A Review of Machine Learning Techniques for Precision Agriculture and Crop Yield Prediction

Muhammad Faraz Manzoor

OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Department of Computer Science, University of Management and Technology, Lahore, Pakistan

Correspondence to: Muhammad Faraz Manzoor, Faraz.manzoor@umt.edu.pk

Additional material is published online only. To view please visit the journal online.

Cite this as: Manzoor MF. A Review of Machine Learning Techniques for Precision Agriculture and Crop Yield Prediction. Premier Journal of Plant Biology 2024;1:100005

DOI: https://doi.org/10.70389/PJPB.100005

Received: 19 October 2024
Revised: 4 December 2024
Accepted: 6 December 2024
Published: 19 December 2024

Ethical approval: N/a
Consent: N/a
Funding: No industry funding
Conflicts of interest: N/a
Author contribution: Muhammad Faraz Manzoor – Conceptualization, Writing – original draft, review and editing
Guarantor: Muhammad Faraz Manzoor
Provenance and peer-review: Commissioned and externally peer-reviewed
Data availability statement: N/a

ABSTRACT

Agriculture plays a critical role in the economic development of nations. However, with a growing global population, shifting climate conditions, and limited natural resources, meeting the increasing food demand is becoming a complex challenge. Precision agriculture, also referred to as smart farming, has emerged as a powerful solution to address these issues by promoting sustainable agricultural practices. At the heart of this innovation is machine learning (ML), ensemble learning, and deep learning, which enable systems to learn and adapt without explicit programming. Together with IoT-enabled farm equipment, ML is driving the next agricultural revolution. In this systematic literature review (SLR), we examine research from 2015 to 2023 on the diverse applications of ML in agriculture. A key area explored in this SLR is disease detection, focusing on identifying diseases. Yield prediction is another prominent area, where ML is used to forecast crop yields based on diverse environmental and agricultural data. The review also examines the prediction of soil properties, such as organic carbon content and moisture levels, which are crucial for optimizing crop management. Lastly, weather prediction is discussed, highlighting its importance in planning and adjusting farming activities in response to climatic conditions. Additionally, challenges such as data availability, model scalability, and the need for regional validation are explored.

Keywords: Precision agriculture, Machine learning in agriculture, Crop yield prediction, Soil property prediction, Disease detection in crops

Introduction

Precision agriculture is a modern farming management concept that leverages advanced technologies to enhance crop production efficiency and sustainability.[1] By utilizing data-driven approaches, farmers can make informed decisions regarding resource allocation, crop management, and environmental conservation. Among the most significant advancements in this field is the application of machine learning (ML), which has emerged as a powerful tool for analyzing complex agricultural data and extracting meaningful insights.[2] The use of ML makes it easier to find patterns and relationships among the big data to build a predictive model to help in different agricultural tasks.

Another key use of ML that has seen significant growth is in the assessment and management of soil profile properties.[3] Understanding the availability of water in the soil, the amount of organic material, and the nutrient regime in the soil is critical for good crop production and soil usage. The literature shows that remote sensing and spectral data-based ML techniques may accurately predict soil parameters.[4,5] These arise to improve effective water and fertilizer management by having an appropriate irrigation schedule, the proper time to apply soil amendments, and, most significantly, the right type of land management that promotes ecological sustainability.

ML is also useful when it comes to yield prediction, which is among the key elements in precision agriculture. Crop yield forecasting is highly relevant, especially in the areas of food security, logistics, and planning.[6] Crop yield prediction models may therefore include all kinds of enhancements, such as climate conditions, type of soil, and previous year production rates. These models are crucial in ensuring that farmers are provided with significant information that can be used to predict output levels; they affect the mode of planting and resource use, thereby boosting agricultural yield.

ML is also reporting immense benefits in the area of the prediction of unfavorable diseases. Production and quality of produce are also affected by pests and diseases resulting in the need to develop early vigilance or diagnostic tools.[7] ML also can analyze photos taken from drones or phones as well as temperature, humidity, and altitude sensors and previous episodes of the illness to find correlations with the breakout. These models predict the likelihood of developing diseases and inform farmers early enough to avoid losses, improving the yields.

To achieve the goal of this systematic literature review, the study will describe some of the uses of ML in precision agriculture especially on soil, yield, and disease. In this article, we will review the current state of the art in many domains, as well as the primary findings, research techniques, and challenges associated with each. Therefore, this literature review seeks to discuss how ML is transforming precision agriculture and to examine emerging techniques that can guide future research in this rapidly growing area.

The rest of the article is structured as follows: The Research Methodology section outlines the approach used for selecting and reviewing primary studies, detailing the criteria and process for data collection. The Analysis of Primary Studies presents a comprehensive review of the research studies, focusing on key applications of ML in agriculture, such as disease detection, yield prediction, soil property prediction, and weather forecasting. In the Discussion section, the implications of the findings are explored, highlighting challenges and opportunities in the field. Finally, the Conclusion and Future Directions offers a summary of the findings and suggests potential avenues for future research and development in precision agriculture.

Research Methodology

To retain the core objective of our work, which is to examine research on ML applications in precision agriculture, we gathered insights and guidance from current approaches outlined in numerous papers[8,9] as depicted in Figure 1. Using this information, we developed clear study objectives, suitable questions for research, and search methodologies to use. This technique enables us to successfully search through and recognize relevant literature on the scale of ML applications in precision agriculture.

Research Objectives

The following are the research objectives of this study:

- To analyze the effectiveness of various ML algorithms in predicting soil properties and their impact on precision agriculture.
- To evaluate the role of ML in enhancing crop yield prediction through data-driven insights and modeling techniques.
- To investigate the applications of ML in early disease detection and management in agricultural practices.

Research Questions

To achieve the objectives of the research, we have presented a set of appropriate research questions addressing particular aspects of ML application in precision agriculture. Table 1 presents these research questions, along with their underlying motivations.

Search Strategy

The following search string is used to find relevant articles to conduct this study.

("machine learning" OR "ML" OR "artificial intelligence" OR "AI") AND ("precision agriculture" OR "smart agriculture" OR "agricultural management" OR "precision farming") AND ("soil properties" OR "soil quality" OR "soil moisture" OR "soil nutrients") AND ("crop yield" OR "crop productivity" OR "yield prediction") AND ("disease detection" OR "disease prediction" OR "crop diseases" OR "plant health")

Articles from a variety of sources, such as Springer, Elsevier, Springer, IEEE, and other respectable publications and conferences, are gathered in the quest for firsthand information on the subject of precision agriculture.

Study Selection

In a systematic literature review process, the study selection is an important step.[10] The review consists of checking the titles and abstracts of the articles retrieved by the search strategy to select the studies that pass the criteria for inclusion and exclusion. This step aims to lessen the number of articles to an achievable volume and, at the same time, maintain those which will best furnish valuable information as presented in Figure 2.

Preliminary research articles on the use of ML in precision agriculture numbering 3480 were gathered for this investigation from different sources. Two writers participated in the selection process, shortlisting

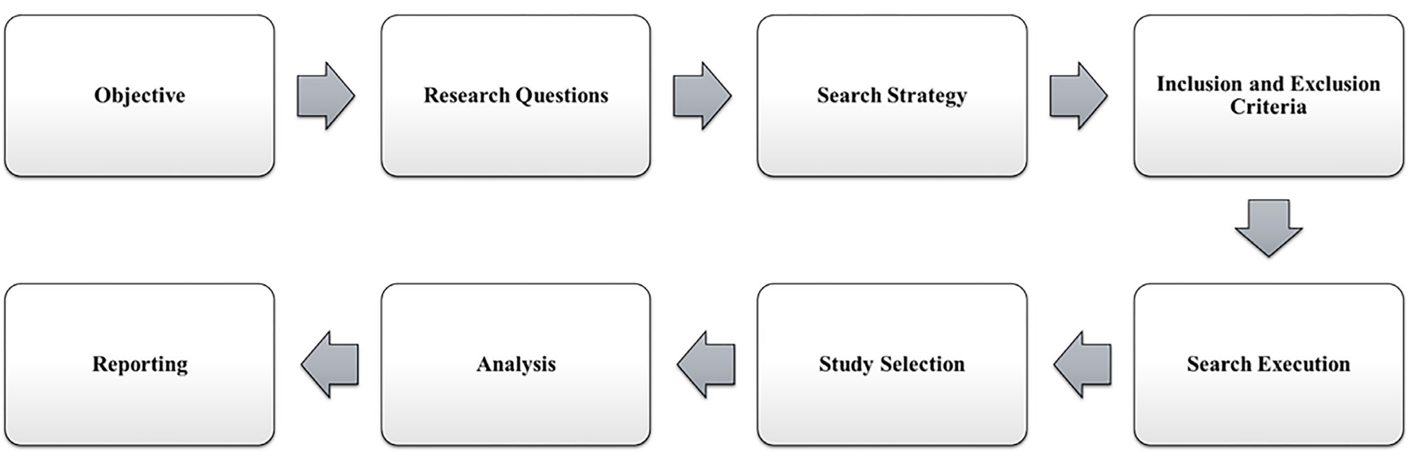

Fig 1 | Research methodology

Table 1 | Research questions and motivations

SR#	Research Question	Motivation
RQ1	What machine learning algorithms demonstrate the highest accuracy in predicting soil properties?	Identifying effective algorithms will help optimize soil management practices, leading to better crop health and yield.
RQ2	How can machine learning models improve the accuracy of crop yield predictions compared to traditional methods?	Improved yield predictions enable farmers to make informed decisions, enhancing productivity and resource management.
RQ3	In what ways can machine learning facilitate early detection and management of diseases in crops?	Early disease detection can significantly reduce crop loss, ensuring sustainable agricultural practices and food security.

 DOI: https://doi.org/10.70389/PJPB.100005 | Premier Journal of Plant Biology 2024;1:100005

the articles according to predetermined inclusion and exclusion criteria. A third author assisted in resolving any issues that arose during this process, and any required adjustments were made to the inclusion/exclusion criteria. With a Cohen's kappa coefficient of 0.86, the inter-rater agreement between the two writers was determined to be almost flawless.[35]

- **Title-based search:** Carefully weeding out papers that are irrelevant based solely on their title is the first stage in the process. At this stage, the number of irrelevant papers was high. There were just 253 papers left after this phase.
- **Abstract-based search:** In this phase, the articles are arranged for analysis and research methods after the abstracts of the articles selected in the previous stage are reviewed. There were just 105 papers remaining after this.
- **Full text-based analysis:** At this point, the articles selected in the previous step are evaluated for their empirical quality. The study's text has been thoroughly analyzed. 43 papers were chosen in total from 105 publications. Table 2 shows the total number of papers for each of the participating portals at various phases of the selection process.

In order to find the answers to the questions outlined above, a total of 43 publications were found and examined. Table 3 displays the source-wise study distribution.

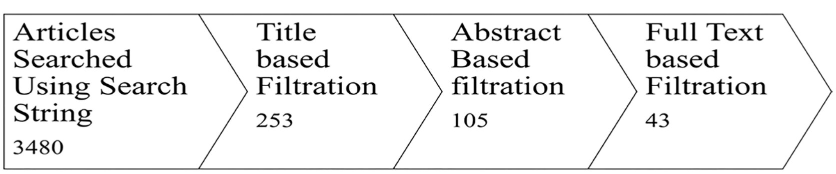

Fig 2 | Study selection process

Analysis of Primary Studies

ML and deep learning (DL) are used for precision agriculture to enhance crop management and maximize productivity.[11] Integrating these advanced algorithms delivers the power of precision agriculture systems, which can order large quantities of complex data about weather, soil health, and crop characteristics. ML and DL models perform disease detection more accurately and timely for intervention early before losses are incurred in crops. Furthermore, these algorithms draw on historical and real-time data to predict crop yields and allow farmers to make data-based decisions on planting strategies, allocation of resources, and pest control. Precision agriculture uses predictive insights to make crop performance better and helps reduce environmental impact and resource wastage.

Soil Properties and Weather Prediction

The prediction of soil properties is an important step in crop selection, land preparation, seed choice, yield, and fertilizer use. More precisely, soil qualities include the ability to anticipate soil nutrients, surface humidity, and weather patterns over the course of a crop's lifecycle. These attributes are correlated with geographic and seasonal circumstances. Crop cultivation has been affected by human activities to significantly change the quality of soil. Farmers are guided by electric and electromagnetic sensors that monitor essential nutrients to the most suitable crops available. Fertilizers and manure can enhance nutrient levels, but it will come at an extra cost and may damage the environment and break the soil cycle.

Determining the qualities of soil requires a thorough investigation of its pH, moisture content (MC), and nutrients. Using data from Radarsat-2, Acar et al[12] tested an extreme learning machine (ELM) regression model to estimate soil surface humidity on two different terrains. After preprocessing and feature extraction, the model achieved the lowest RMSE of 2.19% with a sine kernel function, validated through leave-one-out

Table 2 | Study selection criteria

Criteria #	Inclusion Criteria	Exclusion Criteria
IE1	Studies that focus on machine learning applications in precision agriculture.	Studies not related to agriculture or machine learning.
IE2	Peer-reviewed journal articles published in English.	Non-peer-reviewed articles, conference papers, or book chapters.
IE3	Research conducted within the last 10 years to ensure relevance.	Studies published before the last 10 years.
IE4	Empirical studies that present original research findings.	Review articles, meta-analyses, or theoretical papers without empirical data.
IE5	Studies that provide sufficient methodology and results for analysis.	Studies lacking clear methodology, results, or data on machine learning applications.

Table 3 | Study selection results

Phase	Process	Selection Stage	Elsevier	Springer	IEEE	MDPI	Frontiers	Total
1	Search	Search string	2330	311	517	212	110	3480
2	Screening	Title	57	47	63	39	47	253
3	Screening	Abstract	44	21	19	10	11	105
4	Finalizing	Full text	26	4	9	2	2	43

cross-validation. Wang et al[13] applied ELM-based soft sensors to monitor nutrient solution composition in soilless cultivation, using auxiliary variables like conductivity and pH. The model predicted SO_2^{4-} and H_2PO_4- concentrations with RMSEs of 1.24 and 0.89, respectively. Park et al[14] used random forest (RF) and Cubist algorithms to downscale AMSR2 soil moisture data, achieving an R^2 of 0.96 and RMSE of 0.06, outperforming the ordinary least squares method.

Reda et al[15] applied ML to estimate soil organic carbon (SOC) and total nitrogen (TN) in Moroccan agricultural lands using near-infrared spectroscopy, which reduced computation time compared to traditional chemical methods. Their ensemble model achieved an R^2 of 0.96 with an RMSE of 1.92 for SOC, and an R^2 of 0.94 with an RMSE of 0.57 for TN. Morellos et al[16] also used visible and infrared spectroscopy to predict SOC, TN, and MC in German arable fields, with LS-SVM achieving the lowest RMSE of 0.062 for SOC and 0.457 for MC. Cubist excelled in TN prediction with an RMSE of 0.071. Andrade et al[17] used portable X-ray fluorescence to predict TN, SOM, and CEC in Brazilian soils, with RF outperforming other models. Deiss et al[5] found that partial least squares regression was outperformed by the support vector machines (SVM) model in the prediction of soil parameters (clay, sand, pH, SOC) using spectroscopic data in Tanzania and the United States.

Mahmoudzadeh et al[18] investigated ML algorithms to predict SOC in Iran's Kurdistan province. Their results indicated that RF was the most accurate, achieving an R^2 of 0.60 and an RMSE of 0.35%, outperforming SVM, k-nearest neighbors (kNN), Cubist, and Extreme Gradient Boosting (XGBoost). Key factors influencing SOC included air temperature, annual rainfall, valley depth, and terrain texture. Veres et al[19] utilized DL, specifically convolutional neural networks (CNNs), to predict soil properties from infrared spectroscopy data.

Benke et al[20] used a pedotransfer function that has been based on a generalized linear mixed effects model, combined with residual maximum likelihood for simulation inputs, to predict soil electrical conductivity and SOC across the various regions of Victoria, Australia.

Labrador et al[3] estimated calcium and magnesium content in soil using generalized regression neural networks and genetic algorithms, utilizing digital elevation models and satellite images as inputs. Shin et al[21] developed a seasonal climate forecasting model using a regularized ELM to predict daily mean air temperature for 90 days based on data from the Korea Meteorological Administration. Their model outperformed traditional meteorological data, achieving an RMSE between 1.02 and 3.35, compared to the meteorological data's RMSE range of 1.61 to 3.37.

Soil MC is crucial for precise irrigation scheduling in agriculture. Stamenkovic et al[22] developed a support vector regression (SVR)-based model for soil moisture prediction from hyperspectral images. Song et al[23] demonstrated that a macroscopic cellular automata (MCA) model integrated with a deep belief network can produce a better prediction than MLP-MCA for the soil moisture in cornfields in northwest China in terms of RMSE. To simulate the soil water retention curve of loamy sand, Acheing[24] examined a number of models, including SVR (RBF), artificial neural networks (ANN), and deep neural networks (DNN). Under wet and dry conditions, the predictions provided by the RBF-based SVR model were the best. Feng et al[25] estimated soil temperature at various depths in the Loess Plateau of China using four ML algorithms: For this purpose we use ELM, GRNN, BPNN, and RF. ELM had the lowest RMSE, MAE, NS, and the highest concordance correlation coefficient among other methods. In some Iranian regions, Mohammadi et al[26] showed that ELM could forecast the dew point temperature one day ahead with greater accuracy than SVM and ANN developed.

Zhu et al[27] used a hybrid particle swarm optimization-ELM model to accurately estimate daily evaporated water in northwest China. Alizamir et al[28] predicted soil temperature at depths of 5, 10, 50, and 100 cm, finding that air temperature alone suffices for depths up to 50 cm, while additional data, such as solar radiation and wind speed, is needed for 100 cm.

Rainfall prediction is necessary for water resource and agriculture management. Radial-basis function neural network (RBFNN) had been evaluated by Cramer et al[29] to develop rainfall prediction using seven ML algorithms and found the best evaluation resulted in RBFNN. Atmospheric synoptic patterns are also used to predict rainfall in Tenerife, Spain, by Sierra and Jesus,[30] and neural networks outperformed these criteria. Using neural networks in weather prediction, Kamatchi and Parvathi[31] presented a hybrid recommender system for improving prediction success rate. Lazri et al[32] improved classification standards through the development of a multi-classifier model, combining Meteosat Second Generation images and radar data for estimating precipitation. A comparison of soil properties and weather prediction techniques is shown in Table 4.

The findings from the studies on soil properties and weather prediction using ML highlight significant advancements in predicting various soil characteristics, such as moisture, organic carbon content, and temperature, as well as weather patterns like rainfall and temperature. Several ML models, including ELM, RFs, SVM, and DL techniques have shown strong performance in these applications. For example, ELM models achieved low RMSE values for soil moisture and temperature predictions, while RFs and ensemble models demonstrated high accuracy in estimating soil nutrients and organic content. Additionally, hybrid models and novel approaches, such as PSO with ELM, improved the accuracy of predictions, particularly in complex environmental conditions.

Yield Prediction

Crop production is undergoing a paradigm shift as a result of the application of ML/DL approaches to yield prediction.[33] Farmers can profit from the useful

 DOI: https://doi.org/10.70389/PJPB.100005 | Premier Journal of Plant Biology 2024;1:100005

Table 4 | Comparison of soil properties and weather prediction techniques

Ref	Dataset Name	Soil Surface Humidity	Soil Organic Carbon	Soil Temperature	Soil Moisture	Nutrient Prediction	ML Techniques	EL Technique	DL Technique
12	RADARSAT-2	✓	×	×	×	×	✓	×	×
13	Precipitation, Temperature Inputs	×	✓	×	×	✓	×	✓	×
14	AMSR2, MODIS	×	×	×	✓	×	×	✓	×
15	Southern Italy Agricultural Lands	×	✓	×	×	×	×	✓	×
16	Premislin, Germany	×	✓	×	×	✓	✓	×	×
17	Brazilian Coastal Plains	×	✓	×	×	✓	✓	✓	×
18	Kurdistan Province, Iran	×	✓	×	×	×	✓	✓	×
20	Victoria, Australia	×	✓	×	×	×	×	✓	✓
3	North Western Madison County, Ohio	×	✓	×	×	✓	✓	✓	×
21	GloSea5GC2 Model	×	×	✓	×	×	×	✓	×
25	Maize Field, Northern P.R. China	×	×	✓	×	×	✓	✓	✓
26	Bandar Abbas and Tabas Stations, Iran	×	×	×	×	✓	✓	✓	✓
27	Northwest Region, Arid Area	×	×	×	×	×	×	×	✓
28	Turkey	×	×	✓	×	×	✓	✓	✓
30	Tenerife, Spain	×	×	×	×	✓	✓	✓	
32	Northern Algeria	×	×	×	×	✓	✓	✓	✓

information that ML and DL models offer about how to successfully enhance agricultural management and achieve an increase in yield and security of food in a setting that is changing quickly.

Agricultural scientists have recently found great success with ML techniques for estimating crop output, productivity, and pace. Wang, Shi, and Wen[34] used data from the North China Plain in a recent study to assess the accuracy of five ML models used as predictions of winter season wheat dry matter and yield. Furthermore, Segarra, Araus, and Kefauver[35] mapped the fluctuations in field wheat grain yield in Spain using ML methods in conjunction with Earth Observation Systems (EOS). The model Su, Xu, and Yan[6] created for modeling the growth of rice and predicting yield is known as SBOCM. Next, SVM is used as a basis. Ramos[36] proposed a non-destructive approach to count coffee fruits on trees using machine vision, while Iniyan, Akhil Varma, and Teja Naidu[37] investigated ML techniques to forecast agricultural yields based on characteristics in the soil and surrounding environment. These studies do, however, highlight the value of ML in agricultural decision-making, precision agriculture, and productivity enhancement. Nevertheless, there are issues, such as certain models' limited scalability and non-generalizability.

The most promising area of research, however, is crop output forecasting and the evaluation of numerous aspects that affect agricultural production. DL approaches have been widely applied in agricultural research as the newest technology in the sector. This has been done in an attempt to close the ever-widening gap between production and consumption. By combining geometric parameters and plant indicators utilizing UAVS Multispectral Imaging, it is shown that the approach proposed by Zheng et al[38] is effective in creating a dry biomass weight strawberry model employing canopy geometric factors and spectral indications. It is evident that Chu and Yu[39] also employ the BBI model as a complete prediction model to estimate optimal rice production using neural networks. Similar to this, Ballesteros et al[40] predict vineyard productivity by combining computer vision techniques with vegetation indices from UAV multispectral imagery, and Ali et al[41] use satellite remote sensing data to calculate field grass biomass in the intensively raised pastures. But in addition, the Khaki, Wang, and Archontoulis team[42] created a novel DL method that incorporates agricultural practices and environmental data to precisely estimate productivity. All things considered, the research points to the possibility that ML and DL will revolutionize agricultural decision-making, even though a number of obstacles, such as complexity and generality, remain unresolved. Table 5 presents a comparison of different methods.

Recent studies highlight the significant potential of ML and DL approaches in enhancing crop yield prediction and agricultural management. ML models, such as SVM and machine vision, have been successfully applied to predict wheat, rice, coffee, and other crop yields, leveraging data from sources like EOS and multispectral UAV imagery. DL techniques, especially those incorporating environmental and agricultural data, have further advanced the precision of yield forecasting. Despite their promising results, challenges such as limited scalability, non-generalizability, and model complexity remain.

Disease Detection

The field of disease detection is changing as a result of advances in ML and DL technologies in the spheres of environmental science, healthcare, and agriculture.[1]

Table 5 | Comparison of techniques in crop yield production

Ref	Dataset	Region-Specific	Remote Sensing	ML Techniques	EL Technique	DL Technique	Scalability Issues	Complex Preprocessing
34	Data from 48 peer-reviewed studies in the North China Plain	✓	×	✓	×	×	×	×
35	GPS combine harvester data and Sentinel-2 satellite imagery	×	✓	✓	×	×	×	×
6	Surface weather and soil data from China	✓	×	✓	×	×	×	×
36	1018 coffee branches at varying ripening stages	×	×	✓	×	×	✓	×
37	District-wise crop data from Maharashtra, India	✓	×	✓	×	×	✓	✓
38	UAV multispectral imagery	No	✓	×	×	✓	✓	×
39	Data from 81 counties in the Guangxi Zhuang Autonomous Region	✓	×	×	×	✓	×	✓
40	Multispectral imagery from UAV	No	✓	×	×	✓	✓	✓
41	In situ measurements from two test sites in Ireland	✓	×	×	×	✓	×	×
42	Corn and soybean data across the US Corn Belt	×	×	×	×	✓	×	×

Applications of ML and DL in agriculture are being utilized more and more to diagnose agricultural diseases and make judgments. These algorithms are more accurate and efficient than traditional methods in identifying a pattern or anomaly that indicates a disease by analyzing vast datasets that include data on crop health, environmental conditions, and disease indicators.

Transfer learning techniques, which facilitate the treatment of agronomic diseases, are one of the newer research methodologies. For instance, Morbekar[43] uses the YOLO object identification algorithm, which may quickly identify plant problems in Indian crops like cotton and rice and allow for the quick application of corrective action. The device can only diagnose problems on leaves; it cannot identify diseases in other parts of the plant. Its focus is restricted to main agricultural crops. Similarly, Amarasingam et al[1] used UAV-based remote sensing and DL approaches like YOLOv5, which received high values for precision and recall, to target white leaf disease (WLD) detection in Sri Lankan sugarcane fields. On the other hand, the study focused only on WLD detection in plants. Crop diseases and revalidation in many locations are required to accomplish this. Wang, Liu, and Zhu[7] develop a sophisticated YOLOv3-based system that serves as a real-time guide for the early detection of pests and tomato illnesses, while also addressing challenges such as mutual occlusions and complex backdrops. However, weather and variations in natural light can make it difficult to implement a solution for field or ground missions. Nonetheless, Ferentinos[44] employed convolutional neural network models to detect and diagnose plant illnesses, achieving excellent results but encountering a generalizability issue brought on by insufficient testing on the training set alone and the requirement for more varied data.

Recent studies have looked into the use of different imaging technologies for the detection of crop disorders and stress/distress syndrome. Pantazi, Moshou, and Bochtis[45] use hyperspectral imaging technology to explore the relationship between nitrogen stress and yellow rust disease in crops. Ground-based hyperspectral photography is used to combine ANN with supervised Kohonen networks for mixed categorization. The lack of coverage of a broad range of stress situations and the need for the findings to be confirmed across various biological and agricultural phases are the cited limitations, though. Similarly, Pantazi et al[46] use agronomic technologies such as field spectroscopy and hierarchical self-organizing maps to identify *Silybum marianum* that is systemically infected with the smut fungus *Microbotryum silybum*. Although the method has proven to be accurate in identifying infected individuals during the vegetative development stage, more validation is needed in different environmental settings and on the fieldwork platform.

Recent research has demonstrated the high degree of efficiency that can be achieved in crop disease identification and pest management by combining ML algorithm procedures with remote sensing and machine vision technologies. Chung et al's research[47] uses machine vision and SVM classifiers to develop a non-invasive approach for distinguishing healthy rice seedlings from those infected with *Fusarium fujikuroi*. However, Ebrahimi et al.[50] present a method using automatic image processing and SVM classification to detect thrips in strawberry greenhouses. In order to monitor crops and conduct disease surveillance, Gomez et al[48] investigate the use of remote sensing equipment and ML techniques, primarily with regard to banana plants in African environments. Additionally, Abdulridha et al[49] investigate the use of spectral vegetation indicators and ML models for the diagnosis of DM disease in watermelon plants at varying phases of disease severity. These studies demonstrate the promise of these systems for early illness identification and pest management; however, further validation and improvement are needed before these approaches can be fully implemented. Table 6 presents a comparison of different methods.

Recent advancements in ML and DL have significantly improved disease detection in agriculture, enabling more accurate and efficient identification of plant diseases and pests. Techniques such as YOLO-based object detection, CNNs, and hyperspectral imaging are being employed to detect a variety of crop diseases, including WLD, tomato illnesses, and nitrogen stress. While these methods show promise for early detection and pest management, challenges such as limited generalizability, environmental factors affecting accuracy, and the need for broader validation across different settings remain. Continued research is required to enhance the reliability and applicability of these technologies in real-world agricultural environments.

Discussion

This section will explore the most common techniques used in soil property prediction and the challenges that researchers face in this field. Various ML algorithms, such as SVM, ELM, and RFs, have shown promise in predicting soil attributes like moisture, organic carbon, and nutrient content. However, despite advancements, challenges such as data quality, the need for extensive labeled datasets, and the variability of soil properties across different regions persist.

Common Techniques

In precision agriculture, ML and DL techniques have been widely applied to enhance decision-making processes, optimize resource use, and improve yield prediction and disease detection as shown in Figure 3. ML models like SVM, RF, and boosted regression have proven effective in estimating crop yields and soil

Table 6 | Comparison of techniques in crop disease detection

Ref	Dataset	Uses Remote Sensing	Real-Time Detection	ML Techniques	EL Technique	DL Technique	High Accuracy	Scalability Issues	Requires Validation
43	PlantVillage	×	✓	×	×	✓	✓	✓	×
1	Sugarcane fields at Gal-Oya Plantation, Sri Lanka	×	✓	×	×	✓	✓	✓	✓
7	Tomato greenhouse in Shouguang City	×	✓	×	×	✓	✓	×	✓
44	Open database	×	×	×	×	✓	✓	✓	×
45	Hyperspectral images from Rothamsted Research, UK	×	×	✓	×	✓	✓	✓	✓
46	Experimental field of *S. marianum* plants	×	×	×	×	✓	✓	✓	✓
47	Rice seedlings from 'Tainan 11' and 'Toyonishiki' cultivars	×	×	×	✓	×	✓	×	✓
48	Multi-level satellite and UAV imagery	✓	×	×	✓	✓	✓	✓	✓
49	Hyperspectral images of watermelon leaves	×	×	✓	×	✓	×	✓	×
50	Crop canopy images	×	×	×	✓	×	✓	×	×

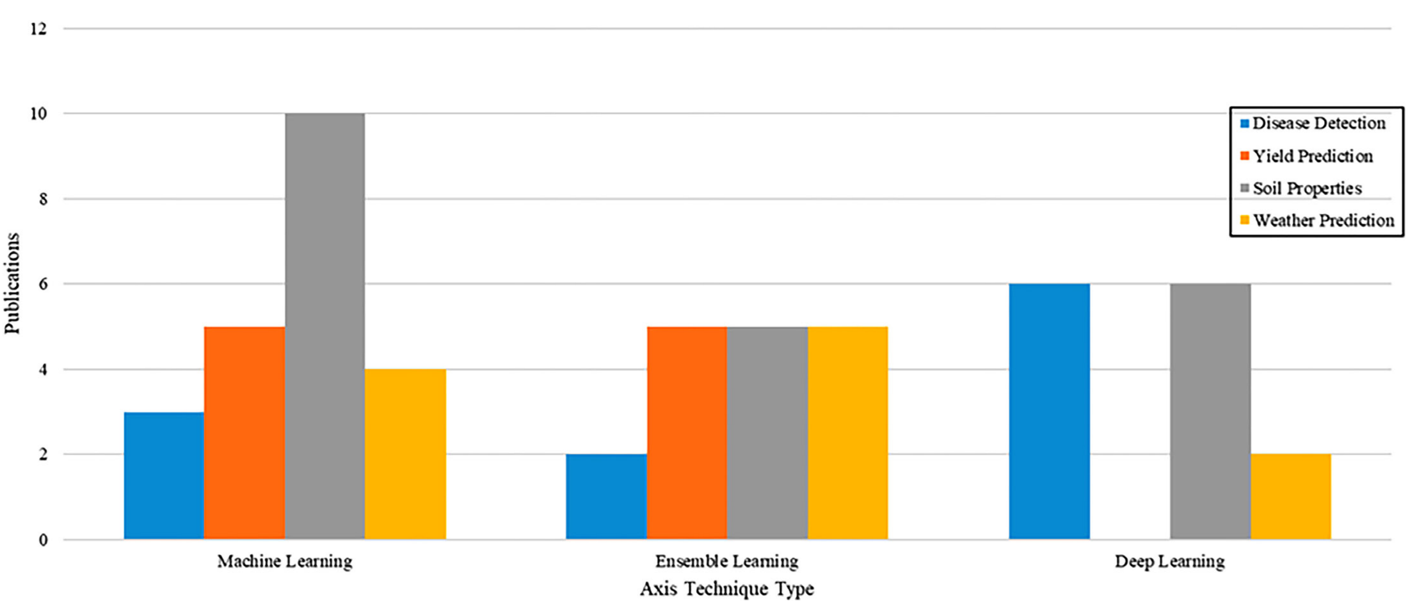

Fig 3 | Most common techniques used in precision agriculture

properties, leveraging data from various sources like satellite imagery, in situ measurements, and multispectral imagery.[35,38] These models offer reliable predictions by analyzing complex agricultural data such as crop growth stages, soil conditions, and weather patterns, making them invaluable tools for farmers to make data-driven decisions. SVM, for example, is often used due to its simplicity and effectiveness in handling small datasets, while RF excels in managing larger and more complex datasets by combining the strengths of multiple decision trees. Additionally, ensemble learning (EL) methods, such as the integration of KNN, SVM, and XGBoost, have shown promise in combining different models to increase accuracy, especially when predicting soil properties and weather conditions.

DL, on the other hand, has gained traction for more intricate tasks like disease detection and biomass prediction. Techniques like YOLO, CNN, and recurrent neural networks are popular for their ability to process large amounts of image and time-series data.[43,51,52] YOLO, for example, is widely used for real-time disease detection on crops, enabling rapid responses to potential issues. Similarly, DBNs and DNNs are applied to assess soil properties, offering enhanced accuracy in capturing complex relationships between soil characteristics and crop performance. The integration of these advanced techniques with traditional ML approaches has significantly improved the precision and efficiency of modern agricultural practices.

Applications of LLM in Precision Agriculture and Crop Yield Prediction

Large language models (LLMs) are increasingly being integrated into precision agriculture to enhance decision-making processes and optimize crop management practices. These models are capable of processing vast amounts of agricultural data, including weather forecasts, soil conditions, satellite imagery, and sensor data. LLMs can provide farmers with valuable insights into the best times to plant, irrigate, and harvest based on real-time data, improving crop health and productivity. Additionally, LLMs facilitate the analysis of historical agricultural data, helping identify patterns and trends that could influence future crop performance. This ability to analyze complex datasets enables precision agriculture tools to recommend specific interventions that optimize resource usage, reduce waste, and increase yields, all while maintaining environmental sustainability.

In the context of crop yield prediction, LLMs play a significant role in forecasting agricultural outputs with high accuracy. These models can predict yield outcomes by synthesizing data from diverse sources, such as weather patterns, crop health indicators, soil quality, and even market trends. LLMs can also simulate the effects of various environmental factors on crop growth, providing farmers with a clear picture of how external conditions could impact their yields. This predictive capability helps farmers make informed

Table 7 | Summary of challenges in precision agriculture and crop yield prediction

Area	Challenge	Description	Impact on Model	Solutions/Considerations
Soil Properties and Weather Prediction	Data Quality and Availability	Lack of accurate and stable data for prediction, especially in soil and climate.	Inaccuracy of model predictions.	Improving data collection methods and access to diverse datasets.
	Complexity of Soil Characteristics	Heterogeneity in soil properties like texture, moisture, and organic content.	Difficulty in building accurate prediction models.	Developing models that can generalize across soil types.
	Environmental Variability	Regional and time-based variations in weather and environmental factors.	Increased complexity in modeling and potential inaccuracies.	Incorporating diverse environmental conditions into models.
	Feature Selection and Dimensionality	Difficulty in selecting relevant features from high-dimensional data.	Overfitting and poor generalization.	Effective feature selection techniques and dimensionality reduction.
Yield Prediction	Temporal Variability	Variability in crop yields due to season, pests, diseases, and weather.	Need to capture dynamic changes accurately in models.	Incorporating time-series data and dynamic modeling techniques.
	Integration of Multisource Data	Difficulty in synchronizing and integrating diverse data sources (e.g., satellite images, climate, soil).	Incompatibility and difficulty in model training.	Data fusion techniques for better integration.
	Model Interpretability	Lack of transparency in deep learning models leading to difficulty in understanding predictions.	Reduced trust and usability for farmers.	Development of explainable AI models for transparency.
	Scalability	Models trained in one environment may not work well in different conditions.	Models need retraining to adjust to new environments.	Scalable and adaptive models for diverse farming systems.
Disease Prediction	Data Scarcity and Quality	Lack of high-quality labeled datasets for disease prediction.	Difficulty in training robust models for disease detection.	Data augmentation and collaboration for better dataset creation.
	Diverse Disease Manifestations	Diseases manifest in various forms depending on species, environment, and pathogen.	Challenge in ensuring data diversity for accurate training.	Gathering diverse and comprehensive training data.
	Dynamic Nature of Pathogens	Pathogens evolve over time, altering virulence and resistance.	Models may become obsolete without frequent updates.	Regular updates and retraining of models with new data.
	Integration of Expert Knowledge	Lack of integration of agronomic and biological expertise into ML models.	Reduced model accuracy and relevance.	Collaboration between AI and domain experts (e.g., agronomists).

decisions regarding crop management, risk mitigation, and resource allocation. Furthermore, LLMs can be used to analyze agricultural research papers and technical documents, extracting knowledge that can be used to refine yield prediction models and enhance their precision over time.

Challenges

ML in agriculture has been successfully implemented in several domains such as soil properties and weather forecasting, yield prediction, and disease forecasting. However, there are a few challenges that reduce the efficiency of the execution and optimization of these models in real-world situations as shown in Table 7. Below are some of the key challenges faced in each of these areas:

Soil Properties and Weather Prediction

- **Data Quality and Availability:** Many a time, the efficacy of the developed ML algorithms totally depends on the quality of data as well as their availability. In the climatic and especially soil conditions prediction, there might not be enough or stable unreliable data, which leads to the inaccuracy of models.
- **Complexity of Soil Characteristics:** Soil is a heterogeneous environment composed of different features including texture, moisture, and organic content. The complex relations between these properties are why it becomes difficult to build highly accurate prediction models that may work well across varying types of soil and other conditions.
- **Environmental Variability:** The weather factors change with regions and time which make the modeling process challenging. Still, more diverse antecedent conditions, including topography and microclimates, may increase variability that models need to consider because of human influences.
- **Feature Selection and Dimensionality:** Another important step of data preprocessing is the selection of features for model training which is often very difficult when working with high-dimensional environmental data. Failure in feature selection leads to overfitting in the models, hence leading to poor generation abilities in field environments.

Yield Prediction

- **Temporal Variability:** Yields in crops can vary depending on the season, pests and diseases, and weather among other factors. This point indicates that it is important to capture the said dynamics appropriately when modeling, to make such models capture actual yield over time.
- **Integration of Multisource Data:** Yield prediction may require other external data such as images from satellites, climate information at a given period, and information on soil. It becomes extremely tough to maintain the compatibility and synchronization of these datasets.
- **Model Interpretability:** Due to the complex hierarchical structures in many deeply learned models, farmers and stakeholders end up having no way of analyzing the predictions beyond relying on the raw output. Such a type of lack of interpretability can create barriers to trust and therefore usage of the models in practice especially in agriculture.
- **Scalability:** When it comes to utilizing ML models across multiple regions and farming systems, scalability-related challenges exist. Each of the trained models may not be very effective when tested in different conditions, adherent to the basic rule which requires reworking through a new set of data to adjust the model to fit the new environment.

Disease Prediction

- **Data Scarcity and Quality:** There are situations when disease-prediction-labeled datasets of high quality are scarce. There is no complete data available for training such robust models that can detect or even predict the disease.
- **Diverse Disease Manifestations:** Crop diseases may manifest themselves in many forms in ways that are dependent on factors such as species, environment, and pathogen. Reflecting diversity in training data is crucial, but not easy at all.
- **Dynamic Nature of Pathogens:** A pathogen develops over some time, which is why features like virulence and resistance change over time. The models learned from previous data might thus be obsolete, thus requiring frequent updating and retraining to meet their purpose.
- **Integration of Expert Knowledge:** Such techniques are not given appropriate attention because they involve interweaving the agronomic and biological concepts into the ML models. This was done by integrating with agronomists and plant pathologists, which can improve the model performance; however, good communication and joint work are also needed.

Conclusion and Future Directions

ML, EL, and DL techniques are revolutionizing traditional farming practices by enabling more accurate detection of diseases, yield prediction, analysis of soil properties, and weather forecasting. This systematic literature review highlights the significant advancements in applying these techniques to optimize agricultural practices, offering promising solutions to enhance productivity and sustainability. However, challenges such as the need for high-quality datasets, the scalability of models across diverse agricultural regions, and the requirement for robust regional validations remain key obstacles. Overcoming these limitations is essential to unlocking the full potential of precision agriculture. By addressing these challenges, future research can lead to more reliable and adaptable models, ultimately benefiting farmers, improving food security, and contributing to sustainable agricultural practices. The findings of this review have the potential to inform policy, guide future research, and enhance the adoption of advanced agricultural technologies, positively impacting both local farming communities and society at large.

Future research should aim to further scale and adapt ML models to generalize over varying agricultural landscapes and climates. Interdisciplinary collaboration is also needed to develop better data collection techniques to train the models on large, high-quality ones. Furthermore, the integration of advanced technology such as IoT and edge computing can further improve real-time monitoring and decision-making in agriculture. Ethical considerations like the privacy of data and the socio-economic contribution of AI in agriculture will also be need to be addressed to ensure the adoption of smart farming technologies sustainably and inclusively.

References

1 Amarasingam N, Gonzalez F, Surantha A, Salgadoe A, Sandino J, Powell K. Detection of white leaf disease in sugarcane crops using UAV-derived RGB imagery with existing deep learning models. Remote Sens. 2022;14(23):6137.

2 Ndlovu HS, Odindi J, Sibanda M, Mutanga O, Clulow A, Chimonyo VGP, et al. A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)-based proximal and remotely sensed data. Remote Sens. 2021;13(20):4091.

3 Labrador Y, Chang C, Viloria J. Prediction of the calcium and magnesium content in soils through a generalized regression neural networks and genetic algorithms. Proceedings of the 2012 11th International Conference on Machine Learning and Applications 2012;(Vol. 1, pp. 661–6). IEEE.

4 Pantazi XE, Moshou D, Alexandridis T, Whetton RL, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric. 2016;121:57–65. http://dx.doi.org/10.1016/j.compag.2015.11.018

5 Deiss L, Margenot AJ, Culman SW, Demyan MS. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. Geoderma. 2020;365:114227.

6 Ying-xue S, Huan X, Li-jiao Y. Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J Biol Sci. 2017;24(3):537–47. http://dx.doi.org/10.1016/j.sjbs.2017.01.024

7 Wang X, Liu J, Zhu X. Early real - time detection algorithm of tomato diseases and pests in the natural environment. Plant Methods. 2021;1–17. https://doi.org/10.1186/s13007-021-00745-2

8 Elberzhager F, Münch J, Nha VTN. A systematic mapping study on the combination of static and dynamic quality assurance techniques. Inf Softw Technol. 2012;54(1):1–15.

9 Ishaq M, Abid A, Farooq MS, Manzoor MF, Farooq U, Abid K, et al. Advances in database systems education: Methods, tools, curricula, and way forward. In Education and Information Technologies 2023;(Vol. 28, pp. 2681–2725). Springer. https://doi.org/10.1007/s10639-022-11293-0

10 Widyassari AP, Affandy F, Noersasongko E, Fanani AZ, Syukur A, Basuki RS. Literature review of automatic text summarization: Research trend, dataset and method. 2019 International Conference on Information and Communications Technology (ICOIACT) 2019;(pp. 491–6). IEEE.

11 Diaz-Gonzalez FA, Vuelvas J, Correa CA, Vallejo VE, Patino D. Machine learning and remote sensing techniques applied to estimate soil indicators – Review. Ecol Indic. 2022;135:108517. https://doi.org/10.1016/j.ecolind.2021.108517

12 Acar E, Ozerdem MS, Ustundag BB. Machine learning based regression model for prediction of soil surface humidity over moderately vegetated fields. 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) 2019;(Vol. 1, pp. 1–4). IEEE.

13 Wang X, Hu W, Li K, Song L, Song L. Modeling of soft sensor based on DBN-ELM and its application in measurement of nutrient solution composition for soilless culture. Proceedings of the 2018 IEEE International Conference of Safety Produce Informatization (IICSPI) 2018;(Vol. 1, pp. 93–7).

14 Park S, Im J, Park S, Rhee J. AMSR2 soil moisture downscaling using multisensor products through machine learning approach. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015;(pp. 1984–7).

15 Reda R, Saffaj T, Ilham B, Issam K, Brahim L, El EM, et al. A comparative study between a newmethod and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy. Chemometr Intell Lab Syst. 2019;195:103873.

16 Morellos A, Pantazi XE, Moshou D, Alexandridis T, Whetton R, Tziotzios G, et al. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosyst Eng. 2016;152:104–16.

17 Andrade R, Silva SHG, Weindorf DC, Chakraborty S, Faria WM, Mesquita LF, et al. Assessing models for prediction of some soil chemical properties from portable X-ray fluorescence (pXRF) spectrometry data in Brazilian Coastal Plains. Geoderma. 2020;357:113957.

18 Mahmoudzadeh H, Matinfar HR, Taghizadeh-Mehrjardi R, Kerry R. Spatial prediction of soil organic carbon using machine learning techniques in western Iran. Geoderma Reg. 2020;21:e00260.

19 Veres M, Lacey G, Taylor GW. Deep learning architectures for soil property prediction. Proceedings of the 2015 12th Conference on Computer and Robot Vision 2015;(pp. 8–15). IEEE.

20 Benke KK, Norng S, Robinson NJ, Chia K, Rees DB, Hopley J. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. Geoderma. 2020;366:114210.

21 Shin JY, Kim KR, Ha JC. Seasonal forecasting of daily mean air temperatures using a coupled global climate model and machine learning algorithm for field-scale agricultural management. Agric For Meteorol. 2020;281:107858.

22 Stamenkovic J, Tuia D, De Morsier F, Borgeaud M, Thiran JP. Estimation of soil moisture from airborne hyperspectral imagery with support vector regression. 2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) 2013;(pp. 1–4). IEEE.

23 Song X, Zhang G, Liu F, Li D, Zhao Y, Yang J. Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. J Arid Land. 2016;8(5):734–48.

24 Achieng KO. Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models. Comput Geosci. 2019;133:104320.

25 Feng Y, Cui N, Hao W, Gao L, Gong D. Estimation of soil temperature from meteorological data using different machine learning models. Geoderma. 2019;338:67–77.

26 Mohammadi K, Shamshirband S, Motamedi S, Petković D, Hashim R, Gocic M. Extreme learning machine based prediction of daily dew point temperature. Comput Electron Agric. 2015;117: 214–25.

27 Zhu B, Feng Y, Gong D, Jiang S, Zhao L, Cui N. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. Comput Electron Agric. 2020;173:105430.

28 Alizamir M, Kisi O, Ahmed AN, Mert C, Fai CM, Kim S, et al. Advanced machine learning model for better prediction accuracy of soil temperature at different depths. PLoS One. 2020;15(4):1–25.

29 Cramer S, Kampouridis M, Freitas AA, Alexandridis AK. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Syst Appl. 2017;85:169–81.

30 Diez-Sierra J, del Jesus M. Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. J Hydrol. 2020;586:124789.

31 Kamatchi SB, Parvathi R. Improvement of crop production using recommender system by weather forecasts. Procedia Comput Sci. 2019;165(2019):724–32.

32 Lazri M, Labadi K, Brucker JM, Ameur S. Improving satellite rainfall estimation from MSG data in Northern Algeria by using a multi-classifier model based on machine learning. J Hydrol. 2020;584:124705.

33 Kung HY, Kuo TH, Chen CH, Tsai PY. Accuracy analysis mechanism for agriculture data using the ensemble neural network method. Sustain. 2016;8(8):1–11.

34 Wang Y, Shi W, Wen T. Prediction of winter wheat yield and dry matter in North China Plain using machine learning algorithms for optimal water and nitrogen application. Agric Water Manag. 2023;277:108140. https://doi.org/10.1016/j.agwat.2023.108140

35 Segarra J, Araus JL, Kefauver SC. Farming and earth observation: Sentinel-2 data to estimate within-field wheat grain yield. Int J Appl Earth Obs Geoinf. 2022;107:102697. https://doi.org/10.1016/j.jag.2022.102697

36 Ramos PJ, Prieto FA, Montoya EC, Oliveros CE. Automatic fruit count on coffee branches using computer vision. Comput Electron Agric. 2017;137:9–22. http://dx.doi.org/10.1016/j.compag.2017.03.010

37 Iniyan S, Akhil Varma V, Teja Naidu C. Crop yield prediction using machine learning techniques. Adv Eng Softw. 2023;175:103326. https://doi.org/10.1016/j.advengsoft.2022.103326

38 Zheng C, Abd-Elrahman A, Whitaker V, Dalid C. Prediction of strawberry dry biomass from UAV multispectral imagery using multiple machine learning methods. Remote Sens. 2022;14(18):4511.

39 Chu Z, Yu J. An end-to-end model for rice yield prediction using deep learning fusion. Comput Electron Agric. 2020;174:105471. https://doi.org/10.1016/j.compag.2020.105471

40 Ballesteros R, Intrigliolo DS, Ortega JF, Ramírez-Cuesta JM, Buesa I, Moreno MA. Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. Precis Agric. 2020;21(6):1242–62. https://doi.org/10.1007/s11119-020-09717-3

41 Ali I, Cawkwell F, Dwyer E, Green S. Modeling managed grassland biomass estimation by using multitemporal remote sensing data-A machine learning approach. IEEE J Sel Top Appl Earth Obs Remote Sens. 2017;10(7):3254–64.

42 Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. Front Plant Sci. 2020;10:1–14.

43 Morbekar A, Parihar A, Jadhav R. Crop disease detection using YOLO. 2020 International Conference for Emerging Technology (INCET) 2020;(pp. 1–5). IEEE.

44 Ferentinos KP. Deep learning models for plant disease detection and diagnosis. Comput Electron Agric. 2018;145:311–8. https://doi.org/10.1016/j.compag.2018.01.009

45 Pantazi XE, Moshou D, Bochtis D. Detection of biotic and abiotic stresses in crops by using hierarchical self organizing classifiers. Precis Agric. 2017;18:383–93.

46 Pantazi XE, Tamouridou AA, Alexandridis TK, Lagopodi AL, Kontouris G, Moshou D. Detection of Silybum marianum infection with Microbotryum silybum using VNIR field spectroscopy. 2017;137:130–7.

47 Chung CL, Huang KJ, Chen SY, Lai MH, Chen YC, Kuo YF. Detecting Bakanae disease in rice seedlings by machine vision. Comput Electron Agric. 2016;121:404–11. http://dx.doi.org/10.1016/j.compag.2016.01.008

48 Gomez M, Vergara A, Montenegro F, Alonso H, Safari N, Raymaekers D, et al. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. ISPRS J Photogramm Remote Sens. 2020;169:110–24. https://doi.org/10.1016/j.isprsjprs.2020.08.025

49 Abdulridha J, Ampatzidis Y, Qureshi J, Roberts P. Identification and Classification of Downy Mildew Severity Stages in Watermelon Utilizing Aerial and Ground Remote Sensing and Machine Learning. Front Plant Sci. 2022;13:1–10.

50 Ebrahimi MA, Khoshtaghaza MH, Minaei S, Jamshidi B. Vision-based pest detection based on SVM classification method. Comput Electron Agric. 2017;137:52–8. http://dx.doi.org/10.1016/j.compag.2017.03.016

51 Ma R, Zhang J, Li M, Chen L, Gao J. Hybrid answer selection model for non-factoid question answering. Proceedings of the 2017 International Conference on Asian Language Processing (IALP) 2018;(pp. 371–3). IEEE.

52 He Z, Jiang T, Jiang Y, Luo Q, Chen S, Gong K, et al. Gated recurrent unit models outperform other Machine learning models in prediction of minimum temperature in greenhouse based on local weather data. Comput Electron Agric. 2022;202:107416. https://doi.org/10.1016/j.compag.2022.107416