not directly linked to observed crop yield data
Sigalla et al. (2023)	Precipitation, maximum/minimum temperature, PET, river discharge	Climate-smart agricultural zoning and yield-risk mapping	Identifies spatial hydro-climatic zones useful for adaptation planning	Requires integration with crop-specific yield, soil, and management data for direct prediction
Remote sensing and time-frequency vegetation response analysis, Batungwanayo et al. (2024)	Rainfall variability, moisture stress, vegetation response	Early warning for rain-fed agriculture and seasonal yield monitoring	NDVI anomalies can provide early signals of crop stress before harvest	NDVI–yield relationship varies by crop, phenological stage, and preprocessing method; association does not prove causation
Evapotranspiration-based climate analysis, Mojid et al. (2024)	Atmospheric water demand, ETo, drought stress	Irrigation scheduling, planting-window optimization, drought assessment	Links hydro-climatic signals to practical farm-management decisions	Does not directly develop a yield prediction model; limited transferability without crop and management data
Arid/semi-arid vegetation–hydroclimate analysis, Dorjsuren et al. (2023)	Rainfall timing, temperature, vegetation dynamics	Dryland cropping, pastoral systems, yield-risk assessment	Highlights the importance of rainfall and temperature during peak growth months in water-limited systems	Basin-level vegetation indices may be affected by grazing, natural vegetation, and land-use change

5.3. Soil Role in Crop Yield Prediction

Different types of soils define the physical, chemical, and biological conditions that dictate crop growth potential. Soils can be classified by various physical and chemical characteristics such as texture (sand/silt/clay), structure, organic matter, pH, nutrient availability, water holding capacity, and drainage characteristics. Classification of soils helps to provide a systematic way to study how different types of soils will help or hinder crop production. The inherent characteristics of soils also affect key agronomic processes, such as root development, water retention, infiltration, aeration, and nutrient cycling, which all directly lead to yield predictions. In predictive modelling, soil classification is both an input and a contextual layer for modifying the impacts of climate and management on crop yield [64].

Soil maps or soil classes are commonly included in machine learning or deep learning models for data-driven systems as either categorical features or encoded features representing the spatial variation in the potential for production. High-resolution soil datasets provide a greater degree of accuracy in prediction models for yields because they help the model distinguish between zones of differing productivity based on their inherent characteristics. When combined with other data sources like climate variables, remote sensing indicators, and management information, soil classification provides greater interpretability of the model and enhances the model's ability to generalize across broader landscapes [65].

In addition, co-locating soil classification with precision agriculture workflows creates zone-specific recommendations such as variable rate fertilization, targeted irrigation, and crop suitability assessments. This directly links soil to optimized productivity. However, due to the limited spatial resolution of soil maps or failure of a classification scheme to capture dynamic soil characteristics (seasonal moisture changes, compaction, salinity) there are limitations to the applications of these approaches. Effective crop yield prediction frameworks leverage a combination of static soil classifications and real-time environmental and sensing data to address the static and dynamic (persistence/temporally) components of soil-related variability [66].

Precision Agriculture includes automated systems, IoT sensors, real-time data collection, a cloud storage solution, and data analysis. Soil characteristics impact soil fertility and cultivating yearly causes a reduction in soil nutrients spent to do so; therefore, this procedure can effectively raise the fertility of the soil to increase crop yield. Concurrently, Big Data, with the inclusion of IoT sensors and smart devices, is being utilized. Throughout nearly every defined phase of Precision Agriculture, there are various sensor devices present that measure moisture, water, solar moisture, pH, and assess mineral deficiencies of the soil so that sensor technology can be used to improve productivity in agriculture. Sensor technologies provide a means to improve soil quality, assure food safety, and increase profit in agriculture [67–69].

With traditional methods, samples are taken in a systematic grid or zone, collected, and then sent to labs for analysis. With new technologies, soil classification methods can use advanced soil and remote sensing tools while integrating with other sources through a GIS system. Also, Artificial Intelligence (AI) and Machine Learning (ML) algorithms are being developed to analyze and classify existing soil data from some of these advanced sources to develop predictions of soil properties and build management zones that account for soil variability. This could lead to sustainable, profit-maximizing, and efficient agricultural and horticultural practices by accurately understanding and addressing the unique characteristics associated with soil and the true levels of necessary care [70].

Both the USDA soil classification system [71] and the FAO (Food and Agriculture Organization) soil classification system [72] have had a profound impact on soil mapping and have provided a standardized framework for understanding, categorizing, and comparing soils worldwide. In the early days of mapping soils, the maps were often broad and qualitative, even in the early days of the USDA system, especially its Soil Taxonomy, now a highly detailed system for soils based on observable (that is, measurable properties of soils. Within U.S. soil taxonomy systems, soils are classified as orders, suborders, great groups, subgroups, families, and series, meaning distinguishing soils based on observable, unique characteristics has a precise, scientific descriptive language. This has resulted in very detailed modern soil maps, like those in the most recent U.S. Soil Survey maps. Whereas the UNESCO/FAO system was developed for a perspective on the world, it aimed to be simpler and have a broader reference for referencing and mapping soils globally. Thus, the FAO system categorizes soils into a small number of major groups as opposed to higher classification levels. In addition, the FAO system was also intended to give soil maps the possibility to be possible from regional to world scale for soils and to try to make the ID range and classification more practical on a smaller scale. Both systems have changed and developed over time [73].

Satellite-based vegetation index analysis is an extremely useful remote sensing application in both precision agriculture and environmental monitoring. It indirectly adds to our understanding of the soil, since it involves observing the plant's health and vigor. Although the vegetation indices directly measure vegetation, they are subject to variation based on the properties of the soil, nutrient availability, and water availability in the soil [74]. The Atmospheric Resistant Vegetation Index (ARVI) [75], is also developed to be less sensitive to atmospheric effects (such as aerosols) than the commonly used Normalized Difference Vegetation Index (NDVI). USDA Soil Taxonomy for Soil Resources to delineate areas that could be managed distinctly from other areas of the field because of similar innate properties. For example, well-drained Alfisols in one zone and poorly drained Vertisols in another zone may now need different crops or drainage in the Vertisols.

Dolas and Joshi (2018) present a soil-classification–based approach in which basic soil attributes, such as pH, moisture, and nutrient indicators, are used to categorize soil types and match them to suitable crops through machine-learning classifiers, demonstrating that soil properties alone hold strong discriminatory power for crop-suitability prediction. While their method is simple, low-cost, and practical for regions with limited data and sensing infrastructure, it primarily addresses crop suitability rather than yield magnitude and does not incorporate temporal dynamics, management practices, or climatic interactions, which constrains its usefulness for real-world yield prediction [76]. Suruliandi et al. (2021) integrate soil characteristics with environmental variables and apply feature-selection techniques to isolate the most influential predictors before modeling crop outcomes, feature-selection methods to identify the most important crop yield predictors before modeling, to improve model performance and explanation of soil nutrients, pH, and moisture; subsequently, the model is reduced to only those variables that are most significant within the data analyzed. Although their methods greatly reduce dimensionality and assist in delineating the degree that each soil variable influences crop yield, they also focus only on crop classification tasks, resulting in limited validation against actual yield data measured in continuous form, plus inadequate assessment of spatial heterogeneity, which presents sizeable limitations regarding fully functioning yield-prediction systems formulated from soil characteristics [77].

Paul et al. (2015) apply data-mining methods to explore how physical and chemical soil properties impact crop yields, revealing a strong relationship between crop yield and soil characteristics, and confirmation of non-linear relationships between soil and crop yields through analyses utilizing non-linear models. This research is among the earliest studies demonstrating an explicit link between soil properties and crop yield, and contributes important data in support of agricultural decision making; conversely, its use of limited datasets, rudimentary modelling methods, and the exclusion of spatial or remote sensing/crop production/climate information greatly impairs generalizability [78]. Khanal et al. (2018) discuss a method for predicting corn yield and soil properties using high-resolution remote sensing combined with machine learning predictive models across multiple fields from spatial predictions of soil properties and corn yield measures. This work demonstrates that there are substantial improvements in maximum yield prediction accuracy, and that soil spatial variability plays a major role in yield variability. This innovative approach provides precision agricultural applications with sample advantages through scalability, explicit spatial structure; however, it also requires significant resources regarding data collection/management and computation, and the transferability of the approach depends on sensor availability, calibration robustness, and region-specific soil/crop relationships [79].

Soil properties are key predictors of crop approval and yield, because they affect nutrient availability, water retention, and root growth. Initially, classification-based approaches were used for predicting yield but were limited in terms of static crop recommendation. Soil-classification models are simple and useful for crop recommendation but provide limited information about expected yield magnitude. Feature-selection-based approaches improve interpretability by identifying the most influential soil variables, but they often remain limited to classification tasks. Data-mining approaches establish a more direct relationship between soil properties and yield, although their predictive power is constrained by limited datasets and lack of spatial or climatic integration. Remote-sensing and machine-learning approaches provide the strongest potential for precision yield prediction because they capture spatial heterogeneity and support field-level decision-making.

From a modeling perspective, soil properties should not be treated as isolated predictors. Instead, effective crop yield prediction models should integrate soil variables with climate, hydrological conditions, remote-sensing indicators, crop-management practices, and temporal growth-stage information. This is particularly important because the influence of soil nutrients, pH, and moisture on yield depends on rainfall distribution, irrigation availability, crop type, fertilizer use, and seasonal temperature patterns. Therefore, future yield prediction frameworks should combine soil-based indicators with spatial data, time-series observations, and explainable machine-learning methods to improve prediction accuracy, model transparency, and practical usefulness for precision agriculture.

6. Precision Agriculture and AI Technologies for Yield Prediction

6.1. Precision Agriculture

Precision Agriculture (PA), referred to as smart farming, and site-specific crop management is the next-generation, data-based approach in agriculture where farmers can monitor, evaluate, and react to both spatial and time-based differences in their farming land to target their resources accordingly to optimize production, enhance productivity, and improve sustainability [80]. A range of advanced technology tools, such as Global Positioning

Systems (GPS) for accurate machine steering and mapping; Internet of Things (IoT) sensors to measure real-time soil moisture, crop quality/life, and environmental conditions; and remote-sensing devices (satellite images and drones) to provide a comprehensive picture of the fields, have all contributed to this scientific paradigm of agricultural production. With the enormous amount of data these devices generate, Artificial Intelligence (AI) and Machine Learning (ML) algorithms can be used to process and analyze the information to forecast and provide optimal planting, fertilization, irrigation, and pest management solutions [81,82].

Talaat (2023) has developed the CYPA framework, a climate-aware crop-yield prediction system that uses IoT-based measurements from the field, and other climate and environmental factors, to yield predictions under the increasing variability brought on by climate change. The CYPA system integrates multiple types of input, including soil conditions, weather indicators, and crop traits; strengthens the accuracy of predictions; and provides a real-time precision agriculture decision support system that uses sensors to measure information. The CYPA model has significant advantages in that it will allow for more accurate predictive modeling of crop yield using modeling approaches aligned with the realities of modern-day PA and will specifically focus on addressing uncertainty due to climate change; However its ability to be generalizable across regions and types of crops will hinge on the availability of sensors; Completeness of data; and On the level of transparency in regards to the entire data pipeline used to develop the models [83].

Burdett and Wellen (2022) complement the study by comparing statistical methods and machine-learning methods for predicting (using high) resolution soil characteristics (and) topography relative to precision agriculture environments, and that there is a strong relationship between fine-scale variability of soil conditions and crop yield, and different modeling methods capture those relationships between soil condition and crop yield in different ways. The work of Burdett and Wellen is meaningful as it provides guidance to those looking to select a method for use in sub-field decision-making [84].

Parashar et al. (2023) employ a big-data and deep-learning framework built around IoT-enabled meteorological measurements to improve yield forecasting, arguing that deep learning's ability to capture complex, nonlinear interactions among hydro-climatic variables leads to more robust predictions under variable weather conditions. Their approach highlights strong scalability and suitability for modeling nonlinear agronomic responses, though it faces typical deep-learning challenges, including limited interpretability, uncertain external validity when trained on narrow datasets, and sensitivity to missing or noisy IoT data streams [85]. Bakthavatchalam et al. (2022) present an IoT-driven measurement framework coupled with supervised machine-learning classifiers, implemented in WEKA, including multilayer perceptron and decision table models, to predict crop or high-yield outcomes from sensor-based field measurements, demonstrating an accessible end-to-end "sense to predict" workflow suited for precision agriculture. Their results indicate that standard ML models can effectively classify crop-related targets when provided with measurable soil and environmental inputs, offering a practical, implementation-friendly approach; however, the study's reliance on limited datasets and categorical prediction tasks means that its classification accuracy may not generalize to robust, continuous yield prediction across seasons and spatially heterogeneous fields [86].

Akintuyi (2024) investigated how self-learning, adaptive AI systems utilize real-time data from sensors, soil and the environment in order to continuously inform decisions for optimum farm operation. He argues that closed-loop, dynamic learning frameworks can not only improve monitoring but will also provide protective measures for crop yield during periods of rapidly changing environmental conditions. Akintuyi effectively conveys that autonomous, feedback-driven systems represent future precision agriculture; however, his conceptual construct has very few empirical benchmarks; the practical constraints of data governance, managing model drift, and deploying computing capabilities at the edge are still major limitations to implementing these systems successfully [87]. Dusadeerungsikul and Nof (2024) present an AI-based responsive monitoring framework using a Dynamic-Adaptive Search (D-AS) algorithm that adjusts inspection intensity based on detected crop-stress signals, enabling earlier problem recognition and more targeted interventions. The primary advantage of their framework is its ability to shift the analytical capabilities of precision agriculture from static yield prediction to actionable operational decision-support, detection of stress, response to stress, and crop yield protection. However, the efficacy of the developed system is dependent upon reliable stress-detection inputs, automatic integration into agricultural operations, and broader validation across diverse crops and field conditions [88].

Peerlinck et al. (2019) develop an innovative method for designing on-farm experiments using a genetic algorithm (GA) approach to create fixed-variable rate nitrogen rate trials across an entire farm field while minimizing overall costs. Their approach creates experimental designs for fixed-variable rate nitrogen rate trials with high coverage of yield response categories, as well as low operational burdens for conducting the trials. The result is an improvement in the overall quality of the agronomic data collected through these experimental designs, which provides for improved farmers' ability to model local response and ultimately supports better yield optimization. Furthermore, since their method focuses specifically on optimizing the overall data collection process, it directly addresses one of the major limitations in precision agriculture today, which is the bottleneck in the overall data collection process [89].

According to recent studies, IoT, AI, and adaptive sensing frameworks contribute to crop yield prediction in different but complementary ways. Climate-aware IoT frameworks such as CYPA improve real-time yield prediction under environmental uncertainty; soil–topographic modeling supports sub-field management; big-data and deep-learning systems enhance nonlinear forecasting capacity; supervised IoT-based classifiers provide practical implementation pathways; adaptive AI frameworks enable closed-loop decision support; and genetic-algorithm-based experimental designs improve the quality of agronomic training data. However, common limitations remain, including limited external validation, weak interpretability, dependence on high-quality sensor streams, lack of continuous yield benchmarking, and uncertain transferability across crops and regions. Therefore, future precision agriculture systems should integrate IoT sensing, remote sensing, soil–topographic data, climate variables, explainable AI, and well-designed on-farm experiments into unified, transparent, and scalable yield prediction frameworks.

6.2. AI Models' Role in Crop Yield Prediction

Progresses made in the use of machine learning (ML) in precision agriculture have led to improvements within multiple applications, including predicting yield, classifying soils accurately, and estimating yields optimally. The impressive power of ML (especially supervised learning) to model complex and typically nonlinear relationships between various environmental factors (e.g., climate attributes such as weather patterns and soil nutrients, precipitation and temperatures) relative to their corresponding agricultural outcomes is evident. Random Forest (RF) is considered one of the most effective ensemble methods in ML techniques due to its robustness, ability to accommodate high dimensions of available information, and immunity to overfitting when predicting yield [90].

Shekhar et al. developed a machine learning-based crop recommendation system designed to enhance agricultural yield by leveraging environmental and soil features such as nutrient content, pH, temperature, and rainfall. They incorporated multiple supervised learning models to produce optimal crop recommendations for a given set of input features. From a practical standpoint, using machine learning to provide intelligent recommendations helps improve agricultural decision-making, resulting in more efficient use of resources and improved yields [91].

Baligar revolutionized traditional agriculture by harnessing machine-led territorial decision aids and timely data analysis via agricultural machinery and real-time data sources. The recommendation system combines data from various sources to produce the best possible recommendations based on existing field data; it collects data from sensors and weather API's, enabling it to quickly develop dynamic adaptive recommendations that are based on current field conditions. The system aims to keep farmers simple and accessible to help farmers make better-quality decisions, reducing their reliance on manual expert knowledge and increasing agricultural production for farmers with smaller land holdings [92]. Bhola and Kumar (2025) presented an intelligent framework for providing farming-level crop recommendations utilizing a combination of machine learning techniques to recommend contextualized crop planting for a specific environment and soil characteristics. It incorporated comprehensive data cleansing and all of the essential data preprocessing, developing of features and validating the respective models; the study considered various real datasets and provided a full validation of the methodology utilized. The work demonstrated opportunities for agriculture to integrate technologically driven agricultural-based practices to promote more appropriate crop selection, higher crop yield, and to provide farmers with valuable information on how to best manage their farms [93].

Mancer et al. proposed an ensemble learning method of improving crop recommendation systems' performance. They combined multiple types of machine learning models (decision trees, random forests, boosted models) in a single system to achieve better accuracy and robustness than using only one type of model. Their research demonstrated that ensemble methods can successfully model complex interactions between agricultural variables, resulting in more reliable recommendations, thus increasing the scalability and adaptability of precision agriculture systems [94]. Mimoune and Lantri developed a region-specific system that used decision trees, support vector machines, and random forests to make recommendations based on local climate and soil characteristics. Data preprocessing (cleaning, normalizing, etc.) is extremely important, and evaluating models thoroughly using various metrics (accuracy, precision, recall) will improve crop recommendation models. While the study found positive results for the area in which the crops were tested, it also noted that the dataset was small and that findings may not apply to other regions. Therefore, validation and real-world application of crop recommendation models are necessary [95].

Choudhary et al. proposed a three-part innovative dual system based on plant detection as well as recommending crops growing in a specific area using machine learning. The first module will use data about soil and climatic conditions to generate crop recommendations, while the second module will use vision systems for image processing to classify identified diseases found on plants. As an integrated approach, the system is designed to assist farmers in producing crops and managing plant diseases by utilizing accurate crop recommendations. The paper lacks detailed performance metrics or evaluation information relating to the crop disease classification portion of the system [96]. Ali et al. created a machine-learning-based crop recommendation system specifically useful to Pakistan's farmers to provide customized crop recommendations for local farmers through the use of environmental, soil, and regional data with the goal to provide farmers with an alternative means to combat specific agricultural problems in the region. Ease of use and regionally relevant customization of the system is emphasized to develop a decision support system designed to improve harvest yield results for farmers working in resource-poor rural communities. The model is practically a useful tool; however, the authors have provided a limited amount of discussion on the need for validation of the system across different climate regions [97].

Kathiria et al. developed a smart crop recommendation system to assist farmers in deciding what crops to grow. The system uses machine learning models to analyze soil conditions and environmental conditions for precision agriculture. Different algorithm types were evaluated; (for example, Decision Tree Algorithms, Support Vector Machines, etc.) to determine the best types of crops for each individual farmer based on specific characteristics of his/her land, such as nutrient content, moisture content, and pH level, etc. The current project provides an overview of all models used and a comparative assessment of their respective performances and concludes with the potential for automation of many agriculture-related decisions through intelligent systems [98]. Chauhan and Chaudhary presented a machine learning-based crop recommendation system that compares the effectiveness of various algorithms, for all of the ML techniques studied, but did not include any information related to Farmers piloting farm equipment or testing their systems before integrating them into standalone devices or using them on a field scale in any type of commercial operation. This makes the results of this project largely academic in nature with neither the potential for commercial application nor much application for practical purposes in real-world situations [99].

Priyadharshini et al. created a crop recommendation system utilizing multiple machine-learning techniques, including Naive Bayes, SVM, KNN, Decision Trees, and ANN. The data is obtained from publicly available sources as well as government agencies. The development of the system involved extensive processes such as preprocessing, model comparison, validation, and selection. The neural-network algorithm achieved an impressive accuracy rate of approximately 90% using the classification system. The research provides a comprehensive comparison of classification models used for crop prediction [100].

Recent research utilized a combination of algorithms ranging from basic classification methods (e.g., Decision Trees and support vector machines) to advanced ensemble methods that analyze environmental, soil, and climate data to generate precise crop recommendations. Many authors, including Mancer et al. (2023) and Shekhar et al. (2025), have focused on the robustness and accuracy of machine learning models. Conversely, authors such as Ali et al. (2021) and Baligar (2025) have attempted to improve the regional applicability and real-time responsiveness of the methods. A major trend among machine-learning crop-recommendation systems that continue to emerge is the transition from single-model systems to hybrid and ensemble systems that provide greater prediction accuracy and generalizability. **Table 2** provides a comparative analysis of machine learning-based crop recommendation and yield prediction systems. The findings indicate that ensemble and hybrid approaches generally achieve higher accuracy and robustness compared to single-model systems.

Table 2. Comparison of Machine Learning-Based Crop Recommendation and Yield Prediction Approaches.

Study	Approach/Model	Data Used	Strengths	Limitations	Key Contribution
Shekhar et al. [91]	Supervised ML models	Soil, climate, nutrients	Improves decision-making and resource efficiency	Limited discussion on scalability	Intelligent crop recommendation system
Baligar [92]	ML + real-time sensor integration	IoT sensors, weather APIs	Real-time adaptive recommendations	Complexity in implementation	Dynamic decision support for farmers
Bhola and Kumar [93]	ML framework with preprocessing & validation	Real-world datasets	Comprehensive pipeline with validation	Limited scalability discussion	Full ML workflow for precision agriculture
Mancer et al. [94]	Ensemble learning (RF, boosted models)	Multi-source agricultural data	High accuracy and robustness	Increased computational complexity	Improved generalization via ensemble methods
Mimoune and Lantri [95]	Decision Trees, SVM, RF	Local soil and climate data	Region-specific optimization	Small dataset, limited generalization	Localized prediction system
Choudhary et al. [96]	ML + computer vision hybrid system	Soil + plant images	Integrated crop recommendation + disease detection	Lack of performance evaluation	Multi-module intelligent farming system
Ali et al. [97]	ML-based regional system	Soil, environmental, regional data	Practical for local farmers	Limited cross-region validation	Customized regional decision support
Kathiria et al. [98]	Multiple ML models comparison	Soil and environmental data	Comparative evaluation of algorithms	Limited real-world deployment	Model benchmarking for crop recommendation
Chauhan and Chaudhary [99]	Comparative ML models	Agricultural datasets	Algorithm comparison	No field validation	Academic evaluation of ML models
Priyadharshini et al. [100]	ML + ANN	Public datasets	High accuracy (~90%)	Limited interpretability	Strong performance of ANN models

7. Prediction Models Architectures

7.1. Field-Level Prediction

The goal of field-level yield prediction systems is to capture spatial variability within a field by incorporating sensor data generated from in situ sensors, proximal sensing systems (e.g., handheld or tractor-mounted sensors, etc.), UAV (drone) imagery, and yield monitoring systems into one model. In addition, yield prediction models utilize high-resolution soil, crop, and management data to model the within-field differences and provide recommendations for variable configurations (irrigation, fertilization, and pest control). Although field-level yield models have high prediction accuracy due to their detailed data collection capabilities, they rely heavily on a dense sensing infrastructure and require strict calibration. To improve sub-field and field-level yield predictability with multi-modal data fusion and machine learning techniques, Dhaliwal and Williams have found that employing field-level management, weather, and soil data to machine learning models improved accuracy, and improved crop-specific yield predictions when combined with field-level management, weather, and soil data [101], while Pathak et al. [102]. and Helber et al. [103] systematically confirm that the prediction performance improves with respect to richer input data provided by remote sensing, weather, soil, and crop management data that have been integrated at finer spatial resolutions. Mena et al. [104] extend these findings by proposing adaptive fusion strategies that dynamically weight multi-modal remote sensing inputs, and indicate that some data sources may not contribute equally to predictive capability across plant development stages or across fields. Filippi et al. [105] provide an early operational perspective, demonstrating that scalable, multi-farm datasets can support transferable yield models when spatial and temporal variability are explicitly modeled. Across all studies, the key advances lie in sub-field resolution, multi-source data fusion, and operational scalability, while common challenges remain in model transferability, data harmonization, and balancing prediction accuracy with deployment complexity in real-world precision agriculture systems.

7.2. Regional-Scale Prediction

Regional-scale yield predictive models utilize and aggregate multi-field (or administrative unit) data sources and forecast trends in crop yields across larger geographic areas of interest. Frequently, they incorporate historical weather records, soils derived from general soil types, aggregated management data, and medium spatial resolution remote sensing data. All of these data sources are helpful and valuable resources for food security planning, market prediction and professional policy analysis; however, as regional models are incorporated into these systems frequently to smooth out local variability at the regional level, and therefore reduce the amount of accuracy that is achieved for individual farms, and introduce more uncertainty into heterogeneous landscape scenarios, the use of integrated remote sensing technology (via NDVI, EVI, SIF, temperature, and moisture indicators) with climate data (e.g., precipitation, temperature) and crop growth models provides the foundation for robust yield predictions at the regional level, and should be conducted to accurately assess the spatial-temporal variability of large land areas used for agricultural production. Kang et al. [106] found that satellite-based physiological indicators (SIF) can be an effective early predictor of regional cotton productivity when utilized via data-driven temporal-spatial frameworks. By using a multi-task learning approach to jointly analyze climate data and remote sensing data to improve forecast accuracy for mid-season yield predictions and consistently predict yield across heterogeneous landscapes.

Wang et al. [107] were able to improve forecast accuracy for mid-season yield predictions, and consistently predict yield across heterogeneous landscapes.

Kirthiga and Patel [108], as well as Zhuo et al., highlighted the value of hybrid approaches that couple regional climate models or numerical weather forecasts with process-based crop models, allowing yield prediction under varying environmental conditions while retaining biological interpretability [109]. Bandaru et al. [110] generalized this concept through a geospatial crop simulation framework that scales crop and water-use assessments regionally. Across these works, the main strength of regional-scale prediction lies in its ability to balance coverage and accuracy for policy, market, and food-security applications, while persistent challenges include spatial aggregation errors, calibration across diverse agro-ecological zones, and the computational complexity of integrating multi-source data streams into operational forecasting systems.

7.3. Satellite-Based Large-Scale Systems

Multi-temporal Earth observation data, such as vegetation indices, surface temperature, and soil moisture products, are used in satellite-based yield prediction systems to assess the growth of crops over very large areas. These systems make it possible to estimate agricultural yields accurately and consistently at global and/or national scales while providing repeatable, cost-effective yield estimates, particularly in areas where little ground data is available. The performance of these systems is influenced by cloud cover, sensor resolution, and revisit frequency, and they generally need crop and area-specific calibrations to convert spectral signal measurements into accurate yield estimates. Previously demonstrated use of satellite-based techniques for predicting crop yield at a large scale demonstrates that multi-temporal Earth observation data is a scalable and reliable source of information for measuring crop growth and estimating yield in very large areas.

Ji et al. demonstrated how the combination of vegetation index (VI) time series with phenological metrics can produce improved generalized models of crop developmental dynamics at very large scales and provide greater robustness to predictions made based solely on single date data [111]. Li et al. illustrated how the utilization of multi-source satellite data, combined with other environmental variables, can improve the accuracy of yield prediction through the incorporation of spatial variability in climatic and land surface conditions into modelled predictions [112]. Tripathy et al., the potential advantages of fusing data from multiple satellite platforms to provide higher-resolution yield estimates for a range of cropping systems, particularly in data-poor areas [113]. Ishaq et al. extend this work by using time-series satellite observations integrated with machine-learning algorithms, as well as crop growth models, to improve the stability and geospatial robustness of large-area wheat yield predictions. Collectively, these studies indicate that satellite-based systems represent a strong option for the coverage, repeatability, and cost effectiveness when forecasting yields over large areas, while their major limitations are their dependence on sensor resolution and frequency of overpass, sensitivity to cloud cover, and the need for region and crop-specific calibration to convert spectral signals into accurate yield estimates [114].

7.4. Edge and Cloud Processing

Edge-based architectures perform data processing and preliminary prediction directly on farm-level devices or gateways, reducing latency, bandwidth usage, and dependence on continuous connectivity. In contrast, cloud-based architectures centralize storage and computation, enabling advanced analytics, model training, and large-scale data fusion. Hybrid edge–cloud designs are increasingly adopted to balance real-time responsiveness with computational scalability, though challenges remain in synchronization, security, and model consistency across distributed environments.

Edge and cloud processing for crop yield prediction collectively highlight a shift toward distributed intelligence architectures that balance real-time responsiveness with scalable analytics. Dey et al. [115] have demonstrated the ability to perform some amount of processing and predictions at the edge using lightweight (6) bi-directional long-short term memory (BiLSTM) models while transferring large amounts of data to the cloud for model training, historical data analysis, and global optimizations, thus lowering latency and reducing total bandwidth consumption [116]. Sharma and Rathore [117] AGRO-Cloud model emphasizes centralized cloud intelligence for large-scale data aggregation and decision support, demonstrating improved prediction accuracy through scalable computation but with higher dependency on connectivity. Bera et al. [118] have further advanced the development of edge-cloud architecture by proposing the Dew-Edge model, which provides local autonomy and resiliency in low-connectivity environments through multi-parametric sensing integrated directly into the on-site recommendations. Edge-cloud computing architecture will serve to increase the performance, reliability, and scalability of predictive models in precision agriculture, thus supporting the advancement of intelligent systems for crop yield prediction.

7.5. Real-Time Yield Forecasting

Yield forecasts that are provided in real-time update during the growing season alone by putting together data from previous growing seasons along with other data collected in the current growing season from control and sensor data, meteorological forecasts, and observations collected through remote sensing. This type of yield prediction model allows yield prediction to be a dynamic and timely tool to make decisions about how to intervene and how to mitigate risk by providing accurate information about the probability of yield before harvesting begins. New yield forecasting models offer better prediction of yield in response to climate instability and offer more resilience. To achieve good yield forecasting through real-time forecasting, the yield prediction systems need reliable data pipelines, adaptive models, and systems that can manage uncertainty, missing data, and rapid change in environmental conditions.

Dhakar et al. [119], and Zare et al. [120] have shown that one can make accurate yield predictions at the time of the growing season by continuously updating yield predictions through real-time data collection and online merging of real-time data with crop simulation models. Chen and Tao [121] investigated yield predictions made early in the growing season are significantly improved by combining remote sensing observations with crop simulation models using forecasts for seasonal weather at large scales. New hybrid crop modeling techniques provide ways for real-time UAV cameras (drone) to provide fast updates in-field imagery and updated information, such as the use of machine learning agricultural helpers, as Jain et al. [122] provide rapid field-level updates that enhance responsiveness to emerging crop stress, while adaptive AI frameworks, Akintuyi [87] emphasizes self-learning models capable of adjusting predictions as new data arrive. Real-time yield forecasting emerges as a shift from static, end-of-season estimation toward iterative, uncertainty-aware prediction systems that support timely management decisions, though persistent challenges remain in data latency, model complexity, uncertainty propagation, and the operational integration of heterogeneous real-time data streams. **Table 3** presents a comparative analysis of crop yield prediction architectures across different spatial scales and system designs. The results highlight a trade-off between accuracy and scalability, where field-level models achieve high precision but require intensive data infrastructure, while satellite-based and regional models provide broader coverage at the cost of reduced granularity.

Table 3. Comparison of Crop Yield Prediction Architectures across Spatial and System Levels.

Architecture	Approach	Data Sources	Strengths	Limitations	Key Studies
Field-Level Prediction	Supervised ML models	IoT sensors, UAV imagery, soil data, management data	High accuracy, captures spatial variability, supports precision farming	Requires dense sensing infrastructure, high cost, and calibration complexity	Dhaliwal and Williams [101]; Pathak et al. [102]; Helber et al. [103]
Regional-Scale Prediction	Multiple ML models	Climate data, aggregated soil data, satellite indices (NDVI, SIF), crop models	Suitable for policy, food security, and market forecasting	Reduced accuracy at the farm level, spatial aggregation errors	Kang et al. [106]; Wang et al. [107]; Kirthiga and Patel [108]; Zhuo et al. [109]
Satellite-Based Large-Scale Systems	ML-based regional prediction system	Multi-temporal satellite data, vegetation indices, surface temperature	Scalable, cost-effective, useful in data-scarce regions	Sensitive to cloud cover, resolution limits, and requires calibration	Ji et al. [111]; Li et al. [112]; Tripathy et al. [113]; Ishaq et al. [114]
Edge Computing Systems	Deep learning models	Local sensor data, IoT devices	Low latency, reduced bandwidth, real-time decisions	Limited computational capacity	Dey et al. [115,116]
Cloud-Based Systems	Comparative ML models	Centralized multi-source data (historical + real-time)	High computational power, advanced analytics, scalable	Dependency on connectivity, latency issues	Sharma and Rathore [117]
Hybrid Edge–Cloud Systems	ML with real-time sensor integration	IoT + cloud + distributed data	Combines real-time processing with scalability	Synchronization and security challenges	Bera et al. [118]
Real-Time Yield Forecasting Systems	ML with real-time sensor integration	Sensor data, weather forecasts, UAV, and remote sensing	Continuous updates support decision-making during the season	Data latency, uncertainty handling, and system complexity	Dhakar et al. [119]; Zare et al. [120]; Chen and Tao [121]; Jain et al. [122]; Akintuyi et al. [87]

8. Data Sources for Crop Yield Prediction

8.1. Ground Sensor Data

Soil and crop conditions, such as soil moisture, temperature, electrical conductivity, nutrient levels, and canopy microclimate, can be measured in real-time and accurately through ground sensors. Measurements made using ground sensors are critical for capturing variability on the field and dynamics that impact the growth and development of crops on a short-term basis, as well as improving the accuracy of forecasts and providing the ability to forecast in real-time; however, some limitations exist with ground sensor data due to cost of deployment, maintenance issues with the sensors, calibration difficulties with each specific sensor, and poor spatial coverage. A summary of the data sources that are used to predict crop yields is provided in **Table 4**.

Table 4. Comparison of Data Sources Used in Crop Yield Prediction.

Data Source	Spatial Scale	Temporal Resolution	Key Variables	Strengths	Limitations	Suitability
Ground Sensor Data	Field/Sub-field	Real-time to hourly	Soil moisture, soil temperature, nutrients, microclimate	High accuracy, real-time monitoring, captures local variability	High cost, limited coverage, maintenance and calibration issues	Field-level prediction, real-time forecasting, edge-based systems
Remote Sensing Data	Field to Global	Daily to seasonal	NDVI, EVI, LAI, canopy temperature, soil moisture proxies	Scalable, spatially continuous, cost-effective	Cloud contamination, resolution limits, indirect yield inference	Regional & large-scale prediction, spatial variability analysis
Historical Yield Records	Field to National	Seasonal/annual	Crop yield statistics	Essential ground truth, supports model training & validation	Inconsistent quality, missing management context	Model calibration, benchmarking, trend analysis
Weather Databases	Local to Global	Hourly to seasonal	Precipitation, temperature, radiation, humidity, wind, ET	Strong physiological relevance, widely available	Spatial interpolation errors, forecast uncertainty	All prediction scales, early- and mid-season forecasting
Farmer Input Logs	Field	Event-based/seasonal	Planting date, cultivar, irrigation, fertilizer, pesticides	Explains management effects, improves interpretability	Incomplete, inconsistent, low availability at scale	Field-level ML models, decision support systems
Public Agricultural Datasets	Regional to Global	Seasonal/annual	Yield statistics, soil maps, land use, RS products	Open access, reproducibility, large-scale coverage	Coarse resolution, limited local specificity	Regional/global forecasting, policy and food security studies

8.2. Remote Sensing Data

Remote sensing data from satellites and UAVs provide spatially uniform data over large areas by collecting data from vegetation indices (NDVI, EVI, LAI), surface temperature, and proxies for soil moisture. The information provided by remote sensing data is necessary to monitor and track crop growth, stress, and phenological stages of crops at different geographic scales. Remote sensing-based predictions are scalable, cost-effective, and can provide data in real-time; however, remote sensing-based predictions may also be affected by clouds, limitations in sensor resolution, and the requirement to calibrate the sensor and model for the specific crop and region to accurately interpret spectral signals to obtain yield estimates for crops.

8.3. Historical Yield Records

Yield data from previous years also captures a basis for developing, validating, verifying, and benchmarking models that predict agricultural yields. It shows the long-term increase or decrease in yield based on various factors such as cropping practices, land management, and how the variability of climate has affected agricultural production. The wear and tear that can affect farmers' yield data has resulted in an inconsistent pattern, which may not be directly applicable to predictive models for crop yield prediction and thus requires preprocessing and/or normalization.

8.4. Weather Databases

Weather databases provide critical inputs to the predictive models by supplying the necessary hydro-climatic variables (precipitation, temperature, solar radiation, relative humidity, wind speed) required to create these predictive models. The hydro-climatic variables strongly influence the crop growth rate and also the amount of stress an individual crop will be under during its growth phase. Therefore, the weather data constitutes the foundation for both empirical yield models and process-oriented yield models. However, there may be issues such as the error associated with spatial interpolation, forecast error or the mismatch between station-based observations and conditions directly next to the station.

8.5. Farmer Input Logs

Farmer logbooks are essential in documenting the management practices used by individual farmers, including when crops are planted, what variety of crops are planted, when irrigation occurs, how much fertilizer is used and when pest control was used. The information within these logbooks creates a very important context or frame of reference when looking at crop yield data and helps to sort out both the farmer management practices and the environmental conditions that occur at a specific time. Although farmer log books are of great relevance in looking at agricultural yield predictions, due to their often incomplete, inconsistent or not available in a systematic manner at a large scale they are not be used in large systemic processes that are involved in agriculture.

8.6. Public Agricultural Datasets

Agricultural data sets, such as national crop statistics, general remote sensing products, soil databases, and crop benchmark yield data sets, allow repeatability in research (and thus usability) for creating large-scale models.

They can also support yield prediction from regional to continental levels and allow for comparisons of agricultural production from different agro-ecological zones. Methods used to gather the data are often through coarse temporal and spatial resolutions, limiting their usability to directly apply to the precision agriculture field without additional data sources. **Table 5** shows a comparison table of selected major public agricultural datasets.

Table 5. Comparison Table of Major Public Agricultural Datasets.

Dataset	Geographic Coverage	Data Types Included	Primary Uses	Key Features
CropHarvest	Global	Sentinel-1 & 2 imagery, ERA5 climate, SRTM DEM, crop-type labels.	Crop type mapping, ML benchmarking, yield-model inputs.	95,186 datapoints; integrates 21 datasets; ML-ready with benchmarks.
NASA Harvest Portal	Global	Satellite imagery, crop datasets, climate data, food-security indicators.	Yield estimation, food-security modeling, agri-monitoring.	Searchable repository of geospatial datasets from multiple providers.
FAO Agro-informatics	Global & Country-level (Kenya, Pakistan, etc.)	Yield data, crop type maps, flood extent, suitability layers, damage assessments.	Agrifood planning, yield monitoring, climate-impact assessment.	Frequent country-specific releases; supports the Hand-in-Hand initiative.
Lacuna Fund ML Datasets	Primarily Africa (Kenya, Rwanda, Tanzania, etc.)	Crop images, field geolocation, yield data, spectrometry, disease/pest images.	ML for yield modeling, crop-health detection, field-boundary mapping.	High-quality labeled data for under-represented regions (Global South).

CropHarvest is the largest accessible (open-source, machine learning-ready) dataset that is intended to assist with global agricultural modeling by providing information to distinguish one crop type from another. The dataset has 21 separate agricultural land-use/remote sensing data types in a single format; the unified database has over 95,000 datapoints, including binary crop and non-crop identification and multiclass crop identification labels (crop types). The dataset is built using satellite images from both Sentinel-1 and -2, location data from SRTM, and climate data from ERA5, all of which make it a very useful resource to predict crop yield, model land cover, or to assist with research in domain shift. The CropHarvest dataset can be downloaded from either GitHub with supporting documentation, benchmarks, and preprocessed data uploaded in a standardized format [123].

The NASA Harvest Portal is the primary catalog for global agricultural geospatial data. Through structured access to satellite imagery, crop data, climate data, and indicators of food security, this portal has been developed to provide researchers and decision makers with all the information they need to successfully use Earth observation data for agricultural monitoring, yield prediction, and planning for risk reduction. In addition, the Harvest Portal contains a searchable catalog of datasets that have been collected from a variety of sources, including satellite imagery, and provides the ability to examine agricultural data in many different dimensions (e.g., crop yield, climate effects, land use change) related to an area of interest. As such, it acts as a centralized, integrated remote sensing tool for agriculture-related decision-making [124].

FAO also provides a variety of open data on agricultural geospatial data through its development of the Agro-informatics platform. This platform contains open geospatial datasets that support the analysis of the global agri-food system and includes crop yield estimation, crop type maps, land suitability assessments, flood extent datasets, and assessments of agricultural damage. The datasets are all updated regularly through various FAO initiatives, such as the Hand in Hand program, and include data on specific countries—for example, the rice yield estimation for Kenya from 2019 to 2022 and crop type data for Pakistan. The platform also contains data on aquaculture, forestry, agricultural vulnerability, and rural development planning. Ultimately, this resource supports evidence-based government agriculture transformation and provides valuable assistance in agriculture policy analysis.

The Lacuna Fund provides multiple high-quality, open datasets designed specifically for machine-learning applications in global agriculture, with a focus on improving food security and supporting smallholder farming systems. Its datasets include Eyes on the Ground, a large collection of labeled crop images with yield and management information from Kenyan smallholder farms; a high-accuracy maize plot geolocation and yield dataset covering Kenya, Rwanda, and Tanzania; a large-scale crop pest and disease imagery and spectrometry dataset with over 127,000 images; and datasets tracking pastoral land-use patterns in Tanzania. These datasets support tasks such as yield modeling, crop-health detection, field-boundary mapping, and community-level land-use planning.

9. Discussion

The discussion is structured in alignment with the PRISMA-based systematic review process and the statistical findings presented in Section 2. Specifically, insights are interpreted in relation to the temporal distribution of publications, methodological trends, and data-source utilization identified across the 82 selected studies. This ensures that the discussion directly reflects the synthesized evidence base rather than providing a general narrative interpretation. The role of soil properties as both a predictive feature and a contextual variable further emphasizes the importance of multi-dimensional data integration. While high-resolution soil datasets enhance model accuracy and interpretability, their static nature and limited temporal resolution restrict their ability to capture dynamic soil processes. This reinforces the need to integrate static datasets with real-time sensor data to improve predictive performance.

From a systems perspective, this review highlights a critical limitation in current research: the lack of standardized datasets, evaluation metrics, and benchmarking frameworks. The absence of such standards makes it difficult to conduct fair comparisons between models and limits the reproducibility of results across studies. This issue is further compounded by the fragmented nature of data sources and preprocessing techniques, which hinders the development of scalable, transferable prediction systems.

The review also reveals a clear difference between prediction models developed for field-level, regional-level, and large-scale applications. Field and sub-field studies generally report higher accuracy because they use high-resolution soil, sensor, management, and proximal-sensing data. These studies are particularly useful for precision agriculture applications such as variable-rate fertilization, irrigation scheduling, and site-specific crop management. However, field-level models often suffer from limited generalizability because they are trained on local datasets and may not transfer well across regions, crops, or management systems. In contrast, regional and large-scale studies rely more heavily on satellite imagery, climate datasets, vegetation indices, and crop-growth models. These approaches are more scalable and useful for food-security monitoring, but they often provide lower spatial precision and require careful calibration against ground-truth yield data.

Hybrid approaches appear to offer the most promising direction because they combine the strengths of multiple modeling paradigms. For example, models that integrate crop-growth simulations with machine learning can combine process-based agronomic knowledge with data-driven pattern recognition. Similarly, remote-sensing and IoT-integrated models can capture both spatial and temporal field variability. These hybrid systems are more suitable for yield prediction under climate uncertainty because they can represent crop physiology, environmental stress, and real-time field conditions more effectively than purely statistical or purely data-driven approaches. However, hybrid models also require more complex data integration, careful calibration, and standardized validation protocols.

The analysis of publication trends also indicates that research in crop yield prediction has experienced rapid growth since 2020, driven by increased global concerns about food security and climate change. This surge reflects the urgency of developing resilient agricultural systems capable of adapting to environmental and socio-economic pressures. However, the concentration of publications within a limited number of publishers suggests potential bias in research dissemination and highlights the need for broader interdisciplinary collaboration. To enhance the transparency and comparability of the reviewed studies, a structured mapping of key attributes was performed. **Table 6** summarizes the main characteristics across the 82 selected studies, including crop type, study region, spatial scale, data sources, modeling approaches, validation strategies, and evaluation metrics.

Table 6. Summary of Key Attributes across Reviewed Studies (N = 82).

Attribute	Categories	Key Observations
Crop Type	Wheat, Rice, Corn, Multi-crop	Majority focus on staple crops
Region	Asia, Europe, Africa, Global	Strong dominance of Asia-based studies
Scale	Field, Regional, Large-scale	Increasing trend toward multi-scale models
Data Sources	Climate, Soil, IoT, Remote Sensing	Multi-modal data integration is dominant
Model Class	ML, DL, Hybrid, Statistical	ML dominates, hybrid models emerging
Validation Type	Cross-validation, Train/Test split, Field validation	Limited real-world validation
Metrics	RMSE, MAE, R^2, Accuracy	RMSE and R^2 most commonly used

10. Comparative Analysis of Existing Studies

A comparison of the studies reviewed shows that the way that the studies conducted their research falls into several different types of methods or methodology categories that were primarily driven by (a) the scale of the predictions to be made (the size of the area), (b) the availability of data to conduct their analyses, and (c) the different philosophies regarding modeling. Studies conducted at the field and sub-field level reveal higher levels of accuracy when high-resolution soil, management, and proximal sensing data were incorporated to produce models to represent the spatial heterogeneity of the study areas in comparison to attempts to provide models for larger regions and in more general terms.

Machine-learning and deep-learning approaches generally outperform traditional statistical models in terms of predictive accuracy because they capture nonlinear relationships among climate, soil, vegetation, and management variables. However, their performance depends on the quality and representativeness of training data. Deep-learning models are especially useful for image-based and time-series prediction tasks, but they introduce challenges related to computational complexity, explainability, and transferability. Ensemble-learning methods provide a useful compromise because they often improve robustness and reduce overfitting compared with individual models.

Hybrid models that integrate crop-growth models, remote sensing, weather forecasts, and machine learning provide the most balanced solution. These models combine the interpretability of process-based approaches with the flexibility of data-driven methods. As a result, they are better suited for predicting crop yield under climate variability and extreme events. However, hybrid models require careful integration of heterogeneous datasets and stronger benchmarking to determine whether their additional complexity leads to consistent improvements across crops and regions.

Using satellite-based products, climate-based data, and crop growth-based data, mainly focused on producing results that can be scaled across larger geographic areas and over time, and generally have lower levels of precision in their results. Satellite-based methodologies focused on providing data (Vegetation Indices or VI and phenological metrics such as the start and end of the growing season) over large extents and at a relatively low cost requires the use of some methodology for calibration to convert the spectral signals into yield predictions. Machine learning and deep learning methodologies consistently produce better-performing results than traditional statistical modeling [84], and multi-modal approaches produce even better-performing results [104]. In contrast, hybrid physics-based (i.e., incorporating both remote sensing and weather forecast data into the crop growth model) machine learning and deep learning-based approaches produce the most stable and interpretable results when compared to the previous methods due to the use of data from different sources; therefore, those methodologies and approaches provide the most stable results with respect to predicting crop yields in a variable climate [120,121].

11. Challenges and Limitations

The existing literature exposes continuing limitations to a variety of other systemic challenges, not able to complete data with modeling, heterogeneity of data, mixed or untrustworthy source data, historical yield, deep learning models, real-time/edge-cloud system constraints/limitations, and ultimately, the lack of standardized benchmark data and evaluation protocols inhibit replacing and comparing the same models across multiple studies. For example, substantial limitations related to model transferability arise when attempting to use all available models in this area. In addition, significant limitations related to the use of IoT sensor(s) exist due to a variety of issues, including missing or corrupted data being collected from networked sensors [83], a potential problem with cloud contamination in satellite imagery , and a lack of reliability with historical yields. One prominent issue concerning using deep learning models is an inability to produce transparent or understandable characteristics to users because of how these types of models are designed and their respective training processes. For example, when trying to make production decisions using deep learning prediction models, production stakeholders need to have transparently defined yield drivers. Finally, the absence of standardized benchmark datasets and unified evaluation protocols significantly limits reproducibility and cross-study comparability.

12. Future Research Directions

Future research should focus on creating standardized, multi-modal benchmark datasets that would combine soil, climate, remote sensing data, management practices, and yield observations from various agro-ecological zones to allow for adequate comparison and repeatability. Developments in explainable AI (XAI) are critical for increasing the transparency of models and gaining the trust of stakeholders [18]. Further exploration of hybrid modelling frameworks that strongly couple crop growth models with machine learning and real-time data assimilation will improve model robustness to climate extremes and uncertainties [110,121]. From a systems perspective, scalable edge-cloud architectures capable of providing real-time forecasting while minimizing latency and energy costs will need to be researched in greater detail [115]. The field needs to move toward developing transferable and adaptable yield prediction models that are generalizable across crops, climates, and management regimes in order to support both precision management and large-scale food security planning. The challenges associated with research, gaps in research and future trends are shown in **Table 7** of the review.

Table 7. Summary of Challenges, Research Gaps, and Future Solutions in Crop Yield Prediction.

Challenge	Observed Research Gap	Proposed Solution/Future Direction
Limited model transferability	Models trained on local datasets fail to generalize across regions and crops	Develop transferable and domain-adaptive models using diverse, multi-region datasets
Data heterogeneity and quality	Inconsistent sensor data, cloud-affected imagery, incomplete yield records	Multi-modal data fusion, robust preprocessing, and uncertainty-aware modeling
Lack of explainability	Deep learning models act as black boxes for decision-makers	Integrate explainable AI (XAI) methods for transparent yield drivers
Poor real-time integration	Static or end-of-season prediction dominates	Real-time data assimilation and adaptive forecasting pipelines
Scalability vs accuracy trade-off	Field-level accuracy does not scale; large-scale models lack precision	Hierarchical and multi-scale prediction architectures
Absence of benchmarks	Inconsistent evaluation metrics and datasets	Standardized benchmark datasets and unified validation protocols
Edge–cloud operational constraints	Latency, energy use, and synchronization issues	Hybrid edge–cloud architectures with lightweight edge inference

13. Conclusions

A systematic review of the literature indicates that crop yield prediction has become more accurate and sophisticated over time, with improvements in the development of precision agriculture practices, which utilize more data, are based on multi-scale systems, and apply adaptive methods. This has resulted in significant increases in research into crop yield prediction, largely as a result of improvements in the fields of machine learning, Internet of Things (IoT) sensing technologies, and remote sensing technology (satellite). Currently, machine learning-based models are the most common approach used; however, their success is linked to not only the quality of the data available to the model but also the spatial scale at which the model is applied and contextual information regarding the soil, climate and agricultural practices related to the crop on which predictions are based. The review also notes that the best approach (with respect to accuracy, robustness, and interpretability), particularly for applications subject to climate variability, is to use hybrid methods that incorporate crop growth models in combination with artificial intelligence and real-time data assimilation. Unfortunately, due to the predominance of single-method studies, lack of standardized benchmarks, and insufficient focus on the in-field implementation of predictive modeling techniques, the ability for researchers to impact real-world applications is limited. Overall, this review indicates that successful yield predictions require a systems-oriented approach to address the complexities related to the integration of data sources, prediction architectures, and decision workflows across all levels of the agricultural system.

Many possible directions for future research can be derived from both qualitative and quantitative syntheses of data trends. The priority is to develop standardized multimodal benchmark datasets that measure soil quality, climate conditions, satellite imagery, agricultural management strategies at the field level, as well as crop yield across multiple agroecological zones. By developing such datasets, researchers can compare results fairly, reproducibly, and with less delay based on available datasets. In addition, future research should focus on developing hybrid/transferable modelling approaches that integrate crop simulation models and machine learning with real-time data assimilation for more broadly applicable modelling of crop yield. Third, the importance of using Explainable AI (XAI) must also be stressed in future research due to the rapid proliferation of black-box AI models, which could limit the extent to which both farmers and policymakers can trust and use AI for their planting decisions. Fourth, more attention needs to be placed on edge-cloud and real-time prediction computing models to provide timely updates of crop yield, while recognizing the limitations of geographic coverage (latency), energy consumption, and access to telecommunications. Finally, the significant lack of literature examining optimization of agricultural management strategies because of crop yield predictions underscores the need for better integration between prediction modelling systems, optimization frameworks for agricultural management practices, and agricultural policy making.

Funding

This research received no external funding.

Institutional Review Board Statement

This study does not involve human or animal subjects. All data were obtained from publicly available sources and used in accordance with ethical research standards.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting this study are derived from publicly available research articles retrieved from major digital libraries (e.g., IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, MDPI). No new datasets were generated. All referenced data are available within the cited literature, and any required data is available upon request.

Acknowledgments

The authors would like to thank the reviewers and editors for their valuable feedback, which significantly improved the quality of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

AI Use Statement

During the preparation of this manuscript, AI tools were used solely for language refinement and proofreading. No AI tools were used for data analysis, interpretation, or generation of scientific content. All outputs were critically reviewed and edited by the authors, who take full responsibility for the integrity and accuracy of the work.

References

1. Sachs, J. *Sustainable Development Report 2022*; Cambridge University Press: Cambridge, UK, 2022.
2. Kuchechuk, L. Global food security: Trends and challenges. *Visn. V. N. Karazin Kharkiv Natl. Univ. Ser. Int. Relat. Econ. Ctry. Stud. Tour.* **2022**, *16*, 34–40.
3. Ben Hassen, T.; El Bilali, H. Impacts of the Russia-Ukraine war on global food security: Towards more sustainable and resilient food systems? *Foods* **2022**, *11*, 2301.
4. Sharma, M.; Joshi, S.; Gupta, P.; et al. Climate, crops, and communities: Modeling the environmental stressors driving food supply chain insecurity. *Earth* **2025**, *6*, 121.
5. Al-Dobashi, H.; Wright, S. Stress-Testing Food Security in a Socio-Ecological System: Qatar's Adaptive Responses to Sequential Shocks. *Systems* **2025**, *14*, 46.
6. Brinkley, C.; Ulimwengu, J.; Raj, S. Real and simulated shocks to the food system: A nested food-economic system model for assessing resilience using a networked input–output approach. *Front. Sustain. Food Syst.* **2025**, *9*, 1608981.
7. Olufemi-Phillips, A.Q.; Ofodile, O.C.; Toromade, A.S.; et al. Strategies for adapting food supply chains to climate change using simulation models. *Int. J. Eng. Res. Dev.* **2024**, *20*, 1021–1040.
8. Singh, K.; Yadav, M.; Barak, D.; et al. Machine-learning-based frameworks for reliable and sustainable crop forecasting. *Sustainability* **2025**, *17*, 4711.
9. Li, X.; Peng, B.; Zhang, R.; et al. Research on a Real-Time Monitoring and Early Warning System for Abnormal Fluctuations in Agricultural Product Prices. *Front. Agric.* **2025**, *2*, 57–68.
10. Ayoub, A. The Architecture of Intelligent Governance (AIG): A Conceptual Framework for Integration AI, Quantum Computing, and Global Resource Resilience. *Sustainability* **2026**, *18*, 2312.
11. Beier, F.D.; Dietrich, J.P.; Heinke, J.; et al. Planetary boundaries under a land-based climate change mitigation scenario with a food demand transformation: A modelling study. *Lancet Planet. Health* **2025**, *9*, 101249.
12. Thingujam, D.; Gouli, S.; Cooray, S.P.; et al. Climate-resilient crops: Integrating AI, multi-omics, and advanced phenotyping to address global agricultural and societal challenges. *Plants* **2025**, *14*, 2699.
13. Raza, A.; Safdar, M.; Shahid, M.A.; et al. Climate change impacts on crop productivity and food security: An overview. In *Transforming Agricultural Management for a Sustainable Future*; Springer: Cham, Switzerland,

2024; pp. 163–186.
14. Nirmal, N.; Anyimadu, C.F.; Khanashyam, A.C.; et al. Alternative protein sources: Addressing global food security and environmental sustainability. *Sustain. Dev.* **2025**, *33*, 3958–3969.
15. Rezaei, E.E.; Webber, H.; Asseng, S.; et al. Climate change impacts on crop yields. *Nat. Rev. Earth Environ.* **2023**, *4*, 831–846.
16. Mehmood, M.; Tanveer, N.A.; Joyia, F.A.; et al. Effect of high temperature on pollen grains and yield in economically important crops: A review. *Planta* **2025**, *261*, 141.
17. Garajeh, M.K.; Kamran, K.V.; Feizizadeh, B.; et al. Assessing the Impacts of Long-Term Weather Variability and Urban Development on Crop Production to Analyze Food Security in Iran. *Adv. Space Res.* **2026**, in press.
18. Mohan, R.; Rayanoothala, P.S.; Sree, R.P. Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions. *Front. Plant Sci.* **2025**, *15*, 1451607.
19. Newell, L.D. Using Machine Learning to Investigate Weather Impacts on Crop Yield for Probabilistic, Climate-Informed Predictive Crop Breeding. PhD Thesis, The University of North Dakota, Grand Forks, ND, USA, 2025.
20. Li, X.; Li, Y.; Sun, Y.; et al. Integrating genetic diversity and agronomic innovations for climate-resilient maize systems. *Plants* **2025**, *14*, 1552.
21. Shafei, A.; Cioffi, F. Designing an Early-Warning System to Forecast Extreme Climate Conditions Using Machine-Learning and Deep-Learning Methods. In Proceedings of the EGU General Assembly 2024, Vienna, Austria, 14–19 April 2024.
22. FAO report: 1.7 billion people experience lower crop yields due to land degradation. Available online: https://www.fao.org/newsroom/detail/fao-report--1.7-billion-people-experience-lower-crop-yields-due-to-land-degradation/en (accessed on 26 May 2025).
23. Mandal, D.; Roy, T. Climate change impact on soil erosion and land degradation. In *Climate Change Impacts on Soil-Plant-Atmosphere Continuum*; Springer: Singapore, 2024; pp. 139–161.
24. Food and Agriculture Organization of the United Nations. *The State of the World's Land and Water Resources for Food and Agriculture 2021: Systems at Breaking Point*; FAO: Rome, Italy, 2022.
25. Jonkman, S.; Curran, A.; Bouwer, L.M. Floods have become less deadly: An analysis of global flood fatalities 1975–2022. *Nat. Hazards* **2024**, *120*, 6327–6342.
26. Jabed, M.A.; Murad, M.A.A. Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon* **2024**, *10*, e40836.
27. Hernández Hernández, G.C.; Gómez Gómez, J.; Jiménez-Cabas, J.J. Predictive Models Based on Artificial Intelligence to Estimate Crop Yield: A Literature Review. *Agriculture* **2025**, *15*, 2438.
28. Sudha, S.; Loret, J. A review on machine learning-based precision agriculture techniques for crop farming monitoring with IOT. *Data Eng.* **2026**, *4*, 10.
29. Lanza, B. Development of Optical Measurement Methodologies for In-Field Estimation of Plant Health and Agricultural Yield. PhD Thesis, University of Brescia, Brescia, Italy, 2025.
30. Shen, Y. Winter Wheat Phenotyping Based on Unmanned Aerial Vehicle Remote Sensing and Machine Learning—From Yield Prediction to Growth Diagnosis. PhD Thesis, University of Liège, Liège, Belgium, 2026.
31. Torres-Quezada, E.; Fuentes-Peñailillo, F.; Gutter, K.; et al. Remote sensing and soil moisture sensors for irrigation management in avocado orchards: A practical approach for water stress assessment in remote agricultural areas. *Remote Sens.* **2025**, *17*, 708.
32. Siyabonga, G. Exploring the Integration of Satellite and UAV Remotely Sensed Data for Enhanced Crop Monitoring in Small-Scale Farming. PhD Thesis, University of KwaZulu-Natal, Pietermaritzburg, South Africa, 2025.
33. Kate, A. Multimodal Fusion of Hyperspectral and RGB Imaging for Early Detection of Nutrient Deficiency and Disease Stress in Tea Plantations. PhD Thesis, Universidade de São Paulo, São Paulo, Brazil, 2025.
34. Li, Y.; Xiao, X. Deep learning-based fusion of optical, radar, and LiDAR data for advancing land monitoring. *Sensors* **2025**, *25*, 4991.
35. Anconitano, G. Soil Moisture Estimation and Crop Area Classification from Multi-Frequency and Multi-Polarization SAR Data. PhD Thesis, Sapienza University of Rome, Rome, Italy, 2025.
36. Saha, S.; Kucher, O.D.; Utkina, A.O.; et al. Precision agriculture for improving crop yield predictions: A literature review. *Front. Agric.* **2025**, *7*, 1566201.
37. Adewumi, I.O.; Adejumo, A.O.D.; Adegbokan, O.; et al. Edge AI–IoT Integration for Real-Time Precision Farm-

ing: A Framework for Adaptive Monitoring, Prediction, and Resource Optimization. *Lond. J. Phys.* **2026**, *3*.
38. Wassay, M.; Khalid, B.; Ashraf, M.A.; et al. Geo-Intelligent Agriculture: Integrating GIS, Remote Sensing, and IoT for Real-Time Soil and Crop Health Monitoring and Predictive Farm Management. *Agric. Res. Rep.* **2025**, *3*, 58–69.
39. Barrile, V.; Maesano, C.; Genovese, E. Optimization of crop yield in precision agriculture using WSNs, remote sensing, and atmospheric simulation models for real-time environmental monitoring. *J. Sens. Actuator Netw.* **2025**, *14*, 14.
40. Ou, G.; Chen, Y.; Han, Y.; et al. Design and experiment of an internet of things-based wireless system for farmland soil information monitoring. *Agriculture* **2025**, *15*, 467.
41. Saha, G.; Shahrin, F.; Khan, F.H.; et al. Smart IoT-driven precision agriculture: Land mapping, crop prediction, and irrigation system. *PLoS One* **2025**, *20*, e0319268.
42. Abedalrhman, K.; Alzaydi, A.J.A. Precision Agriculture 4.0: Integrating advanced IoT, AI, and robotics solutions for enhanced yield, sustainability, and resource optimization-evidence from agricultural practices in Syria. *Appl. Sci. Biotechnol. J. Adv. Res.* **2025**, *4*, 7–27.
43. Usoro, S.O.; Omogiate, I.A.; Felix, R.Z.; et al. AI-Driven Risk Forecasting for Strengthening the United States Food Supply Chain Resilience: Case Study: A National AI-Enabled Food Supply Chain Risk Forecasting Framework and a Data Analytics Approach to Predicting Disruptions. *Int. J. Sci. Res. Arch.* **2026**, *18*, 744.
44. Coluzzi, R.; Lanfredi, M.; Imbrenda, V. Towards a risk-informed land system approach in the age of artificial intelligence and analysis-ready satellite data. *Geomat. Nat. Hazards Risk* **2026**, *17*, 2599487.
45. Screpnik, C.; Zamudio, E.; Gimenez, L.J.I.A. Artificial Intelligence in Agriculture: A Systematic Review of Crop Yield Prediction and Optimization. *IEEE Access* **2025**, *13*, 70691–70697.
46. Bali, N.; Singla, A. Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. *Arch. Comput. Methods Eng.* **2022**, *29*, 95–112.
47. Rivero, R.M.; Mittler, R.; Blumwald, E.; et al. Developing climate-resilient crops: Improving plant tolerance to stress combination. *Plant J.* **2022**, *109*, 373–389.
48. Khan, S.; Irshad, S.; Mehmood, K.; et al. Biochar production and characteristics, its impacts on soil health, crop production, and yield enhancement: A review. *Plants* **2024**, *13*, 166.
49. Rashid, A.; Schutte, B.J.; Ulery, A.; et al. Heavy metal contamination in agricultural soil: Environmental pollutants affecting crop health. *Agronomy* **2023**, *13*, 1521.
50. Bacelar, E.; Pinto, T.; Anjos, R.; et al. Impacts of climate change and mitigation strategies for some abiotic and biotic constraints influencing fruit growth and quality. *Plants* **2024**, *13*, 1942.
51. Yuan, X.; Li, S.; Chen, J.; et al. Impacts of global climate change on agricultural production: A comprehensive review. *Agronomy* **2024**, *14*, 1360.
52. Chandio, A.A.; Jiang, Y.; Amin, A.; et al. Modeling the impact of climatic and non-climatic factors on cereal production: Evidence from Indian agricultural sector. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 14634–14653.
53. Gul, A.; Chandio, A.A.; Siyal, S.A.; et al. How climate change is impacting the major yield crops of Pakistan? An exploration from long- and short-run estimation. *Environ. Sci. Pollut. Res. Int.* **2022**, *29*, 26660–26674.
54. Chen, Z.; Galli, M.; Gallavotti, A. Mechanisms of temperature-regulated growth and thermotolerance in crop species. *Curr. Opin. Plant Biol.* **2022**, *65*, 102134.
55. Dubey, V.; Panigrahi, S.; Vidyarthi, V.K.; et al. Statistical trend analysis of major climatic factors over Chhattisgarh State, India. *Environ. Sustain.* **2023**, *7*, 629–648.
56. Golkarian, A.; Khosravi, K.; Panahi, M.; et al. Spatial variability of soil water erosion: Comparing empirical and intelligent techniques. *J. Geogr. Futures* **2023**, *14*, 101456.
57. Xu, Y.; Liao, J.; Liu, Q.; et al. The interaction between vegetation greenness and hydro-climatic factors in China. *J. Hydrol.* **2026**, *668*, 134956.
58. Kakavand, M.; Haghizadeh, A.; Soleimani-Motlagh, M.J. Spatio-temporal analysis of trend the hydro-climatic time series and identify factors affecting Flood Hazard. *Environ. Syst. Indic.* **2025**, *18*, 177.
59. Alehu, B.A.; Desta, H.B.; Daba, B.I. Assessment of climate change impact on hydro-climatic variables and its trends over Gidabo Watershed. *Model. Earth Syst. Environ.* **2022**, *8*, 3769–3791.
60. Sigalla, O.Z.; Valimba, P.; Selemani, J.R.; et al. Analysis of spatial and temporal trend of hydro-climatic parameters in the Kilombero River Catchment, Tanzania. *Sci. Rep.* **2023**, *13*, 7864.
61. Batungwanayo, P.; Vanclooster, M.; Alonso, A.; et al. Wavelet-based analysis of hydro-climatic and vegetation dynamics in heterogeneous agro-climatic zones of East Africa. *J. Water Clim. Change* **2024**, *15*, 4054–4075.
62. Mojid, M.A.; Mousumi, K.A.; Kamal, M.R. Temporal-Scale Climate Information and Crop Production Management Decisions—A Case Study in the North-central Region of Bangladesh. *Water Conserv. Sci. Eng.* **2024**, *9*,

19.
63. Dorjsuren, B.; Zemtsov, V.A.; Batsaikhan, N.; et al. Hydro-climatic and vegetation dynamics spatial-temporal changes in the great lakes depression region of Mongolia. *Water* **2023**, *15*, 3748.
64. Delgado, A.; Gómez, J.A. The soil: Physical, chemical, and biological properties. In *Principles of Agronomy for Sustainable Agriculture*; Springer: Cham, Switzerland, 2024; pp. 15–30.
65. De Caires, S.A.; Martin, C.S.; Atwell, M.A.; et al. Advancing soil mapping and management using geostatistics and integrated machine learning and remote sensing techniques: A synoptic review. *Discov. Soil* **2025**, *2*, 53.
66. Senapaty, M.K.; Ray, A.; Padhy, N. IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. *Computers* **2023**, *12*, 61.
67. Ahmed, U.; Lin, J.C.-W.; Srivastava, G.; et al. A nutrient recommendation system for soil fertilization based on evolutionary computation. *Comput. Electron. Agric.* **2021**, *189*, 106407.
68. Shaheb, M.R.; Sarker, A; Shearer, S.A. Precision Agriculture for Sustainable Soil and Crop Management. In *Soil Science—Emerging Technologies, Global Perspectives and Applications*; IntechOpen: London, UK, 2022.
69. Belal, A.A.; El-Ramady, H.; Jalhoum, M.; et al. Precision farming technologies to increase soil and crop productivity. In *Agro-Environmental Sustainability in MENA Regions*; Springer: Cham, Switzerland, 2021; pp. 117–154.
70. Lawrence, P.G.; Roper, W.; Morris, T.F.; et al. Guiding soil sampling strategies using classical and spatial statistics: A review. *Agron. J.* **2020**, *112*, 493–510.
71. Garc, A.; Frankenstein, S. *USCS and the USDA Soil Classification System: Development of a Mapping Scheme*; US Army Corps of Engineers: Washington, DC, USA, 2015.
72. Canton, H. Food and Agriculture Organization of the United Nations—FAO. In *The Europa Directory of International Organizations 2021*; Routledge: London, UK, 2021; pp. 297–305.
73. Mantel, S.; Dondeyne, S.; Deckers, S. World reference base for soil resources (WRB). *Encycl. Soils Environ.* **2023**, *4*, 206–217.
74. Radočaj, D.; Šiljeg, A.; Marinović, R.; et al. State of major vegetation indices in precision agriculture studies indexed in web of science: A review. *Agriculture* **2023**, *13*, 707.
75. Parida, P.K.; Somasundaram, E.; Krishnan, R.; et al. Unmanned Aerial Vehicle-Measured Multispectral Vegetation Indices for Predicting LAI, SPAD Chlorophyll, and Yield of Maize. *Agronomy* **2024**, *14*, 1110.
76. Dolas, V.M.; Joshi, P. A novel approach for classification of soil and crop prediction. *Int. J. Comput. Sci. Mob. Comput.* **2018**, *7*, 20–24.
77. Suruliandi, A.; Mariammal, G.; Raja, S.; et al. Crop prediction based on soil and environmental characteristics using feature selection techniques. *Mach. Learn. Data Sci.* **2021**, *27*, 117–140.
78. Paul, M.; Vishwakarma, S.K.; Verma, A. Analysis of soil behaviour and prediction of crop yield using data mining approach. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; pp. 766–771.
79. Khanal, S.; Fulton, J.; Klopfenstein, A.; et al. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225.
80. Raj, N.; Singh, A.; Verma, J. Data-Driven Precision Farming for Optimized Extension Impact. In *New Frontiers in Agricultural Extension Strategies*; Textify Publishers: New Delhi, India, 2025; pp. 131–155.
81. Sharma, K.; Shivandu, S.K. Integrating artificial intelligence and Internet of Things (IoT) for enhanced crop monitoring and management in precision agriculture. *Smart Innov.* **2024**, *5*, 100292.
82. Getahun, S.; Kefale, H.; Gelaye, Y. Application of precision agriculture technologies for sustainable crop production and environmental sustainability: A systematic review. *Sci. World J.* **2024**, *2024*, 2126734.
83. Talaat, F.M. Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. *Neural Comput. Appl.* **2023**, *35*, 17281–17292.
84. Burdett, H.; Wellen, C. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precis. Agric.* **2022**, *23*, 1553–1574.
85. Parashar, V.; Labhade-Kumar, N.; Rajkumar, B.P.; et al. Enhancing crop yield prediction in precision agriculture through sustainable big data analytics and deep learning techniques. *Carpath. J. Food Sci. Technol.* **2023**, *15*, 1–18.
86. Bakthavatchalam, K.; Karthik, B.; Thiruvengadam, V.; et al. IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. *Int. J. Syst. Assur. Eng. Manag.* **2022**, *10*, 13.

87. Akintuyi, O.B. Adaptive AI in precision agriculture: A review: Investigating the use of self-learning algorithms in optimizing farm operations based on real-time data. *Res. J. Multidiscip. Stud.* **2024**, *7*, 16–30.
88. Dusadeerungsikul, P.O.; Nof, S.Y. Precision agriculture with AI-based responsive monitoring algorithm. *Int. J. Prod. Econ.* **2024**, *271*, 109204.
89. Peerlinck, A.; Sheppard, J.; Pastorino, J.; et al. Optimal design of experiments for precision agriculture using a genetic algorithm. In Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 10–13 June 2019; pp. 1838–1845.
90. Gul, D.; Banday, R.U.Z. Transforming crop management through advanced AI and machine learning: Insights into innovative strategies for sustainable agriculture. *Adv. Comput. Sci. Technol.* **2024**.
91. Shekhar, K.; Sravanthi, G.; Sadhashiva, P.; et al. Machine Learning-Based Crop Recommendation System for Enhanced Yield Production. *J. Comput. Appl.* **2025**, *34*, 428–439.
92. Baligar, P. Revolutionizing Agriculture with an Intelligent Crop Recommendation System. *Vidhyayana* **2025**, *10*.
93. Bhola, A.; Kumar, P. Farm-level smart crop recommendation framework using machine learning. *Ann. Data Sci.* **2025**, *12*, 117–140.
94. Mancer, M.H.; Terrissa, L.S.; Ayad, S.; et al. Advancing Crop Recommendation Systems through Ensemble Learning Techniques. In *Innovations in Smart Cities Applications*; Springer: Cham, Switzerland, 2024; pp. 45–54.
95. Mimoune, A.; Lantri, F. Precision Agriculture Using Crop Recommendation Systems. Master's Thesis, Ibn Khaldoun University-Tiaret, Tiaret, Algeria, 2024.
96. Choudhary, M.; Sartandel, R.; Arun, A.; et al. Crop recommendation system and plant disease classification using machine learning for precision agriculture. In *Artificial Intelligence and Communication Technologies*; Soft Computing Research Society (SCRS): Maharashtra, India, 2022; pp. 39–49.
97. Ali, S.M.; Das, B.; Kumar, D. Machine learning based crop recommendation system for local farmers of Pakistan. *Rev. Geintec Gest. Inov. Tecnol.* **2021**, *11*, 5735–5746.
98. Kathiria, P.; Patel, U.; Madhwani, S.; et al. Smart crop recommendation system: A machine learning approach for precision agriculture. In *Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022, Volume 1*; Springer: Singapore, 2023; pp. 841–850.
99. Chauhan, G.; Chaudhary, A. Crop recommendation system using machine learning algorithms. In Proceedings of the 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 10–11 December 2021; pp. 109–112.
100. Priyadharshini, A.; Chakraborty, S.; Kumar, A.; et al. Intelligent crop recommendation system using machine learning. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 843–848.
101. Dhaliwal, D.S.; Williams, M.M. Sweet corn yield prediction using machine learning models and field-level data. *Precis. Agric.* **2024**, *25*, 51–64.
102. Pathak, D.; Miranda, M.; Mena, F.; et al. Predicting crop yield with machine learning: An extensive analysis of input modalities and models on a field and sub-field level. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 2767–2770.
103. Helber, P.; Bischke, B.; Habelitz, P.; et al. Crop yield prediction: An operational approach to crop yield modeling on field and subfield level with machine learning models. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 2763–2766.
104. Mena, F.; Pathak, D.; Najjar, H.; et al. Adaptive fusion of multi-view remote sensing data for optimal sub-field crop yield prediction. *Remote Sens. Environ.* **2024**, *318*, 114547.
105. Filippi, P.; Jones, E.J.; Wimalathunge, N.S.; et al. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precis. Agric.* **2019**, *20*, 1015–1029.
106. Kang, X.; Huang, C.; Zhang, L.; et al. Regional-scale cotton yield forecast via data-driven spatio-temporal prediction (STP) of solar-induced chlorophyll fluorescence (SIF). *Remote Sens. Environ.* **2023**, *299*, 113861.
107. Wang, H.; Dai, Y.; Yao, Q.; et al. Multi-task learning model driven by climate and remote sensing data collaboration for mid-season cotton yield prediction. *Field Crops Res.* **2025**, *333*, 110070.
108. Kirthiga, S.; Patel, N. In-season wheat yield forecasting at high resolution using regional climate model and crop model. *Agric. Food Secur.* **2022**, *4*, 1054–1075.

109. Zhuo, W.; Fang, S.; Gao, X.; et al. Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013. *Agric. Syst.* **2022**, *106*, 102668.
110. Bandaru, V.; Yaramasu, R.; Jones, C.; et al. Geo-CropSim: A Geo-spatial crop simulation modeling framework for regional scale crop yield and water use assessment. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 34–53.
111. Ji, Z.; Pan, Y.; Zhu, X.; et al. A generalized model to predict large-scale crop yields integrating satellite-based vegetation index time series and phenology metrics. *Ecol. Indic.* **2022**, *137*, 108759.
112. Li, Z.; Ding, L.; Xu, D. Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Sci. Total Environ.* **2022**, *815*, 152880.
113. Tripathy, R.; Chaudhari, K.; Bairagi, G.; et al. Towards fine-scale yield prediction of three major crops of India using data from multiple satellite. *J. Indian Soc. Remote Sens.* **2022**, *50*, 271–284.
114. Ishaq, R.A.F.; Zhou, G.; Jing, G.; et al. Geospatial Robust Wheat Yield Prediction Using Machine Learning and Integrated Crop Growth Model and Time-Series Satellite Data. *Remote Sens.* **2025**, *17*, 1140.
115. Dey, T.; Bera, S.; Mukherjee, A.; et al. QRCrop: A QR Code-Based Edge–Cloud Framework for Crop Yield Prediction. In *Proceedings of International Conference on Network Security and Blockchain Technology*; Springer: Cham, Switzerland, 2024; pp. 353–364.
116. Dey, T.; Bera, S.; Paul, B.; et al. Fly: Femtolet-based edge-cloud framework for crop yield prediction using bidirectional long short-term memory. *Softw. Pract. Exp.* **2024**, *54*, 1361–1377.
117. Sharma, A.K.; Rathore, A.S. AGRO-Cloud Model and Smart Algorithm to Increase Crop Yield Prediction to Improve Agriculture Quality. *Curr. Agric. Res. J.* **2024**, 371–388.
118. Bera, S.; Dey, T.; Mukherjee, A.; et al. E-CropReco: A dew-edge-based multi-parametric crop recommendation framework for internet of agricultural things. *J. Supercomput.* **2023**, *79*, 11965–11999.
119. Dhakar, R.; Sehgal, V.K.; Chakraborty, D.; et al. Field scale spatial wheat yield forecasting system under limited field data availability by integrating crop simulation model with weather forecast and satellite remote sensing. *Agric. Syst.* **2022**, *195*, 103299.
120. Zare, H.; Weber, T.K.; Ingwersen, J.; et al. Combining crop modeling with remote sensing data using a particle filtering technique to produce real-time forecasts of winter wheat yields under uncertain boundary conditions. *Remote Sens.* **2022**, *14*, 1360.
121. Chen, Y.; Tao, F. Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. *Field Crops Res.* **2022**, *276*, 108398.
122. Jain, M.; Bajwa, M.S.; Kumar, H. Agriculture assistant for crop prediction and farming selection using machine learning model with real-time data using imaging through UAV drone. In *Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021*; Springer: Singapore, 2022; pp. 311–330.
123. Tseng, G.; Zvonkov, I.; Nakalembe, C.L.; et al. Cropharvest: A global dataset for crop-type classification. In Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), Online, 6–14 December 2021.
124. Becker-Reshef, I.; Bandaru, V.; Barker, B.; et al. The NASA harvest program on agriculture and food security. In *Remote Sensing of Agriculture and Land Cover/Land Use Changes in South and Southeast Asian Countries*; Springer: Cham, Switzerland, 2022; pp. 53–80.