



REVIEW

An Integrated Analysis of Yield Prediction Models: A Comprehensive Review of Advancements and Challenges

Nidhi Parashar¹, Prashant Johri¹, Arfat Ahmad Khan⁵, Nitin Gaur¹ and Seifedine Kadry^{2,3,4,*}

¹School of Computer Science and Engineering, Galgotias University, Greater Noida, 203201, India

²Department of Applied Data Science, Noroff University College, Kristiansand, 4612, Norway

³Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, 346, United Arab Emirates

⁴MEU Research Unit, Middle East University, Amman, 11831, Jordan

⁵Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, 40002, Thailand

*Corresponding Author: Seifedine Kadry. Email: seifedine.kadry@noroff.no

Received: 13 January 2024 Accepted: 11 May 2024 Published: 18 July 2024

ABSTRACT

The growing global requirement for food and the need for sustainable farming in an era of a changing climate and scarce resources have inspired substantial crop yield prediction research. Deep learning (DL) and machine learning (ML) models effectively deal with such challenges. This research paper comprehensively analyses recent advancements in crop yield prediction from January 2016 to March 2024. In addition, it analyses the effectiveness of various input parameters considered in crop yield prediction models. We conducted an in-depth search and gathered studies that employed crop modeling and AI-based methods to predict crop yield. The total number of articles reviewed for crop yield prediction using ML, meta-modeling (Crop models coupled with ML/DL), and DL-based prediction models and input parameter selection is 125. We conduct the research by setting up five objectives for this research and discussing them after analyzing the selected research papers. Each study is assessed based on the crop type, input parameters employed for prediction, the modeling techniques adopted, and the evaluation metrics used for estimating model performance. We also discuss the ethical and social impacts of AI on agriculture. However, various approaches presented in the scientific literature have delivered impressive predictions, they are complicated due to intricate, multifactorial influences on crop growth and the need for accurate data-driven models. Therefore, thorough research is required to deal with challenges in predicting agricultural output.

KEYWORDS

Machine learning; crop yield prediction; deep learning; remote sensing; long short-term memory; time series prediction; systematic literature review

Abbreviations

Acronym	Full Form
R	Rainfall
T	Air Temperature
LST	Land Surface Temperature



RH	Relative Humidity
ET	Evapotranspiration
WS	Wind Speed
Sr	Solar Radiation
VP	Vapor Pressure
EDD	Extreme Degree Days
GDD	Growing Degree Days
VPD	Vapor Pressure Deficit
SOC	Soil Organic Carbon
BD	Bulk Density
N	Nitrogen
P	Phosphorus
K	Potassium
CC	Clay Content
CEC	Cation Exchange Capacity
SM	Soil Moisture
SR	Surface Reflectance
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
NDWI	Normalized Difference Wetness Index
TCI	Temperature Condition Index
VTCI	Vegetation Temperature Condition Index
GCI	Green Chlorophyll Index
VCI	Vegetation Condition Index
RI	Ripening Index
WDVI	Weighted Difference Vegetation Index
SAVI	Soil Adjusted Vegetation Index
PVI	Perpendicular Vegetation Index
GVI	Green Vegetation Index
LAI	Leaf Area Index
TV	Tree Volume
EC	Electrical Conductivity
AWC	Available Water Capacity
NIRv	Near-Infrared Reflectance of Vegetation
SIF	Solar Induced Chlorophyll Fluorescence
SMDI	Soil Moisture Deficit Index
PDSI	Palmer Drought Severity Index
SPI	Standard Precipitation Index
DNN	Deep Neural Network
BPNN	Back Propagation Neural Networks
GRNN	Generalized Regressive Neural Network
RBFNN	Radial Basis Function Neural Networks
GNN	Graph Neural Network
WOFOST	World Food Studies
CERES	Crop Environment Resource Synthesis
ARIMA	Autoregressive Integrated Moving Average
APAR	Absorbed Photosynthetically Active Radiation

3D-CNN	3-Dimensional Convolutional Neural Network
LCCC	Lin's Concordance Correlation Coefficient
APSIM	Agricultural Production Systems Simulator
MSE	Mean Square Error
RMSE	Root Mean Square Error
NRMSE	Normalized Root Mean Square Error
MAE	Mean Absolute Error
RMAE	Root Mean Absolute Error
MSLE	Mean Squared Logarithmic Error
MedAE	Median Absolute Error
RMedSE	Root Median Square Error
R	Correlation Coefficient
MBE	Mean Biased Error
MAPE	Mean Absolute Percentage Error
MFE	Mean Forecasting Error
RRMSE	Relative Root Mean Square Error
RPD	Residual Prediction Deviation
LR	Linear Regression
MLR	Multiple Linear Regression
SLR	Stepwise Linear Regression
ELNET	Elastic Network
UAV	Unmanned Aerial Vehicle
PCA	Principal Component Analysis
KNN	K-Nearest Neighbour
ELM	Extreme Learning Machine
RF	Random Forest
SVM	Support Vector Machines
SVR	Support Vector Regression
DT	Decision Tree
BST	Boosting Tree
BGT	Bagging Tree
SDG	Stochastic Descending Gradient
GPR	Gaussian Process Regression
GBM	Gradient Boosting Machine
XGBOOST	Extreme Gradient Boosting
ADABOOST	Adaptive Boosting
NN	Neural Network
MLP	Multiple-Layer Perceptron
ANN	Artificial Neural Network
GRU	Gated Recurrent Network
RBF	Radial Basis Function
SNN	Spiking Neural Network
XY-FS	XY-Fused Networks
SKN	Supervised Kohonen Networks
SOM	Self-Organizing Maps
RNN	Recurrent Neural Network
BMA	Bayesian Model Averaging

DFNN	Deep Feedforward Neural Network
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
LSTM	Long Short-Term Memory
RPIQ	Ratio of Prediction Performance to Inter-Quartile Range
LASSO	Least Absolute Shrinkage and Selection Operator
CAFFE	Convolutional Structure for Rapid Feature Embedding
DQN	Deep Q Network

1 Introduction

With the world population increasing and farmland becoming scarcer, the agriculture industry has to increase production and maintain an adequate food supply. According to the Food and Agriculture Organization (FAO), agribusiness contributes to about 3.9% of the world's GDP as it is the source of income for more than 40% of the population across the globe. AI for agriculture has transformational benefits across a variety of uses. Precision agriculture employs AI to improve the utilization of resources, increasing crop yields while reducing inputs such as fertilizer and irrigation water. AI offers early disease identification and crop monitoring, leading to reduced crop losses. Weather forecasting helps farmers make accurate choices to plant and harvest times. Automation improves supply-chain operations and lowers labor expenses, resulting in better productivity and earnings in agriculture. The unpredictability of weather, a changing climate, and limited resources pose critical challenges to agricultural output and reliability. Weather change impacts plant growth during different phases, resulting in considerable in-season yield variation. The regional variation in soil characteristics, irrigation regime, pest control, fertilizer usage, rotating crops, and cropland preparation routines add to the challenge of accurately evaluating crop production. On-field operations in cropping systems seem too expensive owing to their complexities. The reality is that they are associated with sensitive plant/edible product interactions [1]. Advanced software tools for crop yield prediction can transform farming methods. These applications use Artificial Intelligence (AI), machine learning, and data analytics to analyze enormous agricultural data, including weather trends, soil conditions, crop conditions, and past yield statistics. Farmers and government may utilize such data to make accurate choices about crop choice, planting strategies, utilization of resources, and handling risks. Accurate crop yield prediction has significant economic implications. The studies show that improved crop yield predictions might result in high-cost savings and enhanced food production worldwide. This research provides insights into the enormous significance of data-driven techniques that affect the agricultural sector through an in-depth analysis of current developments and future directions.

1.1 Crop Yield Prediction Modeling Techniques

The effect of environmental factors on crop production changes depending on the stage of crop development. Mechanistic models such as crop growth models are mathematical depictions of plant growth and development processes that integrate existing knowledge of plant growth with crop physiology, soil science, meteorology, and management practices to produce reliable predictions [2]. Fig. 1 shows the different crop yield prediction models used in the literature.

The Agricultural Production Systems Simulator (APSIM) [3], decision support system for agriculture technology transfer (DSSAT) [4], World Food Studies Simulation Model (WOFOST) [5], Crop Environment Resource Synthesis (CERES) [6], and AquaCrop are some highly utilized crop simulation models in the literature. A study [6] examined two methods for integrating seasonal climate

predictions with crop models to enhance crop production predictability. A study [7] assessed the ability of crop model predictor variables to replicate crop yields variations and showed that they performed better than weather variables to explain yield variability. Weather data from the Australian seasonal climate model is utilized as input to a crop growth simulation model to predict yield [8]. A study [9] forecasted corn and soybean yields and plant phenology on soil nitrogen and water dynamics using the APSIM crop model. However, the requirement of extensive input data makes the crop models expensive for complex calibrations. Additional algorithms for analyzing data have become influential as computational processing power has increased to work on larger datasets [10]. As shown in Fig. 2, another type of yield prediction technique are data-driven techniques based on historical yield data such as statistical models, ML, and DL techniques.



Figure 1: Types of crop yield prediction models

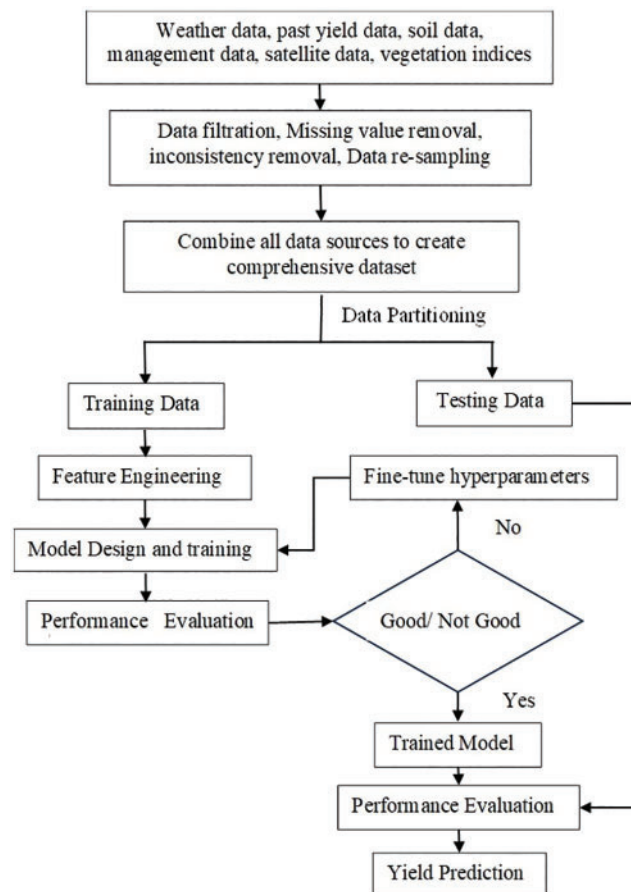


Figure 2: Data-driven crop yield prediction workflow

These techniques consider a set of most effective input parameters as independent variables contributing to the yield variation as the dependent variable. However, statistical methods can model unidentified management strategies, heat, excessive rainfall, periods of drought, pest problems, diseases, and unusual weather. These weather fluctuations necessitate the systems to be loaded with a large volume of information to develop dependable models. The researchers have integrated crop models with ML to add flexibility to the heavy crop simulation models. These meta-models are ML models trained on simulation models to make them run much faster than before. A slightly less accurate meta-model is better than more accurate but computationally slow simulation models. A study [11] explored the potential of LASSO, Ridge Regression, RF, XGBoost, and their ensembles for APSIM for predicting simulated yields of maize and loss of nitrogen with data available at the time of planting. XGBoost was best-performing in yield prediction with an RRMSE of 13.5%, RF was the best-performing in N-loss prediction with an RRMSE of 54%, and the prediction accuracy increased with the inclusion of ML-ensembles. A meta-model is effective for local impact assessment without comprehensive input datasets [12]. There has been little research in this field that has integrated crop models with linear regression. The primary approach is integrating yield trends into crop models using regression analysis [13–16].

1.2 Contribution of Present Study

This review study focuses on various modeling techniques for yield prediction tasks, including crop growth modeling, ML-based yield prediction, DL-based yield prediction, and meta-modeling. The research studies are analyzed based on the crop, input parameters fed into the prediction model, the modeling techniques, and evaluation metrics for assessing the model performance. The focus is on the recent research on crop yield prediction published from January 2016 to March 2024. This study includes a critical analysis of the importance of various input parameters used to carry out crop yield prediction. This study includes research for analyzing the usage and importance of input parameters from 2007 till the present. The summary of yield prediction techniques is also presented in a handy tabular manner so that the researchers can quickly read it before starting their research. This research also discusses the benefits and challenges of using AI for agriculture based on the research conducted on 125 research papers. No existing review paper based on crop yield prediction methodologies covered the comprehensive discussion on meta-modeling techniques used for crop yield prediction tasks. This review paper will be informative for scholars who pursue research in predicting agricultural yields.

1.3 Organization of Paper

The remaining part of the paper is as follows. [Section 2](#) elaborates on the research. This section addresses the methodology employed in the literature selection process. [Section 3](#) discusses the effectiveness of multiple input factors impacting yield. [Section 4](#) presents the reported work on machine learning, meta-modeling, and deep learning perspectives in crop yield prediction. [Section 5](#) provides a comprehensive overview of all the investigations. [Section 6](#) covers the paper's conclusion, and [Section 7](#) presents the future research directions.

2 Research Methodology

This section describes the procedure used for reviewing the crop yield prediction studies. Articles considered in the present literature study are from the most recent innovations done in research undertaken in modern agriculture. This survey article includes four main steps: Set Research Objectives,

search databases for records based on associated keywords, select related research articles for detailed analysis, and finally attain and report on the research objectives, as shown in Fig. 3.

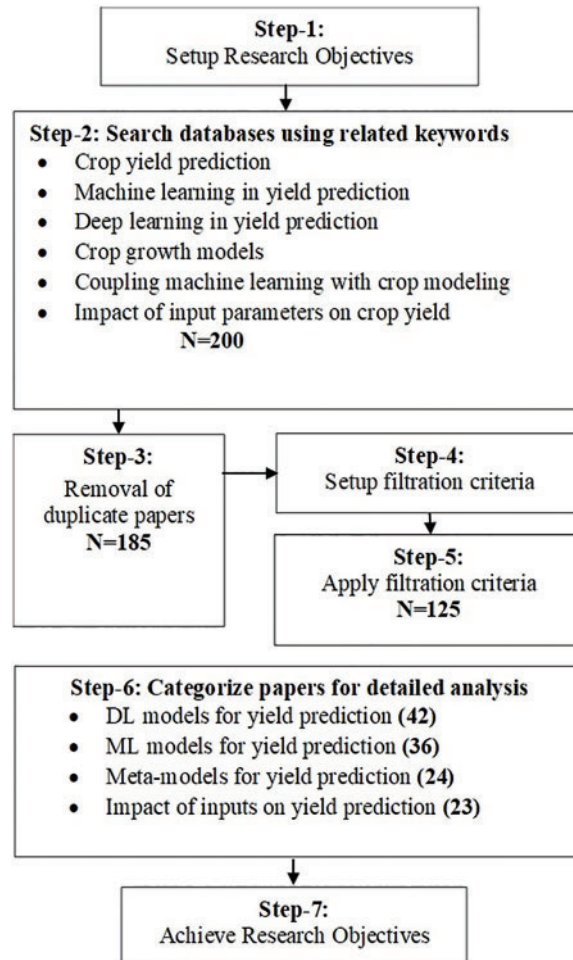


Figure 3: Research methodology

The identified papers are analyzed one by one, considering the following research objectives:

RO1: What input parameters were utilized for yield prediction?

RO2: Which crop was considered to develop the yield prediction model?

RO3: Which modeling techniques were used to predict yield?

RO4: What evaluation metrics were considered to report the accuracy of prediction?

RO5: Which challenges do researchers face in building artificial intelligence-driven prediction models?

Several databases like Scopus, ScienceDirect, Web of Science, Springer, IEEE Explorer, Google Scholar, and Wiley are searched for related work to answer predefined research objectives. Different combinations of keywords used to search for relevant research publications in databases are “Crop yield prediction”, “coupling machine learning with crop modeling,” “machine learning in crop yield prediction”, “deep learning in crop yield prediction”, “impact of input parameters on crop yield”.

We retrieved 200 research papers from the literature. Before studying the research papers in detail, we performed two-phase scanning on the retrieved research papers. During the initial scanning phase, we removed the duplicate papers. In the second phase, we defined some filtration rules for choosing final research papers for our study. Fig. 4 presents the year-wise categorical distribution of research papers based on crop yield prediction considered in this review.

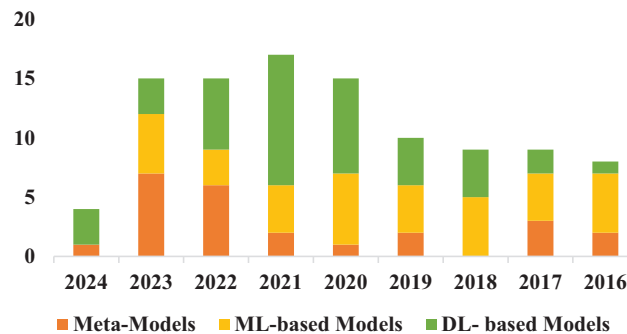


Figure 4: Year-wise categorical distribution of research papers based on crop yield prediction models

The following filtration criteria were applied to select the research papers for review:

Filtration criteria 1: The research article was not related to agriculture.

Filtration criteria 2: The article's publication year is before 2016.

Filtration criteria 3: The article presented a systematic review of any agricultural research.

Filtration criteria 4: The article was not based on the crop yield prediction.

Filtration criteria 5: The article did not use crop growth modeling, meta-modeling, ML, or DL techniques for yield prediction.

Filtration criteria 6: The article text was partially available due to restricted access.

Publication year-based filtration is not applied to select the research papers for input parameters analysis. After filtering, the total number of articles studied thoroughly in this review is 125. The above graph does not include the research on input parameter analysis. The articles that performed prediction based on deep learning are the highest, 42 in total. The articles based on machine learning are 36. The number of articles that used meta-modeling is 24. A total of 23 articles cover the analysis of the effectiveness of input parameters in crop yield prediction. Additionally, they are incorporated and placed according to the research objectives stated in the present research.

3 Efficacy of Yield-Affecting Parameters

To predict crop yields, a variety of available meteorological data, soil data, satellite-based observations, and management practices have been used. This section covers the input parameters presented in Fig. 5 used in crop yield prediction and their effectiveness in predicting crop yields.

Weather data such as rainfall (R), air temperature (T), land surface temperature (LST), relative humidity (RH), evapotranspiration (ET), wind speed (WS), and solar radiation (Sr), and vapor pressure (VP) impact the crop yield in a nontrivial manner. A study [17] found that rainfall, temperature, and Sr were the more influential in yield prediction. Study [18] also assessed the impact of meteorological parameters on the corn yield and presented that R, VP, T, and Sr were the most influential parameters with a relevance factor between 0.021 and 0.033. A study [19] reported that

the correlation between LST and yield is higher than air temperature. Biotic stresses like weeds, pests, and diseases and abiotic stresses like soil salinity, heavy metal content, heat, water, and drought also impacted the yields. To quantify the impact of temperature on yield, heat stress such as extreme degree days (EDD), growing degree days (GDD), and vapor pressure deficit (VPD) have been used during the crop growing season [20,21]. Stable soil properties such as soil depth or granularity, bulk density (BD), clay content, soil organic carbon content (SOC), and dynamic soil properties such as pH, content of nutrients such as nitrogen (N), phosphorus (P), potassium (K), soil moisture (SM) and temperature are considered by the researchers. Remotely sensed surface reflection (SR) has also been an efficient predictor variable. Satellite observations record the yield evolution variability with season advancements, contributing to the peak growing months.

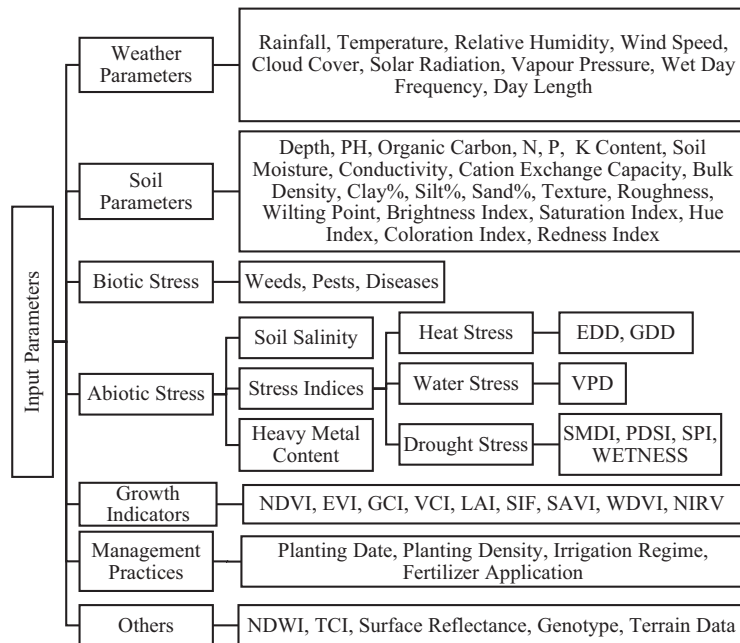


Figure 5: Agricultural yield-affecting parameters

For estimating crop growth and differentiating the crop area, optical bands determine vegetation indices (VIs) such as enhanced VI (EVI), normalized difference VI (NDVI), normalized difference wetness index (NDWI), temperature condition index (TCI), green chlorophyll Index (GCI) and vegetation condition index (VCI) [22–24]. Study [22] used VCI derived from NDVI products to forecast sugarcane production at the county level using Stepwise linear regression. Although these linear models have been very good at simulating seasonality trends, outliers are hard to predict for ARIMA since they are outside the general pattern as caught by the model. Some studies also utilized the Leaf Area Index (LAI) [25,26]. Study [27] predicted future crop yield of wheat, corn, and cotton and yield deficit under soil salinity implications using VIs and ground data. Under different soil salinity, they reported that salt between (8–10) caused 28% losses for wheat, 55% yield loss for corn, and 15% for cotton. Another indicator of plant cover, near-infrared reflectance of vegetation (NIRv), represents the fraction of pixel radiance related to plants in the pixel from the perspective of reality. In some cases, NIRv estimates were less accurate than those calculated by the NIRv or VIs at the growth phase. Soil-adjusted vegetation index (SAVI), weighted difference vegetation index (WDVI), perpendicular vegetation index (PVI), green vegetation index (GVI), and solar-induced chlorophyll fluorescence

(SIF) have been used as fair predictors of vegetation phenomena [28,29]. SIF was directly related to crop yield than VIs [30]. However, EVI performed better than SIF in some cases where the temporal and spatial resolution of the remotely sensed images was low [31].

Satellite-derived soil moisture data provide crucial insights into environmental stress and additional hydrological data such as drought stress and groundwater depth during the growing period. In a study [32], ML models and six ensembles foresee corn yield in three corn-growing states in the US. They showed more accurate predictions for all ML models with APSIM hydrological data. It improved the accuracy of all models in RMSE by 27% [32–36]. Soil moisture deficit index (SMDI), palmer drought severity index (PDSI), and Standard precipitation index (SPI) have been used to assess the impact of drought [37,38]. A study [39] investigated that integrating EVI and LST with VPD and rainfall notably improved the model performance. Management inputs like planting density, planting date, irrigation, and fertilizer composition also provide valuable insights for yield prediction [40,41]. A few studies focused on describing yield as a direct function of genotype, environment, and their ability to interact in selecting the most appropriate cultivars for target regions. A study [42] proposed a stacked LSTM with attention to foreseeing genotype reactions in different settings using per-week climatic variables. To forecast corn yield across the US, a study [43] developed a DNN method using a dataset that included six weather parameters and eight soil characteristics.

3.1 Handling Unbalanced Datasets and Image Pre-Processing

In crop yield prediction, dealing with unbalanced data sets and incorporating image pre-processing are essential for building accurate and robust models. Resolving the data imbalance is crucial since farming data often indicate imbalances in crop yield categories, with some yield levels overrepresented. The oversampling, under-sampling, and cost-sensitive learning techniques guarantee that all yield categories are fairly represented, minimizing the potential risk of biased forecasts by models. Additionally, image pre-processing is critical in maximizing the utility of visual information such as satellite imaging or field images, which provide significant information about crop conditions and external variables influencing yield. Some image pre-processing techniques include image resizing and cropping, normalization, color space transformations, morphological operations, noise reduction, edge detection, texture analysis, histogram equalization, and segmentation to ensure consistency across images, enhance image clarity and to focus on relevant image parts only that indicate crop wellness, crop growth periods, and atmospheric conditions.

4 Crop Yield Prediction

This section covers the research papers related to crop yield prediction using machine learning, meta modeling, and deep learning.

4.1 Machine Learning for Crop Yield Prediction

This section covers the application of several ML methods by various researchers to predict yields for diverse crops. Researchers have taken the task of yield prediction as a time series analysis problem [44]. Certain time functions investigate the aggregate impact of weather on agricultural productivity. Some frequently used parameterized linear time series models for yield prediction are Autoregressive Integrated Moving Average (ARIMA) for stationary time series and seasonal ARIMA for time series with seasonality component [45]. A study [46] used regression analysis and time-series modeling techniques for bajra yield prediction using rainfall, temperature, and past yield. Among all

the implemented models, ARIMAX was best performing with a Root Mean Absolute Error (RMAE) of 2.82%, Root Mean Square Error (RMSE) of 3.86%, and Mean Absolute Error (MAE) of 50.67%.

Study [47] evaluated the yield variation predictions by utilizing a cross-ensemble empirical method to temperature and rainfall results extracted from a cross-ensemble. The Mosaic method gave reliable predictions over 25%–38% of the cultivated area. ML models like support vector machines (SVM) and random forest (RF) are more resistant to overfitting when dealing with nonlinear data. SVM aims to find the optimal hyperplane for classification or regression by optimizing the space between observations and the decision boundary while minimizing classification error. Mathematically, it involves finding the vector of weights (w) and bias (b) that define the hyperplane using $y = w * x + b$. Here, y is the predicted class label, x is the input feature vector, and w and b are learned parameters. LASSO regression has been used in yield prediction by selecting relevant features and mitigating overfitting, thus improving the accuracy of crop yield forecasts. The LASSO objective function is $F = RSS + \lambda * \sum |\beta_i| \lambda$. Here, RSS represents the Residual Sum of Squares, which measures the error between predicted and actual values. β_i represents the regression coefficients for each feature. λ regulates the degree of regularization. Higher λ values lead to more feature selection and sparsity in coefficients.

Study [48] performed the evaluation prediction models such as elastic NET, LASSO, SLR, and PCA-LR based on numerical weather data. Based on the overall average ranking of the models, ELNET was the best-performing. In a study [49], three RF models based on timestamps like Pre-sowing, in-season, and end-season used proximally sensed on-farm data. Study [50] used satellite-based images along with ML models such as LR, RF, NN, SVM, GPR, and Cubist for corn yield prediction at a spatial scale. NN was most accurate for SOM, SVM was most accurate for K, GBM was most accurate for pH, and RF gave the best accuracy. To predict maize and potato yield, reference [51] used RF, SVR, and polynomial regression using rainfall and temperature as predictor variables. Study [52] complemented satellite data (EVI and SIF) by providing weather data (VPD, R, T) to give valuable insights across the whole crop growing season of wheat and evaluated LASSO, SVM, RF, and NN. RF, NN, and SVM performed better than LASSO regression model.

Study [53] developed three ensemble models using a blocked sequential procedure, weather data, soil data, and management data. Compared with LR, LASSO, XGBoost, LightGBM, RF, stacked regression, and stacked versions performed the best with an RMSE of 9.5%. With the least biased predictions, Stacked LASSO topped with an MBE of 53 kg/ha. Study [54] predicted the wheat yields at the county level based upon multiple source datasets and SVM, Gaussian process regression (GPR), RF, KNN, NN, decision tree (DT), boosting tree (BST), and bagging tree (BGT). They used temperature, rainfall, drought index, soil moisture, soil PH, NDVI, and EVI as predictor variables. SVM, GPR, and RF were best-performing with $R^2 > 0.75$ and yield percentage error $< 10\%$. The study [55] assessed DT, RF, stacked sparse autoencoder, and SVM techniques for predicting rice, corn, and soybean yield using Sr, humidity, WS, temperature, rainfall, and vegetation indices from MODIS. SVM was the most effective way for crop yield prediction. ML-based yield prediction system [56] integrated climatic, weather, agricultural yield, and chemical data for making predictions using KNN, MLR, and a DT. The decision tree outperformed other models.

Study [57] utilized 2020 Syngenta Crop Challenge data and ML methods such as DT, gradient boosting machine, RF, adaptive boosting, XGBoost, and neural networks to estimate the corn yield. With an RMSE of 0.0524, the prediction made by XGBoost was more accurate than those made by other models. A study [58] used RF and SVM classifiers using VIs, Sr, humidity, SOC, and WS to

predict wheat yield in China. Employing VIs increased model precision. Predicting yield with near-infrared reflectance of terrestrial vegetation (NIRv) was more accurate than with NDVI and EVI. The RF model using NIRv performed better than the SVM model. Study [59] combined principal component analysis (PCA) and ML with high-order spatial independent component breakdown to improve rice yield prediction accuracy. They integrated shared VCI/TCI geographic diversity into their corresponding subdomains. PCA-ML performed better than LR, SVM, and DT. Study [60] used Cross-validation feature selection algorithms with LR, RT, SVM, NN, Bayesian regression, and KNN to forecast alfalfa yield. The methodology based on correlation was the best performing. Study [61] developed AdaNaive, and AdaSVM ensembles for predicting yields of crops. The results indicated excellent accuracy. In a comparative analysis of RF and MLRs for maize, wheat, and potato yield prediction [62], RF models outperformed all other models. A study [63] used RF to predict crop yields in the kharif and rabi seasons. R-Tool analyzed the data, and the study [64] utilized a spiking neural network (SNN) on NDVI image time series. The study used a 9-feature model to predict crop yield. A study [65] used the RF method and the R package to forecast cotton yield. The predictor variables acquired from multi-sensor satellites such as NDVI, VCI, SPI, GDD, and LST improved the calibration and validation of the RF algorithm. Artificial Neural Network (ANN) offers a prevailing tool for crop yield prediction using historical agricultural data and complex patterns to provide accurate forecasts. ANN is represented mathematically using Eqs. (1) and (2).

$$Z_j^{(i)} = \sum (W_{ij}^{(i)} * a_{i-1}) + b_j^{(i)} \quad (1)$$

$$a_j^{(i)} = f(Z_j^{(i)}) \quad (2)$$

Study [66] applied SVM to analyze weather changes, the K-means method for classifying plants and soil, and used ANN to estimate crop production. Study [67] compared the performance of RF, DT, and ANN to predict the crop yields of staple crops in Morocco. ANN outperformed other models with improved MSE and R^2 values. Study [68] attempted a sugarcane yield prediction task using on-field observation and weather data using ELM to find the sugarcane yield estimates. They used data from the closest operational weather station and field data such as irrigation schedule and soil conductivity. ELM was the best-performing method compared to ANN and genetic models. A study [69] proposed a hybrid MLR-ANN approach for rice crop yield prediction. They used MLR slope and coefficient setting the ANN's input layer weights and biases for estimating accurate yield. The MLR-ANN outperformed conventional models. A study [70] used modular ANNs to predict monsoon rainfall. It used SVR to predict key kharif agricultural yields based on the rainfall data. The MANNs-SVR technique surpassed RF, LR, and KNN.

A study [71] evaluated wheat yield variation using real-time multilayered soil information and satellite imagery. They compared the performance of XY-fused Networks (XY-Fs), counter-propagation ANN, and Supervised Kohonen Networks (SKNs). SKN is an ANN that combines the self-organizing properties of Kohonen Self-Organizing Maps (SOMs) with supervised learning. Building the SOM network involves deciding which neuron is a winner and then adjusting the weights in the layers. A neuron with the smallest Euclidean gap between its weight and independent variables is declared the winner. The weights of the prevailing neuron modify according to $w(t, i, j, k) = w(t-1, i, j, k) + \eta(t) \cdot h(i, j, t) \cdot (o(k) - w(t-1, i, j, k))$. Here $w(t, i, j, k)$ corresponds to the new weight calculated in iteration t , $w(t-1, i, j, k)$ is the current weight from the preceding iteration, $o(k)$ is variable k of the object. The coordinates (i, j) give the neuron position, $\eta(t)$ denotes the learning rate function, and $h(i, j, t)$ denotes the neighborhood function to show how the weight adjustment varies depending on the distance from the winner neuron. SKN had the highest accuracy level of 81.65%.

A study [72] proposed an ANN method for fruit yield prediction using images. Two different BPNN models were built for the beginning and the ripening phase. R^2 , Mean Absolute Percentage Error (MAPE), Mean Forecasting Error (MFE), and RMSE were applied to analyze the results. In the study [73], ANN models forecasted maize yield in South Africa's grain-producing regions using rainfall, temperature, ET, SM, and NDVI. The proposed models gave accuracy-adjusted R^2 values of 0.75, 0.67, 0.86, and 0.82 for four provinces, respectively. Study [74] developed an ANN ensemble to estimate sugarcane crop yield using NDVI time series data. Neural network wrappers with serial reverse feature removal removed unnecessary features. Then, a stacked ensemble with ANN predicted sugarcane production. Stacking improved the results, reducing the RRMSE by 8% and the R^2 up to 0.43. Study [75] compared the effectiveness of ANN, modified ANN, and MLR for wheat yield prediction using precipitation, biomass, soil moisture loss, transpiration, Soil Water, and fertilizer data. C-ANN improved yield prediction, with higher R^2 values and lower % errors.

The impact of fuzzy logic on wheat yield prediction was examined [76]. This study offered a method for estimating wheat crop yields using actual yields and interval-based data division. Study [77] applied Cubist, SVM, DNN, RR, RF, and ensemble learning and used data from multiple sensors for wheat yield prediction. In every ML approach, multi-sensor information-based prediction outperformed individual-sensor data-based predictions. Ensemble learning outperformed all individuals with R^2 of 0.692, RMSE of 0.916 t ha⁻¹, Residual Prediction Deviation (RPD) of 1.771, and the ratio of prediction performance to inter-quartile range (RPIQ) of 2.602. A study [78] investigated a method for estimating coffee yield using Tree volume (TV), Ripening Index (RI), and NDVI and fed into LR, SVR, RF, DT, and Stochastic Descending Gradient (SDG). LR and SDG performed well, with a 56% and 55% correlation. Study [79] used a multiple-layer perceptron (MLP) with optimization methods to estimate MLP parameters. A hybrid gamma test selected the best input. The model reduced the MAE significantly. Table 1 shows the summary of ML techniques considered in this study.

Table 1: Summary of ML techniques considered in this study

Article	Crop	Input data	Algorithm	Compared models	Result
[2]	Potato	Field data, SM, EC, NDVI	LR, SVR ELNET, K-NN	All methods	SVR with RMSE: 4.62–6.60 t/ha
[22]	Sugarcane	NDVI	SLR	No model compared	Predicted~actual yield is <10%
[46]	Bajra	T, R	ARIMAX	ARIMA, LR	RMAE: 2.8 kg/ha RMSE: 3.8 kg/ha MAE: 50.6 kg/ha
[47]	Maize, rice, wheat, soybean	T, R	Ensemble	No model compared	Mosaic accuracy: (25%–38%)
[48]	Coconut	T, R, RH, WS, Sr	ELNET, LASSO, ANN, SLR, PCA-LR, PCA-ANN	All methods	Performance rank: PCA-ANN < PCA-SLR < SLR < ANN < LASSO < ELNET
[49]	Wheat, barley, canola	EVI, R, EC, on-farm data	RF models	All methods	RMSE: 0.36–0.42 t/ha LCCC: 0.89–0.92

(Continued)

Table 1 (continued)

Article	Crop	Input data	Algorithm	Compared models	Result
[50]	Corn	SOM, PH, CEC, NDVI	LR, RF, NN, SVM, GPR, Cubist	All methods	RF best with $R^2 = 0.5$ RMSE = 0.9
[51]	Irish potatoes, maize	R, T	RF, SVR, polynomial regression	All models	RF outperformed For potato: RMSE: 510.8 kg/ha R^2 : 0.875 For maize: RMSE: 129.9 kg/ha R^2 : 0.817
[52]	Wheat	EVI, SIF, R, T, VPD	LASSO, SVM, RF, NN	All methods	ML models better with $R^2 \sim 0.75$
[53]	Corn	Weather, soil, planting data	Three ensembles	LR, LASSO, XGBoost, LightGBM, RF	RMSE: 9.5% MBE: 53 kg/ha
[54]	Wheat	T, R, SM, PH, EVI, NDVI,	SVM, GPR, RF	KNN, NN, DT, BST, BGT, RF, GPR, SVM	$R^2 > 0.75$ PE < 10%
[55]	Rice, corn, soybean	VIs, Sr, RH, WS, T, R	SVM	DT, RF, ANN, LSTM, autoencoder	RRMSE: 0.56 t/ha
[56]	Rice, yam, corn, cassava, cotton, banana	T, R, pesticide usage, NO_2	DT	KNN, LR	$R^2 = 0.95$
[57]	corn	Genotype	XGBoost	DT, RF, NN, AdaBoost, XGBoost	RMSE: 0.0524 t/ha
[58]	Wheat	VIs, Sr, RH, SOC, WS	RF	SVM	R^2 : 0.74 RMSE: 758 kg/ha
[59]	Rice	VCI, TCI	PCA-ML	LR, SVM, DT	Average accuracy: 8.5%–45%
[60]	Alfalfa	T, SM, R, day length	RF, KNN	LR, RT, SVM, NN, Bayesian regression	MAE < 200 lbs./acre R^2 : 0.95
[61]	Rice, gram cotton, sugarcane, groundnut	T, CC, ET, VP, WDF, R	AdaNaive, AdaSVM	Naïve Bias, SVM	AdaSVM performed the best accuracy: 89.42% to 93.72%
[62]	Maize, wheat, potato	R, T, day length, FA, ET	RF	MLR	RMSE: 6% to 14%
[63]	Rice	R, T	RF	No model compared	RF performed good in visualization

(Continued)

Table 1 (continued)

Article	Crop	Input data	Algorithm	Compared models	Result
[64]	Wheat	NDVI, planting area	NN	Different feature size	$r = 0.81$ Accuracy: 95.64% Average error: 0.236 t/ha
[65]	Cotton	NDVI, VCI, SPI, GDD, LST	RF	All methods	RMSE: 62.77 kg/ha MAPE: 0.32 t/ha
[66]	Kharif, rabi, summer crops	T, R, RH, PH	ANN, SVM, K-Means	No model compared	ANN performed good in visualization
[67]	Wheat, apple, olive, date, almond	R, T, SM	ANN, RF, DT,	All models	MSE: 0.10 t/ha R^2 : 0.90
[68]	Sugarcane	Management data, soil data, T, R, RH, Sr	ANN, ELM, GP	ANN, ELM, GP	ELM best performing r : 0.47 R^2 : 0.89 RMSE: 0.8 t/ha
[69]	Rice	R, T, Sr, planting area, FP, IR	MLR-ANN	MLR, ANN, KNN, SVR, RF	MLR-ANN with RMSE: 0.051 t/ha MAE: 0.041 t/ha r : 0.99
[70]	Maize, bajra, rice, ragi	R	SVR, MANN	RF, LR, KNN	SVR best performing RMSE: $1696^{-1}t - 1709^{-1}t$
[71]	Wheat	NDVI, soil data	CP-ANNs, XY-Fs, SKNs	CP-ANNs, XY-Fs, SKNs	SKN performed the best with accuracy: 81.65%
[72]	Apple	Fruit images	ANN, BPNN	No model compared	R^2 : 0.83 MAPE: 8.9% MFE: 0.03 kg/tree RMSE: 2.3 kg/tree
[73]	Maize	NDVI, R, T, ET, SM	ANN	Different weights feature analysis	Adjusted R^2 : 0.75, 0.67, 0.86, 0.82 for four different provinces
[74]	Sugarcane	NDVI	ANN ensembles	Different stack size	RRMSE: 8% R^2 : 0.43
[75]	Wheat	R, SM, ET, SM, FP	C-ANN	ANN, MLR	The R^2 of C-ANN: 97% Average prediction error: 0.5275%
[76]	Wheat	Yield data	Fuzzy logic	No model compared	Low MSE and MFE

(Continued)

Table 1 (continued)

Article	Crop	Input data	Algorithm	Compared models	Result
[77]	Wheat	UAV multi-sensor data	Cubist, SVM, DNN, RR, RF, and their ensembles	All models	Ensemble learning outperformed R ² = 0.692 RMSE = 0.916 t/ha RPD = 1.771 RPIQ = 2.602
[78]	Coffee	TV, NDVI, RI	LR, SDG	SVR, RF, DT	R ² for TV: 71%, R ² for NDVI: 55%, R ² for RI: 63%,
[79]	Wheat	T, WS, RH, ET, R	MLP-HBA, MLP-SCA, MLP-SA, MLP	All models	MAE reduced by MLP: 61%

ML has found widespread application in environmental predictive modeling for its ability to deal with linear or nonlinear relationships, unusual data, and superior performance. A solitary ML model, on the other hand, can be underperformed by a group of modeling techniques (ML ensembles), which could decrease variance, prediction bias, or both and identify the implicit data distribution effectively [53].

4.2 Meta Modeling Based Crop Yield Prediction (Coupling ML and Crop Modeling)

The researchers speculated that combining prediction techniques, particularly crop modeling and ML models, would improve agricultural prediction. Certain studies employed crop model simulation outcomes as input to a multiple-regression model, creating an integrated simulated crops-regression framework to forecast crop yields [80–83]. Recent research has developed hybrid simulation crop modeling-ML models for crop yield prediction. Study [84] used simulations that included biomass from the APSIM, precipitation, temperature, and Sr to an RF model to predict yearly variations in local sugarcane yield. The combined model could forecast yields quite well. Study [85] developed a system that combined APSIM model results with drought, frost, and heat stress to forecast wheat production using the RF model. The hybrid model exceeded the hybrid of APSIM and MLR and the APSIM model alone in terms of R² and RMSE. They extended the work by combining APSIM, simulated biomass, NDVI, and SPEI. Study [86] developed a hybrid model for wheat yield prediction at the plot level using RF and MLR. The RF-based model predicted the yields with LCCC of 0.53, r of 0.62, RMSE of 1.01 t ha⁻¹, MAPE of 27.1%, and ROC score of 0.90.

In a study [87], the WOFOST crop model integrated with ridge regression, SVM, and gradient Boosting to forecast crop yields at subnational scales. They found that including crop simulation data in ML models improved prediction outcomes significantly. In hybrids, the prediction accuracy increased by 8%–9%. A stacked ensemble performed the best with the incorporation of APSIM inputs. They extended the concept and developed ML models and their ensembles, considering contexts including and without including crop model factors using soil, weather conditions, management practices, and past yield information [88]. The results showed that combining crop modeling with ML improved the accuracy.

Study [89] aimed to give a comparative evaluation of the approaches for tea crop yield forecasting using the UN's AquaCrop simulation framework and ML methods (SVR, AdaBoost, Automatic Relevance Determination, DT, MLP, MLR, Random Sample Consensus, LR, XGBoost, and SVM). They obtained the lowest MAE, MSE, and RMSE values using 10-fold cross-validation and XGBoost. Study [90] used a dynamic CERES-Maize to simulate the long-term maize yield and evapotranspiration with management practices and climatic conditions. Six ML models used the results from three growing seasons in experiments and the CERES model outputs. The XGBoost model performed better than other regression techniques in estimating evapotranspiration and yield based on $R^2 > 0.82$ and $RRMSE < 9\%$. A study [91] combined ANN, KNN, RF, and SVR with multiple crop process models of the DSSAT platforms. The findings showed that ANN and RF performed accurate wheat yield prediction. ML methods lowered uncertainty to 7.2% and the yield variation to 8.1%.

In a study [92], the predictive input data from GCM's prediction propelled the APSIM model. The LAI of the rice crop was estimated using air temperature, water vapor, aerosol, and NDVI data in ML and DNN modeling. RF was more efficient than DNN in LAI estimation. In a study [93], total biomass simulated by the APSIM and weather indicators estimated wheat production using the regression models. RF approach outperformed gradient boosting and MLR. RF predicted wheat yield accurately with r of 0.86, MAE of 498 kg ha^{-1} , and RMSE of 683 kg ha^{-1} . Study [94] examined ANN, multivariate adaptive regression splines, and RF for reproducing the APSIM chickpeas crop growth model. All the ML models performed well with $R^2 > 0.95$ at estimating outcomes for each region of the training data set. A study [95] conducted regional simulations using ML and the physiological model APSIM. To emulate APSIM predictions, they trained multivariate regression splines, RF, and boosting trees to forecast SOC variability and crop yield.

In a study [96], a hybrid strategy for predicting corn yield examined yield prediction capabilities at various stages of growth using the WOFOST model and GRU. The findings indicated that the primary features of the maize growth stage were its growth condition and water-related characteristics. The best model achieved an RMSE of 102.65–554.84 kg/ha. In a study [97], APSIM for maize monitors the daily process of biomass accumulation throughout the maize growth period and uses the amount of biomass produced daily to estimate the final grain yield. The findings indicate that the proposed model accomplished prediction with an RRMSE of 7.16%. Study [98] used simulated sunflower and wheat data from DSSAT to examine the impact of data volume, splitting strategies, and prediction technique choice on the accuracy of predictions. ANN and RF analyzed agricultural yields as a function of soil, management, and weather conditions. RF performed better, exhibiting an RMSE of 35%–38%. Study [99] proposed a hybrid framework for winter wheat yield forecasting by assimilating LAI and SM into a crop model and then combining it with ML models. The proposed model achieved an ACC of 0.97 and MAPE of 1.74%. To improve oil seed rape and wheat yields, a study [100] proposed a meta-model using RF and light use efficiency (LUE). They developed four RF models individually with NDVI, weather parameters, NDVI with weather parameters, and LUE with weather parameters. The final LUE coupled with the RF model reduced the RMSE by 8% and R^2 by 14.3%.

4.3 Deep Learning-Based Crop Yield Prediction

This section covers deep learning models to predict yields for diverse crops. A similar idea from ANN goes into Deep Neural Networks (DNN) algorithms. DL provides multiple levels of abstraction by adding more hidden layers to the model and transforming the data with many operations. With significant developments in the classification of images using Convolutional Neural Networks (CNN), deep DL has gained importance in crop management, crop type categorization, and crop yield assessment applications. CNN, as defined in the study [101], specializes in handling gridded datasets

and investigated the ability of RGB images captured by unmanned aerial vehicles (UAV) along with weather data to explain the yields of Wheat, barley, and oats using CNN-LSTM, Conv-LSTM, 3D-CNN. 3D-CNN was the best performing for a full-length series. Study [102] proposed a CNN and used satellite-derived features relating to topography, fertilization, and precipitation as input data. 3D-CNN involves the 3D convolution operation, which computes an output feature map O by convolving an input volume X with a 3D kernel K at each spatial location (x, y, z) using Eq. (3).

$$O(x, y, z) = \sum_i \sum_j \sum_k X(x+i, y+j, z+k) \cdot K(i, j, k) \quad (3)$$

CNN [103] for image classification and Recurrent Neural Networks (RNNs) have been effectively used for their capability to learn long-term dependencies [104]. Some extensions of RNN, such as Long Short-Term Memory (LSTM) [105] and Gated Recurrent Unit (GRU) have also recently demonstrated state-of-the-art outcomes in many applications requiring time series data analysis. LSTM structures exist in a variety of forms, including CNN-LSTM, encoder-decoder LSTM, bidirectional LSTM, and stacked LSTM. A CNN-LSTM model combines CNN and LSTM networks to analyze sequential data with spatial features. The key mathematical operation involves passing CNN feature maps C_i into LSTM cells and the calculation of the LSTM cell states and hidden states at each time step. The mathematical calculations of the CNN-LSTM model are done using Eqs. (4) to (7).

$$I_t = C_i \quad (4)$$

$$F_t = \sigma(W_f * I_t + U_f * H_{t-1} + b_f) \quad (5)$$

$$C_t = F_t * C_{t-1} + I_t * \tanh(U_c * H_{t-1} + W_c * I_t + b_c) \quad (6)$$

$$H_t = O_t * \tanh(C_t) \quad (7)$$

Here C_t represents the LSTM cell state and H_t represents the hidden state, and F_t represents the forget gate state at time step t .

Study [106] used CNN and LSTM with Gaussian procedure to foresee the crop yield using 3D histograms produced from images acquired from remotely sensed SR, land cover, and LST. Study [107] extended the work by testing the ability of a trained model in one area to be transferred to another region. They initialized the LSTM model with NN parameters trained on the Argentina dataset and replaced the last dense layer with the untrained dense layer before training the model again on Brazil data. Study [108] presented a CNN-LSTM model to predict soybean yield using weather data. The data was converted into vectors based on the histogram for training. Study [109] preserved the spatial properties of images by using the entire image rather than averaged pixels or the histograms of pixel values. Their model combined ConvLSTM layers with the 3D-CNN to extract spatiotemporal features. The model was trained using MODIS LST, SR, and land cover data. The model outperformed DT, the histogram-based CNN + GP, and the histogram-based CNN-LSTM, ConvLSTM and 3D-CNN.

Subsequently, the study [110] integrated the outputs of 3D-CNN and ConvLSTM to provide a probability-based forecast of soybean crops using Bayesian Model Averaging (BMA), SR, Land cover, and LST as predictors. The estimates of soybean crops using BMA were more precise than those obtained by the 3D-CNN and ConvLSTM. For wheat yield estimation, a study [111] developed a CNN-LSTM-12. The method operated on raw satellite imagery such as SR and thermal product land cover data. CNN-LSTM-12 outperformed DT, ridge regression (RR), SLR, and LSTM+GP. To formally articulate the temporal connection between many temporal imageries, a study [112] proposed Spectral-Spatial-Temporal neural networks, which combined 3D-CNN and LSTM to

predict corn and wheat yields. Multiple-spectral images obtained spectral-spatial features. Then, a spectral-spatial feature learning component concatenated with the temporal dependency is acquired to take the temporal connection from the continuous imagery. The approach offered greater accuracy in predicting yield.

The authors in [113] used LSTM to predict soybean yield using satellite-driven VIs, LST, and rainfall. The LSTM models outperformed the other linear methods for soybean yield forecasts. A study [114] proposed a CNN to predict wheat and barley yields using RGB images and remotely sensed NDVI. To add temporal and spatial knowledge into the model and then further increase prediction power, study [115] developed a unique graph-based RNN called GNN-RNN for predicting the corn and soybean crop yields using rainfall, temperature, VPD, mean dewpoint temperature, SM, soil temperature, available water capacity (AWC), BD, electrical conductivity (EC), pH, SOM, raw sand, silt, and clay % as input parameters fed to the model. GNN-RNN outperformed LSTM, GRU, and 1-D CNN. A study [116] proposed a smoothening function to predict wheat yield using food production data and MATLAB for simulation.

A study [117] proposed a CNN-RNN to prove its generalizability in Corn and soybean yield prediction. CNN-RNN was best performing for Soybean yield prediction among RF, LASSO, and DFNN. Study [118] proposed new activation functions named Dharasigm, SHBsig, and DharaSig for achieving improved accuracy in wheat yield prediction. New activation functions performed better than conventional sigmoid. The study [119] built a CNN-LSTM to prepare the static and dynamic components of the wheat yield prediction model using soil and weather data, respectively, and tested on different datasets for accuracy evaluation. The proposed model outperformed RF, SVM, and LASSO. A study [104] built an LSTM to predict corn yield using temperature, WS, soil root space, GDD, rainfall, and PDSI. The study [120] employed the deep learning method LSTM to analyze production and tomato yield growth variability in monitored greenhouse settings. LSTM gave more accurate results as compared to SVR and RF.

Study [121] built a reinforcement learning-based model using temperature, rainfall, evapotranspiration, humidity, ground frost frequency (GFF), diurnal temperature range (DTR), and wind speed, soil density, PH and the amount of N, P, K, transmissivity (Tr), permeability (Pr), EC, ground truth data. The proposed model outperformed Deep LSTM, Gradient Boost, ANN, RF, Bernoulli Deep Belief Network (DBN), Interval Deep Generative Artificial Neural Networks (IDANN), Bayesian Artificial Neural Networks (BAN), and Rough Auto Encoders (RAE). Study [122] proposed a CNN model that forecasts winter wheat production using a 1-D convolution process and WS, Sr, temperature, RH, and rainfall, soil data as predictors for wheat yield. A novel CNN model [123] shared the weight values of the core network feature extractor. The proposed method outperformed RF, DFNN, 3D-CNN, RT LASSO, and Ridge regression.

Study [124] built the Bayesian Neural Net model (BNN) for corn yield prediction using EVI, NDWI, GCI, and weather data such as daily temperature, VPD, daily rainfall, evapotranspiration, water stress, and soil properties like soil moisture (SM), organic matter (SOM), and cation exchange capacity (CEC). Study [125] merged satellite VIs, weather indices, and soil features to create LASSO regression, RF, and LSTM to forecast rice yield at regional levels. The findings indicated that LSTM outperformed RF and LASSO. Study [126] created an LSTM model to estimate the wheat grain yield in China by combining weather information with LAI and the vegetation temperature condition index (VTCI). The proposed model outperformed the BPNN and SVM models. An RNN-LSTM [127] estimated wheat crop yield using rainfall, temperature, humidity, ET, WS and direction, and Sr as predictors. The model outperformed ANN, MLR, and RF. Study [128] proposed a data-augmentation

technique for wheat yield prediction on two independent Algerian provinces' smaller data sets. They conducted trials using elementary data sets, data sets of added features, and augmented datasets, employing SVR, RF, ELM, ANN, and DNN to check the effectiveness of data-augmentation methods.

Study [129] proposed an LSTM-RF model for estimating wheat yield utilizing multi-spectral VIs and canopy water stress indices (CWSI) from many growth phases as predictors. A study [130] developed bidirectional LSTM and bi-directional GRU models using temperature, RH, Sr, WS, rainfall, irrigation schedule, and soil water content to predict end-of-season tomato yields. The Bidirectional LSTM outperformed GRU, LSTM, and Bidirectional GRU. Another study [131] developed an LSTM-based model to develop a target-based rice yield prediction model. They used multi-spectral VIs collected with the help of drones in Taiwan. The proposed model outperformed traditional LSTM with a significant improvement. Study [132] developed an ensemble of LSTM, bidirectional LSTM, and GRU and utilized red fox optimization for tuning the hyperparameters during model training. The ensemble model outperformed all three standalone models.

A study [133] proposed an attention-based CNN model with bidirectional LSTM to predict brinjal yield. They used shuffling shepherd optimization for tuning hyperparameters during model training. For prediction, they used daily brinjal prices in Odisha compiled for 6 years with an assumption that prices of a commodity directly depend on its available quantity. The proposed model with optimization outperformed the CNN and bidirectional LSTM. A hybrid yield prediction model based on LSTM and DBN [134] used statistical, correlation, entry, and data extraction for feature extraction. They employed an enhanced feature fusion process by including the results of three statistical feature selection methods. In a study [135], SVR, generalized regressive neural network (GRNNs), Radial Basis Function Neural networks (RBFNNs), and backpropagation Neural networks (BPNNs) estimated the rice crop yield using rainfall, temperature, P, N, K, fertilizer, pH value.

The formula for RBF used in neural networks is $RBF(x) = e^{-\frac{\|x-c\|^2}{2\sigma^2}}$. Here $RBF(x)$ represents the value of the radial basis function for the input x , c is the center of RBF, σ controls the width of the function. The GRNN involves a weighted sum RBF outputs for regression prediction calculated as $Output = \sum_{i=0}^n W_i \cdot RBF_i(Input)$. $Output$ represents the predicted output, W_i is the weight of i^{th} RBF, $RBF_i(Input)$ is the output of the i^{th} RBF for a given input. The results indicated the GRNN algorithm's superior efficiency over SVR, BPNN, and RBFNN. Study [136] explored Convolutional Structure for Rapid Feature Embedding (Caffe) using NDVI, APAR (Absorbed Photosynthetically Active Radiation), canopy surface temperature, and water stress. The correlation's value was 0.810, and the RMSE was 6.298 with two InnerProductLayer models. Study [137] provided a dimensionality-reducing method, a self-organizing Map with Latent Dirichlet Allocation, to predict the best time and crop to be grown. An appropriate season for a suitable crop was determined using a DL method. A study [138] used deep LSTM to predict corn production. A large quantity of data, combining corn production and per-hour data on the weather, demands a DL model.

To explore the fruit-carrying potential of bitter melon plants based on the color and shape of the leaves, a study [139] used the CNN model. A study [140] suggested a yield prediction technique that evaluates crop yield optically at different growth phases. A deep CNN(DCNN) created object detection systems using InceptionV3 as an image feature extraction tool. UAV-based multi-spectral images were collected during four development phases of wheat [141]. CNN models were compared to LR based on EVI as a predictor variable. The CNN model developed for the heading phase had the least RMSE of 0.94 t ha⁻¹. Study [142] analyzed XGBoost, CNN hybrids with DNN, XGBoost, RNN, and LSTM using public soybean data with climate and soil conditions. CNN-DNN outperformed with an MAE of 0.199, RMSE of 0.266, and MSE of 0.071. [Table 2](#) shows the summary of DL techniques considered in this study.

Table 2: Summary of DL techniques considered in this study

Article	Crop	Input data	Model	Compared models	Result
[101]	Wheat, barley, oats	UAV RGB images, T	CNN-LSTM, conv-LSTM, 3D-CNN	CNN-LSTM, conv-LSTM	3D-CNN best performing MAE: 218.9 kg/ha MAPE: 5.51% MAE: 292.8 kg/ha MAPE: 7.17%
[102]	Wheat	R, N, slope, elevation, topographic position index	CNN	Bayesian MLR, RF, AdaBoost, stacked autoencoder, and two CNNs	Best performing with low RMSE and RMedSE
[104]	Corn	T, WS, GDD, R, PDSI	LSTM	RF, SVR, LASSO, RNN AdaBoost	MAE: 0.83 t/ha
[106]	Soybean	SR, T, LC	LSTM-GP, CNN-GP	CNN, LSTM	MAPE: 0.48 t/ha CNN-GP best with RMSE: 5.55 t/ha
[107]	Soybean	SR, T, LC	LSTM, transfer Learning	Location-wise analysis	RMSE: 0.38 t/ha R ² : 0.66
[108]	Soybean	SR, LST, weather data	CNN-LSTM	CNN, LSTM	RMSE: 3.29%
[109]	Soybean	SR, LC, LST	(3D-CNN-Conv-LSTM)	CNN + GP, CNN-LSTM, 3D-CNN, Conv-LSTM	RMSE: 4.79 t/ha
[110]	Soybean	SR, LC, LST	BMA	3D-CNN, Conv-LSTM	RMSE: 0.38 R ² : 0.80
[111]	Wheat	SR, LC	CNN-LSTM-12	DT, RR, SLR, LSTM-GP, CNN-LSTM-9	RMSE: 48–161 kg/ha
[112]	Corn, wheat	Land cover, SR	3D-CNN- LSTM	CNN, LSTM	MAPE reduced by: 12.2%, RMSE reduced by: 20.2%
[113]	Soybean	Weather data, remote sensing data	LSTM	Multivariate OLS	MAE: 0.24 t/ha
[114]	Wheat and barley	NDVI, RGB images	CNN	No model	MAE: 484.3 kg/ha
[115]	Corn and soybean	R, T, VPD, SM, soil T, AWC, BD, EC, pH, SOM, sand, silt, clay %	GNN-RNN	compared LSTM, GRU, CNN	MAPE: 8.8% RMSE: 0.48–0.56 R ² : 0.67–0.75 Corr: 0.82–0.87

(Continued)

Table 2 (continued)

Article	Crop	Input data	Model	Compared models	Result
[116]	Wheat	Food production data	LSTM	ARIMA, RNN	RMSE: 7.29 kg/ha MAE: 7.29 kg/ha r: 0.81
[117]	Corn, soybean	Weather, soil, management data	CNN-RNN	RF, DFNN, LASSO	RMSE: 4.5–16.48 kg/ha r: 0.84–0.85
[118]	Wheat	Weather data	MLP	Different activation functions	New activations performed better
[119]	Wheat	SR, NDVI, soil data	CNN-LSTM	RF, SVM, LASSO	RMSE: 721 kg/ha R ² : 0.77% NRMSE, MAPE < 0.2 t/ha
[120]	Tomato	Weather data, stem diameter variation	LSTM	RF, SVR	MSE: 0.04 t/ha MAE: 0.03 t/ha MSE: 0.002 t/ha
[121]	Rice	T, R, ET, RH, GFF, DTR, WS, PH N, P, K, Tr, Pr, EC, ground data	RNN, deep Q-network	Deep LSTM, GB, ANN, RF, DBN, BAN, RAE, IDANN	Deep Q-network best with MAE: 0.13 t/ha, MSE: 0.03 t/ha RMSE: 0.17 t/ha R ² : 0.87 MedAE: 0.11 t/ha, MSLE: 0.002 t/ha MAPE: 17% Accuracy: 93.7%
[122]	Wheat	WS, Sr, T, RH, R, Soil BD, wilting point, SM	CNN	RF, KNN, LASSO, RT, SVR, DNN XGBoost	RMSE reduced: 7%–4% MAE reduced: 3%–15% r increased by: 4%–50%
[123]	Corn and soybean	T, SR, LC	Transfer learning	RF, DFNN, 3D-CNN, RT LASSO, RR	Corn MAE: 8.74% Soybean MAE: 8.70%
[124]	Corn	EVI, GCI, NDWI, LST, T, R, VPD, ET, SM, SOM	BNN	RR, RF, SVR, MLP, LSTM	R ² = 0.77
[125]	Rice	SIF and EVI	LSTM	LASSO, RF	R ² : 0.77–0.87 RMSE: 298.11–724 kg/ha
[126]	Wheat	VTCL, LAI	LSTM	BPNN, SVM	RMSE: 357.7 kg/ha R ² : 0.83
[127]	Wheat	R, T, RH, ET, WS, SR	RNN-LSTM	ANN, RF, MLR	RMSE: 147.12 kg/ha MAE: 60.50 kg/ha

(Continued)

Table 2 (continued)

Article	Crop	Input data	Model	Compared models	Result
[128]	Wheat	T, R, WS, SR	RF, DNN	SVR, ANN, ELM	Province1: DNN with RMSE: 0.05 q/ha Province2: RF with RMSE: 0.04 q/ha R ² : 0.96
[129]	Wheat	VIs including NDVI, CWVI	LSTM-RF	LSTM	R ² : 0.78 RMSE: 684.1 kg/ha
[130]	Tomato, Potato	T, RH, SR, WS, R, IR, SM	bidirectional LSTM, bidirectional GRU	LSTM, GRU, RF, CNN, MLP,	bidirectional LSTM with R ² : 0.97 to 0.99 MSE: 0.017 to 0.039 q/ha
[131]	Rice	Multi-spectral VIs	Target-based LSTM	LSTM	RMSE: Improved by 6.1% and R ² : Improved by 13.0%
[133]	Brinjal	Brinjal prices over a period of more than 6 years	Attention-based CNN with bidirectional LSTM	Attention-based CNN, bidirectional LSTM	Accuracy: 97.60% Precision: 96.60% Recall: 95.60% Specificity: 96.60% F1-score: 97.60%
[135]	Rice	R, T, N, P, K, pH	GRNN	SVR, BPNN, RBFNN	R ² : 0.9863 RMSE: 0.22 t/ha MAE: 0.12 t/ha MSE: 0.05 t/ha MAPE: 1.34 t/ha CV: 0.02 NSME: 0.0136 t/ha
[136]	Corn	T, R, RH, Sr, snow water, EVI	DNN	SVR, Autoencoder	RMSE: 18.5 kg/ha R ² : 0.773
[137]	12 staple crops	Soil type, R, T, RH	Self-organizing Map, DNN	Performance comparison between different train test split ratios	Specificity: 82.113% Sensitivity: 83.19% Precision: 0.8319 Recall: 0.9
[138]	Corn	R, T, WS GDD, PDSI, soil root space	Deep LSTM	RNN	LSTM performed better than RNN
[139]	Bitter gourd	On-farm image data	CNN	No model compared	Acceptable accuracy
[140]	Apple	Apple image data	DCNN (Inception V3)	No model compared	Accuracy: 90%
[141]	Wheat	EVI	CNN	LR	RMSE: 0.94 t/ha
[142]	Soybean	R, Sr, snow water, VP, T, BD, clay %, SM, EC, SOM, pH, sand %	XGBoost, CNN hybrids	All models	CNN-DNN outperformed with MAE: 0.199 q/ha RMSE: 0.266 q/ha MSE: 0.071 q/ha

5 Discussion

This section presents the discussion related to all five research objectives set up to be answered in this study and key challenges of using AI in agriculture.

5.1 RO1 Discussion

As shown in Fig. 6, rainfall and temperature are the most used weather parameters, followed by solar radiation, relative humidity, and wind speed. Soil moisture, bulk density, electrical conductivity, and soil PH are the most used parameters for soil properties, usually accompanied by soil nutrients (N, P, K). GDD, VPD, and SMDI are the most widely utilized stress indices. Remotely sensed crop growth indicators, such as NDVI and EVI, are the most accurate for yield prediction. Only a few studies considered topography factors such as slope and terrain for crop yield prediction tasks because the areas chosen for research studies were mostly flat. The planting date has been used as a predictor in the studies as a management parameter. Genotype is used less frequently due to less available data.

5.2 RO2 Discussion

However, the choice of crop consideration largely depends upon the geographical position of the study area and data availability. Mainly, the researchers considered the staple crops grown in the study area. The crops considered for yield prediction studies are corn, maize, rice, soybean, wheat, cotton, sugarcane, potato, barley, rapeseed, bajra, coconut, canola, cassava, alfalfa, ragi, oats, and tomato. Most research studies focused on wheat, rice, maize, corn, sugarcane, and soybean.

5.3 RO3 Discussion

In crop yield prediction research, RF, SVM, Gradient Boosting algorithms, and ANN have been the go-to choices. As shown in Fig. 7, LSTM and RNN models are popular in crop yield prediction research.

Furthermore, integrating CNN hybrids with LSTM has gained significant popularity among researchers, proving its capability to successfully manage the complex dynamics of spatial and temporal data, leading to improved prediction accuracy. However, the frequency of application does not mean that the frequently used models are superior in their prediction ability.

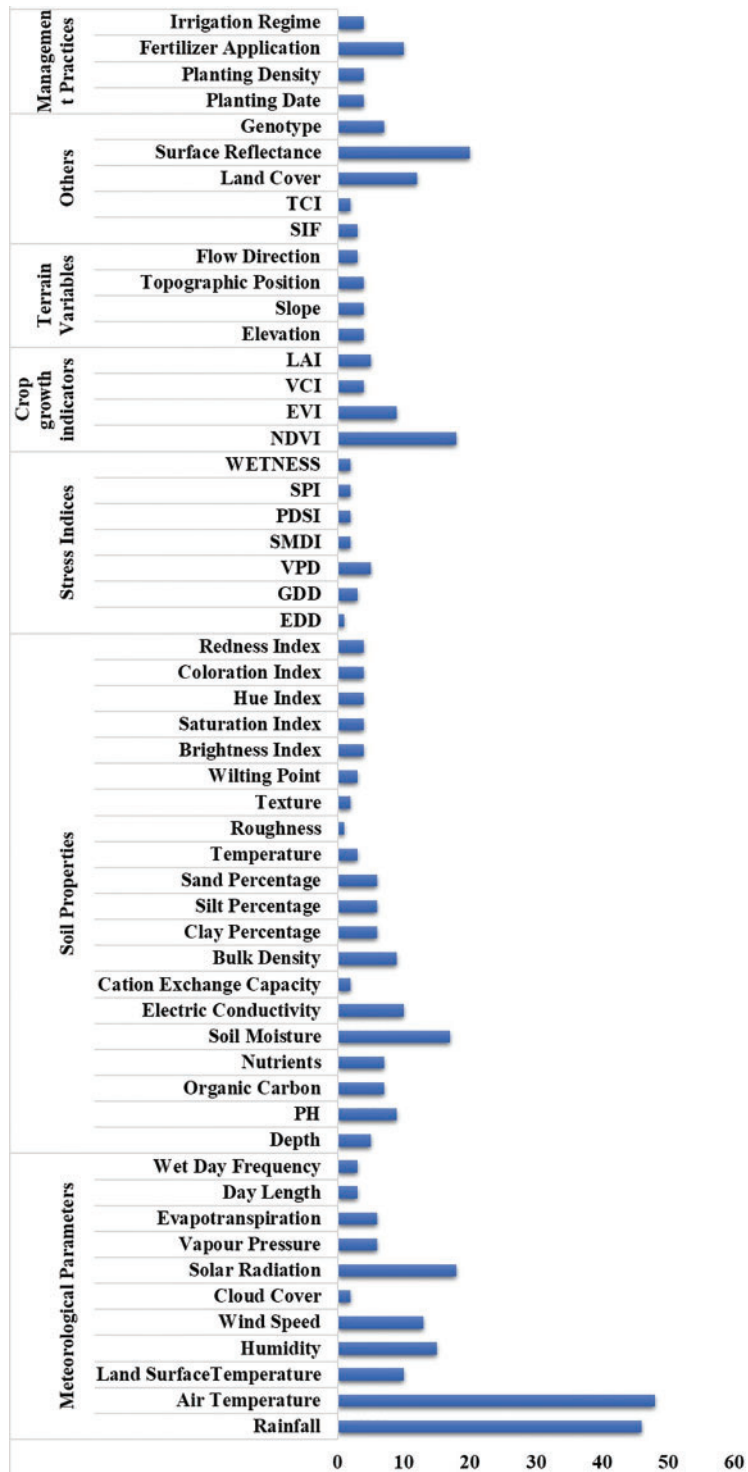


Figure 6: Frequency of usage of input parameters in yield prediction

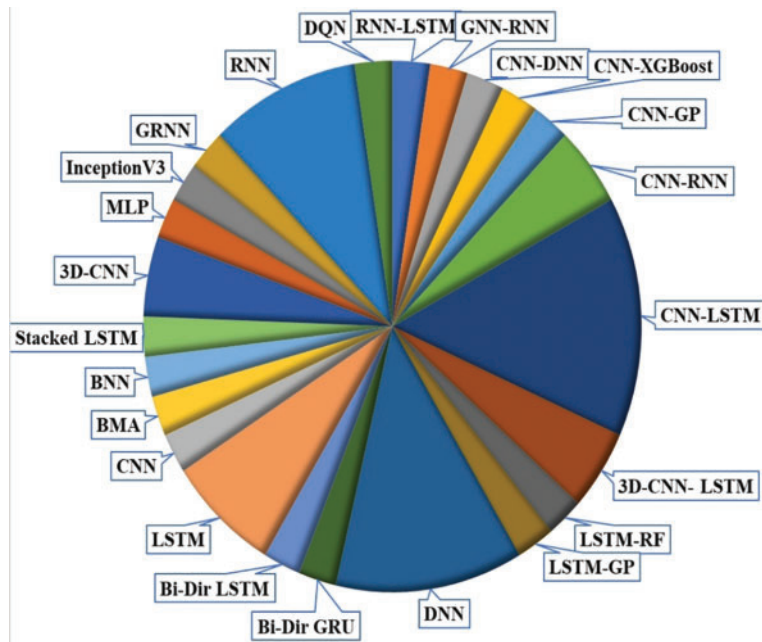


Figure 7: Frequently used DL models for crop yield prediction

5.4 RO4 Discussion

To assess predictive performance, most of the studies used RMSE as the evaluation metric to measure the model quality. Other evaluation parameters are MAE, R^2 , and MSE. RMSE is the most used, followed by R^2 and MAE, as shown in Fig. 8. Some evaluation parameters like Relative Absolute Error (RAE), Adjusted R^2 , LCCC, Mean Squared Logarithmic Error (MSLE), Median Absolute Error (MedAE), MFE, Coefficient of variance (CV), Median Absolute Error (MedAE), Percent Error (PE), and Fractional Bias (FB) are also used to assess model predictive performance in certain studies. Most of these parameters were adjustments on earlier stated parameters. Researchers also used 5-fold and 10-fold validations as an evaluation approach to test their models.

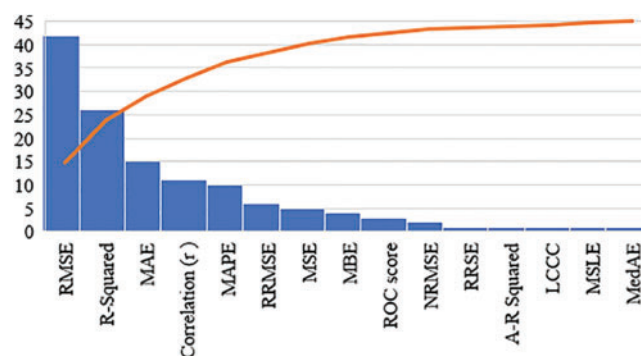


Figure 8: Frequency of accuracy evaluation metrics

5.5 RO5 Discussion

Researchers building ML and DL-based crop yield prediction models encounter several significant challenges. Agricultural data can be sparse, heterogeneous, and subject to measurement errors. So, acquiring reliable training data for ML and DL models is challenging. Crop growth also depends on dynamic factors, like climatic conditions, soil conditions, and farming practices. These require complex models to capture the spatiotemporal variability accurately. It is a critical ethical challenge to ensure that ML and DL models do not introduce or maintain biases, especially in resource allocation or policy decisions. Training DL models often requires substantial computational resources, which can be a limitation for researchers or farmers in resource-constrained regions. Acquiring reliable ground-truth data for model validation can be challenging, which is essential for assessing the model's predictive accuracy. Crop yield prediction also involves expertise from various fields, such as agriculture, remote sensing, and data science, necessitating effective interdisciplinary collaboration. These challenges are reported based on the articles' stated statements. However, other challenges may exist based on the crop or region of study.

5.6 Discussion Related to the Benefits and Challenges of Using AI in Agriculture

AI for agriculture has transformational benefits across a variety of uses. Precision agriculture employs AI to improve the utilization of resources, increasing crop yields while reducing inputs such as fertilizer and irrigation water. AI offers early disease identification and crop monitoring, leading to reduced crop losses. Weather forecasting helps farmers make accurate choices to plant and harvest times. Automation improves supply-chain operations and lowers labor expenses, resulting in better productivity and earnings in agriculture. Although crop evaluation and forecasting may benefit substantially from machine learning, still some challenges have been identified in recent studies [143–146]. The key issues include the following.

5.6.1 Data Quality and Volume

ML model reliability and precision are affected by the quality of the training data. Acquiring high-quality data in agriculture can be difficult due to variations in the weather, soil, topography, and other factors affecting the environment. As a result, obtaining and cleaning data may be a challenge. Study [147] discussed the problems associated with DL-based fruit identification and classification. They determined that the lack of a high-quality fruit dataset was the primary cause of errors and low classification speed. ML models also require a large amount of data for making classifications and predictions. Obtaining and managing such data is complex, especially in the case of small-size farms. Study [148] identified data volume, velocity, diversity, and validity as the primary challenging circumstances of big data.

5.6.2 Interpretability and Accessibility

It can be complex to analyze the results of ML and DL models. As a result, farmers may find it challenging to understand the factors contributing to producing a specific crop prediction. In instances when resources are constrained, acquiring access to the software and hardware components necessary for constructing and running ML models can be difficult.

5.6.3 Data Privacy and Human Considerations

These issues revolve around the difficulties in data gathering, storing, and utilizing confidential data related to agriculture. It can be challenging to balance privacy and security while making it

possible to access data for ML development. Ethical concerns are related to guaranteeing equal access to these innovations, offering assistance, and overcoming a technological gap to avoid further deprivation of specific farming populations. Also, the farmers and other stakeholders may require more time to adapt to new techniques and technological devices, such as AI-based systems. To be widely used, technologies must become available, convenient, and capable of offering real-life benefits.

Overcoming these problems needs collaboration among statisticians, farmers, and all other stakeholders to guarantee that ML methods are efficient, ethical, and affordable. Despite our thorough examination of agricultural yield prediction methodologies, certain drawbacks of this study still exist. This review may not include all relevant material, leaving out newer techniques or geographic-specific approaches.

6 Conclusion

This literature review explores many research paths for developing crop yield prediction models. The study offers a critical and mathematical analysis of weather, soil, and management-related input parameters and the current cutting-edge methods for predicting crop yields—a detailed analysis of the ML, DL, and meta-models used in the timely prediction of crop yield using weather, soil, management information and satellite imagery. Most scientific studies used DL techniques, especially CNN and LSTM, and their hybrid models for crop yield prediction. The researchers have also integrated crop models with ML to add flexibility to the heavy crop simulation models. Most of the current studies have utilized remotely sensed data. Models with extra features do not necessarily provide the best yield prediction. The emphasis must be on studying the optimal deduction of input features. The researchers differ in their crop consideration, region of study, crop type, number of input parameters, data sources, and prediction methodologies. The variations shown by the same modeling technique in various studies are due to diversified input parameters fed into the prediction model and different crop considerations by the studies. This study will lay the foundation for scientific investigations on crop yield prediction challenges.

7 Future Directions in Crop Yield Prediction Research

Despite the significant strides made in crop yield prediction research using AI-based models, several research gaps persist, necessitating further exploration and innovation. One of the notable gaps lies in integrating various data sources, like remote sensing, climate data, and soil information, to enhance the accuracy and robustness of prediction models. A large amount of input data must be dealt with properly to build a practically deployable and cost-effective crop yield prediction model. Some feature ranking approaches, such as statistical correlation-based and tree-based feature ranking, may optimize feature extraction from the input images. Principal component analysis may also help to extract the most relevant input features to prevent overfitting and the generation of enormous parameters. Particular feature sets may be cross-validated to ensure accurate model performance.

Hyperspectral imagery is still not widely utilized for accurate crop yield prediction. It captures images in several hundred small spectrum bands, presenting comprehensive spectrum information. Integrating this data with the data provided by IOT devices set up in the fields can provide crucial information about crop physicochemical properties such as the amount of nutrients, stress due to water, and disease appearance significantly affecting crop yield. Successful integration of the models into the farming systems needs more standardized approaches in handling varied crop types and agroecological conditions, as existing models often exhibit limitations in adaptability. Future directions should also focus on developing dynamic models that adapt to rapidly changing environmental conditions and

incorporate real-time data streams. In the future, a performance comparison of all ML and DL prediction models would determine the best technique and modeling techniques for predicting crop yield using a fully integrated data set. Utilizing transfer learning and Multimodal learning can manage data scarcity and the presence of multiple modalities in the dataset. Addressing these challenges is essential to developing accurate and practical ML and DL-based crop yield prediction models to enhance food security and improve agricultural resource management.

Acknowledgement: The authors want to express their gratefulness to Dr. Tarun, for his insightful advice and practical recommendations during this investigation.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Nidhi Parashar; data collection: Prashant Johri and Nitin Gaur; analysis and interpretation of results: Nidhi Parashar and Arfat Ahmad Khan; draft manuscript preparation: Seifedine Kadry and Nidhi Parashar. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable. All references are from Google Scholar.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Fountas, N. Mylonas, I. Malounas, E. Rodias, C. H. Santos and E. Pekkeriet, "Agricultural robotics for field operations," *Sensors*, vol. 20, no. 9, pp. 2672, May 2020. doi: [10.3390/s20092672](https://doi.org/10.3390/s20092672).
- [2] F. Abbas, H. Afzaal, A. A. Farooque, and S. Tang, "Crop yield prediction through proximal sensing and machine learning algorithms," *Agronomy*, vol. 10, no. 7, pp. 1046, Jul. 2020. doi: [10.3390/agronomy10071046](https://doi.org/10.3390/agronomy10071046).
- [3] B. A. Keating *et al.*, "An overview of APSIM, a model designed for farming systems simulation," *Eur. J. Agron.*, vol. 18, no. 3–4, pp. 267–288, Jan. 2003. doi: [10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9).
- [4] J. W. Jones *et al.*, "The DSSAT cropping system model," *Eur. J. Agron.*, vol. 18, no. 3–4, pp. 235–265, Jan. 2003. doi: [10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7).
- [5] C. A. van Diepen, J. van Wolf, H. van Keulen, and C. Rappoldt, "WOFOST: A simulation model of crop production," *Soil Use Manag.*, vol. 5, no. 1, pp. 16–24, Mar. 1989. doi: [10.1111/j.1475-2743.1989.tb00755.x](https://doi.org/10.1111/j.1475-2743.1989.tb00755.x).
- [6] M. Capa-Morocho, A. V. M. Ines, W. E. Baethgen, B. Rodríguez-Fonseca, E. Han and M. Ruiz-Ramos, "Crop yield outlooks in the Iberian Peninsula: Connecting seasonal climate forecasts with crop simulation models," *Agric Syst.*, vol. 149, no. 4, pp. 75–87, Nov. 2016. doi: [10.1016/j.agry.2016.08.008](https://doi.org/10.1016/j.agry.2016.08.008).
- [7] R. Lecerf, A. Ceglar, R. López-Lozano, M. van der Velde, and B. Baruth, "Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe," *Agric. Syst.*, vol. 168, no. 1293, pp. 191–202, Jan. 2019. doi: [10.1016/j.agry.2018.03.002](https://doi.org/10.1016/j.agry.2018.03.002).
- [8] J. N. Brown, Z. Hochman, D. Holzworth, and H. Horan, "Seasonal climate forecasts provide more definitive and accurate crop yield predictions," *Agric. For. Meteorol.*, vol. 260, no. 3, pp. 247–254, Oct. 2018. doi: [10.1016/j.agrformet.2018.06.001](https://doi.org/10.1016/j.agrformet.2018.06.001).
- [9] S. V. Archontoulis, M. J. Castellano, M. A. Licht, V. Nichols, and M. Baum, "Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt," *Crop Sci.*, vol. 60, no. 2, pp. 721–738, Jan. 2020. doi: [10.1002/csc2.20039](https://doi.org/10.1002/csc2.20039).

- [10] M. D. Johnson, W. W. Hsieh, A. J. Cannon, A. Davidson, and F. Bédard, “Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods,” *Agric. For. Meteorol.*, vol. 218, no. 9, pp. 74–84, Mar. 2016. doi: [10.1016/j.agrformet.2015.11.003](https://doi.org/10.1016/j.agrformet.2015.11.003).
- [11] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, “Maize yield and nitrate loss prediction with machine learning algorithms,” *Environ. Res. Lett.*, vol. 14, no. 12, pp. 124026, Dec. 2019. doi: [10.1088/1748-9326/ab5268](https://doi.org/10.1088/1748-9326/ab5268).
- [12] R. J. Brooks, M. A. Semenov, and P. D. Jamieson, “Simplifying sirius: Sensitivity analysis and development of a meta-model for wheat yield prediction,” *Eur. J. Agron.*, vol. 14, no. 1, pp. 43–60, Jan. 2001. doi: [10.1016/S1161-0301\(00\)00089-7](https://doi.org/10.1016/S1161-0301(00)00089-7).
- [13] I. Supit, “Predicting national wheat yields using a crop simulation and trend models,” *Agric. For. Meteorol.*, vol. 88, no. 1–4, pp. 199–214, Dec. 1997. doi: [10.1016/S0168-1923\(97\)00037-3](https://doi.org/10.1016/S0168-1923(97)00037-3).
- [14] A. S. Nain, V. K. Dadhwal, and T. P. Singh, “Real time wheat yield assessment using technology trend and crop simulation model with minimal data set,” *Curr. Sci.*, vol. 82, pp. 1255–1258, May 2002. Accessed: Apr. 19, 2024. [Online]. Available: <http://www.jstor.org/stable/24107049>.
- [15] A. S. Nain, V. K. Dadhwal, and T. P. Singh, “Use of CERES-wheat model for wheat yield forecast in central Indo-Gangetic Plains of India,” *J. Agric. Sci.*, vol. 142, no. 1, pp. 59–70, Feb. 2004. doi: [10.1017/S0021859604004022](https://doi.org/10.1017/S0021859604004022).
- [16] A. Chipanshi *et al.*, “Evaluation of the integrated Canadian crop yield forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape,” *Agric. For. Meteorol.*, vol. 206, no. 2, pp. 137–150, Jun. 2015. doi: [10.1016/j.agrformet.2015.03.007](https://doi.org/10.1016/j.agrformet.2015.03.007).
- [17] A. Ceglar, A. Toreti, R. Lecerf, M. van der Velde, and F. Dentener, “Impact of meteorological drivers on regional inter-annual crop yield variability in France,” *Agric. For. Meteorol.*, vol. 216, no. 9, pp. 58–67, Jan. 2016. doi: [10.1016/j.agrformet.2015.10.004](https://doi.org/10.1016/j.agrformet.2015.10.004).
- [18] B. L. Sierra-Forero, J. Baron-Velandia, and S. C. Vanegas-Ayala, “Assessment of the relevance of features associated with corn crop yield prediction in Colombia, a country in the Neotropical zone,” *Int. J. Inf. Technol.*, vol. 16, no. 4, pp. 1–10, Mar. 2024. doi: [10.1007/s41870-024-01762-9](https://doi.org/10.1007/s41870-024-01762-9).
- [19] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain and M. Anderson, “Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest,” *Environ. Res. Lett.*, vol. 15, no. 6, pp. 064005, 2020. doi: [10.1088/1748-9326/ab7df9](https://doi.org/10.1088/1748-9326/ab7df9).
- [20] M. J. Roberts, W. Schlenker, and J. Eyer, “Agronomic weather measures in econometric models of crop yield with implications for climate change,” *Am. J. Agric. Econ.*, vol. 95, no. 2, pp. 236–243, Jan. 2013. doi: [10.1093/ajae/aas047](https://doi.org/10.1093/ajae/aas047).
- [21] D. B. Lobell and C. B. Field, “Global scale climate-crop yield relationships and the impacts of recent warming,” *Environ. Res. Lett.*, vol. 2, no. 1, pp. 014002, Mar. 2007. doi: [10.1088/1748-9326/2/1/014002](https://doi.org/10.1088/1748-9326/2/1/014002).
- [22] S. K. Dubey, A. S. Gavli, S. K. Yadav, S. Sehgal, and S. S. Ray, “Remote sensing-based yield forecasting for sugarcane (*Saccharum officinarum* L.) crop in India,” *J. Indian Soc. Remote Sens.*, vol. 46, no. 11, pp. 1823–1833, Nov. 2018. doi: [10.1007/s12524-018-0839-2](https://doi.org/10.1007/s12524-018-0839-2).
- [23] S. M. Quiring and S. Ganesh, “Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas,” *Agric. For. Meteorol.*, vol. 150, no. 3, pp. 330–339, Mar. 2010. doi: [10.1016/j.agrformet.2009.11.015](https://doi.org/10.1016/j.agrformet.2009.11.015).
- [24] S. S. Panda, D. P. Ames, and S. Panigrahi, “Application of vegetation indices for agricultural crop yield prediction using neural network techniques,” *Remote Sens.*, vol. 2, no. 3, pp. 673–696, Mar. 2010. doi: [10.3390/rs2030673](https://doi.org/10.3390/rs2030673).
- [25] A. A. Gitelson, B. D. Wardlow, G. P. Keydan, and B. Leavitt, “An evaluation of MODIS 250-m data for green LAI estimation in crops,” *Geophys. Res. Lett.*, vol. 34, no. 20, pp. 309, Oct. 2007. doi: [10.1029/2007GL031620](https://doi.org/10.1029/2007GL031620).
- [26] K. Guan *et al.*, “Mapping paddy rice area and yields over Thai Binh Province in Viet Nam from MODIS, Landsat, and ALOS-2/PALSAR-2,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 7, pp. 2238–2252, Jun. 2018. doi: [10.1109/JSTARS.2018.2834383](https://doi.org/10.1109/JSTARS.2018.2834383).

- [27] O. Satir and S. Berberoglu, "Crop yield prediction under soil salinity using satellite derived vegetation indices," *Field Crops Res.*, vol. 192, no. 18, pp. 134–143, Jun. 2016. doi: [10.1016/j.fcr.2016.04.028](https://doi.org/10.1016/j.fcr.2016.04.028).
- [28] C. L. Wiegand, A. J. Richardson, D. E. Escobar, and A. H. Gerbermann, "Vegetation indices in crop assessments," *Remote Sens. Environ.*, vol. 35, no. 2–3, pp. 105–119, Feb. 1991. doi: [10.1016/0034-4257\(91\)90004-P](https://doi.org/10.1016/0034-4257(91)90004-P).
- [29] J. Liu, E. Pattey, and G. Jégo, "Assessment of vegetation indices for regional crop green LAI estimation from Landsat images over multiple growing seasons," *Remote Sens. Environ.*, vol. 123, no. 2, pp. 347–358, Aug. 2012. doi: [10.1016/j.rse.2012.04.002](https://doi.org/10.1016/j.rse.2012.04.002).
- [30] K. Guan *et al.*, "Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence," *Glob. Chang. Biol.*, vol. 22, no. 2, pp. 716–726, Feb. 2016. doi: [10.1111/gcb.13136](https://doi.org/10.1111/gcb.13136).
- [31] G. Badgley, C. B. Field, and J. A. Berry, "Canopy near-infrared reflectance and terrestrial photosynthesis," *Sci. Adv.*, vol. 3, no. 3, pp. e1602244, Mar. 2017. doi: [10.1126/sciadv.1602244](https://doi.org/10.1126/sciadv.1602244).
- [32] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, Jan. 2021. doi: [10.1038/s41598-020-80820-1](https://doi.org/10.1038/s41598-020-80820-1).
- [33] M. C. Anderson *et al.*, "An intercomparison of drought indicators based on thermal remote sensing and NLDAS-2 simulations with US drought monitor classifications," *J. Hydrometeorol.*, vol. 14, no. 4, pp. 1035–1056, Aug. 2013. doi: [10.1175/JHM-D-12-0140.1](https://doi.org/10.1175/JHM-D-12-0140.1).
- [34] Y. Yang *et al.*, "Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE," *USA Remote Sens. Environ.*, vol. 210, no. 5, pp. 387–402, Jun. 2018. doi: [10.1016/j.rse.2018.02.020](https://doi.org/10.1016/j.rse.2018.02.020).
- [35] I. E. Mladenova *et al.*, "Intercomparison of soil moisture, evaporative stress, and vegetation indices for estimating corn and soybean yields over the US," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 4, pp. 1328–1343, Jan. 2017. doi: [10.1109/JSTARS.2016.2639338](https://doi.org/10.1109/JSTARS.2016.2639338).
- [36] S. Siebert, H. Webber, and E. E. Rezaei, "Weather impacts on crop yields-searching for simple answers to a complex problem," *Environ. Res. Lett.*, vol. 12, no. 8, pp. 081001, Aug. 2017. doi: [10.1088/1748-9326/aa7f15](https://doi.org/10.1088/1748-9326/aa7f15).
- [37] L. Łabędzki, "Estimation of local drought frequency in central Poland using the standardized precipitation index SPI, irrigation and drainage," *Irrig. Drain.: J. Int. Comm. Irrig. Drain.*, vol. 56, no. 1, pp. 67–77, Feb. 2007. doi: [10.1002/ird.285](https://doi.org/10.1002/ird.285).
- [38] S. Khan, H. F. Gabriel, and T. Rana, "Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas," *Irrig. Drain. Syst.*, vol. 22, no. 2, pp. 159–177, Jun. 2008. doi: [10.1007/s10795-008-9049-3](https://doi.org/10.1007/s10795-008-9049-3).
- [39] Y. Li *et al.*, "Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US," *Field Crops Res.*, vol. 234, pp. 55–65, Mar. 2019. doi: [10.1016/j.fcr.2019.02.005](https://doi.org/10.1016/j.fcr.2019.02.005).
- [40] K. D. Subedi, B. L. Ma, and A. G. Xue, "Planting date and nitrogen effects on grain yield and protein content of spring wheat," *Crop Sci.*, vol. 47, no. 1, pp. 36–44, Jan. 2007. doi: [10.2135/cropsci2006.02.0099](https://doi.org/10.2135/cropsci2006.02.0099).
- [41] M. Hu and P. Wiatrak, "Effect of planting date on soybean growth, yield, and grain quality," *Agron. J.*, vol. 104, no. 3, pp. 785–790, May 2012. doi: [10.2134/agronj2011.0382](https://doi.org/10.2134/agronj2011.0382).
- [42] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar and A. K. Singh, "Crop yield prediction integrating genotype and weather variables using deep learning," *PLoS One*, vol. 16, no. 6, pp. e0252402, Jun. 2021. doi: [10.1371/journal.pone.0252402](https://doi.org/10.1371/journal.pone.0252402).
- [43] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Front. Plant Sci.*, vol. 10, pp. 621, May 2019. doi: [10.3389/fpls.2019.00621](https://doi.org/10.3389/fpls.2019.00621).
- [44] A. Choudhury and J. Jones, "Crop yield prediction using time series models," *J. Econ. Econ. Educ. Res.*, vol. 15, no. 3, pp. 53–67, Sep. 2014.
- [45] K. K. Suresh and S. R. Krishna Priya, "Forecasting sugarcane yield of Tamilnadu using ARIMA models," *Sugar Tech.*, vol. 13, no. 1, pp. 23–26, Mar. 2011. doi: [10.1007/s12355-011-0071-7](https://doi.org/10.1007/s12355-011-0071-7).
- [46] S. Dharmaraja, V. Jain, P. Anjoy, and H. Chandra, "Empirical analysis for crop yield forecasting in India," *Agric. Res.*, vol. 9, no. 1, pp. 132–138, Mar. 2020. doi: [10.1007/s40003-019-00413-x](https://doi.org/10.1007/s40003-019-00413-x).

- [47] T. Iizumi, Y. Shin, W. Kim, M. Kim, and J. Choi, "Global crop yield forecasting using seasonal climate information from a multi-model ensemble," *Clim. Serv.*, vol. 11, pp. 13–23, Aug. 2018. doi: [10.1016/j.cliser.2018.06.003](https://doi.org/10.1016/j.cliser.2018.06.003).
- [48] B. Das *et al.*, "Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the West Coast of India," *Int. J. Biometeorol.*, vol. 64, no. 7, pp. 1111–1123, Jul. 2020. doi: [10.1007/s00484-020-01884-2](https://doi.org/10.1007/s00484-020-01884-2).
- [49] P. Filippi *et al.*, "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning," *Precis. Agric.*, vol. 20, no. 5, pp. 1015–1029, Oct. 2019. doi: [10.1007/s11119-018-09628-4](https://doi.org/10.1007/s11119-018-09628-4).
- [50] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, and S. Shearer, "Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield," *Comput. Electron. Agric.*, vol. 153, pp. 213–225, Oct. 2018. doi: [10.1016/j.compag.2018.07.016](https://doi.org/10.1016/j.compag.2018.07.016).
- [51] M. Kuradusenge *et al.*, "Crop yield prediction using machine learning models: Case of Irish potato and maize," *Agriculture*, vol. 13, no. 1, pp. 225, Jan. 2023. doi: [10.3390/agriculture13010225](https://doi.org/10.3390/agriculture13010225).
- [52] Y. Cai *et al.*, "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," *Agric. For. Meteorol.*, vol. 274, no. 60, pp. 144–159, Aug. 2019. doi: [10.1016/j.agrformet.2019.03.010](https://doi.org/10.1016/j.agrformet.2019.03.010).
- [53] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting corn yield with machine learning ensembles," *Front. Plant Sci.*, vol. 11, pp. 1120, Jul. 2020. doi: [10.3389/fpls.2020.01120](https://doi.org/10.3389/fpls.2020.01120).
- [54] J. Han *et al.*, "Prediction of winter wheat yield based on multi-source data and machine learning in China," *Remote Sens.*, vol. 12, no. 2, pp. 236, Jan. 2020. doi: [10.3390/rs12020236](https://doi.org/10.3390/rs12020236).
- [55] S. Ju *et al.*, "Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models," *Agric. For. Meteorol.*, vol. 307, no. 10, pp. 108530, Sep. 2021. doi: [10.1016/j.agrformet.2021.108530](https://doi.org/10.1016/j.agrformet.2021.108530).
- [56] L. S. Cedric *et al.*, "Crops yield prediction based on machine learning models: Case of West African countries," *Smart Agric. Technol.*, vol. 2, no. 15, pp. 100049, Dec. 2022. doi: [10.1016/j.atech.2022.100049](https://doi.org/10.1016/j.atech.2022.100049).
- [57] F. B. Sarijaloo, M. Porta, B. Taslimi, and P. M. Pardalos, "Yield performance estimation of corn hybrids using machine learning algorithms," *Artif. Intell. Agric.*, vol. 5, pp. 82–89, Jan. 2021. doi: [10.1016/j.aiaa.2021.05.001](https://doi.org/10.1016/j.aiaa.2021.05.001).
- [58] L. Li *et al.*, "Developing machine learning models with multi-source environmental data to predict wheat yield in China," *Comput. Electron. Agric.*, vol. 194, no. 11, pp. 106790, Mar. 2022. doi: [10.1016/j.compag.2022.106790](https://doi.org/10.1016/j.compag.2022.106790).
- [59] H. T. Pham, J. Awange, M. Kuhn, B. van Nguyen, and L. K. Bui, "Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices," *Sensors*, vol. 22, no. 3, pp. 719, Jan. 2022. doi: [10.3390/s22030719](https://doi.org/10.3390/s22030719).
- [60] C. D. Whitmire, J. M. Vance, H. K. Rasheed, A. Missaoui, K. M. Rasheed and F. W. Maier, "Using machine learning and feature selection for alfalfa yield prediction," *AI*, vol. 2, no. 1, pp. 71–88, Feb. 2021. doi: [10.3390/ai2010006](https://doi.org/10.3390/ai2010006).
- [61] N. Balakrishnan and G. Muthukumarasamy, "Crop production-ensemble machine learning model for prediction," *Int. J. Computer Sci. Softw. Eng.*, vol. 5, no. 7, pp. 148, Jul. 2016.
- [62] J. H. Jeong *et al.*, "Random forests for global and regional crop yield predictions," *PLoS One*, vol. 11, no. 6, pp. e0156571, Jun. 2016. doi: [10.1371/journal.pone.0156571](https://doi.org/10.1371/journal.pone.0156571).
- [63] P. Priya, U. Muthaiah, and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," *Int. J. Eng. Sci. Res. Technol.*, vol. 7, no. 1, pp. 1–7, Apr. 2018.
- [64] P. Boste, N. K. Kasabov, L. Bruzzone, and R. N. Hartono, "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6563–6573, Jul. 2016. doi: [10.1109/TGRS.2016.2586602](https://doi.org/10.1109/TGRS.2016.2586602).
- [65] N. R. Prasad, N. R. Patel, and A. Danodia, "Crop yield prediction in cotton for regional level using random forest approach," *Spat. Inf. Res.*, vol. 29, no. 2, pp. 195–206, Jul. 2021. doi: [10.1007/s41324-020-00346-6](https://doi.org/10.1007/s41324-020-00346-6).

- [66] P. P. Bhangale, P. Y. S. Patil, and D. D. Patil, "Improved crop yield prediction using neural network," *IJARIIIE*, vol. 3, no. 2, pp. 2395–4396, 2017.
- [67] R. Ed-Daoudi, A. Alaoui, B. Ettaki, and J. Zerouaoui, "Improving crop yield predictions in morocco using machine learning algorithms," *J. Ecol. Eng.*, vol. 24, no. 6, pp. 392–400, 2023. doi: [10.12911/22998993/162769](https://doi.org/10.12911/22998993/162769).
- [68] P. Taherei Ghazvinei *et al.*, "Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 738–749, Jan. 2018. doi: [10.1080/19942060.2018.1526119](https://doi.org/10.1080/19942060.2018.1526119).
- [69] P. S. M. Gopal and R. Bhargavi, "A novel approach for efficient crop yield prediction," *Comput. Electron. Agric.*, vol. 165, no. 2, pp. 104968, Oct. 2019. doi: [10.1016/j.compag.2019.104968](https://doi.org/10.1016/j.compag.2019.104968).
- [70] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environ. Dev. Sustain.*, vol. 22, no. 6, pp. 5687–5708, Aug. 2020. doi: [10.1007/s10668-019-00445-x](https://doi.org/10.1007/s10668-019-00445-x).
- [71] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agric.*, vol. 121, no. 5, pp. 57–65, Feb. 2016. doi: [10.1016/j.compag.2015.11.018](https://doi.org/10.1016/j.compag.2015.11.018).
- [72] H. Cheng, L. Damerow, Y. Sun, and M. Blanke, "Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks," *J. Imaging*, vol. 3, no. 1, pp. 6, Jan. 2017. doi: [10.3390/jimaging3010006](https://doi.org/10.3390/jimaging3010006).
- [73] O. M. Adisa *et al.*, "Application of artificial neural network for predicting maize production in South Africa," *Sustainability*, vol. 11, no. 4, pp. 1145, Feb. 2019. doi: [10.3390/su11041145](https://doi.org/10.3390/su11041145).
- [74] J. L. Fernandes, N. F. F. Ebecken, and J. C. D. M. Esquerdo, "Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble," *Int. J. Remote Sens.*, vol. 38, no. 16, pp. 4631–4644, Aug. 2017. doi: [10.1080/01431161.2017.1325531](https://doi.org/10.1080/01431161.2017.1325531).
- [75] K. A. Shastry, H. A. Sanjay, and A. Deshmukh, "A parameter based customized artificial neural network model for crop yield prediction," *J. Artif. Intell.*, vol. 9, no. 1–3, pp. 23–32, 2016. doi: [10.3923/jai.2016.23.32](https://doi.org/10.3923/jai.2016.23.32).
- [76] B. Garg, S. Aggarwal, and J. Sokhal, "Crop yield forecasting using fuzzy logic and regression model," *Comput. Electr. Eng.*, vol. 67, pp. 383–403, Apr. 2018. doi: [10.1016/j.compeleceng.2017.11.015](https://doi.org/10.1016/j.compeleceng.2017.11.015).
- [77] S. Fei *et al.*, "UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," *Precis. Agric.*, vol. 24, no. 1, pp. 187–212, Feb. 2023. doi: [10.1007/s11119-022-09938-8](https://doi.org/10.1007/s11119-022-09938-8).
- [78] J. Bolaños, J. C. Corrales, and L. V. Campo, "Feasibility of early yield prediction per coffee tree based on multispectral aerial imagery: Case of Arabica coffee crops in Cauca-Colombia," *Remote Sens.*, vol. 15, no. 1, pp. 282, Jan. 2023. doi: [10.3390/rs15010282](https://doi.org/10.3390/rs15010282).
- [79] F. Soroush, M. Ehteram, and A. Seifi, "Uncertainty and spatial analysis in wheat yield prediction based on robust inclusive multiple models," *Environ. Sci. Pollut. Res.*, vol. 30, no. 8, pp. 20887–20906, Feb. 2023. doi: [10.1007/s11356-022-23653-x](https://doi.org/10.1007/s11356-022-23653-x).
- [80] T. Mavromatis, "Spatial resolution effects on crop yield forecasts: An application to rainfed wheat yield in north Greece with CERES-wheat," *Agric. Syst.*, vol. 143, no. 3, pp. 38–48, Mar. 2016. doi: [10.1016/j.agry.2015.12.002](https://doi.org/10.1016/j.agry.2015.12.002).
- [81] L. Busetto *et al.*, "Downstream services for rice crop monitoring in Europe: From regional to local scale," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 12, pp. 5423–5441, Apr. 2017. doi: [10.1109/JSTARS.2017.2679159](https://doi.org/10.1109/JSTARS.2017.2679159).
- [82] V. Pagani *et al.*, "Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil," *Agric. Syst.*, vol. 154, pp. 45–52, Jun. 2017. doi: [10.1016/j.agry.2017.03.002](https://doi.org/10.1016/j.agry.2017.03.002).
- [83] M. J. Roberts, N. O. Braun, T. R. Sinclair, D. B. Lobell, and W. Schlenker, "Comparing and combining process-based crop models and statistical models with some implications for climate change," *Environ. Res. Lett.*, vol. 12, no. 9, pp. 095010, Sep. 2017. doi: [10.1088/1748-9326/aa7f33](https://doi.org/10.1088/1748-9326/aa7f33).

- [84] Y. Everingham, J. Sexton, D. Skocaj, and G. Inman-Bamber, "Accurate prediction of sugarcane yield using a random forest algorithm," *Agron. Sustain. Dev.*, vol. 36, no. 2, pp. 1–9, Jun. 2016. doi: [10.1007/s13593-016-0364-z](https://doi.org/10.1007/s13593-016-0364-z).
- [85] P. Feng, B. Wang, D. L. Liu, C. Waters, and Q. Yu, "Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia," *Agric. For. Meteorol.*, vol. 275, no. 2, pp. 100–113, Sep. 2019. doi: [10.1016/j.agrformet.2019.05.018](https://doi.org/10.1016/j.agrformet.2019.05.018).
- [86] P. Feng *et al.*, "Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique," *Agric. For. Meteorol.*, vol. 285, no. 3, pp. 107922, May 2020. doi: [10.1016/j.agrformet.2020.107922](https://doi.org/10.1016/j.agrformet.2020.107922).
- [87] D. Paudel *et al.*, "Machine learning for large-scale crop yield forecasting," *Agric. Syst.*, vol. 187, pp. 103016, 2021. doi: [10.1016/j.agry.2020.103016](https://doi.org/10.1016/j.agry.2020.103016).
- [88] S. S. Sajid, M. Shahhosseini, I. Huber, G. Hu, and S. V. Archontoulis, "County-scale crop yield prediction by integrating crop simulation with machine learning models," *Front. Plant Sci.*, vol. 13, pp. 1000224, Nov. 2022. doi: [10.3389/fpls.2022.1000224](https://doi.org/10.3389/fpls.2022.1000224).
- [89] D. Batool *et al.*, "A hybrid approach to tea crop yield prediction using simulation models and machine learning," *Plants*, vol. 11, no. 15, pp. 1925, Jul. 2022. doi: [10.3390/plants11151925](https://doi.org/10.3390/plants11151925).
- [90] A. Attia *et al.*, "Coupling process-based models and machine learning algorithms for predicting yield and evapotranspiration of maize in arid environments," *Water*, vol. 14, no. 22, pp. 3647, Nov. 2022. doi: [10.3390/w14223647](https://doi.org/10.3390/w14223647).
- [91] A. M. S. Kheir, K. A. Ammar, A. Amer, M. G. M. Ali, Z. Ding and A. Elnashar, "Machine learning-based cloud computing improved wheat yield simulation in arid regions," *Comput. Electron. Agric.*, vol. 203, no. 3, pp. 107457, Dec. 2022. doi: [10.1016/j.compag.2022.107457](https://doi.org/10.1016/j.compag.2022.107457).
- [92] S. Jeong, J. Ko, T. Shin, and J. Yeom, "Incorporation of machine learning and deep neural network approaches into a remote sensing-integrated crop model for the simulation of rice growth," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, May 2022. doi: [10.1038/s41598-022-13232-y](https://doi.org/10.1038/s41598-022-13232-y).
- [93] Y. Zhao *et al.*, "The prediction of wheat yield in the North China plain by coupling crop model with machine learning algorithms," *Agriculture*, vol. 13, no. 1, pp. 99, Dec. 2023. doi: [10.3390/agriculture13010099](https://doi.org/10.3390/agriculture13010099).
- [94] D. B. Johnston, K. G. Pembleton, N. I. Huth, and R. C. Deo, "Comparison of machine learning methods emulating process driven crop models," *Environ. Model. Softw.*, vol. 162, no. 7, pp. 105634, May 2023. doi: [10.1016/j.envsoft.2023.105634](https://doi.org/10.1016/j.envsoft.2023.105634).
- [95] L. Xiao, G. Wang, H. Zhou, X. Jin, and Z. Luo, "Coupling agricultural system models with machine learning to facilitate regional predictions of management practices and crop production," *Environ. Res. Lett.*, vol. 17, no. 11, pp. 114027, Nov. 2022. doi: [10.1088/1748-9326/ac9c71](https://doi.org/10.1088/1748-9326/ac9c71).
- [96] Y. Ren *et al.*, "Analysis of corn yield prediction potential at various growth phases using a process-based model and deep learning," *Plants*, vol. 12, no. 3, pp. 446, Jan. 2023. doi: [10.3390/plants12030446](https://doi.org/10.3390/plants12030446).
- [97] Y. Chang, J. Latham, M. Licht, and L. Wang, "A data-driven crop model for maize yield prediction," *Commun. Biol.*, vol. 6, no. 1, pp. 439, Apr. 2023. doi: [10.1038/s42003-023-04833-y](https://doi.org/10.1038/s42003-023-04833-y).
- [98] A. Morales and F. J. Villalobos, "Using machine learning for crop yield prediction in the past or the future," *Front. Plant Sci.*, vol. 14, pp. 1128388, Mar. 2023. doi: [10.3389/fpls.2023.1128388](https://doi.org/10.3389/fpls.2023.1128388).
- [99] H. Zhuang *et al.*, "Integrating data assimilation, crop model, and machine learning for winter wheat yield forecasting in the North China plain," *Agric. For. Meteorol.*, vol. 347, pp. 109909, Mar. 2024. doi: [10.1016/j.agrformet.2024.109909](https://doi.org/10.1016/j.agrformet.2024.109909).
- [100] M. S. Dhillon *et al.*, "Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape," *Front. Remote Sens.*, vol. 3, pp. 1010978, Jan. 2023. doi: [10.3389/frsen.2022.1010978](https://doi.org/10.3389/frsen.2022.1010978).
- [101] P. Nevavuori, N. Narra, P. Linna, and T. Lipping, "Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models," *Remote Sens.*, vol. 12, no. 23, pp. 4000, Dec. 2020. doi: [10.3390/rs12234000](https://doi.org/10.3390/rs12234000).

- [102] G. Morales, J. W. Sheppard, P. B. Hegedus, and B. D. Maxwell, "Improved yield prediction of winter wheat using a novel two-dimensional deep regression neural network trained via remote sensing," *Sensors*, vol. 23, no. 1, pp. 489, Jan. 2023. doi: [10.3390/s23010489](https://doi.org/10.3390/s23010489).
- [103] V. R. R. Kolipaka and A. Namburu, "An automatic crop yield prediction framework designed with two-stage classifiers: A meta-heuristic approach," *Multimed. Tools Appl.*, vol. 83, no. 10, pp. 28969–28992, Mar. 2024. doi: [10.1007/s11042-023-16612-2](https://doi.org/10.1007/s11042-023-16612-2).
- [104] Z. Jiang, C. Liu, B. Ganapathysubramanian, D. J. Hayes, and S. Sarkar, "Predicting county-scale maize yields with publicly available data," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Sep. 2020. doi: [10.1038/s41598-020-71898-8](https://doi.org/10.1038/s41598-020-71898-8).
- [105] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [106] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," *Thirty-First AAAI Conf. Artif. Intell.*, vol. 31, no. 1, pp. 4559–4566, 2017. doi: [10.1609/aaai.v31i1.11172](https://doi.org/10.1609/aaai.v31i1.11172).
- [107] A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in *Proc. 1st ACM SIGCAS Conf. Comp. Sustain. Soc.*, 2018, pp. 1–5. doi: [10.1145/3209811.3212707](https://doi.org/10.1145/3209811.3212707).
- [108] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," *Sensors*, vol. 19, no. 20, pp. 4363, Oct. 2019. doi: [10.3390/s19204363](https://doi.org/10.3390/s19204363).
- [109] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting," *Expert Syst. Appl.*, vol. 184, no. 4, pp. 115511, Dec. 2021. doi: [10.1016/j.eswa.2021.115511](https://doi.org/10.1016/j.eswa.2021.115511).
- [110] P. Abbaszadeh, K. Gavahi, A. Alipour, P. Deb, and H. Moradkhani, "Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction," *Agric. For. Meteorol.*, vol. 314, no. 1, pp. 108773, Mar. 2022. doi: [10.1016/j.agrformet.2021.108773](https://doi.org/10.1016/j.agrformet.2021.108773).
- [111] S. Sharma, S. Rai, and N. C. Krishnan, "Wheat crop yield prediction using deep LSTM model," Nov. 2020. doi: [10.48550/arXiv.2011.01498](https://doi.org/10.48550/arXiv.2011.01498).
- [112] M. Qiao *et al.*, "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 102, no. 12, pp. 102436, Oct. 2021. doi: [10.1016/j.jag.2021.102436](https://doi.org/10.1016/j.jag.2021.102436).
- [113] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," *Agric. For. Meteorol.*, vol. 284, pp. 107886, Apr. 2020. doi: [10.1016/j.agrformet.2019.107886](https://doi.org/10.1016/j.agrformet.2019.107886).
- [114] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Comput. Electron. Agric.*, vol. 163, no. 11, pp. 104859, Aug. 2019. doi: [10.1016/j.compag.2019.104859](https://doi.org/10.1016/j.compag.2019.104859).
- [115] J. Fan, J. Bai, Z. Li, A. Ortiz-Bobea, and C. P. Gomes, "A GNN-RNN approach for harnessing geospatial and temporal information: Application to crop yield prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 11873–11881. doi: [10.1609/aaai.v36i11.21444](https://doi.org/10.1609/aaai.v36i11.21444).
- [116] S. A. Haider *et al.*, "LSTM neural network-based forecasting model for wheat production in Pakistan," *Agronomy*, vol. 9, no. 2, pp. 72, Feb. 2019. doi: [10.3390/agronomy9020072](https://doi.org/10.3390/agronomy9020072).
- [117] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction," *Front. Plant Sci.*, vol. 10, pp. 1750, Jan. 2020. doi: [10.3389/fpls.2019.01750](https://doi.org/10.3389/fpls.2019.01750).
- [118] S. H. Bhojani and N. Bhatt, "Wheat crop yield prediction using new activation functions in neural network," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13941–13951, Sep. 2020. doi: [10.1007/s00521-020-04797-8](https://doi.org/10.1007/s00521-020-04797-8).
- [119] X. Wang, J. Huang, Q. Feng, and D. Yin, "Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches," *Remote Sens.*, vol. 12, no. 11, pp. 1744, May 2020. doi: [10.3390/rs12111744](https://doi.org/10.3390/rs12111744).

- [120] B. Alhnaity, S. Pearson, G. Leontidis, and S. Kollias, "Using deep learning to predict plant growth and yield in greenhouse environments," in *Int. Symp. Adv. Technol. Manag. Innov. Greenhouses: GreenSys2019*, 2019, pp. 425–432. doi: [10.17660/ActaHortic.2020.1296.55](https://doi.org/10.17660/ActaHortic.2020.1296.55).
- [121] D. Elavarasan and P. M. D. Vincent, "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications," *IEEE Access*, vol. 8, pp. 86886–86901, May 2020. doi: [10.1109/ACCESS.2020.2992480](https://doi.org/10.1109/ACCESS.2020.2992480).
- [122] A. K. Srivastava *et al.*, "Winter wheat yield prediction using convolutional neural networks from environmental and phenological data," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, Feb. 2022. doi: [10.1038/s41598-022-06249-w](https://doi.org/10.1038/s41598-022-06249-w).
- [123] S. Khaki, H. Pham, and L. Wang, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, May 2021. doi: [10.1038/s41598-021-89779-z](https://doi.org/10.1038/s41598-021-89779-z).
- [124] Y. Ma, Z. Zhang, Y. Kang, and M. Özdoğan, "Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach," *Remote Sens. Environ.*, vol. 259, no. 8, pp. 112408, Jun. 2021. doi: [10.1016/j.rse.2021.112408](https://doi.org/10.1016/j.rse.2021.112408).
- [125] J. Cao *et al.*, "Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches," *Agric. For. Meteorol.*, vol. 297, no. 6, pp. 108275, Feb. 2021. doi: [10.1016/j.agrformet.2020.108275](https://doi.org/10.1016/j.agrformet.2020.108275).
- [126] H. Tian, P. Wang, K. Tansey, J. Zhang, S. Zhang and H. Li, "An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China," *Agric. For. Meteorol.*, vol. 310, no. 6, pp. 108629, Nov. 2021. doi: [10.1016/j.agrformet.2021.108629](https://doi.org/10.1016/j.agrformet.2021.108629).
- [127] N. Bali and A. Singla, "Deep learning-based wheat crop yield prediction model in Punjab region of North India," *Appl. Artif. Intell.*, vol. 35, no. 15, pp. 1304–1328, Dec. 2021. doi: [10.1080/08839514.2021.1976091](https://doi.org/10.1080/08839514.2021.1976091).
- [128] N. Chergui, "Durum wheat yield forecasting using machine learning," *Artif. Intell. Agric.*, vol. 6, pp. 156–166, Jan. 2022. doi: [10.1016/j.aiaa.2022.09.003](https://doi.org/10.1016/j.aiaa.2022.09.003).
- [129] Y. Shen *et al.*, "Improving wheat yield prediction accuracy using LSTM-RF framework based on UAV thermal infrared and multispectral imagery," *Agriculture*, vol. 12, no. 6, pp. 892, Jun. 2022. doi: [10.3390/agriculture12060892](https://doi.org/10.3390/agriculture12060892).
- [130] K. Alibabaei, P. D. Gaspar, and T. M. Lima, "Crop yield estimation using deep learning based on climate big data and irrigation scheduling," *Energies*, vol. 14, no. 11, pp. 3004, May 2021. doi: [10.3390/en14113004](https://doi.org/10.3390/en14113004).
- [131] Y. J. Chang, M. H. Lai, C. H. Wang, Y. S. Huang, and J. Lin, "Target-Aware Yield Prediction (TAYP) model used to improve agriculture crop productivity," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–11, Mar. 2024. doi: [10.1109/TGRS.2024.3376078](https://doi.org/10.1109/TGRS.2024.3376078).
- [132] P. S. S. Gopi and M. Karthikeyan, "Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 13159–13179, Feb. 2024. doi: [10.1007/s11042-023-16113-2](https://doi.org/10.1007/s11042-023-16113-2).
- [133] M. V. Rao, Y. Sreeraman, S. V. Mantena, V. Gundu, D. Roja and R. Vatambeti, "Brinjal crop yield prediction using Shuffled shepherd optimization algorithm based ACNN-OBDLSTM model in Smart Agriculture," *J. Integr. Sci. Technol.*, vol. 12, no. 1, pp. 710, 2024. Accessed: Apr. 19, 2024. [Online]. Available: <https://pubs.thesciencein.org/journal/index.php/jist/article/view/a710/423>
- [134] S. Boppudi and S. Jayachandran, "Improved feature ranking fusion process with hybrid model for crop yield prediction," *Biomed. Signal Process. Control*, vol. 93, no. 27, pp. 106121, Jul. 2024. doi: [10.1016/j.bspc.2024.106121](https://doi.org/10.1016/j.bspc.2024.106121).
- [135] V. Joshua, S. M. Priyadharson, and R. Kannadasan, "Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu," *Agronomy*, vol. 11, no. 10, pp. 2068, Oct. 2021. doi: [10.3390/agronomy11102068](https://doi.org/10.3390/agronomy11102068).
- [136] K. Kuwata and R. Shibasaki, "Estimating corn yield in the united states with modis EVI and machine learning methods, ISPRS annals of the photogrammetry," *Remote Sens. Spat. Inf. Sci.*, vol. 3, pp. 131–136, Jun. 2016. doi: [10.5194/isprs-annals-III-8-131-2016](https://doi.org/10.5194/isprs-annals-III-8-131-2016).

- [137] P. Mohan and K. K. Patil, "Deep learning based weighted SOM to forecast weather and crop prediction for agriculture application," *Int. J. Intell. Eng. Sys.*, vol. 11, no. 4, pp. 167–176, Aug. 2018. Accessed: Apr. 19, 2024. [Online]. Available: <https://www.inass.org/2018/2018083117.pdf>
- [138] Z. Jiang, C. Liu, N. P. Hendricks, B. Ganapathysubramanian, D. J. Hayes and S. Sarkar, "Predicting county level corn yields using deep long short term memory models," May 2018. doi: [10.48550/arXiv.1805.12044](https://doi.org/10.48550/arXiv.1805.12044).
- [139] M. B. Villanueva and M. L. M. Salenga, "Bitter melon crop yield prediction using machine learning algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, Mar. 2018. doi: [10.14569/IJACSA.2018.090301](https://doi.org/10.14569/IJACSA.2018.090301).
- [140] J. Fourie, J. Hsiao, and A. Werner, "Crop yield estimation using deep learning," in *7th Asian-Australas. Con. Precis. Agri.*, Oct. 2017, pp. 1–10.
- [141] R. Tanabe, T. Matsui, and T. S. T. Tanaka, "Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery," *Field Crops Res.*, vol. 291, no. 1, pp. 108786, Feb. 2023. doi: [10.1016/j.fcr.2022.108786](https://doi.org/10.1016/j.fcr.2022.108786).
- [142] A. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid deep learning-based models for crop yield prediction," *Appl. Artif. Intell.*, vol. 36, no. 1, pp. 2031822, Dec. 2022. doi: [10.1080/08839514.2022.2031823](https://doi.org/10.1080/08839514.2022.2031823).
- [143] A. Tzachor, M. Devare, B. King, S. Avin, and S. ÓhÉigeartaigh, "Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities," *Nat. Mach. Intell.*, vol. 4, no. 2, pp. 104–109, Feb. 2022. doi: [10.1038/s42256-022-00440-4](https://doi.org/10.1038/s42256-022-00440-4).
- [144] S. O. Araújo, R. S. Peres, J. Barata, F. Lidon, and J. C. Ramalho, "Characterising the Agriculture 4.0 landscape—emerging trends, challenges and opportunities," *Agronomy*, vol. 11, no. 4, pp. 667, Apr. 2021. doi: [10.3390/agronomy11040667](https://doi.org/10.3390/agronomy11040667).
- [145] S. Qazi, B. A. Khawaja, and Q. U. Farooq, "IoT-equipped and AI-enabled next generation smart agriculture: A critical review, current challenges and future trends," *IEEE Access*, vol. 10, pp. 21219–21235, Feb. 2022. doi: [10.1109/ACCESS.2022.3152544](https://doi.org/10.1109/ACCESS.2022.3152544).
- [146] M. Javaid, A. Haleem, I. H. Khan, and R. Suman, "Understanding the potential applications of artificial intelligence in agriculture sector," *Adv. Agrochem*, vol. 2, no. 1, pp. 15–30, Mar. 2023. doi: [10.1016/j.aac.2022.10.001](https://doi.org/10.1016/j.aac.2022.10.001).
- [147] F. Xiao, H. Wang, Y. Xu, and R. Zhang, "Fruit detection and recognition based on deep learning for automatic harvesting: An overview and review," *Agronomy*, vol. 13, no. 6, pp. 1625, Jun. 2023. doi: [10.3390/agronomy13061625](https://doi.org/10.3390/agronomy13061625).
- [148] A. Cravero, S. Pardo, S. Sepúlveda, and L. Muñoz, "Challenges to use machine learning in agricultural big data: A systematic literature review," *Agronomy*, vol. 12, no. 3, pp. 748, Mar. 2022. doi: [10.3390/agronomy12030748](https://doi.org/10.3390/agronomy12030748).