INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

REVIEW ARTICLE

AI-based Yield Prediction: A Thorough Review

Soma Gupta[1]*, **Satarupa Mohanty**[1], **Dayal Kumar Behera**[1]

1 School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India

OPEN ACCESS

Received: 19-02-2025

Accepted: 19-03-2025

Published: 01-04-2025

Citation: Gupta S, Mohanty S, Behera DK (2025) AI-based Yield Prediction: A Thorough Review. Indian Journal of Science and Technology 18(10): 822-838. https://doi.org/10.17485/IJST/v18i10.175

* **Corresponding author**.

somagupta29@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2025 Gupta et al. 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.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: Traditional farming practices often rely on conventional methods and anecdotal knowledge, but accurate crop yield prediction is crucial for strategic decision-making in agriculture, including import-export strategies and financial planning. Machine learning, a subset of artificial intelligence, offers a data-driven approach that can improve yield prediction by considering multiple factors. ML models can be either explanatory, analyzing past events, or predictive, forecasting future outcomes. Effective feature selection and data preprocessing are essential for accurate ML-based yield prediction. **Objectives:** The primary objective of this review is to analyze the existing research to outline the crucial components used for crop yield prediction techniques. The review focuses on comparative and thorough assessment of the crop yield predictions methods and feature sets used. It examines the benefits and challenges of using various algorithms and features for yield prediction **Methods:** The study employed a systematic literature review methodology. A comprehensive search query was used across five databases, initially retrieving 450 articles. After applying inclusion and exclusion criteria, 40 articles were selected for in-depth review. The selected studies were analyzed to identify the ML and DL algorithms used, the features employed, and the overall findings. **Findings:** This research introduces several novel aspects to the field of crop yield prediction. The review revealed that various ML models, including Decision Trees, Random Forests, Support Vector Machines, Bayesian Networks, and Artificial Neural Networks, are used for crop yield prediction. Within DL, Convolutional Neural Networks, Long Short-Term Memory networks, and Deep Neural Networks were identified as frequently used algorithms. The review also highlights the importance of feature selection techniques in preprocessing raw data. It investigates combining information from diverse sources like images, text, and sensor readings to create richer representations for yield prediction. By addressing these directions, future research can contribute to more sustainable, resilient, and productive agricultural systems, enhancing global food security in the face of growing challenges.

Keywords: Yield prediction; Feature Selection; Machine Learning (ML); Deep Learning (DL)

1 Introduction

Agriculture has a notable role in every country's economic growth. In earlier times the decision of the crop for cultivation, its development, and its date of reaping could all be determined by farmers. The farming sector now finds it harder to carry on owing to the quick changes in ecological conditions.

Precision agriculture, often known as smart farming, has surfaced as a novel strategy for handling sustainability challenges in agriculture. Precision Agriculture is primarily concerned with increasing crop yield and quality production while lowering production costs and environmental impacts to improve the viability of the farm. Due to resource shortages, climatic variability, and increasing numbers of people, feeding current populations has become more challenging. To boost food production globally, greater emphasis should be given to agricultural yield predictions, protection of crops, and land assessment.

Digital innovations like IoT, AI, Data Analytics, Cloud Computing, and Blockchain Technology greatly aid precision agriculture. Numerous production factors affect potential growth and output including weather, soil characteristics, terrain, irrigation, and fertilizer management and thus traditional approaches for yield prediction are unreliable. The data is analyzed and assessed using ML algorithms to enhance agricultural activities. By leveraging numerous features, ML which concentrates on learning, is a practical technique that improves yield estimation accuracy.

Machine Learning (ML) is often preferred over traditional statistical approaches in yield prediction because of several advantages such as ability to handle non-linearity & complex relationships, better performance with large & high-dimensional data. ML can incorporate satellite images, sensor data, weather forecasts, and genomic data, while traditional approaches struggle with such multi-source data. Traditional statistical models often require manual feature engineering and selection, but ML models can automatically learn important patterns and interactions from the data. ML algorithms, particularly deep learning, consistently outperform statistical models in predictive tasks due to their ability to generalize complex patterns. ML models can be updated continuously with new data, improving predictions over time, whereas traditional models may require frequent re-calibration.

1.1 Research Gap

A thorough evaluation of the literature is being done to give a research summary on the use of ML in yield prediction. The systematic review of the literature finds the potential gaps in research in a specific area of study. The review outlines several limitations of existing research in yield prediction.

i) Existing research has not fully explored the potential of hybrid models.

ii) The focus of articles on a specified timeframe may exclude more recent advancements and limit the scope of review.

iii) Existing research has not adequately addressed the integration of real time data from sensors, satellites despite its potential to improve predictions.

iv) More research is needed on transfer learning which adapts pretrained models to adapt new datasets or regions.

v) Integrating ML models into farm management systems remains unexplored.

1.2 Research Contributions

The major contributions of the paper are summarized as below:

i) Outlining the crucial components of the agricultural yield prediction methods.

ii) Comparative and thorough assessment of the crop yield predictions methods and feature sets used.

iii) Examines the benefits and challenges of using various algorithms and features for yield prediction.

iv) Explores hybrid models combining process-based models with machine learning algorithms.

v) Focuses on the development and applications of Explainable-AI techniques to enhance the transparency and interpretability of deep learning models.

Subsequent sections of the paper are structured as follows: Section 2 provides an overview of related work and an organized methodology for searching, selecting, and excluding articles. Section 3 provides an overview of the various models, features and feature selection techniques used for yield prediction. Section 4 discusses results and finding by comparing the features, ML algorithms, and performance metrics used in yield prediction. In Section 5 the paper reaches its conclusion.

2 Review Methodology

2.1 Related Work

For quick decision-making at the democratic and regional levels, crop production forecasting is a crucial task. The literature on machine learning-based yield prediction was analyzed in this review.

A review study on using various ML approaches for crop yield prediction and the estimate of nitrogen status was carried out by [1]. The study results indicate that ML and sensing technology developments will soon help the agriculture industry. It also includes various ML methods used to accomplish the same goal in Precision Agriculture. A review study on machine learning's use in agriculture was published by [2].

A systematic review explored ML algorithms for predicting crop production based on meteorological conditions [3]. The research suggests doing an extensive search to uncover other factors that affect crop productivity. In agriculture, the problems with automated tracking, cause identification, decision support, and classification were emphasized [4]. A Bayesian network can be utilized to centralize and solve many agricultural issues. [5] introduces an outstanding systematic review that looks at several features and prediction techniques. However, the paper focuses more on information extraction than critical analysis, research gap assessment, or recommendations. The objective of this work [6] is to document the new and developing characteristics and ML algorithms in the agricultural sector and promote the recent implementation of ML-driven yield prediction in the palm oil industry. The authors have provided an in-depth analysis of the ML application in this study [7]. This review article will give readers an understanding of how the research society views implementing digital methods in agriculture systems management.

The objective of the study [8] is to predict yield using factors like weather, crop, rainfall, and area. Machine learning such as DT, RF, XG Boost, deep learning such as CNN, and LSTM techniques are used in the study for yield prediction. This research [9] combines data from numerous sources to forecast wheat yield In Multan, Pakistan. This study [10] compared supervised and unsupervised machine learning models in forecasting yield. The study shows that classification outperforms regression and unsupervised in terms of accuracy for yield prediction with numerous factors. This work [11] uses 2 different models 2D-CNN, skip connection, LSTM attentions, and 3D-CNN skip connections and ConvLSTM attention for MODIS products like land cover, land surface, and surface temperature for prediction. Using machine learning techniques, pesticide records, and yield data this study [12] proposes an innovative yield forecast system.

The purpose of this systematic evaluation of the review is to discover research gaps in the applications of machine learning approaches and to provide guidance for analysis of the various elements that affect the yield of crops. In addition to being evaluated, all studies from conferences, journals, and other online resources are combined and presented in a way that answers the research questions for the systematic review of the literature.

The appropriate research articles are chosen using several sources, including IEEE Explorer, ScienceDirect, MDPI, CABI Digital Library, and Taylor & Francis. These studies are evaluated and sorted according to the quality standards.

First, the research questions were defined, and then the review methodology was created. Data that are taken out and examined from the primary studies is be used to address the research questions. Consequently, a search strategy was developed that contained resources to review and search terms for analysis. In addition, research selection and quality evaluation criteria were established to evaluate studies and decide which ones should be excluded from the review. Additionally, a plan for extracting data is established to acquire the necessary information.

2.2 Research Questions

Initially, the aim is to investigate the rationale behind using machine learning to forecast crop yields. It was also crucial to determine which machine learning algorithms were applied to which crops and which attributes were employed to estimate crop productivity. For our research, identifying the data sources was crucial.

We thus sought to characterize them by responding to the appropriate research questions for this literature review. This literature review aims to comprehend existing research on machine learning techniques for yield prediction. Studies have been explored to understand a variety of aspects.

The research questions (REQs) listed below have been established for this literature review study.

REQ1- What was the primary motivation behind using machine learning for predicting crop yields?

REQ2- In the literature, which machine learning techniques have been used to forecast crop yield?

REQ4- What techniques and standards have been used to assess and forecast agriculture yield?

REQ5- Which crops were used in machine learning algorithms to estimate yield?

REQ6- Which algorithm was more effective at predicting crop yield?

REQ7- What challenges arise when estimating yields using machine learning algorithms?

2.3 Criteria for Searching

For this review, digital databases such as IEEE Explorer, Science Direct, MDPI, CABI Digital Library, and Taylor & Francis were utilized as resources. These databases were chosen because they index publications subject to peer review. Since machine learning and agricultural yield prediction are vast areas, the search approach for the mentioned databases considered this study's scope. Because of this, a precise search query was created to prevent finding results that might be beyond the limits of this study.

All databases, however, use distinct search strategies. Because of this, the query string was initially defined exactly as it is displayed below, with a few modifications. The query string, which included machine learning and crop yield prediction or estimation, was utilized to search for content in the Title, Abstract, and Keywords fields. The search approach was used, and the results showed 450 publications.

2.4 Criteria for Exclusion

The studies were evaluated according to criteria for exclusion to filter out irrelevant findings and to establish the parameters for the review. Below is a list of exclusion criteria:

EXC 1 – Articles published beyond the agriculture sector and yield prediction using machine learning.

EXC2 – Publication is not written in English.

EXC3 – Duplicate publication or one that was previously obtained from a different database.

EXC 4 – Nonavailability of full-text publication.

EXC 5 – Publication is a paper review or survey.

EXC 6 – The article was published before 2014.

Merely 90 studies were left for further review after the EXC1, EXC2, and EXC3 were applied. After applying all six exclusion criteria, 40 studies were selected for further research. Per the databases we searched, Figure 1 displays the flowchart for the included articles for the review after applying inclusion and exclusion criteria. As indicated by Table 1, the majority of the papers were obtained from the databases of IEEE Explorer, ScienceDirect, MDPI, CABI Digital Library, and Taylor & Francis. It shows the number of articles retrieved initially, and the number remaining after the exclusion criteria were applied. Data have been gathered and compiled from the chosen papers to address the research issues. The information received was mostly used to address the research questions and determine if the studies satisfied the criteria for exclusion.

Table 1. Articles distribution across several databases

Database Name	Number of articles retrieved after the database search	Number of articles relevant to the study	Number of articles after exclusion criteria
IEEE Explore	63	25	6
CABI Digital Library	122	51	9
MDPI	56	18	7
Science Direct (Elsevier)	140	59	10
Taylor and Francis	69	27	8
Total	450	180	40

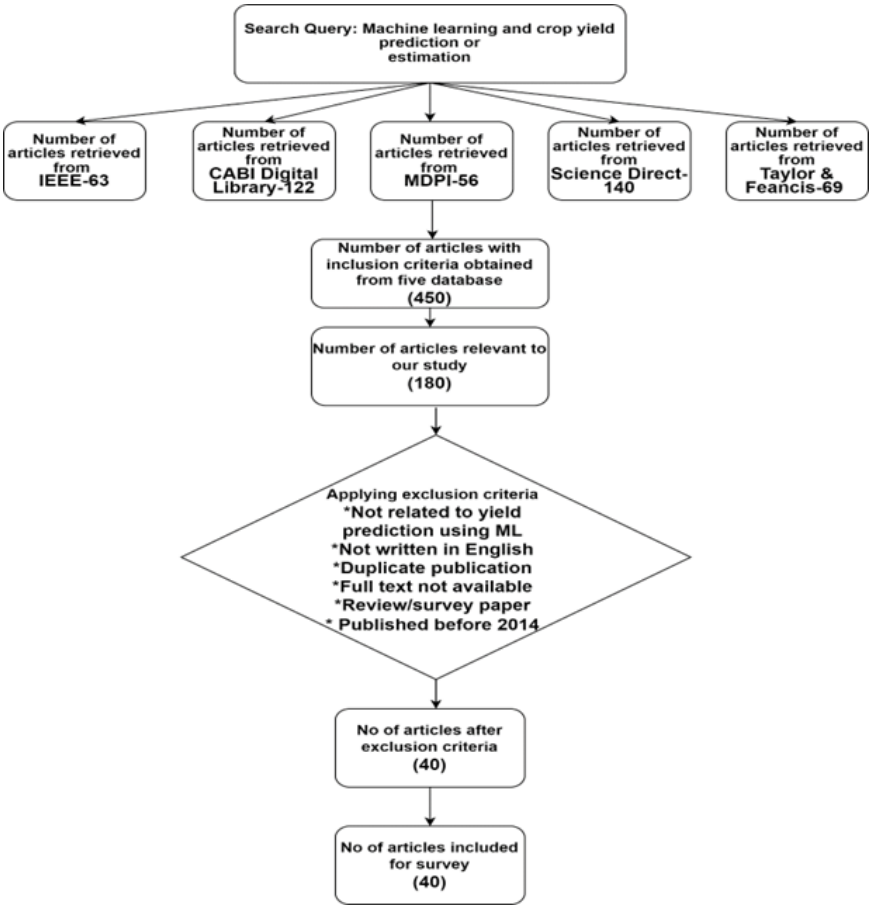


Fig 1. Flowchart for the searching, selection, and exclusion of articles

3 Overview

3.1 Models for forecasting crop yield

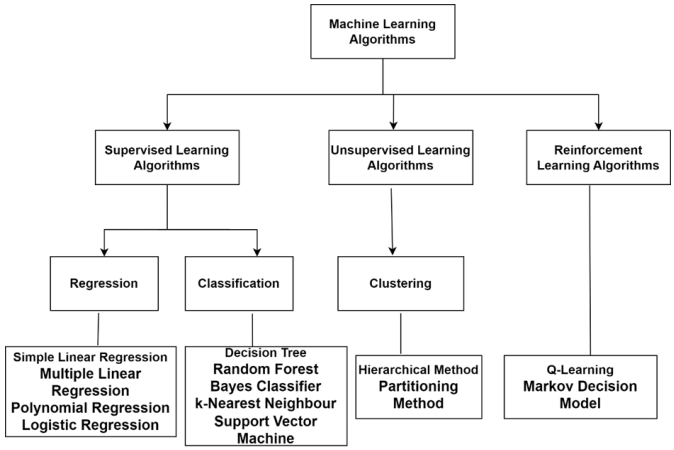


Fig 2. Machine Learning Algorithms Classification

Supervised, Unsupervised, and Reinforcement learning are the three categories of machine learning algorithms described in Figure 2.

3.1.1 Machine Learning Algorithm

The study[13] investigated silage maize yield prediction using a time series NDVI dataset acquire from Landsat8 OLI. The performance of the advanced ML techniques used in this study- RFR, SVR, GPR, and BRT- was compared to that of some suggested conventional regression techniques. A straightforward method was created[14] to forecast corn yields utilizing the MODIS data output from two geographically distinct significant corn crop-producing regions. Corn yield was predicted using two straightforward linear regression models YP and YF. The (YP) used leaf area duration from predicted emergence to maturity, and the (YF) yield model used leaf area duration for a specified time from DOY 89 to a specific EOD (end of DOY (EOD)). RF is a potent instrument that has been used in agricultural studies to forecast yield[15]. RF performs more accurately than any other decision tree, the bias is balanced off by the single decision tree because of randomness[16]. A similar ensemble approach to RF is ERT uses decision trees without pruning. The node partitioning variables and tree count are set up to match RF specifications.

The most popular machine learning method for predicting agricultural yield[2] is an ANN, which can represent the complex nonlinear relationship between input and output. The variables and number of trees used to partition the nodes are set up per Random Forest. The most popular machine learning method for predicting agricultural yield[17] is an ANN, which represents the input and output intricate nonlinear relationship. The performance of an ANN is influenced by several factors, including the number of nodes in hidden layers, the rate of learning, and the tolerance for training[18]. Multiple iterations are needed for the projected value to be within a reasonable range of what has been seen, and the learning rate determines this. K-NN is used for classification and regression, giving close neighbors greater weight in the prediction process such that they relate more to the average than distant neighbors[19]. The distance between two neighbors can be calculated using several distance formulae, including the Euclidean, Manhattan, and Minkowski. MLR and three more ML techniques BP, SVM, and RF were used to simulate rice yields in this study[20]. They used site-based phenology, climate, geography, and yield data for early rice.

3.1.2 Deep Learning Algorithm

A novel hybrid deep learning model was proposed, utilizing both recurrent and convolutional neural networks (RNN) and (CNN). Deep Yield, a hybrid structure that integrates the ConvLSTM layers with the 3DCNN to extract spatiotemporal characteristics more accurately and consistently, was introduced by[21]. To increase accuracy and decrease complexity, this work[22] develops an LSTM-RNN model that uses the temporal pattern of crops. The primary objectives of this work[23] is to assess the impact of in-season multitemporal photography for predicting grain yield using hand-crafted features and raw input imagery and to suggest an end-to-end deep learning technique for field-scale yield prediction based on raw imagery. To predict the yield of crops, the proposed study[18] develops a Deep Recurrent Q-Network model, an RNN algorithm over the Q-Learning reinforcement learning technique. Using NDVI time series data the study[24] provides the first SNN computational model for estimating crop yield. The research also examines the ideal number of features required to enhance experimental data collection outcomes. In this study[17] NDVI and RGB data acquired by UAVs are used to develop a model for crop yield prediction using CNN. This deep learning technology excels at image categorization challenges. This article[25] discusses a unique spatiotemporal remote sensing crop classification approach based on 3DCNN. To extract the essential features of rice grain yields from low-altitude remotely sensed data this study[26] suggested using CNN architecture. CNN comprises convolutional, pooling, and activation layers between the input and output layer[27]. In contrast to a pooling layer, which uses max and average pooling to extract low-dimensional data from the input data, a convolution layer applies many local filters to perform convolution on the input data. The nonlinear procedures in the activation layer can improve CNN's nonlinear fitting capabilities**.** CNN and BPN use BP to modify the weight. A two-branch deep learning model was suggested in this study[28] to predict the winter wheat output. With inputs from meteorological and remote sensing data, the initial branch of the model was built using LSTM. CNN was used to create a different branch to model static soil properties.

3.1.3 Hybrid

In addition to these, a variety of regression and classifier methods, such as Machine Learning Ensembles[29], XGBoost[30], and SNN[24], are used for yield prediction. Crop yield prediction has also used some hybrid approaches, such as LSTM-RNN[22], CNN-RNN[31], LSTM-3DCNN[21] and CNN-LSTM[32].

3.1.4 Transfer Learning

[33] Proposes a novel approach to crop yield prediction by combining transfer learning and LSTM networks. The outcome is more effective solution for optimizing yield predictions contributing to sustainable agriculture and precision farming. The

study[34] demonstrates that transfer learning with careful hyperparameter tuning and a combination of several techniques can significantly enhance crop yield prediction. This research[35] highlights the potential of combining deep learning with attention mechanism and transfer learning for accurate and efficient crop yield prediction. The study[36] highlights the limitations of traditional agricultural practices and emphasizes the importance of accurate yield prediction for informed decision making.

3.1.5 Explainable AI (XAI)

The study[37] uses an Explainable Boosting machine to predict cotton yield using a variety of datasourses.it emphasizes the importance of using multiple data sources and providing an interpretable model to understand the complex factors influencing cotton yield. The study[38] integrates XAI techniques specifically saliency maps into ML algorithm. It aims to bridge the gap between AI and agriculture by making AI driven insights more accessible and understandable to farmers. The study[39] introduces an XAI powered smart agriculture system designed to provide holistic recommendations for precision farming and thus enhancing productivity and minimizing environmental impact.

Crop yield prediction has made use of a variety of regression- and classification-based prediction techniques. Features, advantages, and disadvantages of several methods, used for yield prediction, including ANN, CNN, DNN, DT, RF, Extra tree, LSTM, MARS, K-NN, SVM SVR, and LR described in Table 2.

Table 2. Features, Advantages, and Disadvantages of Popular Prediction Algorithms

Algorithm	Features	Advantages	Disadvantages
DT	Leaf, Decision, and Root make up nodes the decision tree	Can be applied to problems related to both classification and regression. Data normalization and scaling are not required	The data overfitting issue in this model may lead to inaccurate predictions.
RF	RF includes ensemble learning that makes use of the bootstrap and bagging processes.	Less likely to have overfitting issues	Even a small modification in the data could have a major impact.
ANN	The brain inspires the input, output, and hidden layers of ANN.	Complex nonlinear interactions between independent and dependent variables can be handled using it.	Has a higher processing burden and a black-box nature.
CNN	CNN consists of three layers: input, hidden, and output layer Convolutional, ReLU, pooling, and fully linked layers are also included in the hidden layer.	A fully linked layer that replaces the convolution layer shows a significant parameter drop.	Convolution is a far slower method than ANN.
DNN	To progressively retrieve attributes from the raw input, many layers are used.	The problem of overfitting and local minima can be resolved through a thorough optimization and activation process.	A high-end computer is required
LSTM	It is a feed-forward neural network (NN) having input layers that provide feedback to neurons' output layers.	RNN represents the data sequence in such a way that it is possible to assume that each sample depends on the ones before it.	Large sequences cannot be processed by this network when ReLU or tanh is utilized as the activation function.
SVM	In the feature space, the SVM creates multidimensional borders between the data points.	It facilitates efficient data grouping by using kernel functions to optimize the margin between categories	Susceptible to the overfitting issue that depends on the kernel function used for the best classification
MARS	A non-parametric regression technique for dealing with nonlinearity and correlations between variables that combines many linear models	It provides a flexible model that can accommodate both linearities and nonlinearities.	It is limited to handling massive data and is vulnerable to overfitting problems.

Figure 3 explains the process involved in forecasting yield using machine learning. The agricultural dataset is first used to predict crop yield. After that, pre-processing is applied to the data to eliminate noisy data. The pre-processed data is passed through a feature extraction process to extract features utilized in machine learning-based classification. These features include information about nutrients, soil, field management, and other factors. In Machine Learning, selecting the right train-test split is crucial since it influences the model's accuracy and generalization. The findings produced by the current models employing

machine learning algorithms are explained in this section.

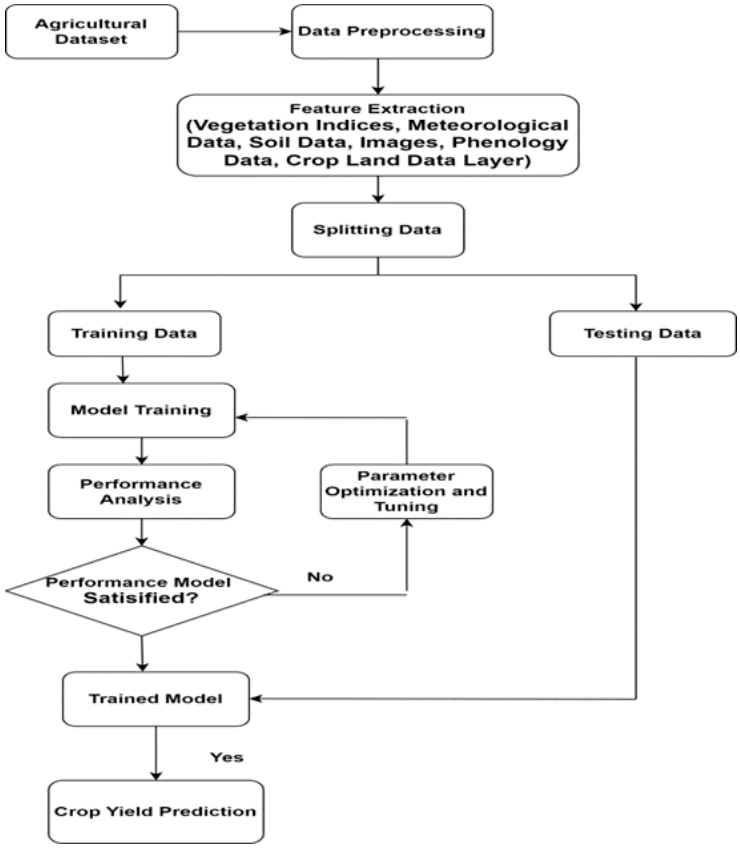


Fig 3. ML framework for yield prediction

3.2 Crop features

Several factors influencing the yield of crops and the uncertainties associated with cultivation are described in Figure 4. According to the recent literature, the most significant factors related to yield are fruit, leaves, irrigation and fertilization information, climate, cropland, vegetation index, satellite, crop management, phenology and historical yield data, soil and groundwater characteristics, and socio-economic factors.

3.2.1 Climatic feature

[40] Performed EDA to decide which climate variables to include, and spatiotemporal factor analysis was used to determine how various climate variables and satellite data are related to yield.[41] SCYM interprets vegetation indices from satellite images using region-specific crop model simulations and climatic data, making it possible to estimate sub-field yields across vast areas with high accuracy. With a temporal resolution of three and a spatial resolution of 0.1 degrees, seven meteorological variables were used as drivers by[28] to replicate the crop growth.[42] The precipitation, monthly drought index, and monthly max-min temperature were used to estimate the winter wheat yield.

The historical weather data in[43] is derived from METDATA and average wind speed, incoming shortwave radiation (sunlight), relative humidity, precipitation, and minimum and maximum air temperatures were observed daily. LST and precipitation were chosen as the two variables to be assessed in the models[44]. LST is a measurement comparable to air temperature but not precisely the same.

3.2.2 Satellite Data and Vegetation Data

The first SNN computational model for crop yield estimation using NDVI time series is introduced in the work[24]. CNN a deep learning technology that performs very well on image classification tasks, is used in this study[17] for creating a model for crop yield prediction based on NDVI and RGB data obtained from UAVs. In this article[45], agricultural yields are estimated using MODIS satellite data and the SAFY model while considering the impacts of drought stress on crop development. Utilizing satellite data and climatic data at the statistical division level, this study[40], constructed wheat yield forecast statistical models RF, SVM, NN, and LASSO i.e. a linear method. Based on an NDVI time-series dataset, this study[13] effectively incorporates BRT, RFR, SVR, and GPR ML algorithms in forecasting silage maize yield. The goals of this study[14] were to create a straightforward method for predicting crop phenology and yield utilizing the bare minimum of remote sensing data products and to assess how portable it is in terms of space. To extract the key characteristics associated with rice grain production from remotely sensed images taken from low elevations, an effective CNN architecture was suggested in[26]. Using weekly NDVI data from the MODIS, this research[46] aims to identify and further quantify flood effects on corn.

As remote sensing input data, the NDVI and the bands "SREFL_CH1", "SREFL_CH2, and "SREFL_CH3" were used[28]. The EVI and the NDVI are two different types of VI that we[42] gathered. An index of vegetation can track how the vegetation is changing dynamically.

3.2.3 Soil Properties

A unique hierarchical DL model integrating RNN and CNN was suggested in this work for extracting both spatial and temporal properties. Remote sensing time-series and soil property data are included as inputs and the model outputs yield. A recurrent neural network that processes enormous amounts of information was inspired by the DQN [Deep Q Network] concept. The suggested method[18] uses DNN for forecasting crop production using several environmental, soil, and groundwater variables. The soil variables used are organic carbon content and pH of the topsoil and subsoil layer in this study[42].

3.2.4 Cropland Data Layer (CDL)

This work[22] prepared training samples using data from ground truth, time series Cropland Data Layer (CDL), and Landsat. LSTM RNN model using time series CDL, and ground-measured data was developed to categorize Landsat images in this work. This study[46] aims to evaluate the possibility of detecting flood occurrences and if the relationship between the NDVI and change in corn yield at the county level is consistent. Higher resolution spatial MODIS weekly NDVI and CDL are used.

3.2.5 Phenology Data

The goals of this work[14] were to create a straightforward method for predicting crop phenology and yield using a minimal collection of remote sensing data products and to assess spatial portability. To distinguish between several crop species using Landsat 8 OLI images, this research[47] investigates a novel, site-specific classification method based on phenology. In particular, the phenology of the characteristic crops of the area of interest and agroecological zoning were obtained using the FAO sources. Different combinations of phenology, weather, and geographic data were examined in this work[20] to forecast the rice yields using regression-based and three machine-learning techniques.

Early mature rice phenology data for each location comprised the SO, EM, TL, TR, TG, TI, BO, HE, FL, MI, and MA. Weather data includes Tmean, Tmax, Tmin, Dh, Pcp, Rhmean, Rhmin, Wsmean, and Wsmax, respectively.

3.3 Feature selection

Weather, soil, availability of irrigation, the need for fertilizers, farm capacity, and other factors all affect crop productivity. Machine learning methods analyze the data accumulated over multiple years to generate early predictions.

Before utilizing machine learning algorithms for the prediction process, it is crucial to identify relevant and valuable features (feature selection) by collecting all available variables that affect the yield. Choosing the most significant characteristics from a specific collection of data set features involves eliminating superfluous, redundant, and noisy features. The process of choosing features is different from that of extracting features. By applying a series of selection functions to convert the initial high-dimensional feature space into a low-dimensional space, feature extraction produces a new feature space.

The primary concern in feature selection is the identification of the ideal subsets and assess the flawlessly generated subsets. In any machine learning application, sensitive feature selection is crucial because the right number of features can improve performance while lowering computational and model complexity.

Table 3 describes various feature selection techniques with its pros and cons.

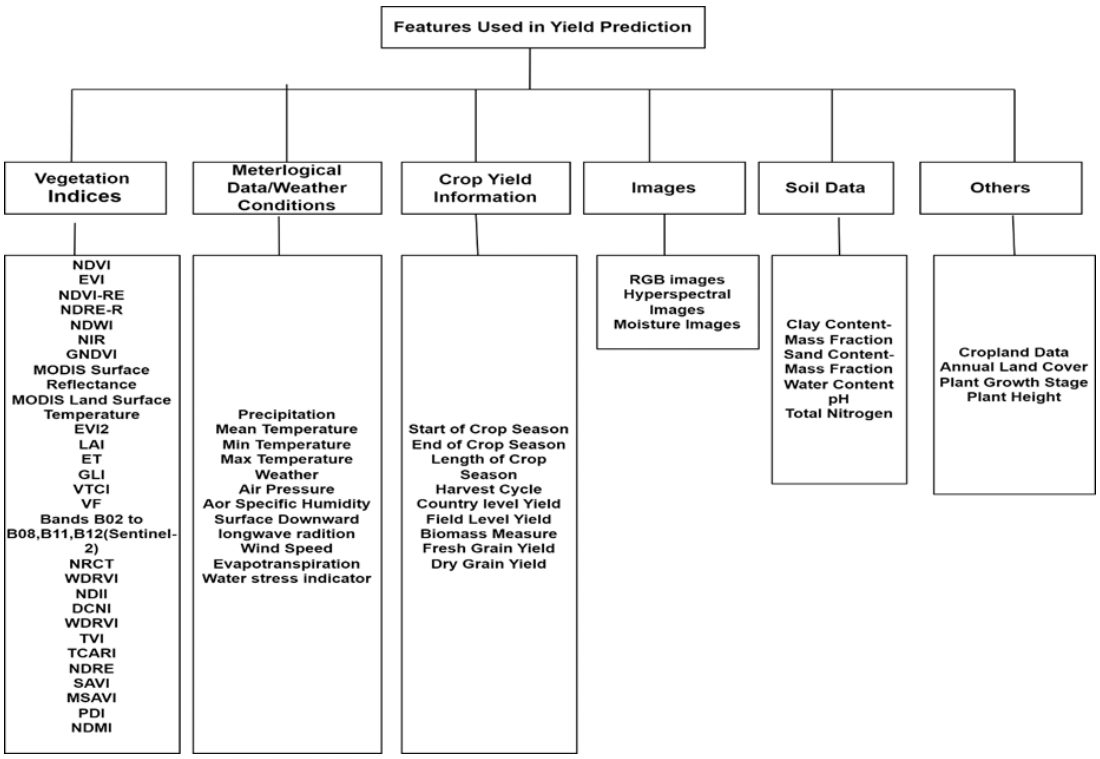


Fig 4. Some categories of features used for crop yield prediction

Table 3. Comparative study of various feature selection techniques

Feature Selection Technique	Method	Advantages	Disadvantages	Use Case
Forward Selection	Iterative method. Adds features one by one based on model performance	Simple, interpretable	Computationally expensive	When you have a small number of features
Backward Elimination	Starts with all features, removes least significant ones iteratively	Effective for removing irrelevant features	Slow for large feature sets	When many features are irrelevant
Recursive Feature Elimination (RFE)	Iteratively removes least important features based on model ranking	Robust, works well with tree-based models	Expensive, needs retraining	When model-based ranking of features is needed
LASSO (L1 Regularization)	Shrinks coefficients, forcing some to zero	Efficient, handles collinearity	May remove too many features	When automatic feature selection is needed in regression
PCA (Principal Component Analysis)	Transforms features into principal component	Reduces redundancy, handles multicollinearity	Loses interpretability	When reducing dimensionality in high-dimensional data

3.4 Crops

3.4.1 Corn

The goals of this study[14] was to create a straightforward method for predicting crop phenology and yield using a minimal collection of remote sensing data products and to assess the spatial portability. After examining the robustness and dependability of LAI and NDVI, Lykhovyd[48] they concluded that LAI is better and more effective than the NDVI-based method. However, using other agricultural variables compared to those based on vegetation indices could improve the forecast capability.

3.4.2 Soyabean

The end-of-season production of soybeans has been predicted using remote sensing data from the MODIS satellite. The soybean yields for 2018 and 2019 were predicted using the models, and the results indicated that the forecasts were quite accurate[21]. The aim of the study[44] is to assess how soybean yield can be estimated using different algorithms using NDVI,EVI vegetation indices and weather-related data like precipitation and land surface temperature as independent variables. It evaluates how early soybean yield can be predicted in the growing season using this method. Prediction of soybean yield both in-end of the season in CONUS at the country level, we propose a deep CNN-LSTM model in the study[32]. To create machine learning models for forecasting China's winter wheat production, we[42] combined 21 variables from data on soil parameters, climate, and remote sensing, based on the GEE platform.

3.4.3 Wheat

One major winter wheat-growing region in China, Shandong Province, used the suggested method[24] to calculate the winter wheat yield. Building and testing the SNN model for agricultural production prediction with NDVI data is the goal of this paper. The goal of this study[45] is to confirm the yield estimation results at the country/regional scale and present the estimation yields from the SAFY-V model. This study[28] uses multi-source data and long-term statistical yields to calculate the country-level winter wheat yield in China's key wheat-producing region by deep learning techniques.

3.4.4 Maize

The performance of several conventional regression methods was compared with advanced machine learning techniques which included BRT, RFR, SVR, and GPR approaches in the study[13]. This study evaluates the viability of these techniques for predicting silage maize yield and how well they perform compared to traditional crop yield estimating methodologies. It also suggested using the trained machine learning models for locations and times where no NDVI time series match. To assess the Scalable Crop Yield Mapper (SCYM) this study[41] used a yield monitor dataset that included more than a million field observations of maize collected throughout the US Corn Belt between 2008 and 2018. SCYM employs climatic data and crop model simulations appropriate for a given location to interpret satellite-observed vegetation indices. This eliminates the need for ground data calibration and allows for the efficient prediction of subfield yield across vast areas over a long period.

3.4.5 Rice & Paddy

To improve the accuracy of yield estimation by including spatial-spectral information and exploring the time for forecasting yield, this study[26] aims to explore CNN algorithms' capacity to determine rice grain yield using UAV-based RGB images and multi-spectral images. In this study[20] MLR, a more sophisticated machine learning method called BP, SVM, and RF were tested along with different combinations of phenology, weather, and geographic data for rice prediction. The study[18] focuses on the forecast of rice crop yield for the southern Indian district of Vellore.

4 Results and Findings

Metrics for assessment can be used to define a model's efficacy. Because evaluation metrics can distinguish between the results of various learning models, they are essential. Performance measures of varying prediction algorithms for different crop yield predictions are shown in Tables 4, 5 and 6.

Table 4. Performance Measure of Machine Learning Algorithms for Various Crop

References	Crop	Attribute	Algorithm	Assessment Metric
[40]	Wheat	Satellite data and climate data	RF, SVM, and NN, LASSO	R^2 ~0.75
[16]	Wheat	Climate data, satellite data & Socio-Economic (SC) factors	RR, RF, and Light-GBM	R^2 - 0.05 to 0.25
[42]	Wheat	NDVI	DT, RF, BST, BGT, KNN, NN, SVM, GPR	
[14]	Corn	MODIS	Linear regression models	
[48]	Corn	NDVI, LAI		R^{2-} 0.85

Continued on next page

Table 4 continued

[20]	Rice	Phenology, Climate and Geography data	BP, MLR, SVM and RF	BP: RMSE-800, (R^2) (0.24) SVM: RMSE-737, (R^2) (0.33) RF: RMSE-744, (R^2) (0.31) kg/ha
[13]	Maize	NDVI	RFR, BRT, Support Vector Regression, GPR	BRT - average R-value higher than 0.87
[29]	Corn	Environmental and management variables	Ensemble models	RMSE - 9.5%.
[15]	Cotton	Long-term agromet-spectral variables	RF	RMSE - 62.77 kg/ha, MAPE - 0.32
[19]	Potato	NDVI, Soil electrical conductivity, moisture content, slope, and chemistry	LR, EN, k-NN, and SVR	RMSE 5.97 (NB2017), 4.62 (NB-2018), 6.60 (PE-2017) and 6.17 (PE-2018)

Table 5. Performance Measure of Deep Learning Algorithms for various crops

References	Crop	Attribute	Algorithm	Assessment Metric
[24]	Wheat	NDVI image	SNN	RMSE 0.29, MAE 0.24
[28]	Wheat	NDVI, air-specific humidity, air pressure, temperature, precipitation,	LSTM and CNN	R^2 - 0.77 and RMSE - 721 kg/ha
[19]	Soybean	SR, LC, LST	3DCNN	RMSE - 4.803
[44]	Soybean	Satellite imagery NDVI, EVI, Land Surface temperature, precipitation, and Weather data.	NN, LSTM	MAE-0.24 Mg ha
[26]	Rice	UAV data	CNNs	
[22]	Paddy	Environmental, soil, and groundwater	DQN (Deep Q Network)	Accuracy - 93.7%
[20]	CDL products	Landsat, CDL, and ground truth	LSTM RNN	Accuracy - 97.33% to 99.21%
[24]	Wheat & Malting barley	NDVI and RGB data	CNN	Early stages of growth (MAPE: 8.8%) Later growth stage MAPE: 12.6%
[23]	Corn and Soybean	Multitemporal Imagery and Hand-Crafted Features	CNN	$R^2 > 0.8$ RMSE% < 10

Table 6. Performance Measure of Other Algorithms for Various Crop

References	Crop	Attribute	Algorithm	Assessment Metric
[41]	Maize and soybeans	Climate data and region-specific crop model	SCYM	R^2 - 0.40
[45]	Wheat	Vegetation Temperature Condition Index (VTCI), a Time-series remotely sensed drought monitoring index	Simple Algorithm for Yield Estimate (SAFY-V) model integrating LAIs, VTCI	

Many metrics of performance are employed to evaluate the efficacy of the regression model, including determination coefficient (R-squared), mean absolute percentage error, MSE, MAPE, and RMSE. The mean absolute variance between the predicted and actual results is measured using the arithmetic mean (MAE) method. This information is utilized to determine the average significance of the errors associated with a collection of forecasts. MSE measures how closely the regressor line tracks the dataset points, providing information about the estimator's performance. When estimating the residuals or the standard deviation of the predicted error, RMSE shows how effectively the data is centered on the best-fit line. The determination coefficient illustrates how the developed framework outperforms the original framework and is used to determine the accuracy of the fit of the regression framework. MAPE function is utilized to compute the mean percentage error, which indicates the degree to which the model's forecast differs from its corresponding results.

On the other hand- Precision, Accuracy, Sensitivity, Recall, Specificity, and F1 score are the metrics used by classification models to predict crop yield. For classification problems, the most popular and useful metric is the classification accuracy.

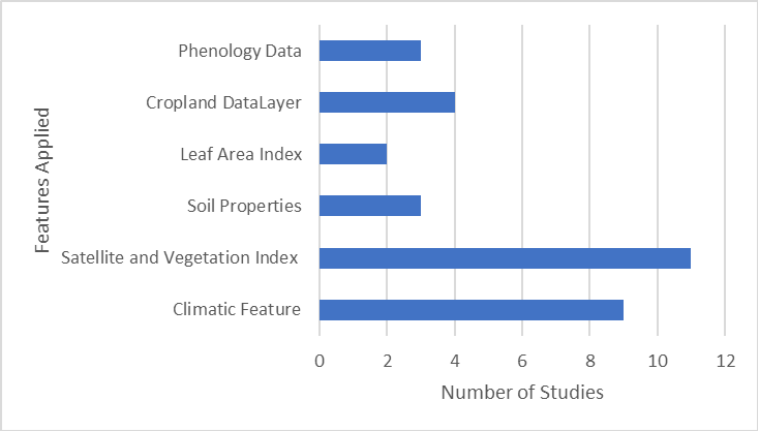


Fig 5. Feature Used v/s Number of Studies for Yield Prediction

The feature lists are divided into six groups for this review. The most popular features are shown in Figure 5 along with a selection of studies that used them. The most used features are vegetation index and satellite data, 11 studies use them followed by climatic features (9 studies). In addition, phenology data (3 studies), leaf area index (2 studies), cropland information (4 studies), soil properties (3 studies), and phenology data (3 studies) are moderately used in crop yield prediction. Precipitation, humidity, solar radiation, temperature, rainfall, vapor pressure deficit, and precipitation are some of the sub-features that researchers have applied most frequently.

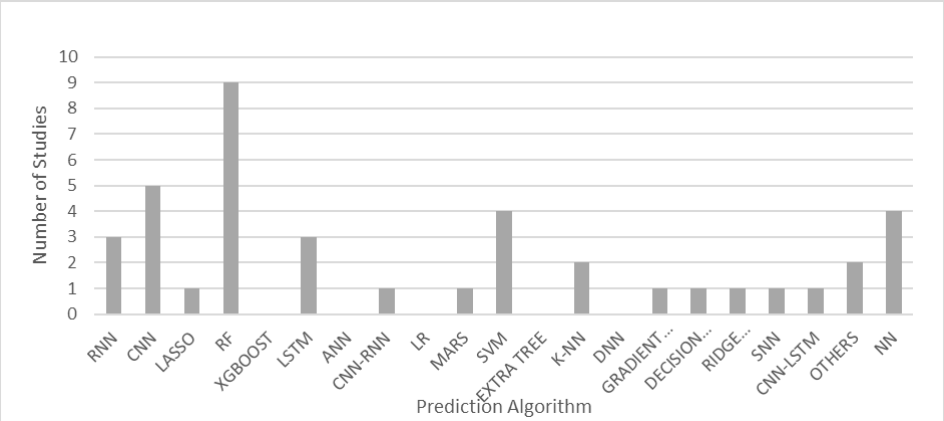


Fig 6. Prediction Algorithms V/S Number of Studies

Several prediction algorithms have been used for the study of yield prediction. Figure 6 lists the prediction algorithm names and the number of studies using them. According to the review, the most widely used prediction algorithm is RF followed by the second CNN used in 9 and 5 studies respectively. SVM, NN, RNN, and LSTM are the other widely used algorithms used in 4, 4, 3, 3 studies. Other algorithms such as KNN, LASSO, SNN, CNN-LSTM are also used in various studies.

Crop yield prediction algorithms use a variety of performance evaluation criteria. RMSE is the most often used metric in crop yield prediction according to Figure 7. In particular, 6, 5, 3 studies have used R-squared, MAE, and accuracy as performance metrics. It is nearly hard to compare prediction models assessed using different performance metrics. As a result, a standardized and systematic method or one metric should be used for yield prediction. It is feasible to compare several yield prediction algorithms if each algorithm uses similar performance metrics.

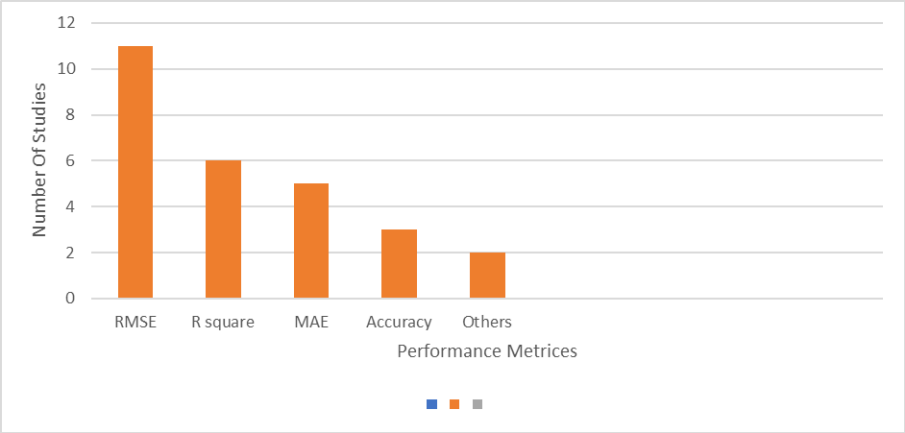


Fig 7. Performance Metrics V/S Number of Studies

5 Conclusion

Farmers require appropriate instructions on time to enable them to predict crop yields and develop tactics that would optimize those yields. Over the past few years, there has been a significant advancement in both sensing technologies and machine learning approaches. More complex algorithmic methods and more affordable, comprehensive information may be available in the future, facilitating improved assessment and decision-making regarding crop and environmental conditions.

Based on the research scope and data availability, this review suggested that the chosen articles utilize a variety of features. The studies differ in crop, scale, and geographical location. The dataset's availability and the study goal determine the features chosen. Studies also demonstrated that models with a higher feature count did not always produce the most accurate yield prediction.

A comparison of models with fewer and more features is required to determine which performs the best. The findings show that although the optimal machine learning model cannot be determined, but some models were used more often than others.

To know which machine learning model provides the best predictions, several models were tested by numerous studies. NN, GB, RF, and Linear Regression are the most often applied models. Based on the study, the most recommended deep learning algorithms are CNN, LSTM, and DNN.

This research introduces several novel aspects to the field of crop yield prediction. It explores deep learning to extract more meaningful features from existing data. It focuses on developing more comprehensive evaluation metrics that go beyond simple accuracy, incorporating measures of uncertainty, bias, and fairness and investigates combining information from diverse sources like images, text, and sensor readings to create richer representations for yield prediction.

By addressing these directions, future research can contribute to more sustainable, resilient, and productive agricultural systems, enhancing global food security in the face of growing challenges.

List of Abbreviations

- Artificial Intelligence [AI]
- Turning Green [TG]
- Machine Learning [ML]
- Tiller [TI]
- Deep Learning [DL]
- Booting [BO]
- Artificial Neural Network [ANN]
- Heading [HE]
- Convolution Neural Network [CNN]
- Flowering [FL]
- Deep Neural Network [DNN]
- Milky [MI]
- Decision Tree [DT]

- Mature [MA]
- Random Forest [RF]
- Mean Temperature [Tmean]
- Extremely Randomized Tree [Extra Tree/ERT]
- Maximum Temperature [Tmax]
- Long Short-Term Memory [LSTM]
- Minimum temperature [Tmin]
- Multivariate Adaptive Regression Splines [MARS]
- Daily Sunshine Hours [Dh]
- k-Nearest Neighbors [KNN]
- Precipitation [Pcp]
- Support Vector Machine [SVM]
- Mean relative humidity [Rhmean]
- Support Vector Regression [SVR]
- Minimum relative humidity [Rhmin]
- Linear Regression [LR]
- Mean wind speed [Wsmin]
- Multiple Linear Regression [MLR]
- Maximum Wind Speed [Wsmax]
- Random Forest Regression [RFR]
- Surface Reflectance [SR]
- Support Vector Regression [SVR]
- Land Cover [LR]
- Gaussian Process Regression [GPR]
- Mean Squared Error [MSE]
- Boosted regression Tree [BRT]
- Mean Absolute Error [MAPE]
- Back Propagation Neural Network [BP]
- Root Mean Square Error [RMSE]
- Moderate Resolution imaging spectroradiometer [MODIS]
- Transplanting [TR]
- 3-dimensional CNN [3DCNN]
- Date of Sowing [SO]
- Recurrent Neural Network [RNN]
- Emergence [EM]
- Normalized difference Vegetation Index [NDVI]
- Three Leaves [TL]
- Vegetation Index [VI]
- Enhanced Vegetation Index [EVI]
- Land Surface Daytime Temperature [LST]
- Precipitation [Prec]
- Monthly Mean Temperature [Tair]
- Vapor Pressure Deficit [VPD]
- Mean Decrease Impurity [MDI]
- Boost Tree [BST]
- Bagging Tree [BGT]
- Soil Adjusted Vegetation Index [SAVI]
- Perpendicular Vegetation Index [PVI]
- Green Vegetation Index [GVI]
- Leaf Area Index [LAI]
- Elastic Net [EN]

References

1) Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*. 2018;151:61–69. Available from: https://doi.org/10.1016/j.compag.2018.05.012.
2) Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine Learning in Agriculture: A Review. *Sensors*. 2018;18(8):1–29. Available from: https://doi.org/10.3390/s18082674.
3) Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*. 2018;155:257–282. Available from: https://doi.org/10.1016/j.compag.2018.10.024.
4) Drury B, Valverde-Rebaza J, Moura MF, Lopes ADA. A survey of the applications of Bayesian networks in agriculture. *Engineering Applications of Artificial Intelligence*. 2017;65:29–42. Available from: https://doi.org/10.1016/j.engappai.2017.07.003.
5) Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*. 2020;177:1–18. Available from: https://doi.org/10.1016/j.compag.2020.105709.
6) Rashid M, Bari BS, Yusup Y, Kamaruddin MA, Khan N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. *IEEE Access*. 2021;9:63406–63439. Available from: https://doi.org/10.1109/ACCESS.2021.3075159.
7) Sharma A, Jain A, Gupta P, Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*. 2021;9:4843–4873. Available from: https://doi.org/10.1109/ACCESS.2020.3048415.
8) Sharma P, Dadheech P, Aneja N, Aneja S. Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*. 2023;11:111255–111264. Available from: https://doi.org/10.1109/ACCESS.2023.3321861.
9) Ashfaq M, Khan I, Alzahrani A, Tariq MU, Khan H, Ghani A. Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion. *IEEE Access*. 2024;12:40947–40961. Available from: https://doi.org/10.1109/ACCESS.2024.3376735.
10) Sunil GL, Nagaveni V, Shruthi U. A Review on Prediction of Crop Yield using Machine Learning Techniques. In: 2022 IEEE Region 10 Symposium (TENSYMP). IEEE. 2022. Available from: https://doi.org/10.1109/TENSYMP54529.2022.9864482.
11) Nejad SMM, Abbasi-Moghadam D, Sharifi A, Farmonov N, Amankulova K, Laszlz M. Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2023;16:254–266. Available from: https://doi.org/10.1109/JSTARS.2022.3223423.
12) Hoque MJ, Islam MS, Uddin J, Samad MA, De Abajo BS, Vargas DLR. Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning. *IEEE Access*. 2024;12:47768–47786. Available from: https://doi.org/10.1109/ACCESS.2024.3383309.
13) Aghighi H, Azadbakht M, Ashourloo D, Shahrabi HS, Radiom S. Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2018;11(12):4563–4577. Available from: https://doi.org/10.1109/JSTARS.2018.2823361.
14) Ban HY, Kim KS, Park NW, Lee BW. Using MODIS data to predict regional corn yields. *Remote Sensing*. 2017;9(1):1–19. Available from: https://doi.org/10.3390/rs9010016.
15) Prasad NR, Patel NR, Danodia A. Crop yield prediction in cotton for regional level using random forest approach. *Spatial Information Research*. 2021;29(2):195–206. Available from: https://doi.org/10.1007/s41324-020-00346-6.
16) Cao J, Zhang Z, Tao F, Zhang L, Luo Y, Han J, et al. Identifying the Contributions of Multi-Source Data for Winter Wheat Yield Prediction in China. *Remote Sensing*. 2020;12(5):1–22. Available from: https://doi.org/10.3390/rs12050750.
17) Nevavuori P, Narra N, Lipping T. Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*. 2019;163:104859. Available from: https://dx.doi.org/10.1016/j.compag.2019.104859.
18) Elavarasan D, Vincent PMD. Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. *IEEE Access*. 2020;8:86886–86901. Available from: https://doi.org/10.1109/ACCESS.2020.2992480.
19) Abbas F, Afzaal H, Farooque AA, Tang S. Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*. 2020;10(7):1–16. Available from: https://doi.org/10.3390/agronomy10071046.
20) Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, et al. Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators*. 2021;120:1–11. Available from: https://doi.org/10.1016/j.ecolind.2020.106935.
21) Gavahi K, Abbaszadeh P, Moradkhani H. DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert Systems with Applications*. 2021;184:115511. Available from: https://dx.doi.org/10.1016/j.eswa.2021.115511.
22) Sun Z, Di L, Fang H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International Journal of Remote Sensing*. 2019;40(2):593–614. Available from: https://dx.doi.org/10.1080/01431161.2018.1516313.
23) Sagan V, Maimaitijiang M, Bhadra S, Maimaitiyiming M, Brown DR, Sidike P, et al. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021;174:265–281. Available from: https://doi.org/10.1016/j.isprsjprs.2021.02.008.
24) Bose P, Kasabov NK, Bruzzone L, Hartono RN. Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series. *IEEE Transactions on Geoscience and Remote Sensing*. 2016;54(11):6563 –6573. Available from: https://doi.org/10.1109/TGRS.2016.2586602.
25) Ji S, Zhang C, Xu A, Shi Y, Duan Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sensing*. 2018;10(1):1–17. Available from: https://dx.doi.org/10.3390/rs10010075.
26) Yang Q, Shi L, Han J, Zha Y, Zhu P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*. 2019;235:142–153. Available from: https://dx.doi.org/10.1016/j.fcr.2019.02.022.
27) Yuan Q, Shen H, Li T, Li Z, Li S, Jiang Y, et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*. 2020;241:111716. Available from: https://dx.doi.org/10.1016/j.rse.2020.111716.
28) Wang X, Huang J, Feng Q, Yin D. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches. *Remote Sensing*. 2020;12(11):1–20. Available from: https://dx.doi.org/10.3390/rs12111744.
29) Shahhosseini M, Hu G, Archontoulis SV. Forecasting Corn Yield With Machine Learning Ensembles. *Frontiers in Plant Science*. 2020;11:1–16. Available from: https://dx.doi.org/10.3389/fpls.2020.01120.
30) Zhang L, Zhang Z, Luo Y, Cao J, Tao F. Combining Optical, Fluorescence, Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches. *Remote Sensing*. 2019;12(1):1–20. Available from: https://dx.doi.org/10.3390/rs12010021.
31) Sun J, Lai Z, Di L, Sun Z, Tao J, Shen Y. Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* . 2020;13:5048–5060. Available from: https://doi.org/10.1109/JSTARS.2020.3019046.

32) Sun J, Di L, Sun Z, Shen Y, Lai Z. County-level soybean yield prediction using deep CNN-LSTM model. *Sensors [Switzerland]*. 2019;19(20):1–21. Available from: https://doi.org/10.3390/s19204363.
33) Swanth B, Sheela J. Agricultural Yield Forecasting with the Fusion of Transfer Learning and LSTM Approach. In: 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). IEEE. 2024;p. 1268–1277. Available from: https://doi.org/10.1109/IDCIoT59759.2024.10467515.
34) Sabahath A, Begum A, Kiran A, Abdullaeva F, Reddy PCS, Dey S. An Efficient Transfer-Learning based Approach for Crop Yield Prediction using Remote Sensing Data. In: 2024 First International Conference on Software, Systems and Information Technology (SSITCON). IEEE. 2024. Available from: https://doi.org/10.1109/SSITCON62437.2024.10796392.
35) Zhang S, Qi X, Duan J, Yuan X, Zhang H, Feng W. Comparison of Attention Mechanism-Based Deep Learning and Transfer Strategies for Wheat Yield Estimation Using Multisource Temporal Drone Imagery. *IEEE Transactions on Geoscience and Remote Sensing*. 2024;62:1–23. Available from: https://doi.org/10.1109/TGRS.2024.3401474.
36) Sapthami I, Krishna BM. Transfer Learning Techniques For Improving The Quality Of Prediction In Crop Yield. In: 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE. 2023. Available from: https://doi.org/10.1109/ICCCI56745.2023.10128344.
37) Celik MF, Isik MS, Taskin G, Erten E, Camps-Valls G. Explainable Artificial Intelligence for Cotton Yield Prediction With Multisource Data. *IEEE Geoscience and Remote Sensing Letters* . 2023;20:1–5. Available from: https://doi.org/10.1109/LGRS.2023.3303643.
38) Konda O, Mohammad RAS, Mishra S, Rajeev N, Verma A. Harvesting Insights: Leveraging Explainable AI to Optimize Farming Practices. In: 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET). IEEE. 2024. Available from: https://doi.org/10.1109/ACROSET62108.2024.10743799.
39) Martin RJ, Mittal R, Malik V, Jeribi F, Siddiqui ST, Hossain MA. XAI-Powered Smart Agriculture Framework for Enhancing Food Productivity and Sustainability. *IEEE Access*. 2024;12:168412 –168427. Available from: https://doi.org/10.1109/ACCESS.2024.3492973.
40) Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*. 2019;274:144–159. Available from: https://doi.org/10.1016/j.agrformet.2019.03.010.
41) Deines JM, Patel R, Liang SZ, Dado W, Lobell DB. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*. 2021;253:1–16. Available from: https://doi.org/10.1016/j.rse.2020.112174.
42) Han J, Zhang Z, Cao J, Luo Y, Zhang L, Li Z, et al. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sensing*. 2020;12(2):1–22. Available from: https://doi.org/10.3390/rs12020236.
43) Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*. 2018;13(11):1–13. Available from: https://doi.org/10.1088/1748-9326/aae159.
44) Schwalbert RA, Amado T, Corassa G, Pott LP, Prasad PVV, Ciampitti IA. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*. 2020;284:107886. Available from: https://doi.org/10.1016/j.agrformet.2019.107886.
45) Han D, Wang P, Tansey K, Zhang S, Tian H, Zhang Y. Improving Wheat Yield Estimates by Integrating a Remotely Sensed Drought Monitoring Index into the Simple Algorithm for Yield Estimate Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2021;14:10383–10394. Available from: https://doi.org/10.1109/JSTARS.2021.3119398.
46) Shrestha R, Di L, Yu EG, Kang L, Shao YZ, Bai YQ. Regression model to estimate flood impact on corn yield using MODIS NDVI and USDA cropland data layer. *Journal of Integrative Agriculture*. 2017;16(2):398–407. Available from: https://doi.org/10.1016/S2095-3119(16)61502-2.
47) Luciani R, Laneve G, Jahjah M. Agricultural monitoring, an automatic procedure for crop mapping and yield estimation: The great rift valley of Kenya case. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2019;12(7):2196–2208. Available from: https://doi.org/10.1109/JSTARS.2019.2921437.
48) Lykhovyd PV. Sweet corn yield simulation using normalized difference vegetation index and leaf area index. *Journal of Ecological Engineering*. 2020;21(3):228–236. Available from: https://doi.org/10.12911/22998993/118274.