

Article – e007

A Comprehensive Review of Machine Learning and Deep Learning Approaches for Crop Yield Prediction

Pavan Sahu¹ , **Prof. O.P. Karada²**  

¹ Student, Sage University, Indore, India

² Professor, Sage University, Indore, India

Received: 22/06/2026

Revision Received: 14/06/2026

Accepted: 17/08/2026

ABSTRACT

Crop yield prediction is essential for improving agricultural productivity, resource management, and food security in the face of climate change and increasing population demands [1], [5]. Recent advances in Machine Learning (ML) and Deep Learning (DL) have significantly enhanced the accuracy of crop yield forecasting by analyzing complex agricultural data [1], [3].

This paper presents a comprehensive review of ML- and DL-based approaches for crop yield prediction. It examines widely used algorithms such as Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Artificial Neural Network (ANN), XGBoost, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) models [2], [3], [9]. The review also highlights the contributions of remote sensing, Unmanned Aerial Vehicles (UAVs), and Internet of Things (IoT) technologies in precision agriculture [4], [7], [11].

The analysis indicates that deep learning models generally achieve higher prediction accuracy with large datasets, while machine learning techniques remain effective for structured data. However, challenges such as limited datasets, data heterogeneity, computational complexity, and model generalization still exist [5], [6]. This review identifies current research gaps and discusses future directions for developing robust and sustainable crop yield prediction systems for smart agriculture.

KEYWORDS: Crop Yield Prediction, Machine Learning, Deep Learning, Precision Agriculture, Remote Sensing, UAV, IoT

1. INTRODUCTION

Agriculture is one of the most important sectors supporting the global economy and ensuring food security for a rapidly growing population. However, agricultural productivity is significantly influenced by various factors such as climate change, rainfall patterns, soil properties, temperature, humidity, irrigation practices, and pest infestations. The increasing uncertainty of these factors has made accurate crop yield prediction a challenging task for farmers and policymakers [1], [5], [6].

Traditionally, crop yield estimation relied on historical records, field surveys, and statistical models. Although these approaches provide useful insights, they often fail to capture the complex and nonlinear relationships among agricultural variables. Recent developments in Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), have introduced advanced data-driven techniques capable of analyzing large and heterogeneous datasets to improve prediction accuracy [1], [3], [8].

Machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Artificial Neural Network (ANN), and XGBoost have been widely applied for crop yield prediction because of their ability to process climatic, soil, and crop-related data effectively [2], [9]. Similarly, deep learning models including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Deep Neural Networks (DNN) have demonstrated superior performance in extracting complex spatial and temporal patterns from large-scale agricultural datasets [1], [3], [13].

The advancement of precision agriculture has further accelerated the adoption of emerging technologies such as Remote Sensing, Unmanned Aerial Vehicles (UAVs), Internet of Things (IoT) sensors, and multispectral imaging systems. These technologies provide real-time information about crop health, soil moisture, vegetation indices, and environmental conditions, thereby enhancing the effectiveness of ML and DL models for crop yield prediction [4], [7], [11], [12].

Despite considerable progress, several challenges remain, including limited availability of high-quality datasets, data heterogeneity, computational complexity, model generalization, and the integration of multimodal data sources. Addressing these issues is essential for developing robust and scalable crop yield prediction systems that support sustainable agriculture and global food security [5], [6], [12].

1.1 Research Gap and Novelty of the Present Review

Recent studies have provided valuable insights into the application of machine learning (ML) and deep learning (DL) techniques for crop yield prediction. Existing research has examined a wide range of prediction algorithms, including Random Forest, Support Vector Machine, Artificial Neural Network, hybrid machine learning models, and deep learning approaches [1]–[3], [9], [13]. Other studies have explored the role of remote sensing, Unmanned Aerial Vehicles (UAVs), precision agriculture, and Internet of Things (IoT) technologies in agricultural monitoring and crop-related prediction systems [4], [6], [7], [11], [12].

However, several aspects require further integrated analysis. First, existing literature often discusses prediction algorithms separately from the agricultural data sources and sensing technologies used to support them. Consequently, the relationship between prediction models, input features, datasets, and agricultural technologies is not always presented within a unified framework. Second, the increasing integration of remote sensing, UAVs, and IoT sensors has introduced heterogeneous data sources with different spatial and temporal characteristics, creating additional challenges for data integration and model development [4], [7], [11], [12]. Third, the growing use of hybrid and advanced learning approaches has increased the need to consider practical factors such as computational requirements, data availability, interpretability, scalability, and model generalization [2], [9], [13].

Therefore, the novelty of the present review does not lie in proposing a new prediction algorithm. Instead, it provides an **integrated synthesis of crop yield prediction research**

from multiple complementary perspectives. The review jointly examines (i) classical machine learning methods, (ii) deep learning architectures, (iii) hybrid and ensemble approaches, (iv) remote sensing and UAV-based prediction systems, and (v) IoT-enabled precision agriculture applications. These approaches are further examined in relation to their datasets, input features, application contexts, strengths, limitations, and computational requirements.

The review also identifies important challenges affecting the development and deployment of crop yield prediction systems, including limited availability of high-quality datasets, data heterogeneity, climatic and environmental variability, computational complexity, model generalization, interpretability, and integration of multiple agricultural data sources. Based on these observations, the review discusses emerging directions such as multimodal learning, lightweight deep learning, explainable artificial intelligence, transfer learning, edge computing, and integrated ML-IoT frameworks. Thus, the present review aims to provide a structured and integrated perspective connecting **prediction algorithms, agricultural data sources, enabling technologies, practical limitations, and future research opportunities** for reliable and scalable crop yield prediction.

1.2 Research Questions

To establish a clear connection between the review methodology, literature analysis, identified research gaps, and future directions, the present study addresses the following research questions:

RQ1: What machine learning and deep learning techniques have been applied to crop yield prediction?

RQ2: What datasets, data sources, and input features are commonly used in machine learning- and deep learning-based crop yield prediction?

RQ3: How do classical machine learning, deep learning, hybrid, remote sensing/UAV-based, and IoT-enabled approaches differ in terms of their applicability, strengths, limitations, and computational requirements?

RQ4: What major challenges affect the accuracy, generalization, scalability, and real-world deployment of crop yield prediction models?

RQ5: What emerging research directions can improve the robustness, interpretability, efficiency, and practical applicability of AI-based crop yield prediction systems?

These research questions guide the subsequent analysis of the selected studies. RQ1 is addressed primarily through the literature review and taxonomy of prediction techniques, RQ2 through the analysis of datasets and input features, RQ3 through the comparative analysis of prediction approaches, RQ4 through the discussion of current challenges, and RQ5 through the identification of future research directions.

1.3 Contributions of the Review

The main contributions of this review are as follows:

- **Integrated taxonomy:** The study provides a structured taxonomy of crop yield prediction approaches, covering classical machine learning, deep learning, hybrid

models, remote sensing and UAV-based approaches, and IoT-enabled agricultural systems.

- **Data-centric analysis:** The review examines the major agricultural data sources, datasets, and input features used for crop yield prediction, including climatic, soil, historical yield, satellite, UAV, vegetation-index, and IoT sensor data.
- **Comparative synthesis:** The study compares the major prediction approaches according to their typical applications, strengths, limitations, data requirements, and computational considerations rather than assigning unsupported universal performance rankings.
- **Technology integration:** The review analyzes how remote sensing, UAVs, and IoT technologies complement ML and DL models and contribute to the development of intelligent and precision agriculture systems.
- **Research gap identification:** The study identifies persistent challenges related to dataset availability, data heterogeneity, climatic variability, model generalization, computational complexity, interpretability, and real-time deployment.
- **Future research directions:** Based on the reviewed literature, the study highlights promising directions including multimodal learning, lightweight deep learning, explainable AI, transfer learning, edge computing, and integrated ML-IoT frameworks.
- Overall, the review connects prediction algorithms with their underlying agricultural data sources, technological environments, practical limitations, and emerging research opportunities, thereby providing an integrated perspective for researchers working on AI-based crop yield prediction.

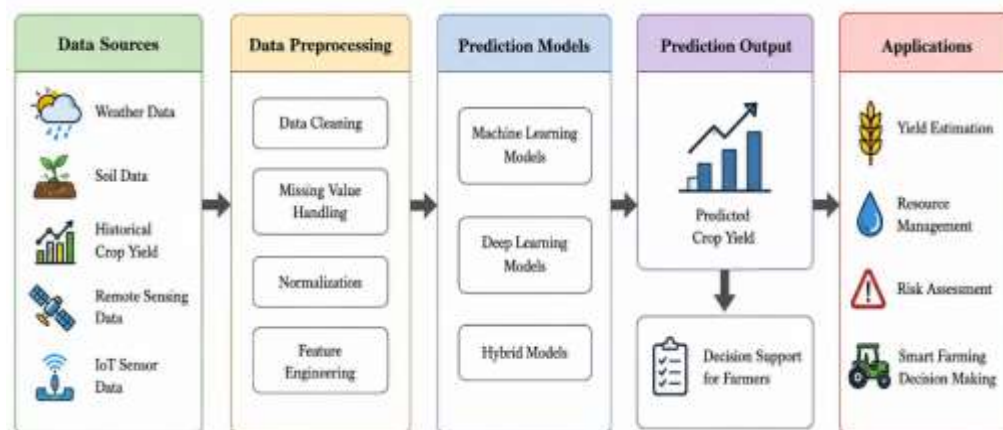


Fig. 1. General framework of crop yield prediction systems.

2. REVIEW METHODOLOGY

This study adopts a Systematic Literature Review (SLR) approach to systematically examine research on machine learning (ML) and deep learning (DL) techniques for crop yield prediction. The review focuses on predictive models, agricultural datasets, input features, precision agriculture technologies, model evaluation practices, challenges, and emerging

research directions. A structured literature-selection procedure was followed to improve the transparency and consistency of the review.

The methodology consists of five main stages: (1) formulation of the search strategy, (2) identification of relevant studies, (3) screening and removal of irrelevant or duplicate records, (4) full-text eligibility assessment, and (5) final selection of studies for detailed analysis. The study-selection procedure was organized according to the PRISMA framework.

2.1 Literature Search Strategy

A structured literature search was conducted using major academic databases and digital libraries, including IEEE Xplore, ScienceDirect, SpringerLink, MDPI, Frontiers, ACM Digital Library, and Google Scholar. These sources were selected to provide coverage of research in machine learning, deep learning, artificial intelligence, remote sensing, precision agriculture, and agricultural information systems.

The search strategy used combinations of keywords related to crop yield prediction, artificial intelligence, precision agriculture, remote sensing, UAVs, and IoT. The principal search queries included:

- ("Crop Yield Prediction" AND "Machine Learning")
- ("Crop Yield Prediction" AND "Deep Learning")
- ("Precision Agriculture" AND "Artificial Intelligence")
- ("Smart Agriculture" AND "IoT")
- ("Remote Sensing" AND "Crop Yield")
- ("UAV" AND "Agriculture")
- ("Machine Learning" AND "Precision Farming")

The search terms were selected to capture studies addressing both predictive modelling techniques and the agricultural technologies used to generate or enhance prediction data. Database-specific search syntax was adapted where necessary while maintaining the same conceptual keywords.

2.2 Time Range and Database Coverage

The review primarily considered studies published between 2020 and 2026. This period was selected to capture recent developments in machine learning, deep learning, remote sensing, UAV-based agricultural monitoring, IoT-enabled precision agriculture, and multimodal agricultural prediction systems.

The selected databases collectively provide coverage of computer science, engineering, agricultural science, remote sensing, artificial intelligence, and interdisciplinary research. Both journal articles and conference papers were considered where they satisfied the predefined selection criteria.

2.3 Inclusion Criteria

Studies were considered eligible for inclusion when they satisfied the following criteria:

- The study was published between 2020 and 2026.
- The study was a peer-reviewed journal article or conference paper.
- The study focused on crop yield prediction, yield estimation, or closely related agricultural yield forecasting.
- Machine learning, deep learning, or artificial intelligence techniques were applied as part of the prediction methodology.
- The study provided sufficient information about the dataset, prediction approach, or experimental methodology.
- The study included quantitative evaluation or performance analysis.
- Studies involving precision agriculture technologies such as remote sensing, UAVs, satellite imagery, or IoT were included when they contributed to crop yield prediction.
- The publication was available in English.

2.4 Exclusion Criteria

Studies were excluded when they met one or more of the following conditions:

- Duplicate records identified across the searched databases.
- Studies unrelated to crop yield prediction.
- Studies focused exclusively on crop disease, pest, or plant classification without a crop-yield prediction component.
- Articles without sufficient methodological or experimental information.
- Studies without quantitative or experimental evaluation.
- Non-peer-reviewed editorials, short abstracts, and other publications that did not provide sufficient research details.
- Studies outside the defined publication period.
- Publications not available in English.

2.5 Study Selection Process

The study-selection process followed the PRISMA framework and consisted of identification, screening, eligibility assessment, and final inclusion stages.

During the identification stage, a total of 95 research records were collected from the selected academic databases using the predefined search queries. Following the initial identification, duplicate and irrelevant records were removed. This resulted in 68 records being retained for the screening stage.

During screening, the titles and abstracts of the 68 records were examined to determine their relevance to machine learning- and deep learning-based crop yield prediction. Following this assessment, 32 studies were shortlisted for full-text evaluation.

The full texts of the 32 shortlisted studies were subsequently assessed against the predefined inclusion and exclusion criteria. The assessment considered the relevance of the research objective, prediction methodology, dataset usage, model implementation, and availability of experimental or quantitative evaluation.

Finally, 13 unique research studies were selected for detailed analysis in the present review. These studies collectively cover machine learning and deep learning approaches, hybrid prediction models, remote sensing, UAV-based agricultural monitoring, IoT-enabled precision agriculture, and emerging directions in crop yield prediction.

The complete study-selection process is illustrated in Fig. 2 using the PRISMA framework.

2.6 Quality Assessment Criteria

The selected studies were assessed using a set of quality-related criteria to ensure that the studies included in the detailed synthesis provided meaningful methodological and research information. The assessment considered the following characteristics:

- Clarity of the research objective.
- Clarity of the methodology and prediction approach.
- Quality and diversity of the dataset.
- Description of the machine learning or deep learning model.
- Reporting of performance evaluation metrics.
- Experimental validation or performance analysis.
- Practical relevance to crop yield prediction and precision agriculture.

These criteria were used to support consistent interpretation and comparison of the selected literature. Studies with insufficient methodological information or inadequate experimental evidence were excluded during the eligibility assessment.

2.7 Study Selection and Bias Reduction

Several measures were adopted to reduce potential selection bias. Multiple academic databases were considered rather than relying on a single source, allowing research from different disciplines and publication venues to be identified. Predefined search terms and eligibility criteria were used throughout the selection process to maintain consistency.

Duplicate records were removed before detailed screening, and titles and abstracts were assessed before full-text evaluation. The same inclusion and exclusion criteria were applied during the eligibility stage. The final set of 13 studies was then examined systematically according to research methodology, datasets, prediction models, evaluation practices, technological integration, limitations, and future research opportunities.

The selected studies provide the basis for the taxonomy, comparative analysis, dataset analysis, challenges, and future research directions presented in the subsequent sections.

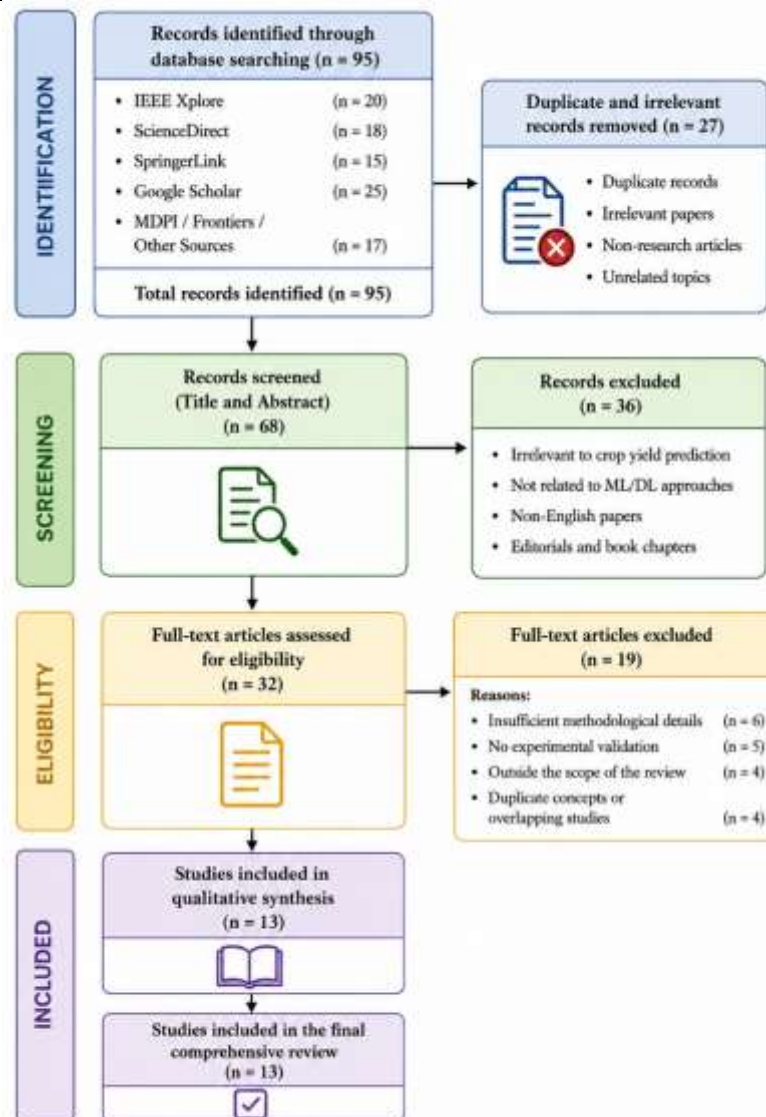


Fig. 2. PRISMA flow diagram illustrating the study selection process for the systematic literature review on machine learning and deep learning approaches for crop yield prediction.

3. LITERATURE REVIEW

3.1 Recent Primary Research in Crop Yield Prediction

Recent primary studies demonstrate a shift from conventional single-source prediction models toward data-intensive, multimodal, and increasingly explainable crop yield prediction systems. These studies provide experimental evidence that complements the findings reported in earlier review-based research.

De Clercq and Mahdi investigated district-level rice yield prediction in India using 20 years of climate, satellite, and historical yield data across 247 rice-producing districts. Nineteen machine learning models, including CatBoost, LightGBM, Orthogonal Matching Pursuit, and Extremely Randomized Trees, were evaluated. The study reported out-of-sample performance of up to ($R^2=0.82$), with MAE and MAPE values of 0.29 and 0.16, respectively. The study also demonstrated that model reliability can vary across geographical regions, highlighting the importance of spatial variability and regional generalization in agricultural prediction systems.

Khan et al. investigated the use of MODIS-derived gross primary production data and deep transfer learning for crop yield prediction in the US Corn Belt. Their study evaluated deep learning models using time-series remote-sensing information for county-level corn and soybean yield prediction. The findings demonstrate the potential of combining vegetation productivity information with deep learning and transfer learning to support early crop yield estimation.

Toledo et al. examined multimodal remote sensing and deep learning for yield prediction in plant breeding experiments. Their work investigated attention-based deep learning architectures, including stacked LSTM models, for maize grain yield prediction. The study emphasizes the value of combining heterogeneous remote-sensing information with temporal deep learning and also highlights interpretability as an important requirement for AI-based agricultural prediction.

Cheng et al. proposed a hybrid approach for in-season wheat yield forecasting by integrating UAV remote-sensing data with the AquaCrop crop model and ensemble learning. The study combined crop-model simulation variables with UAV-derived information and demonstrated the potential of integrating process-based crop models and data-driven learning for field-scale yield forecasting. The work illustrates how hybrid systems can combine complementary information sources rather than relying exclusively on either remote sensing or conventional crop modelling.

Joshi et al. developed an explainable Bidirectional Long Short-Term Memory (Bi-LSTM) model for regional winter wheat yield prediction. The study incorporated explainability techniques, including SHAP, LIME, and Integrated Gradients, to improve the interpretability of deep learning predictions. This research demonstrates the increasing importance of explainable artificial intelligence in crop yield prediction, particularly where model transparency is required for agricultural decision-making.

Another recent study integrated deep learning, remote sensing, and process-based crop modelling for rice yield prediction. Four deep learning architectures—feed-forward neural networks, LSTM, GRU, and Bi-LSTM—were evaluated. The study reported strong predictive performance but also observed variations when the models were applied to regional datasets that were not represented during training. This finding reinforces the importance of diverse training datasets and cross-regional validation for improving model robustness.

UAV-based research has also demonstrated the usefulness of high-resolution aerial information for crop yield prediction. A study on wheat phenotyping used multispectral UAV imagery and machine learning regression approaches to predict yield under water-stress conditions. Four regression approaches were evaluated, including linear regression, support vector machine, random forest, and a deep learning model. The results demonstrated that UAV-derived vegetation and stress-related features can support yield prediction and large-scale crop phenotyping.

Recent research has further examined the effect of validation strategy on model transferability. A soybean yield prediction study using UAV-based remote sensing showed that random data splitting can produce optimistic estimates of model performance, while spatially aware cross-validation can provide a more realistic assessment of independent-field prediction. This finding is particularly relevant to the generalization challenge because high predictive performance within a dataset does not necessarily imply reliable performance in geographically independent fields.

Similarly, Sentinel-2-based research has investigated early-season rice yield prediction for precision nitrogen management. The study evaluated upscaling and downscaling strategies and found that machine learning approaches could support within-field yield variability prediction, demonstrating the potential of remote sensing and ML integration for site-specific agricultural management.

Overall, these recent primary studies demonstrate several important trends. First, remote sensing and UAV data are increasingly being combined with ML and DL models rather than being used only as independent data sources. Second, hybrid systems are integrating process-based crop models with data-driven algorithms. Third, transferability, spatial validation, and model generalization are receiving greater attention. Finally, explainability and multimodal learning are emerging as important components of practical AI-based crop yield prediction systems. These developments complement the broader findings of earlier review studies and strengthen the evidence base for the research gaps and future directions identified in this review.

Table 2. Summary of Selected Studies in Crop Yield Prediction

Author Study /	Study Type	Method Technology /	Main Contribution	Key Limitation
Jabed and Murad [1]	Review	ML and DL	Comprehensive analysis of AI techniques for crop yield prediction	Generalization challenges across crops and regions
Manjunath and Palayyan [2]	Primary Research	Hybrid ML	Developed a hybrid machine learning framework for crop yield prediction	Performance depends on dataset characteristics
Anand and Jhajharia [3]	Review	ML and DL	Compared machine learning and deep learning approaches for yield prediction	Limited development of a generalized prediction framework
Hussain et al. [4]	Review	UAV + ML	Examined UAV-based approaches for crop-yield prediction	Limited availability and diversity of UAV datasets
Cravero et al. [5]	Review	ML and Big Data	Examined intelligent agricultural ecosystems and AI-based agricultural applications	Infrastructure, integration, and scalability challenges
Mohyuddin et al. [6]	Review	ML for Precision	Reviewed machine learning applications in smart and precision	Data heterogeneity and implementation

		Agriculture	agriculture	challenges
Saha et al. [7]	Review	Precision Agriculture + Remote Sensing	Examined the role of precision agriculture and remote sensing in crop-yield prediction	Need for more integrated and hybrid prediction systems
Sharma et al. [8]	Review	ML for Precision Agriculture	Provided a broad review of machine learning applications in precision agriculture	Computational and implementation challenges
Kariya and Mudholkar [9]	Review	ML and DL	Reviewed machine learning models used to improve crop-yield prediction	Environmental and regional variability
Mustard Crop Study [10]	Primary Research	ANN and KNN	Investigated soil-based crop-yield prediction for mustard	Crop-specific and limited generalization
Guan et al. [11]	Primary Research	RF + UAV Multispectral Data	Used Random Forest with UAV and geophysical data for agricultural soil-property prediction	Site dependency and sensor-data requirements
Sudha and Loret [12]	Review	ML + IoT	Reviewed ML-based IoT applications for precision agricultural monitoring	Cybersecurity, scalability, and sensor reliability
Parashar et al. [13]	Review	Integrated AI Models	Reviewed developments and challenges in crop-yield prediction models	Need for robust multimodal and generalized models

Source: Compiled from the selected studies included in the systematic literature review.

4. TAXONOMY OF CROP YIELD PREDICTION TECHNIQUES

Crop yield prediction techniques have evolved significantly from traditional statistical methods to advanced Artificial Intelligence (AI)-based approaches. Modern prediction systems utilize Machine Learning (ML), Deep Learning (DL), Remote Sensing technologies, Unmanned Aerial Vehicles (UAVs), and Internet of Things (IoT) frameworks to improve prediction accuracy and support precision agriculture. Based on the selected literature, crop yield prediction techniques can be categorized into the following major groups.

4.1 Classical Machine Learning Approaches

Traditional machine learning algorithms have been extensively used for crop yield prediction because of their simplicity, interpretability, and ability to process structured agricultural data. Commonly used algorithms include Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Linear Regression (LR), and XGBoost.

These methods use climatic, soil, rainfall, humidity, temperature, and crop-related features to estimate crop production. Random Forest and XGBoost are widely used because of their ability to model nonlinear relationships and handle large feature sets [1], [2], [9]. However, classical ML methods often require manual feature engineering and may struggle with highly complex and nonlinear agricultural data.

4.2 Deep Learning-Based Approaches

Deep learning approaches are particularly applicable to large and complex agricultural datasets because they can automatically learn spatial and temporal representations. However, their effectiveness depends on dataset size, data quality, model architecture, and available computational resources.

CNN models are effective for processing satellite and image-based agricultural data, while LSTM networks are suitable for time-series prediction using weather and historical yield data [1], [3].

4.3 Hybrid Machine Learning and Deep Learning Models

Hybrid approaches combine multiple algorithms to leverage their individual strengths and improve prediction performance. Examples include CNN-LSTM, ensemble learning methods, and hybrid ML frameworks integrating Decision Trees, Random Forest, and XGBoost models.

These architectures can combine spatial feature extraction with temporal sequence modelling and may be useful when both spatial and temporal information are available [2], [13]. However, hybrid models generally involve greater architectural complexity and may require additional computational resources and longer training times.

4.4 Remote Sensing and UAV-Based Approaches

Remote sensing technologies and Unmanned Aerial Vehicles (UAVs) have become important components of precision agriculture. Satellite imagery, multispectral cameras, and UAV platforms provide high-resolution information about vegetation health, soil moisture, crop stress, and environmental conditions.

Vegetation indices such as NDVI, EVI, NDWI, and LAI are commonly integrated with ML and DL models to improve crop yield estimation [4], [7], [11]. UAV-based monitoring enables real-time data collection over large agricultural fields; however, the effectiveness of these systems depends on data quality and sensor availability.

4.5 IoT-Based Precision Agriculture Systems

The Internet of Things (IoT) has transformed modern agriculture by enabling continuous monitoring through smart sensors and connected devices. IoT sensors collect real-time information related to soil moisture, temperature, humidity, rainfall, nutrient levels, and irrigation status.

Machine learning and deep learning algorithms process these data streams to support intelligent decision-making and accurate crop yield prediction [6], [12]. IoT-enabled precision agriculture improves resource utilization and farming efficiency, but challenges such as data heterogeneity, sensor reliability, and cybersecurity remain important research issues.

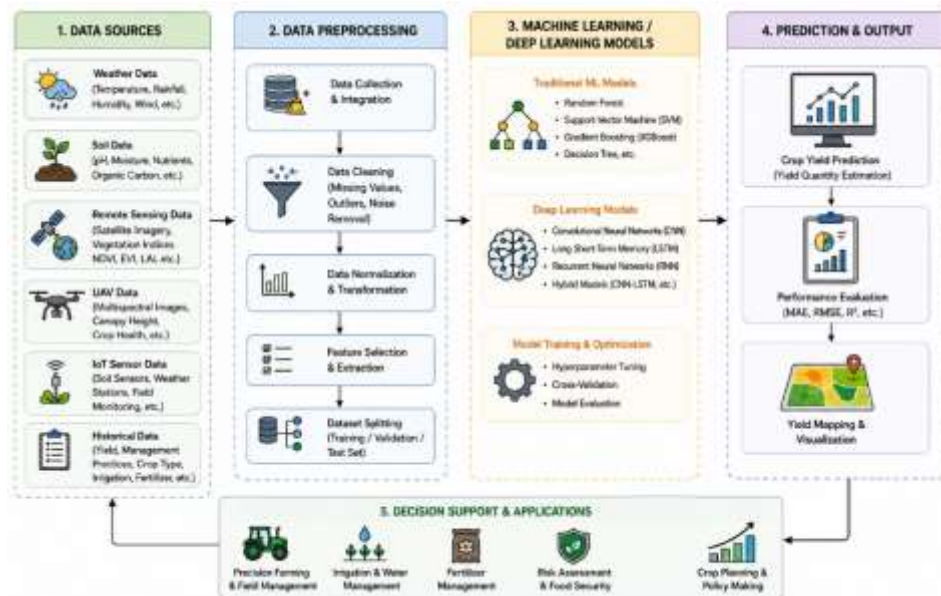


Fig. 3. Architecture of a crop yield prediction system using machine learning and deep learning approaches.

Fig. 3. Architecture of a crop yield prediction system using machine learning and deep learning approaches.

The proposed architecture illustrates the overall workflow of a crop yield prediction system. Initially, data are collected from multiple sources, including weather records, soil properties, remote sensing images, UAV data, IoT sensors, and historical crop information. The collected data undergo preprocessing steps such as data cleaning, normalization, feature extraction, and dataset splitting. Subsequently, Machine Learning and Deep Learning models are trained and optimized to predict crop yield. The prediction results are evaluated using standard performance metrics and utilized for various precision agriculture applications, including irrigation management, fertilizer optimization, crop planning, and decision support systems.

The taxonomy illustrates the evolution of crop yield prediction techniques from traditional machine learning methods to advanced intelligent agricultural ecosystems that integrate AI, remote sensing, UAVs, and IoT technologies.

5. COMPARATIVE ANALYSIS OF CROP YIELD PREDICTION TECHNIQUES

Crop yield prediction techniques have evolved from traditional machine learning algorithms to advanced deep learning and hybrid models integrated with precision agriculture technologies. Early approaches primarily relied on statistical methods and classical machine learning algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Networks (ANN). These methods effectively process structured agricultural data but often require manual feature engineering and have limited capability to model complex nonlinear relationships [2], [10].

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNNs), have improved

prediction accuracy by automatically extracting spatial and temporal features from large agricultural datasets [1], [3]. CNN models are particularly suitable for satellite and image-based analysis, whereas LSTM networks are suitable for time-series weather and crop-yield forecasting. Hybrid models such as CNN-LSTM and ensemble learning approaches combine the strengths of multiple algorithms to improve prediction performance under varying environmental conditions [2], [13]. However, these models require greater computational resources and longer training times.

The integration of Remote Sensing, Unmanned Aerial Vehicles (UAVs), and Internet of Things (IoT) technologies has further enhanced crop yield prediction by providing real-time information about soil conditions, vegetation health, and climatic factors [4], [7], [11], [12]. Despite these advancements, challenges related to data heterogeneity, computational complexity, and model generalization continue to limit large-scale deployment.

Overall, there is a trade-off between prediction accuracy, computational requirements, and real-time applicability among different crop yield prediction approaches.

Table 3. Comparison of Crop Yield Prediction Techniques

Technique	Advantages	References
Machine Learning (RF, SVM, DT, ANN)	Simple implementation and suitability for structured datasets	[2], [9], [10]
Deep Learning (CNN, LSTM, DNN)	Automatic feature extraction from complex spatial and temporal data	[1], [3], [13]
Hybrid Models (CNN-LSTM, Ensemble)	Combines spatial and temporal learning	[2], [13]
Remote Sensing and UAV-Based Models	Provides high-resolution field monitoring and vegetation analysis	[4], [7], [11]
IoT-Based Precision Agriculture	Enables real-time monitoring and intelligent decision-making	[6], [12]

Machine learning approaches are suitable for applications involving structured agricultural datasets with moderate computational resources. Deep learning techniques perform better when large-scale datasets and image-based information are available. Hybrid models provide improved prediction accuracy but increase computational complexity. Remote sensing, UAVs, and IoT technologies strengthen prediction systems by supplying real-time environmental and crop-related information, making them essential components of modern precision agriculture.

6. PERFORMANCE EVALUATION OF CROP YIELD PREDICTION MODELS

Performance evaluation plays a critical role in assessing the effectiveness of crop yield prediction models. Various Machine Learning (ML) and Deep Learning (DL) techniques are evaluated using different statistical measures to determine their prediction capability under diverse agricultural conditions. These performance metrics help researchers compare different models and identify the most suitable approach for precision agriculture applications.

Early crop yield prediction systems primarily relied on traditional machine learning algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Networks (ANN). Although these models achieved satisfactory results, their effectiveness was highly dependent on the quality of the input data and feature engineering techniques. With the advancement of deep learning, architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid models have demonstrated improved prediction accuracy by automatically extracting complex spatial and temporal features from agricultural datasets [1], [3], [13].

6.1 Evaluation Metrics

Several evaluation metrics are commonly employed to measure the performance of crop yield prediction models. The **Mean Absolute Error (MAE)** calculates the average absolute difference between actual and predicted crop yield values and is expressed as

$$\text{MAE} = (1/n) \times \sum |Y_i - \hat{Y}_i|$$

where Y_i represents the actual crop yield, $\hat{Y}_i$ denotes the predicted crop yield, and n is the total number of observations.

Another widely used metric is the **Root Mean Square Error (RMSE)**, which measures the standard deviation of prediction errors and is given by

$$\text{RMSE} = \sqrt{(1/n) \times \sum (Y_i - \hat{Y}_i)^2}$$

A lower RMSE value indicates better prediction performance.

The **Mean Absolute Percentage Error (MAPE)** expresses the prediction error as a percentage and can be calculated using

$$\text{MAPE} = (100/n) \times \sum |(Y_i - \hat{Y}_i) / Y_i|$$

Lower MAPE values indicate higher forecasting accuracy.

The **Coefficient of Determination (R^2)** is commonly used to evaluate how well a prediction model explains the variability in crop yield data. It is defined as

$$R^2 = 1 - [\sum (Y_i - \hat{Y}_i)^2 / \sum (Y_i - \bar{Y})^2]$$

where $\bar{Y}$ is the mean of the observed crop yield values. An R^2 value closer to 1 indicates better model performance.

These statistical measures provide a comprehensive assessment of prediction accuracy and are extensively used for comparing machine learning and deep learning models in agricultural applications.

6.2 Comparative Analysis of Crop Yield Prediction Models

The performance of crop yield prediction models depends strongly on the characteristics of the dataset, crop type, geographical region, input variables, prediction horizon, and evaluation methodology. Therefore, direct performance ranking of different algorithms using generalized labels such as “High,” “Very High,” or “Excellent” may be misleading when the models are evaluated under different experimental conditions. Instead, this review compares the approaches according to their typical data requirements, application contexts, strengths, limitations, and computational considerations.

Table 4. Comparative Analysis of Crop Yield Prediction Approaches

Model Approach	Typical Data Type	Typical Application	Major Strengths	Major Limitations
Linear Regression (LR)	Historical yield, climate, soil and agricultural variables	Baseline yield estimation	Simple, computationally efficient, and easy to interpret	Limited ability to represent complex nonlinear relationships
Support Vector Machine (SVM)	Climate, soil and structured agricultural data	Yield prediction with small- to medium-sized datasets	Effective in high-dimensional feature spaces and flexible through kernel functions	Sensitive to kernel and hyperparameter selection; computational cost can increase with dataset size
Decision Tree (DT)	Structured agricultural data	Rule-based yield prediction and feature analysis	Easy to interpret and capable of modelling nonlinear relationships	Susceptible to overfitting without appropriate pruning or regularization
Random Forest (RF)	Soil, weather, crop and environmental variables	Structured agricultural yield prediction	Handles nonlinear relationships, multiple features, and provides relatively robust predictions	Requires greater memory and computational resources than individual decision trees
Artificial Neural Network (ANN)	Climate, soil, crop and environmental variables	Nonlinear agricultural prediction	Learns complex relationships between input variables and yield	Requires appropriate parameter tuning and sufficient training data
Convolutional Neural Network (CNN)	Satellite, UAV and other image-based agricultural data	Spatial crop monitoring and image-based yield estimation	Automatically extracts spatial features from agricultural imagery	Requires substantial image data and computational resources
Long Short-Term Memory (LSTM)	Weather, historical yield and other time-series data	Temporal yield forecasting	Capable of learning temporal dependencies in sequential agricultural data	Training can be computationally demanding and performance depends on sequence quality

				and length
CNN-LSTM	Image and time-series or multimodal agricultural data	Spatial-temporal yield prediction	Combines spatial feature extraction with temporal modelling	More complex architecture and higher computational requirements
Ensemble / Hybrid ML Models	Heterogeneous agricultural datasets	Integrated yield prediction	Combines complementary models and can improve robustness	Increased model complexity, training cost, and implementation requirements
Remote Sensing + ML/DL	Satellite and multispectral imagery with agricultural variables	Large-area crop monitoring and yield estimation	Provides spatially distributed information over agricultural fields	Performance depends on image quality, spatial/temporal resolution, and availability of appropriate remote-sensing data
UAV + ML/DL	UAV RGB, multispectral or hyperspectral imagery	Field-level crop monitoring and yield estimation	Provides high-resolution field observations and detailed crop information	Data acquisition can require specialized equipment, flight planning, and preprocessing
IoT-based ML/DL	Soil, weather and environmental sensor data	Real-time agricultural monitoring and prediction	Supports continuous data collection and decision support	Sensor reliability, data synchronization, connectivity, and cybersecurity remain important concerns

Source: Synthesized from the studies reviewed in this paper.

The comparison shows that no single prediction approach is universally optimal for all crop-yield prediction scenarios. Classical machine learning methods are particularly suitable for structured agricultural datasets containing climatic, soil, historical yield, and crop-related variables. Among these approaches, tree-based and ensemble methods can model nonlinear relationships and interactions among multiple agricultural variables without requiring the same degree of manual feature transformation as simpler statistical models.

Deep learning approaches are particularly relevant when large-scale, complex, or image-based datasets are available. CNNs are suitable for extracting spatial information from

satellite and UAV imagery, whereas LSTM networks are designed to capture temporal dependencies in sequential weather and yield data. Hybrid architectures such as CNN-LSTM can combine these capabilities when both spatial and temporal information are available. However, their increased architectural complexity can result in greater computational requirements.

Remote sensing and UAV-based approaches provide additional spatial information that can improve crop monitoring and yield estimation when integrated with ML or DL models. Similarly, IoT-based systems provide continuous measurements of environmental and soil conditions, enabling near-real-time agricultural monitoring. However, these approaches introduce additional challenges related to sensor reliability, data quality, communication, synchronization, and infrastructure.

Consequently, model selection should be determined by the characteristics of the prediction problem rather than by assigning a universal performance category to each algorithm. Factors such as dataset size, data modality, spatial and temporal resolution, computational resources, prediction objective, interpretability requirements, and deployment environment should be considered when selecting an appropriate crop yield prediction approach.

6.3 Discussion

The comparative analysis indicates that the suitability of a crop yield prediction model is strongly dependent on the type and availability of agricultural data. Classical ML approaches remain useful for structured datasets, particularly when climate, soil, historical yield, and crop-related variables are available. Deep learning approaches provide greater flexibility for complex spatial and temporal information, particularly when satellite, UAV, or long-term time-series data are available.

Hybrid and ensemble approaches provide opportunities to combine complementary learning capabilities, but their increased complexity may make implementation and deployment more challenging. Similarly, the integration of remote sensing, UAV, and IoT technologies can provide richer agricultural information but also introduces additional requirements related to data acquisition, preprocessing, infrastructure, and data integration.

An important observation from the reviewed literature is that reported performance values should be interpreted within the experimental context of each study. Differences in crop type, geographical location, dataset size, prediction target, input variables, validation strategy, and evaluation metrics can substantially influence reported results. Therefore, numerical performance values from different studies should not be interpreted as directly comparable unless the studies use sufficiently similar experimental conditions.

The evidence synthesized in this review therefore supports a context-dependent approach to model selection. Future crop yield prediction systems should aim to balance predictive capability with computational efficiency, interpretability, data availability, scalability, and practical deployment requirements.

7. DATASETS USED IN CROP YIELD PREDICTION SYSTEMS

Datasets play a vital role in the development and evaluation of crop yield prediction systems. The performance of Machine Learning (ML) and Deep Learning (DL) models largely depends on the quality, quantity, and diversity of agricultural data available for training and testing. Accurate crop yield prediction requires the integration of multiple data sources,

including historical crop records, climatic conditions, soil characteristics, remote sensing imagery, and real-time sensor data.

Agricultural datasets can generally be classified into historical agricultural datasets, weather datasets, soil datasets, satellite and UAV datasets, and IoT sensor datasets. Historical datasets provide information about crop production and harvested area over several years, while climatic datasets contain environmental parameters such as rainfall, temperature, humidity, solar radiation, and wind speed. Soil datasets include information regarding soil moisture, pH, nitrogen, phosphorus, potassium, and organic carbon content. Recent developments in precision agriculture have also encouraged the use of multispectral satellite imagery, UAV-based aerial data, and IoT-enabled smart sensors for crop monitoring and yield estimation [4], [7], [11], [12].

Deep learning techniques require large and diverse datasets to effectively learn spatial and temporal patterns associated with crop growth. The integration of multimodal datasets can provide complementary information and may improve crop yield prediction when the data sources are appropriately aligned and validated.

7.1 Publicly Available Datasets for Crop Yield Prediction

Several publicly available datasets are widely used for training and evaluating crop yield prediction models. These datasets vary in terms of data type, scale, geographical coverage, and agricultural applications.

Table 5. Commonly Used Datasets for Crop Yield Prediction

Dataset	Data Type	Description
FAOSTAT	Agricultural Statistics	Global crop production and harvested area records
USDA-NASS	Historical Agricultural Data	Crop yield and acreage information
NASA POWER	Weather Dataset	Temperature, rainfall, humidity, and solar radiation
MODIS	Satellite Imagery	Vegetation indices and land surface information
Sentinel-2	Multispectral Satellite Data	Crop monitoring and precision agriculture
UAV Multispectral Dataset	Aerial Image Data	Crop health assessment and yield estimation
IoT Sensor Dataset	Real-time Sensor Data	Soil moisture, temperature, humidity, and nutrient monitoring

7.2 Characteristics of Crop Yield Prediction Datasets

Crop yield prediction datasets differ in several important aspects, including dataset size, temporal resolution, spatial coverage, and feature diversity. Large datasets collected over multiple years generally improve the predictive capability of machine learning and deep learning models.

Weather datasets provide environmental information that influences crop growth, whereas soil datasets describe nutrient availability and moisture conditions. Remote sensing and UAV datasets capture vegetation characteristics through multispectral images and vegetation indices such as NDVI, EVI, and LAI. IoT sensor datasets offer continuous monitoring of agricultural fields, enabling real-time prediction and precision farming applications.

The combination of these heterogeneous data sources enables intelligent systems to model complex interactions between environmental and agricultural factors more effectively.

7.3 Input Features Used in Crop Yield Prediction

Various input parameters are considered while developing crop yield prediction models. The most used features include:

- Rainfall
- Temperature
- Relative Humidity
- Soil Moisture
- Soil pH
- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)
- Solar Radiation
- Vegetation Indices (NDVI, EVI)
- Historical Crop Yield
- Irrigation Data
- Weather Forecast Data

The selection of appropriate features significantly affects the accuracy and reliability of prediction models.

7.4 Challenges Related to Agricultural Datasets

Despite the availability of various agricultural datasets, several challenges remain in developing robust crop yield prediction systems.

One of the major challenges is the limited availability of high-quality and labelled datasets, especially for specific crops and geographical regions. Agricultural data are often collected from different sources and formats, leading to data heterogeneity and integration difficulties.

Another challenge is the presence of missing and incomplete records, which can negatively affect model performance. Many publicly available datasets are also region-specific and may not generalize well to different climatic and agricultural conditions.

Furthermore, collecting UAV imagery and IoT sensor data requires significant infrastructure investment, making large-scale deployment difficult. Variations in weather conditions,

farming practices, and soil properties further complicate the development of generalized prediction models.

Table 6: Summary of Key Dataset Challenges

Challenge	Impact on Crop Yield Prediction
Limited datasets	Reduces model training capability
Data heterogeneity	Complicates data integration
Missing values	Decreases prediction accuracy
Region-specific datasets	Limits model generalization
High data collection cost	Restricts large-scale implementation
Lack of benchmark datasets	Makes model comparison difficult

To address these issues, future research should focus on developing standardized, multimodal, and large-scale agricultural datasets that integrate climatic, soil, remote sensing, UAV, and IoT information.

Dataset Name	Crop(s)	Data Type / Features	Source	Spatial / Temporal Resolution	Usage in Crop Yield Prediction
 NASA POWER	Multiple (corn, wheat, rice, soybean, etc.)	Weather data (temperature, rainfall, humidity, wind speed, solar radiation, etc.)	NASA Prediction of Worldwide Energy Resources	0.5° spatial Daily	Provides long-term historical weather data for crop yield modeling and analysis.
 FAO FAOSTAT	Multiple (global crops)	Crop production, yield, harvested area, land use, inputs, population, etc.	Food and Agriculture Organization (FAO)	Country level Annual	Widely used for statistical analysis and baseline yield estimation.
 MODIS (MOD13Q1)	Multiple	Vegetation indices (NDVI, EVI), land surface temperature, reflectance	NASA MODIS Satellite	250 m spatial 16-day	Useful for monitoring crop health, vegetation, and yield estimation.
 Landsat 8 Surface Reflectance	Multiple	Surface reflectance, NDVI, land cover, temperature	USGS Earth Explorer	30 m spatial 16-day	High-resolution imagery for crop monitoring and yield prediction.
 UAV-based Datasets	Wheat, Rice, Maize, Cotton, Soybean, etc.	Multispectral images, conventional images, vegetation indices, canopy height	Various (Field Experiments, Research Institutions)	cm-level spatial Variable (Flight-based)	Provides high-resolution field data for crop growth monitoring and yield prediction.
 IoT Sensor Datasets	Multiple	Soil moisture, temperature, pH, humidity, rainfall, nutrient levels, etc.	Open IoT Platforms, Research Projects	Point-level Real-time	Used for real-time monitoring and data-driven yield prediction models.
 AgMIP Datasets	Wheat, Maize, Rice, Soybean, etc.	Climate, soil, crop management, simulation outputs, yield data	AgMIP Projects	Variable (Regional to Global)	Provides harmonized multi-model datasets for crop yield simulations and analysis.
 National Agricultural Datasets (e.g., India, USA, China)	Major crops (region-specific)	Weather, soil, crop production, irrigation, fertilizer use, yield data	National Agricultural Departments, Government Portals	Region specific Annual / Seasonal	Important for region-specific yield prediction and agricultural planning.

Fig. 4. Commonly used datasets for crop yield prediction systems.

8. CHALLENGES IN CROP YIELD PREDICTION

Despite advances in machine learning, deep learning, remote sensing, UAVs, and IoT-enabled agricultural systems, several challenges continue to affect the reliability and practical deployment of crop yield prediction models. The reviewed literature indicates that these challenges are primarily associated with data availability and quality, heterogeneous data

integration, environmental variability, computational requirements, model generalization, and the integration of emerging sensing technologies.

8.1 Limited Availability of High-Quality Datasets

The availability of sufficiently large, diverse, and well-annotated agricultural datasets remains a major limitation. Many existing studies use datasets associated with particular crops, locations, or growing conditions, which can restrict the applicability of the resulting models to other agricultural environments. This limitation is particularly relevant for deep learning approaches because their effectiveness generally depends on the availability of adequate training data.

8.2 Data Heterogeneity

Crop yield prediction increasingly relies on multiple sources, including weather observations, soil measurements, satellite imagery, UAV data, and IoT sensors. These sources can differ considerably in format, spatial and temporal resolution, measurement quality, and sampling frequency. Integrating such heterogeneous information requires appropriate preprocessing and data-fusion strategies and remains a challenge for unified prediction systems.

8.3 Climatic and Environmental Variability

Crop yield is affected by environmental conditions such as rainfall, temperature, drought, flooding, and pest-related disturbances. Models trained using historical observations may therefore experience reduced reliability when environmental conditions differ substantially from those represented in the training data. This issue highlights the importance of evaluating prediction models under diverse climatic and agricultural conditions.

8.4 Computational Complexity

Advanced deep learning and hybrid architectures can require considerable computational resources for model training and inference. Such requirements may limit their deployment in agricultural environments where computing resources, connectivity, or energy availability are constrained. Therefore, practical crop yield prediction systems must consider the balance between model complexity and deployment requirements.

8.5 Limited Model Generalization

A further challenge is the transferability of prediction models across crops, geographical regions, and environmental conditions. Models developed using a particular crop or region may not maintain the same level of reliability when applied to previously unseen conditions. Consequently, cross-regional validation and evaluation using independent datasets are important for determining whether a model can be reliably deployed beyond its original training environment.

8.6 Integration of Emerging Technologies

Remote sensing, UAVs, and IoT technologies can provide valuable agricultural information, but integrating these technologies into a unified prediction framework introduces additional technical requirements. Sensor calibration, communication reliability, data synchronization, and the management of heterogeneous data streams must be considered when developing integrated agricultural prediction systems.

Table 7. Summary of Key Challenges

Challenge	Primary Impact
Limited availability of high-quality datasets	Restricts model training and broader validation
Data heterogeneity	Increases preprocessing and data-integration requirements
Climatic and environmental variability	Can reduce prediction reliability under unseen conditions
Computational complexity	Limits deployment on resource-constrained platforms
Limited model generalization	Restricts transferability across crops and regions
Integration of UAV, remote sensing, and IoT data	Increases system and data-management complexity

The challenges identified above indicate that improving crop yield prediction requires attention not only to model architecture but also to dataset diversity, validation strategy, computational efficiency, and reliable integration of multiple agricultural data sources.

9. FUTURE RESEARCH DIRECTIONS

The challenges identified in the reviewed literature suggest several directions for future research. Rather than focusing exclusively on increasing predictive performance, future studies should consider data diversity, generalization, computational efficiency, interpretability, and practical deployment.

9.1 Large-Scale and Standardized Agricultural Datasets

Future research should prioritize the development of larger and more diverse agricultural datasets covering multiple crops, geographical regions, climatic conditions, and growing seasons. Integrating climatic, soil, historical yield, remote sensing, UAV, and IoT information could provide a stronger basis for developing and validating generalizable prediction models.

9.2 Lightweight Models and Efficient Deployment

The computational requirements of complex deep learning and hybrid architectures create challenges for real-world agricultural deployment. Future research should therefore investigate lightweight architectures, model compression, efficient training strategies, and edge-computing solutions. These approaches could enable prediction systems to operate more effectively on resource-constrained agricultural devices.

9.3 Multimodal Data Integration

Future crop yield prediction systems can benefit from the joint use of weather, soil, satellite, UAV, and IoT information. Research should investigate robust data-fusion and multimodal learning strategies capable of handling differences in data format, spatial resolution, temporal frequency, and data quality. Such approaches could provide a more comprehensive representation of crop and environmental conditions.

9.4 Explainable Artificial Intelligence

Interpretability remains important when AI-based predictions are used to support agricultural decisions. Future studies should therefore investigate explainable AI techniques that can identify the variables and patterns influencing model predictions. Improved interpretability could help agricultural researchers and practitioners better understand model outputs and assess their reliability.

9.5 Transfer Learning and Cross-Regional Validation

The limited generalization of models across crops and geographical regions indicates the need for stronger transfer-learning and cross-regional evaluation strategies. Transfer learning can potentially reduce the amount of training data required when adapting an existing model to a new crop or geographical environment. Future studies should also employ independent regional datasets to evaluate whether improvements observed during training are maintained under different agricultural conditions.

9.6 Integration of AI, IoT, UAV, and Remote Sensing

Future intelligent agricultural systems are likely to combine AI models with continuous sensor observations, UAV imagery, satellite data, and other remote-sensing sources. IoT and edge-computing technologies can support continuous agricultural monitoring, while ML and DL models can transform these observations into prediction and decision-support information. However, future research should address data synchronization, communication reliability, system scalability, and interoperability when designing such integrated systems.

Overall, future research should move toward data-diverse, interpretable, computationally efficient, and transferable prediction systems rather than focusing only on increasing model complexity.



Fig. 5. Future research roadmap for crop yield prediction.

10. CONCLUSION

Crop yield prediction is an important application of artificial intelligence in precision agriculture because reliable yield estimates can support agricultural planning, resource management, and decision-making. This review systematically examined machine learning

and deep learning approaches used for crop yield prediction, together with the datasets and technologies that support these approaches.

The reviewed literature was organized into classical machine learning, deep learning, hybrid approaches, remote sensing and UAV-based systems, and IoT-enabled agricultural applications. The analysis shows that different approaches are suited to different data and application contexts. Classical machine learning methods remain relevant for structured agricultural datasets, while deep learning approaches are particularly applicable to complex spatial, temporal, and image-based information. Hybrid and integrated approaches provide opportunities to combine complementary information but may introduce additional computational and implementation requirements.

The review also identified several persistent challenges, including limited availability of diverse agricultural datasets, heterogeneous data sources, climatic and environmental variability, computational complexity, limited model generalization, and difficulties associated with integrating remote sensing, UAV, and IoT data. These challenges indicate that reported model performance should be interpreted within the context of the dataset, crop, geographical region, prediction objective, and evaluation methodology rather than through universal rankings of algorithms.

Based on these findings, future research should focus on developing large-scale and standardized agricultural datasets, efficient and lightweight prediction models, multimodal data-integration strategies, explainable AI, transfer learning, and cross-regional validation. Greater integration of ML and DL with remote sensing, UAV, and IoT technologies may further support scalable agricultural monitoring and decision-support systems.

Overall, the reviewed evidence indicates that the future of crop yield prediction depends not only on developing increasingly sophisticated prediction models but also on improving data quality, model transferability, interpretability, computational efficiency, and integration with agricultural sensing technologies. These factors are essential for translating AI-based crop yield prediction from experimental research into reliable and practical precision-agriculture applications.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all researchers and organizations whose valuable contributions in the field of crop yield prediction and precision agriculture have supported this review study.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this review paper.

References

- [1] M. A. Javed and M. A. A. Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, 2024, Art. no. e40836.

-
- [2] M. C. Manjunath and B. P. Palayyan, "An efficient crop yield prediction framework using hybrid machine learning model," *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 1049–1058, 2023.
- [3] A. Anand and K. Jhaharia, "Advanced crop yield prediction using machine learning and deep learning: A comprehensive review," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 23, no. 2, pp. 402–415, 2025.
- [4] N. Hussain, S. Sarfraz, and S. Javed, "A systematic review on crop-yield prediction through unmanned aerial vehicles," 2023.
- [5] A. Cravero, S. Sepúlveda, F. Gutiérrez, and L. Muñoz, "From precision agriculture to intelligent agricultural ecosystems: A systematic review of machine learning and big data applications," *Agronomy*, vol. 16, no. 5, p. 516, 2026.
- [6] G. Mohyuddin *et al.*, "Evaluation of machine learning approaches for precision farming in smart agriculture system: A comprehensive review," *IEEE Access*, vol. 12, pp. 60155–60184, 2024.
- [7] S. Saha, O. D. Kucher, A. O. Utkina, and N. Y. Rebouh, "Precision agriculture for improving crop yield predictions: A literature review," *Frontiers in Agronomy*, vol. 7, Art. no. 1566201, 2025.
- [8] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021.
- [9] J. Kariya and P. Mudholkar, "Machine learning models for enhancing accuracy in crop yield prediction: A critical review," in *Proc. 3rd IEEE Int. Conf. Knowledge Engineering and Communication Systems (ICKECS)*, 2025.
- [10] "Machine learning techniques for mustard crop yield prediction from soil analysis," *Journal of Scientific Research*, vol. 64, no. 2, pp. 394–398, 2020.
- [11] Y. Guan, K. Grote, J. Schott, and K. Leverett, "Prediction of soil water content and electrical conductivity using random forest methods with UAV multispectral and ground-coupled geophysical data," *Remote Sensing*, vol. 14, no. 4, p. 1023, 2022.

[12] S. P. Sudha and J. B. Shajilin Loret, "A review on machine learning-based precision agriculture techniques for crop farming monitoring with IoT," *Discover Environment*, vol. 4, Art. no. 10, 2026.

[13] N. Parashar, P. Johri, A. A. Khan, N. Gaur, and S. Kadry, "An integrated analysis of yield prediction models: A comprehensive review of advancements and challenges," *Computers, Materials & Continua*, vol. 2024, 2024.