Review article

Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability

Md. Abu Jabed [a,b], Masrah Azrifah Azmi Murad [a,*]

[a] *Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia*
[b] *Department of Computer Science and Engineering, University of Creative Technology Chittagong, Chattogram, Bangladesh*

ARTICLE INFO

ABSTRACT

The agriculture sector is confronted with numerous challenges in the quest for accurate crop yield estimation, which is essential for efficient resource management and mitigating food scarcity in a rapidly growing global population. This research paper delves into the application of advanced Artificial Intelligence (AI) techniques to enhance crop yield estimation in the context of diverse agricultural challenges. Through a systematic literature review and analysis of relevant studies, this paper explores the role of AI methods, such as Machine Learning (ML) and Deep Learning (DL), in addressing the complexities posed by geographical variations, crop diversity, and cultivation areas. The review identifies a wealth of AI-powered solutions employed in crop yield prediction, emphasizing the importance of precise environmental and agricultural data. Key factors contributing to accurate estimation include temperature, rainfall, soil type, humidity, and various vegetation indices, such as NDVI, EVI, LAI, and NDWI. The research paper also examines the algorithms frequently utilized in the machine learning domain, including Random Forest (RF), Artificial Neural Networks (ANN), and Support Vector Machine (SVM). In the realm of deep learning, Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), and Deep Neural Networks (DNN) emerge as promising candidates. The findings of this study shed light on the transformative potential of advanced AI techniques in improving crop yield estimation accuracy, ultimately enhancing agricultural planning and resource management. By addressing the challenges posed by geographical diversity, crop heterogeneity, and changing environmental conditions, AI-driven models offer new avenues for sustainable agriculture in an ever-evolving world. This research paper provides valuable insights and directions for future studies, highlighting the critical role of AI in ensuring food security and sustainability in agriculture.

1. Introduction

Agriculture, a focus of scientific and technological innovation, has seen significant transformations due to technology (Montero et al., 2020). Crop yield estimation is a critical aspect, depending on various factors, including agrometeorological variables like soil properties, climate, and irrigation [1]. Potential and actual yields, as well as the yield gap, are essential concepts in this context [1]. Traditionally, crop performance assessments relied on observations by experts, influencing decision-making and livelihoods [2].

However, errors in these predictions can lead to food shortages. Modern precision agriculture faces challenges in predicting crop production, with many effective models proposed [3]. Addressing these challenges requires extensive datasets [4] and a continuous drive for improved accuracy [5].

Machine learning techniques are being widely adopted in agriculture to monitor and predict environmental factors [6,7]. These techniques build prediction models using data and have applications across various domains (Udousoro, 2020). However, developing accurate and interpretable models in crop production prediction remains challenging due to multifaceted influences [8]. Various methods, including field surveys, statistical models, crop growth models, and remote sensing, have been employed in crop yield prediction [9].

Accurate crop yield prediction is essential at national and regional levels, enabling informed decisions by farmers [4,10]. Multiple methods exist for forecasting agricultural yields, necessitating extensive datasets due to the complex nature of agricultural production influenced by climate and soil factors [4]. The choice between predictive and descriptive machine learning models depends on the research context [3,11].

This study conducted a systematic literature review (SLR) to explore the use of machine learning and deep learning in predicting crop yields. The SLR aims to identify research gaps and provide guidance to practitioners and academics interested in this field [3]. Relevant papers were retrieved from electronic databases and synthesized following established SLR methodology to address the research questions. The organization of this study is as follows: Section II delves into the background, Section III offers an overview of crop yield prediction, Section IV presents the results and discussion, Section V introduces practical applications, section VI focuses on recommendations for future research, and Section VII serves as the conclusion of the paper.

2. Section II. Related work

In this section, we summarize the existing review articles on yield estimation. Prioritizing crop production enhancement is vital for efficient decision-making at national and regional levels. Various methods exist for forecasting agricultural yields, and it is a challenging task in precision agriculture, with several effective models introduced [4]. A comprehensive understanding of crop production growth patterns is essential for farmers. This review article delves into the research on crop production prediction utilizing machine learning and deep learning techniques found in the literature.

Chlingaryan et al. [12] reviewed nitrogen status estimation in agriculture using machine learning to improve crop yield predictions. They anticipate the development of cost-effective agricultural solutions through advances in sensing technologies and machine learning, with the emergence of hybrid systems based on machine learning approaches.

Liakos et al. [13] examined a review of the application of machine learning in agriculture, analyzing publications related to water resources, livestock, crop, and land management. Their study highlights the potential benefits of machine learning technologies for the agricultural sector.

Young [14] examined essential methods in official statistics, remote sensing, and surveys for crop yield forecasting, highlighting prediction uncertainties and research gaps. However, it lacks coverage of prevalent machine learning techniques and specific crop yield models for different crops, potentially limiting its relevance to some readers.

Elavarasan et al. [15] conducted a review of papers focusing on machine learning models for forecasting agricultural production using meteorological factors. The research indicates the need for further investigation into additional factors influencing crop productivity.

Kamilaris & Prenafeta-Boldú [16] surveyed 40 research initiatives using deep learning in agriculture and food production. They found that deep learning is highly effective, outperforming traditional image processing methods, and can address diverse agricultural challenges with high accuracy.

Beulah [17] assessed various data mining approaches used in crop production prediction and concluded that data mining techniques can be applied to address this issue.

Koirala et al. [18] focused on deep learning for fruit identification and localization in agriculture, advocating for standardized metrics in model comparison. They highlighted the efficacy of deep learning methods and recommended CNN detectors, deep regression, and LSTM for fruit load assessment.

van Klompenburg et al. [3] conducted a systematic review of crop yield prediction techniques, focusing on information extraction without in-depth analysis or recommendations. They identified common deep learning algorithms used in the field, with CNN, LSTM, and DNN being the most preferred methods for crop yield prediction.

Häni et al. [19] presented a system for apple yield estimation through deep learning-based fruit detection and counting. Their study compared semi-supervised and deep learning methods, revealing that Gaussian mixture models outperformed deep learning methods like U-NET, Fast R-CNN, and CNN in most datasets for yield detection.

Zhang et al. [20] explored Deep Learning (DL) for agricultural tasks in dense scenes, showing its efficiency and accuracy in activities like recognition and yield estimation. DL, particularly Convolutional Neural Networks (CNNs), outperformed alternative methods in dense agricultural scenes, emphasizing its importance in computer vision tasks.

Dharani et al. [21] explore the use of deep learning methods, emphasizing their role in crop yield prediction for agriculture. They propose that artificial intelligence can improve crop management and yield forecasting, with recurrent neural networks and hybrid networks like RNN-LSTM showing superior accuracy in predictions compared to other networks.

Darwin et al. [22] review deep learning and computer vision in smart agriculture for crop yield estimation, highlighting automation benefits in image analysis and remote sensing. They conclude that using deep learning with machine vision improves the accuracy of automated agricultural systems.

Maheswari et al. [23] investigate deep learning-based semantic segmentation for fruit yield estimation in orchards, emphasizing its advantages over traditional methods. They address challenges in fruit yield estimation and suggest that DL-based techniques outperform traditional approaches by incorporating human cognition into the architecture.

Monteiro et al. [24] provide a brief overview of scientific and technological tools in precision agriculture and their applications in crop and livestock farming. They emphasize resource optimization and how precision agriculture can meet food demand while ensuring sustainability.

Rashid et al. [25] conducted a comprehensive review on machine learning-based palm oil yield prediction, highlighting the need for diverse features and prediction techniques in future research. They note the common use of ML algorithms like LR, RF, and NN, as well as some DL models in crop yield forecasting.

Benos et al. [26] assess current scholarly literature on machine learning in agriculture, providing insights for stakeholders interested in its potential advantages. The study highlights the use of various sensors on satellites, ground vehicles, and aerial vehicles for gathering reliable input data for machine learning in agriculture.

Hasan et al. [27] provide an overview of deep learning-based algorithms for weed identification and classification in agriculture, covering data collection, dataset preparation, deep learning techniques, and assessment metrics approaches.

Modi et al. [1] conduct a comprehensive investigation into non-destructive techniques for yield prediction, focusing on data collection, pre-processing, properties, techniques, and outcomes, while identifying commonly used methods and recommending model integration for improved accuracy.

Muruganantham et al. [28] highlight the benefits of using deep learning in crop yield forecasting and recommend remote sensing technologies based on data requirements and influencing factors. The study acknowledges challenges like enhancing model accuracy, practical application, and addressing the opacity of deep learning models.

Oikonomidis et al. [29] conduct a systematic review on deep learning applications in crop yield prediction, analyzing motivations, target crops, methods, features, and data sources in the identified papers.

Bouguettaya et al. [30] review the use of deep learning techniques for crop and plant classification from UAV-based remote sensing imagery, emphasizing the importance of robust tools like Convolutional Neural Networks (CNN) and the potential of combining UAV data and deep learning for accurate crop classification.

Ojo & Zahid [31] explore the recent developments, challenges, and future prospects of using deep learning in controlled environment agriculture (CEA). They highlight DL applications in CEA, challenges, research directions, and commonly used models, focusing on CNNs and RNN-LSTM for time series forecasting in CEA.

The purpose of this systematic literature review is to analyze the influence of environmental factors on crop growth and identify research gaps in machine learning and deep learning technology. It explores the benefits of using machine learning and deep learning for crop yield prediction, identifies appropriate remote sensing technologies, and considers factors affecting crop yield, offering fresh insights into current research.

3. Section III. Methodology

3.1. Review protocol

This study conducts a systematic literature review (SLR) on crop yield estimation, focusing solely on journal articles. Its purpose is to identify research gaps in the machine learning and deep learning methods within this specific field [28]. The review not only encompasses all journal research but also aligns with our study's research questions. Such comprehensive literature reviews are vital for evaluating theories and data accuracy within a field [32]. Article selection adheres to PRISMA guidelines [33], using databases like Scopus, Google Scholar, Science Direct, PubMed, Web of Science, Mendeley Research Networks, and Wiley. Pertinent research is screened and evaluated based on inclusion and quality standards [3]. All pertinent data is collected from these studies and synthesized to address our research questions.

3.2. Research questions

The following research questions are developed to guide the systematic review:

1. What machine learning and deep learning techniques are employed for crop yield prediction?
2. What features or variables are utilized for crop yield prediction through machine learning and deep learning methods?
3. What criteria and methodologies are applied to evaluate crop yield prediction?
4. What obstacles exist in crop yield prediction using machine learning and deep learning techniques?

Q1 helps us assess both the benefits and drawbacks of employing machine learning and deep learning techniques for crop yield prediction. Q2 assists us in gaining insight into the diverse factors that impact the utilization of machine learning and deep learning methods in crop yield prediction. Q3 provides valuable insights into the assessment criteria and methods applied in the context of crop yield prediction. Q4 facilitates our comprehension of the constraints and difficulties associated with existing approaches.

3.3. Technique for article searches

The systematic literature review's objective forms the foundation for the article search methodology. The search is carried out with a specific emphasis on the key concepts relevant to this review's parameters. The primary focus of this review is shaped by the terms "machine learning or deep learning" and "crop yield prediction." The scope is restricted to publications from January 2018 to April 2023. Upon identifying a pertinent study, an examination of its references was conducted to identify any additional studies missed in the initial search. This iterative process continued until no more relevant studies were uncovered.

The exclusion criteria employed during the analysis of retrieved publications included the determination of whether they were survey papers or general review articles. The subsequent section addresses the topics associated with the excluded publications.

Exclusion Criteria The studies underwent a comprehensive assessment and were systematically categorized according to pre-defined exclusion criteria. These criteria were set to delineate the boundaries of the systematic review and to eliminate studies that did not meet the relevant criteria. The subsequent list outlines the exclusion criteria (EC):

First criterion Articles pertaining to agriculture but not specifically focused on crop yield prediction.

Second criterion Publications that have been previously acquired or are duplicates.

Third criterion Studies without full-text accessibility.

Fourth criterion Articles authored in languages other than English, conference papers, book chapters, reviews, surveys, Master's theses, or PhD dissertations.

Fifth criterion Publications published before the year 2018.

Following the application of the exclusion criteria, a total of 176 articles were ultimately chosen. Subsequently, the removal of duplicate entries from the selected databases led to the evaluation of 115 unique articles. It is worth emphasizing that only journal articles met the inclusion criteria for this review. The methodology employed in this review adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, as illustrated in Fig. 1 (PRISMA, accessed on February 1, 2021).

Table 1 displays the initial tally of retrieved papers, the quantity that met the selection criteria, and the total number of articles remaining after the removal of duplicates. Notably, a substantial portion of the papers originated from the Scopus, Science Direct, and Google Scholar databases. The primary focus of data analysis revolved around assessing adherence to the exclusion criteria. In the data synthesis stage, all accumulated information was compiled and integrated.

3.4. Approaches used in literature discussion

We have identified and highlighted a total of 115 research papers published in various journals from 2018 to April 2023, all of which are relevant to our study. These papers have been meticulously compiled into a comprehensive Literature Review Table. This table encompasses essential details, including the databases utilized, author names, publication years, data sources, features employed, methodologies applied, measured attributes, research objectives, findings, and any limitations noted in these articles.

Furthermore, we have dedicated an Appendix section that presents a detailed analysis of the 20 most closely related papers from the aforementioned selection. These papers offer a more in-depth exploration of the subject matter for reference and further examination.

The Appendix (Literature Review Table) offers an overview of the various approaches employed for predicting crop yields, encompassing machine learning, deep learning, and hybrid methods. Unique approaches like Extreme Learning Machine (ELM), Stochastic Gradient Descent (SGD), Sequential Minimal Optimization Regression (SMOreg), Elastic Net (EN), Interaction Regression (IR), Deep Recurrent Q-Network (DRQN), Cubist, and YieldNet are discussed. Frequently used machine learning algorithms include Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN). K-Nearest Neighbors (K-NN), Decision Trees (DT), and Multi Linear Regression (MLR) are also common choices.

In the realm of deep learning, CNN, DNN, and LSTM are frequently employed for crop yield prediction. Deep learning falls under the broader umbrella of machine learning and is gaining significant attention for addressing crop yield prediction challenges. Additionally, these deep learning approaches are often combined, such as in CNN-LSTM, RNN-LSTM, and CNN-RNN multilevel deep learning systems with multiple layers, along with multimodal fusion techniques. Transfer learning (TL) is another method by which pre-trained deep learning models can be adapted for accurate crop yield prediction across various agricultural settings.
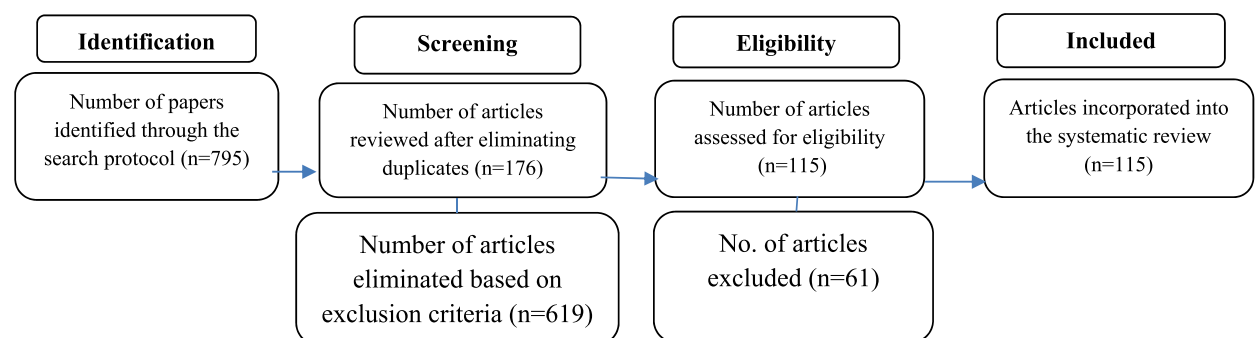

Identification	**Screening**	**Eligibility**	**Included**
Number of papers identified through the search protocol (n=795)	Number of articles reviewed after eliminating duplicates (n=176)	Number of articles assessed for eligibility (n=115)	Articles incorporated into the systematic review (n=115)
	Number of articles eliminated based on exclusion criteria (n=619)	No. of articles excluded (n=61)	

Fig. 1. Flowchart displaying search results.

Table-1
Distribution of articles is based on the databases.

Database	Number of articles that were first retrieved	Number of papers remaining after exclusion criteria	Number of articles after removing the repeated articles
Science Direct	87	38	26
Web of Science	68	12	8
Scopus	255	72	53
Google Scholar	294	33	20
PubMed	36	9	5
Mendeley Research Networks	20	5	1
Wiley	35	7	2
Total	**795**	**176**	**115**

Notably, when remote sensing data is incorporated, simple neural networks are less commonly used for agricultural yield forecasting. Data pre-processing is a prevalent component integrated into the majority of these methodologies.

3.4.1. Machine learning approach

In contrast to rule-based methods, machine learning utilizes statistical analysis to detect data patterns. Traditional machine learning approaches like SVM, LR, RF, K-NN, and DT are trained on labeled datasets for predicting new labels. RF, SVM, and ANN have shown excellent performance in crop yield prediction [34]. RF is an efficient method widely used in agricultural studies [35–40]. SVM is suitable for yield prediction using weather and MODIS-based vegetation indices [41]. ANNs are often used for crop production prediction models [42]. Machine learning predictors for classification include PSO-SVM, KNN, and RF [43].

Conventional machine learning methods require pre-processing stages, including data pre-processing, feature extraction, and feature selection, to effectively train algorithms [44]. A study [45] combined six algorithms to enhance crop identification based on soil type. ML models demonstrate strong performance using 10-fold cross-validation [46].

3.4.2. Deep learning-based crop yield prediction

Deep learning (DL), a subset of machine learning, excels at analyzing labeled and unstructured data [47,48]. It's widely applied in agriculture due to its ability to handle large datasets, discover variable correlations, and employ nonlinear functions [49,50]. Deep learning outperforms traditional methods in feature extraction, crucial for agricultural yield prediction [48].

Notably, Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM) and Deep Neural Network (DNN) are the most widely used deep learning methods in crop yield prediction. LSTM, a variant of Recurrent Neural Networks (RNN), is particularly valuable for capturing time-dependent information [51]. Deep neural networks consist of multiple nonlinear layers, extracting information at each level [52]. Discovering nonlinear associations between input and response variables is a key role of deep neural networks, often with various hidden layers [52].

Enhancing deep learning algorithms' performance involves using techniques like stochastic gradient descent (SGD), batch normalization, and dropout. Some of these deep learning methods are briefly outlined below:

Deep Neural Networks (DNN) DNN methods, apart from the number of hidden layers, closely resemble traditional ANN algorithms [53]. Both DNN and ANN networks typically feature multiple fully connected hidden layers [54]. In contrast, other deep learning algorithms like CNN incorporate various types of layers, including convolutional and pooling layers [3].

Convolutional Neural Networks (CNN) CNN comprises fundamental units arranged between input and output layers, including convolutional, pooling, and activation layers [30,55]. In the convolution layer, local filters perform convolution operations on input data, while the pooling layer generates reduced-dimensional data through operations like max-pooling and average-pooling. The activation layer's nonlinear operations enhance CNN's ability for nonlinear fitting [55]. CNN updates weights using Backpropagation (BP), similar to the Backpropagation Neural Network (BPNN) approach.

Long-Short Term Memory (LSTM) LSTM is a type of Recurrent Neural Network (RNN) that uses gradient-based algorithms to learn time-dependent data. LSTM consists of a chain structure consisting of an input layer, one or more LSTM layers, and an output layer [25,28].

Recurrent Neural Networks (RNN) Recurrent Neural Networks (RNNs) are a type of neural network that processes data in sequences [56]. It is a sort of artificial neural network in which the temporal dependencies of nodes are represented by a directed graph [57]. RNN is useful for sequence modeling; This approach makes RNN more efficient because the sequence data used is better understood [50].

Transfer Learning (TL) Transfer learning is a machine learning technique where a model trained on one task is repurposed on a second related task. It is particularly useful when working with limited labeled data because it allows us to leverage knowledge from one domain and apply it to another [58]. In deep learning, transfer learning involves taking a pre-trained model (trained on a large dataset) and fine-tuning it on a new dataset or task [58,59]. This process can significantly speed up training and improve the performance of the model, especially when the new task has limited available data [59].

3.4.3. Miscellaneous approaches in crop yield prediction

Various classification and regression techniques have proven successful in agricultural production prediction, including Linear

Regression (LR) [60], Naive Bayes (NB) [61], Extreme Gradient Boosting (XGBoost) [62–65], Gradient boosting (GB) (Anbananthen et al., 2022), Interaction Regression [8], AdaBoost [66], Logistic Regression (LR) [61], Levenberg–Marquardt (LM) [67], Extreme Learning Machine (ELM) [68], Deep Learning Multi-Layer Perceptron (DLMLP) [69], Deep recurrent Q-learning (DRQN) [70], General Regression Neural Networks (GRNN) (S. V. [71]; V. [72]), Deep Convolutional Regression Network (DCRN) [73], Light Gradient Boosting Machine (LightGBM) [74], Linear Discriminant Analysis (LDA) (Mupangwa et al., 2020), Multi-parametric Deep Neural Network (MDNN) (Kalaiarasi and Anbarasi, 2021), Special-Spectral-Temporal Neural Network (SSTNN) [75], YieldNet [76], Ensemble methods [36,45,77–81], Stochastic Gradient Descent (SGD) [82], and Cubist [83], all of which have been effectively used in crop yield prediction.

3.4.4. Hybrid approach to crop yield prediction

Combining the strengths of various machine learning and deep learning algorithms is a common practice. Anbananthen et al. [36] integrated gradient boosting, random forest, and LASSO regression for localized crop yield forecasting. Raja et al. [80] integrated RF, Bagging, K-NN, SVM, DT, and NB to improve prediction accuracy for cereals, potatoes, and energy crops. Sajid et al. [84] combined LR, Lasso Regression, RF, XG Boost, and Light GBM for maize yield prediction. Srivastava et al. [85] integrated CNNs and FCNNs for winter wheat yield prediction. Oikonomidis et al. [65] combined CNN-LSTM, CNN-DNN, CNN-XGBoost, and CNN-RNN for soybean yield prediction. Khaki et al. [57] integrated CNN-RNN for corn and soybean yield prediction. Olofintuyi et al. [50] used CNN-RNN with LSTM for cocoa yield prediction. Bali & Singla [49] combined RNN-LSTM for wheat yield prediction. Shook et al. [86] integrated SVR-RBF for soybean yield prediction. Jui et al. [87] used DRS-RF for tea yield prediction. Neural network-based crop yield prediction models have shown promising accuracy ([47,65,81,88,89]; Wang et al., 2018). However, their black-box nature hinders interpretability [90]. To address the interpretability problem in the future, deep learning-based techniques can be utilized to solve fractional models [91,92] and differential equations [93] in the field of crop yield prediction.

3.5. Performance evaluation metrics used in crop yield prediction analysis

The evaluation of model outputs against actual data is a crucial step in estimation. This section discusses the commonly used evaluation metrics and techniques.

Performance Evaluation Metrics Evaluation metrics are essential for assessing model performance, distinguishing between different learning models [70]. Key performance metrics for regression models include the mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), determination coefficient (R-squared), and mean absolute percentage error (MAPE). The average importance of errors with a given array of forecasts is calculated using the MAE, which is defined as an arithmetic mean of the absolute deviations between the predicted observations and the actual observations [94]. MSE measures how closely the regressor line resembles the dataset points, which is used to assess the estimator's performance [70]. The RMSE measures how well the data are focused on the best fit line and used to calculate the standard deviation of the residuals or projected error [95]. The superiority of the generated framework over the baseline framework is demonstrated by the coefficient of determination (R-Squared), which is used to assess how well the regression framework fits the data [96]. The average of the percentage errors, or how far the model's prediction deviates from the related results, is calculated using MAPE [70].

For machine learning-based classification algorithms in crop yield prediction, evaluation includes accuracy [40,80], precision (Li et al., 2018; Sivanantham et al., 2022), recall [59,97], sensitivity [34,43], specificity [80,97], and F1 Score (Kalaiarasi and Anbarasi, 2021; Mupangwa et al., 2020). Classification accuracy remains the most widely used and informative metric for classification problems.

4. Section IV results and discussion

The papers chosen for the review are evaluated and condensed.

Fig. 2 exhibits the number of articles published spanning from 2018 to April 2023. Evidently, there has been notable advancement in research related to predicting crop yields.
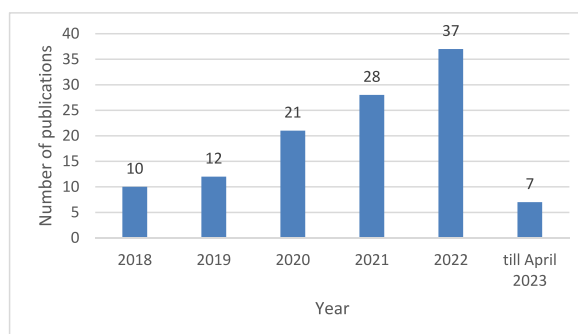

Fig. 2. Distribution of articles.

4.1. Q1: approaches used in literature discussion

For the first research question (Q1) Previous studies have employed a range of classification and regression algorithms for crop yield prediction. Among the selected 115 articles, various machine learning and deep learning methods have been summarized. Unique methods include Interaction Regression (IR), Elastic Net (EN), Extreme Learning Machine (ELM), Cubist, and YieldNet. Commonly used machine learning algorithms are Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN), while Extreme Gradient Boosting (XGBoost), Decision Trees (DT), and Multi Linear Regression (MLR) are also prevalent, as depicted in Fig. 3. In contrast, deep learning algorithms like Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Deep Neural Network (DNN) are frequently utilized for crop yield prediction. Deep learning falls under the umbrella of machine learning and is increasingly promising in this field. These methods are often combined, such as CNN-LSTM and multilevel deep learning systems, particularly when utilizing remote sensing data. Notably, simple neural networks are less commonly used for crop production prediction. Many of these approaches incorporate data pre-processing. Fig. 3 illustrates the distribution of crop yield prediction publications based on machine learning and deep learning methods.

4.1.1. Approaches discussion

Fig. 3 presents the most commonly employed methods as identified from the selected articles.

Random forest (RF) is one of the most popular machine learning techniques for predicting crop yields, as seen in Fig. 3. Because of its unique capacity to locate the crucial details inside the data, it is widely used. The advantages of adopting the RF technique include the ability to resolve the variable collinearity problem, which is typically addressed by using conventional LR models. According to Fig. 3, the second most used algorithm is the Artificial Neuron Network (ANN). Artificial Neurons allows them to process huge volumes of data, identify a pattern from it, and draw conclusions. ANN was utilized extensively in agricultural yield prediction. The third most used algorithm is support vector machine (SVM). SVR performed significantly better than K-NN, BNN, and LR in research [98].

Additionally, depicted in Fig. 3 is the convolutional neural network (CNN), which stands as one of the most widely used deep learning techniques for forecasting agricultural productivity. It has a unique ability to locate the crucial details in the data, which makes it quite popular. Nevavuori et al. [89] investigated crop yield prediction using a distinctive profile of temperature and photoperiod, creating a region-specific deep learning model. For agricultural production modeling, Oikonomidis et al. [65] created CNN-DNN, CNN-RNN, CNN-LSTM, and CNN-XGBoost. The CNN-DNN model outperformed the others in terms of performance.

Results from prior crop yield prediction by Z. Zhang et al. [99] showed that the ML and DL approaches definitely outperformed LR, mainly because LASSO was able to extract the dynamic correlations between the variables and the target predictor [100,101].The ability of LSTM to learn time-dependent information distinguishes it from other deep learning techniques, which were also employed extensively in crop production prediction. X. Wang et al. [102] studied the model performance accuracy employed various combinations of soil, meteorological, and remote sensing data. It is not always true that the most popular algorithm is also the best one. Both classifiers and clustering methods are employed in the chosen papers. In those publications, images are utilized for clustering; hence the publication is related to machine vision rather than ML utilizing a numerical dataset.

ML and DL techniques such as RF, SVM, DT, ANN, XGBoost, Classification, LSTM, CNN and DNN have been applied for yield estimation. These approaches are based on statistical analysis and regression. It is quite clear that regression-based methods are mainly used for yield estimation [1].

Strengths and Weaknesses of Some Popular Algorithms Agricultural data often contain noisy features and outliers, which regularization can help mitigate. Regularized classifiers like linear SVM outperform non-regularized methods like LDA, and SVM's decision rule is a reliable and low-variance linear function in kernel space [103–105]. Low variability is crucial for minimizing classification errors in the inherently unstable agricultural data over time [53]. SVM performs well with small training sets and high-dimensional feature vectors [12,106]. Random Forest (RF) offers high-speed operation, generalization performance, and an
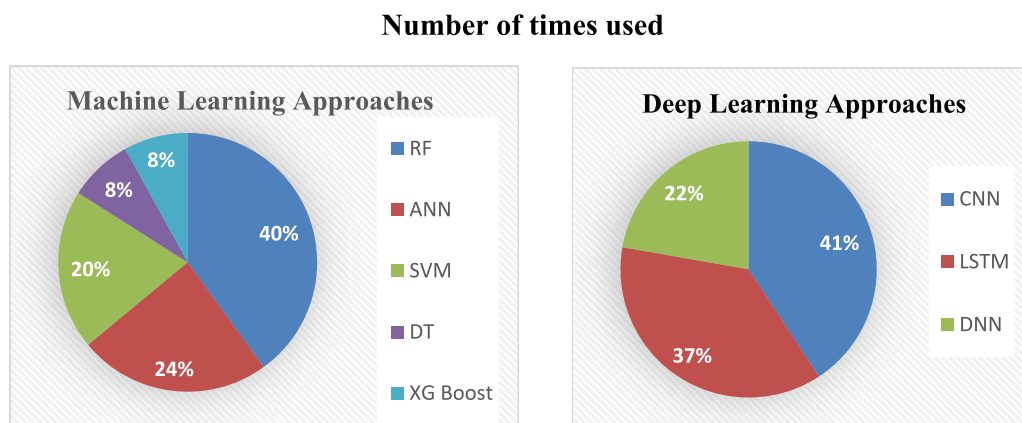
Number of times used

Fig. 3. Most Used Machine Learning and Deep Learning approaches.

ensemble of tree-structured classifiers, making it ideal for finding non-linear patterns [107]. RF requires minimal data preprocessing and excels in feature selection for superior yield results. Convolutional Neural Networks (CNNs) are essential in computer vision and remote sensing, capable of handling various retrieval tasks, including soil moisture retrieval (T. [55,108]). Unlike conventional neural networks, CNNs have a fully connected layer with local connectivity, making them highly effective for processing spatial data.

Algorithm's ability to Cope with Specific Problems High-dimensional feature vectors in agricultural data can be effectively classified using techniques like Support Vector Machines (SVM). To handle sequences of high-dimensional data, dynamic classifiers can be beneficial, particularly when dealing with multiple time segments. Other methods like Random Forest (RF), Principal Component Analysis (PCA), and correlation coefficients can also manage high-dimensional datasets. For improved performance, combining classifiers, especially when dealing with non-stationary datasets, can reduce variance. Combining Linear Discriminant Analysis (LDA) and SVM often leads to better results. Additionally, Artificial Neural Networks (ANN) are frequently used in agricultural production and vegetation prediction due to their ability to extract complex, dynamic, and non-linear patterns, particularly with remote sensing data [12]. Convolutional Neural Networks (CNN) are ideal for image data with substantial training samples, enabling tasks like nutrient level estimation, crop yield mapping, and disease identification. For object recognition applications, advanced CNN architectures such as Faster R-CNN, R-FCN, and SSD are highly effective. Regression-based algorithms like RF, ANN, and Support Vector Regression (SVR) are successful for crop yield prediction, and model predictability can be enhanced by ensembling multiple algorithms [84].

4.2. Q2: features used in crop yield prediction

Regarding research question two (Q2), an exploration and synthesis of the features integrated into the machine learning and deep learning algorithms showcased in the papers was executed. Table 2 provides an overview of the features that have been examined for predicting crop yield. Since crop yield data is a crucial component of crop yield prediction, it is used exclusively in all publications.

Among the features considered, soil type, maximum and minimum temperature, humidity, and rainfall are consistently observed in the majority of studies, highlighting their prevalence. Nevertheless, it's important to recognize that certain studies introduce unique features tailored to their specific contexts. Included in these variables are precipitation forecasts, normalized difference vegetation index (NDVI), humidity, fertilizer, soil pH, enhanced vegetation index (EVI), normalized difference water index (NDWI), irrigation, photoperiod, gamma radiation, MODIS-EVI, leaf area index (LAI), and crop data. In addition, some studies have included extra nutrients like calcium, magnesium, potassium, sulfur, nitrogen, and boron as part of their feature selection. It is worth highlighting that the frequently employed features may not always rely on the same types of information. For instance, when examining temperature, researchers may consider factors such as average temperature, maximum temperature, and minimum temperature, as illustrated in the work of van Klompenburg et al. [3]. These studies primarily aim to predict the specific type of crop.

In Table 3, we have structured feature groups to represent essential attributes, providing a more organized view of the independent variables. These attributes span multiple domains, covering moisture, nutrients, field management, soil characteristics, and crop-related details. Notably, the most frequently utilized feature categories revolve around soil properties, meteorological data,

Table 2

Features used for crop yield prediction.

Feature	No. of times used	Feature	No. of times used
Rainfall	50	Soil fertility	1
Temperature (max-min)	69	Snow water equivalent	3
Humidity	27	Yield	22
Soil Moisture	5	APAR	1
NDVI	40	GNDVI	3
EVI	24	GCVI	2
LAI	12	LSWI	2
FPAR	3	SAVI	6
Solar radiation	10	Irrigation	11
NIR	2	Number of tanks	1
DVI	1	number of tube wells and open wells	2
Average temperature	4	Wind Speed	6
Precipitation	19	SPI/GDO/LST	1
Soil type	34	Soil pH	11
Season	4	RGB	7
Area	10	Cultivation practices	1
TPI	1	Wind direction	2
Vapor pressure	5	Crop	8
Wet day frequency	1	Sample ration (SR)	2
Sentinel-2/2Y-102D/GRD/SAR (Remote sensing)	1	WDRVI/GPP/GRVI/ARD	1
Evapotranspiration	10	Satellite images	1
Seed variety	3	CI	1
Fertilizers (Nitrogen, Phosphorus, Potash)	23	Crop Phonology	1
Pesticides	1	NDWI	11
DESIS	1	SWIR-1, SWIR-2	3
SIF	2		

Table 3
Grouped features.

Group	Name of the features	No. of times used (GroupWise)
Soil Data	Soil Moisture (5), Soil type (34), Area (10), Soil pH (11), Soil fertility (1)	61
Meteorological Data/Weather Conditions	Rainfall (50),Temperature (max-min) (69), Humidity (27),Average temperature (4), Precipitation (19), Wet day frequency (1), Evapotranspiration (10), Vapor pressure (5), Wind speed (6), Wind direction (2), Solar radiation (10), Snow water equivalent (3)	206
Crop yield Information	Pesticides (1), Crop (8), Yield data (22), Crop Phonology (1),Season (4)	36
Images	RGB (7), Satellite images (1)	8
Vegetation Indices	NDVI—normalized difference vegetation index (40), EVI—enhanced vegetation index (24), LA I – leaf area index (12), FPAR – fraction of photosynthetically active radiation (3), Sentinel-2 generated VIs and topographic data (1), NIR – near-infrared (2),DVI – difference vegetation index (1), TPI – Topographic Position Index (1),FPAR – fraction of photosynthetically active radiation (3) APAR – absorbed photosynthetically active radiation(1), GNDVI – Green Normalized Difference Vegetation Index (3), GCVI – green chlorophyll vegetation index (2), LSWI – Land Surface Water Index (2), SAVI – Soil Adjusted Vegetation Index (6), DESIS – DLR Earth Sensing Imaging Spectrometer (1), SIF - Stress intensification factor (2), CI - Composite Index (1), SPI/GDO/LST (1), WDRVI/GPP/GRVI/ARD (1), NDWI (11), SWIR-1, SWIR-2 (3)	121
Farm Management	Fertilizers (Nitrogen, Phosphorus, Potash) (Kg) (23), Seed variety (3), Number of tanks (1), number of tube wells and open wells (2), Cultivation practices (1), Irrigation (11)	41

weather conditions, and crop yield information.

Within the 'Soil Information' attribute group, you'll come across elements like soil moisture, soil type, pH value, and production area. The 'Humidity' category includes various aspects related to precipitation, humidity levels, expected precipitation, and precipitation forecasts. 'Field management' deals with the decisions made by farmers regarding field modifications and encompasses nutrient management, including elements such as irrigation and fertilizer management. Additionally, data from solar sources can be used to derive parameters related to irradiance or temperature.

Environmental factors play a crucial role in crop yield prediction as they directly influence plant growth, development, and productivity. Understanding the complex interactions between environmental variables and crop yields is essential for developing accurate prediction models. Several key environmental factors include weather variables such as temperature, precipitation, humidity, and solar radiation, which significantly impact crop growth and development. Soil properties like texture, nutrient content, pH, and water-holding capacity also affect crop growth by influencing root development, nutrient uptake, and water availability. Additionally, topography, including geographical features such as elevation, slope, and aspect, can influence microclimates and water drainage, thereby affecting local growing conditions and crop yields.

Data Collection using Remote Sensing Crop production is influenced by environmental factors, weather patterns, diseases and other parameters [28]. Ground observation, remote sensing, global positioning system, and on-site surveys serve as methods for monitoring environmental conditions, additional parameters, and crop growth. Obtaining data over a large area through ground observation or traditional methods poses challenges, resulting in less precise and variable outcomes [12]. To overcome this constraint, remote sensing is increasingly being used for crop monitoring. Remote sensing technology comprises the process of collecting and analyzing information about the Earth and its components via instruments placed in the atmosphere or on satellites, eliminating the necessity for physical engagement [57]. Compared to approaches like field surveys, remote sensing can generate a substantial quantity of data [109]. Computation of vegetation index (VI) is one of the most important reasons for using optical remote sensing to acquire agricultural data [28]. Normalized difference vegetation index (NDVI), normalized difference water index (NDWI), enhanced vegetation index (EVI), green normalized difference vegetation index (GNDVI), fraction of photosynthetically active radiation (FPAR), soil adjusted vegetation index (SAVI) and various other indices are instances of vegetation indices (VI).

4.2.1. Effect of vegetation indices and environmental factors

Vegetation indices are designed to enhance sensitivity to vegetation characteristics while minimizing interference from factors like soil reflectance and atmospheric conditions. Zhao et al. [110] established a function linking vegetation indices to crop coefficients (Kc) for efficient irrigation management and water conservation. Vegetation indices find extensive use in crop yield estimation, protein analysis, biomass assessment, weed control, and fertilizer management (X. [102]). Normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), leaf area index (LAI), green normalized difference vegetation index (GNDVI), normalized difference water index (NDWI) and soil adjusted vegetation index (SAVI) are frequently employed vegetation indices. Research has explored how the relationship between remotely sensed data and crop yield changes over the growing season, with varying associations at different growth stages (Shiu and Chuang, 2019; Lin et al., 2019). Earlier studies often concentrated on specific crops and years, with some vegetation indices showing high accuracy in predicting crop yields, particularly for corn [94,111–113]. Crop yield prediction combines vegetation indices with environmental variables like canopy temperature, water stress, humidity, nutrients, and soil data (X. [102]). The complexity of crop yield prediction is noted, with limited research on individual features that significantly impact predictions.

In order to calculate the corn yield, X. Wang et al. [102] combined vegetation indices like the normalized difference vegetation index (NDVI) and the absorbed photosynthetically active radiation (APAR) with environmental variables included canopy surface temperature and the water stress index. In addition, additional characteristics like humidity, nutrients, and soil data are also utilized to

forecast crop yields. There has been little research done to identify individual features that have a significant impact on crop yield prediction because so many features are already used in agricultural yield prediction. In light of this, a comprehensive exploration is required to gain a deeper insight into the variables and factors influencing crop production predictions.

4.3. Q3: third research question

In previous studies on crop yield prediction, a wide array of performance evaluation metrics has been employed, as depicted in Fig. 4. Remarkably, RMSE stands out as the most frequently utilized metric in crop yield prediction algorithms. Specifically, RMSE, R-squared, MAE, and classification accuracy were utilized as performance evaluation metrics in the papers by Joshua et al. [71], Pham et al. [114], Sajid et al. [84], and Shahhosseini et al. [81] respectively. However, the diverse use of evaluation metrics across these studies has posed challenges in comparing similar crop yield prediction models. As noted by Rashid et al. [25], this diversity complicates direct model comparisons. To address this issue, it is recommended that a consistent and systematic approach or a single standardized metric be adopted to quantify specific crop production prediction models. This would enable more straightforward and meaningful comparisons between different crop yield prediction algorithms.

4.4. Q4: challenges discussion

This summary discusses the challenges and findings related to crop yield prediction based on a review of 115 articles, including 68 using machine learning (ML) methods and 47 using deep learning (DL) methods. Key points include:

Challenges in Crop Yield Prediction The primary challenge identified is the complexity of creating a functional model. Model accuracy tends to improve with larger and more diverse datasets. Implementing these models in farm management systems poses another difficulty, and accuracy increases when local parameters are incorporated.

Data Insufficiency Several studies mentioned the problem of insufficient data availability. Despite limited data, some methods still showed promise, but further testing with more diverse data is recommended. Integrating data from various sources is also suggested for improvement.

Variability in Features and Scope The selected publications vary in the features they use, which depends on the depth and quantity of investigation, study scope, region, and crop. The choice of attributes is influenced by dataset availability and research objectives. Notably, more features do not always lead to better yield prediction performance.

Algorithm Diversity Various studies employed a range of algorithms, making it challenging to determine an optimal model. However, some machine learning models were more commonly used than others.

Data Integration and Environmental Factors The consensus among studies is that incorporating appropriate data, especially environmental factors, is crucial for accuracy. Most articles used data from limited sources, and few considered environmental factors, which raised questions about their accuracy.

5. Section V: Practical applications

Crop yield prediction models have numerous practical applications in agriculture. Farmers can utilize these models to optimize the allocation of resources such as water, fertilizer, and pesticides, thereby maximizing yields and minimizing input costs. Additionally, by anticipating potential yield fluctuations due to environmental factors, farmers can implement risk management strategies to mitigate the impact of adverse weather conditions or other environmental stressors. Accurate yield predictions also enable stakeholders to make informed decisions regarding market planning, trade policies, and commodity pricing, contributing to more efficient and resilient agricultural markets. Furthermore, policymakers can use crop yield prediction models to develop evidence-based policies and programs aimed at promoting food security, sustainable agriculture, and rural development.
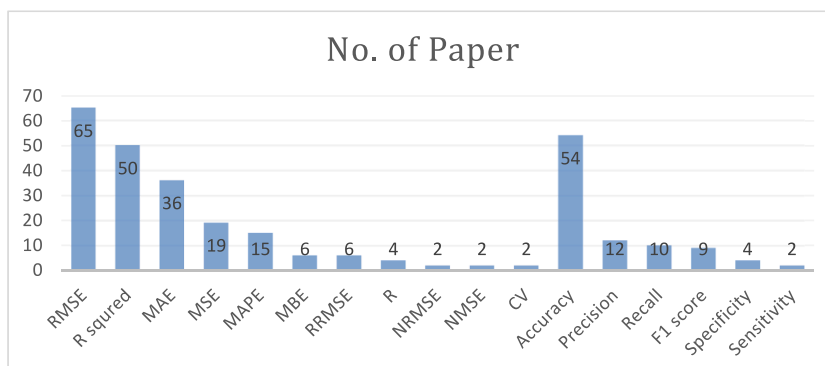

Fig. 4. Metrics for assessing performance.

6. Section VI: recommendations for future research

The need for further research is emphasized, particularly in identifying the most accurate approaches or methods for crop yield prediction by considering environmental factors. Selecting suitable datasets, thorough pre-processing, and efficient regression model training are crucial for obtaining the best results, especially given the continuously changing environmental conditions in different districts. Transfer learning and domain adaptation techniques can help overcome the challenge of limited labeled data by transferring knowledge from related domains or pre-trained models [58,59]. Fine-tuning pre-trained models on agricultural datasets can significantly improve model performance, especially in cases where labeled data is scarce [115]. Developing techniques to improve the interpretability and explainability of DL models is crucial for gaining trust and acceptance among farmers and stakeholders. Integrating data from multiple sources, including remote sensing, IoT devices, satellite imagery, and weather data, can provide a more comprehensive understanding of agricultural systems. Multimodal data fusion techniques can help extract valuable insights and improve the performance of ML and DL models. Collaboration among researchers, farmers, agricultural extension services, and policymakers is essential for advancing the field of agricultural ML and DL. Open data sharing initiatives can help address the challenge of data scarcity by making agricultural datasets more accessible to the research community. Developing robust ML and DL models that can generalize well across different environmental conditions, cropping systems, and geographic regions is crucial for their widespread adoption. Robustness can be improved through techniques such as data augmentation, ensemble learning, and model regularization. Designing ML and DL models with the end-user in mind is essential for their adoption and impact. Engaging with farmers and stakeholders throughout the model development process can help ensure that the models address real-world challenges and provide actionable insights.

Overall, the summary highlights the challenges and areas for improvement in crop yield prediction, emphasizing the importance of data quality, feature selection, and environmental factors in enhancing prediction accuracy. Furthermore, ML and DL have the potential to revolutionize agriculture, enabling more sustainable and efficient farming practices while ensuring food security and environmental operation.

7. Section VII: Conclusion

To address the global challenge of feeding a growing population, adopting advanced agricultural technology is crucial. Agricultural practitioners require timely and accurate guidance for crop yield prediction. Our review reveals that selected articles emphasize different features, particularly data accessibility and research coverage extent in yield prediction using machine learning and deep learning techniques. Models with more attributes do not consistently perform better, and feature selection depends on dataset availability and research objectives. Different algorithms have been used across studies, with variations in usage. Crop type, geographical location, and intensity levels vary among conducted studies. To determine the most effective model, specific feature selection algorithms should be tested with high-performing models, with significant differences in machine learning model usage. Traditional machine learning frameworks like random forest (RF), support vector machine (SVM), and neural network (NN) show promise.

The review explores 47 articles on deep learning algorithms. Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Deep Neural Networks (DNN) are commonly adopted. These techniques, along with pixel categorization and remote sensing data, show promise in crop yield prediction.

Combining quantitative and qualitative data sources, such as temperature, soil type, and remote sensing indices like NDVI, enhances prediction accuracy. Integrating remote sensing data, image processing, and deep learning can improve accuracy in predicting crop yields over large regions.

Various models, including crop models, machine learning, and remote sensing models, employ weight-dependent yield estimation. Count-based assessments using image processing and remote sensing modeling face challenges like object clutter. Common performance metrics include RMSE, R2, MAE, MSE, and MAPE, which measure prediction accuracy. The study sets the foundation for future advancements in crop yield prediction, with a focus on developing deep learning-based models. It emphasizes the need for testing specific feature selection algorithms with high-performing models and addressing the challenges of dynamic variables in crop prediction.

Overall, the review highlights the importance of advanced technology, data integration, and machine learning and deep learning techniques in improving crop yield prediction accuracy, contributing to addressing global food security challenges.

CRediT authorship contribution statement

Md. Abu Jabed: Writing – review & editing, Writing – original draft. **Masrah Azrifah Azmi Murad:** Writing – review & editing, Supervision.

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

APPENDIX. Literature Review Table (Selected Publications)

Databases	Authors	Data Source	Features	Methods	Measure attributes	Objectives	Findings	Limitations
Web of Science	[85]	271 counties in Germany from 1999 to 2019	Weather, soil, and crop phenology, yield value	CNN, CNN-FCNN, RF, KNN, LASSO and Ridge Regression, SVR, XGBoost, DNN, Regression tree.	MAE, RMSE and correlation coefficient matrices (r)	To benchmark winter wheat yield prediction in Germany with advanced machine learning techniques. To introduce a 1-D convolutional neural network (CNN) for enhanced accuracy.	CNN model outperformed all other baseline models used for winter wheat yield prediction	Sophisticated models are required. Black box nature;
PubMed.	[65]	**9 states of the United States of America (USA)** Khaki and Wang [88], and Khaki, Wang, and Archontoulis [57]	weather, soil, and farm management data	XGBoost, CNN-XGBoost, CNN-DNN, CNN-RNN, and CNN-LSTM	Efficiency (Features Selection) R2, RMSE, MSE	To leverage the strengths of each approach within a hybrid model to optimize the effectiveness of crop yield prediction.	The hybrid CNN-DNN model performs exceptionally well with an R2 of 0.87, RMSE of 0.266, MSE of 0.071, and MAE of 0.199, surpassing all other models. XGBoost achieved the second-highest performance while executing faster than deep learning models.	The Corn Belt soybean crop dataset in the U. S. is a commonly used resource in various studies. Results may differ with other datasets when applying the proposed models.
Scopus	[116]	Google Earth Engine (GEE) cloud, and spectral matching techniques (SMTs).	Landsat-8 30 m and MODIS 250 m data, (ARD for Landsat 30 m data, ARD for MODIS 250 m data, SWIR 1 & 2, Thermal Infrared (TIR), RGB, NDVI, NDVIMVCi, EVI, NDWI)	RF	Accuracy R2, RMSE	To develop cropland products in South Asia using Landsat-8 30 m, MODIS 250, and Google Earth Engine (GEE) with machine learning and spectral matching techniques (SMTs) for food and water security.	The irrigated vs. rainfed 30m product showed 79.8 % overall accuracy, with 79 % for irrigated and 74 % for rainfed cropland. The cropping intensity product had an 85.3 % overall accuracy, with 88 % for single cropping, 85 % for double cropping, and 67 % for triple cropping. Crop type mapping ranged from 72 % to 97 % accuracy, explaining 63–98 % variability when compared to	Absence of reference data for specific croplands Complex mapping over vast areas

(*continued*)

Databases	Authors	Data Source	Features	Methods	Measure attributes	Objectives	Findings	Limitations
Science Direct	[69]	Sentinel-1 satellite, Sentinel-2 satellite of the European Space Agency (ESA)	SAR, optical and field data.	Ordinary Least Squares Regression (OLS), RF, DT, KNN, Ridge regressor, SVM and DLMLP model	Accuracy and performance evaluation (R^2, MAE and RMSE)	To predict soil moisture, EC (salinity), and SOC (organic carbon) levels through diverse machine learning methods and subsequently assessing their respective accuracies for comparison.	national statistics. Significant correlations exist between soil health parameters and crop yield (R2 values: 0.79 for soil moisture, 0.88 for EC, and 0.51 for SOC). The DLMLP model achieved good accuracy, even without the need to validate soil health parameters.	Limited to soil data only.
Web of Science	[86]	GitHub link (https://github.com/tryambakganguly/Yield-Prediction-Temporal-Attention).	Weather data	Stacked LSTM and Temporal Attention model, Support Vector Regression with Radial Basis Function kernel (SVR-RBF) and least absolute shrinkage and selection operator (LASSO) regression	RMSE, MAE, and coefficient of determination or R-squared (*R*2)	To enhance the interpretability of prediction results in multivariate time-series forecasting without compromising accuracy.	The deep learning model demonstrated significantly superior performance when contrasted with the USDA model, which relies on domain knowledge.	Less factors used Limited data
Scopus	[87]	POWER Data Access Viewer v2.0.0 MERRA-2 database, https://power.larc.nasa.gov/data-access-viewer/ (accessed on 1 January 2022).	satellite-derived hydro-meteorological data between 1981 and 2020	DRS–RF hybrid model and SVM as feature selection approach	Prediction Performance Correlation coefficient (C), RRMSE, MAPE	To design an innovative hybrid machine learning model that combines Random Forest (RF) with Dragonfly Optimization (DR) and Support Vector Regression (SVR) to predict tea yield in Bangladesh.	The DRS–RF hybrid model outperformed standalone machine learning methods, yielding the lowest relative error rate of 11 % in tea yield forecasting.	Advanced model can be applied for better prediction.
Scopus	[79]	Orchard, Google Earth Engine [Sentinel 2 and Landsat satellite images, including the normalized difference vegetation index (NDVI) and normalized difference water index (NDWI)]	**Field Data Collection**: irrigation, fertilization, phytosanitary treatment, daily data on temperature, precipitation, humidity, wind speed, and solar radiation, **Spectral data acquisition:** NDVI, NDWI (2015–2019)	linear algorithms such as linear regression, lasso regression, ridge regression. Ensemble learning techniques such as CatBoost Regressor, Light Gradient Boosting Machine, Random Forest Regressor,	Accuracy MAE, MSE	To develop a state-of-the-art solution for forecasting citrus fruit crop yields prior to harvest, integrating machine learning algorithms trained on historical agricultural data with spectral data obtained from satellite imagery.	The orthonormal automatic pursuit algorithm achieved favorable prediction scores with MAE of 0.2489 and MSE of 0.0843.	Field data collection in precision agriculture is challenging, requiring specialized sensors and technical teams for data quality control. Historical and open-source databases are scarce.

(*continued*)

Databases	Authors	Data Source	Features	Methods	Measure attributes	Objectives	Findings	Limitations
Scopus	[40]	Paprika smart farm (greenhouse paprika data in the year October to December 2019 and January to July 2020) in South Korea	Environmental and solar energy factors	Extreme Gradient Boosting, etc RF, SVM, and gradient boosting machine (GBM),	Accuracy precision, recall, f1-score kappa	To assess and contrast linear classification methods and various machine learning models in terms of their predictive accuracy for paprika growth, taking into account environmental factors and solar energy data across two beds.	RF can effectively predict paprika growth with an accuracy of 0.88.	Limited data Small area
Scopus	[88]	Syngenta and the Analytics Society of INFORMS, United States and Canada	Crop genotype, yield performance, and environment (weather and soil)	DNN, Lasso, shallow neural networks (SNN), and regression tree (RT)	Accuracy RMSE (Reduced to 11 % of the average yield and 46 % of the standard)	To conduct feature selection using the trained DNN model, effectively reducing the input space dimension without a substantial decrease in prediction accuracy.	DNN(W) and DNN(S) exhibited comparable strong performance, surpassing DNN (G), underscoring the greater influence of environmental factors (weather and soil) on crop yield variability than genotype.	Advanced models (Hybrid) needed with explanations
Scopus	[57]	Corn Belt in the United States., National Agricultural Statistics Service of the United States (USDA-NASS, 2019), Daymet (Thornton et al., 2018) and Gridded Soil Survey Geographic Database for the United States (gSSURGO, 2019).	Yield performance, management, weather, and soil	CNN, RNN, RF, deep fully connected neural networks (DFNN) and the least absolute shrinkage and selection operator (LASSO)	Performance RMSE	To introduce a deep learning framework that employs Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to predict crop yields using environmental data and agricultural management practices.	The proposed method (CNN-RNN) outperformed than other Methods. The new model (CNN-RNN) achieved a root-mean-square-error (RMSE) 9 % and 8 % of their respective average yields.	More accurate weather prediction models need to be considered
Scopus	[117]	Sentinel-1 satellites.	Nitrogen rate applied, precipitation, slope, elevation, topographic position index (TPI), aspect, and two radar backscatter coefficients	AdaBoost stacked autoencoder (SAE), three-dimensional CNN (3D-CNN), CNN-Late Fusion (CNN-LF), RF, Bayesian multiple linear regression (BMLR), and multiple linear regression (MLR).	Accuracy RMSE, RMedSE, Pearson correlation coefficient (r) and structural similarity (SSIM)	To propose a new method for predicting crop yield using Convolutional Neural Networks (CNNs) and multi-channel input raster.	The proposed approach yields superior predictions when compared to five other methods.	Data collected during early winter wheat season. Proposed method may demand substantial computational resources. No permission to share farmers' data.

(*continued*)

Databases	Authors	Data Source	Features	Methods	Measure attributes	Objectives	Findings	Limitations
Scopus	[95]	SRDI, BARC, BBS, and BMD (Aus, Aman, Boro, jute, wheat, and potato)	rainfall, max-min temperatures, and humidity, TSP, diammonium phosphate (DAP), and MOP, HL, MHL, MLL, LL, and VLL. soil types, soil moisture, texture, consistency, and reaction	Logistic regression (LR), SVM, RF, DNN	Accuracy (MSE)	To identify the optimal approach for selecting crop yield effectively.	The Neural Network excels over other methods, maintaining an average prediction accuracy of 96.06 % across six different crops.	Extend the model in different region
Science Direct	[118] (2)	Google Earth Engine platform, Agricultural Yearbook of the provinces and county level statistics bureaus (http://www.stats.gov. cn)	Crop planting areas, county-level and field-level yield, climate, satellite, soil properties, and spatial information	RF, DNN, ID-CNN, LSTM	Accuracy RMSE, R2	To showcase a scalable, straightforward, and cost-effective method for accurately estimating crop yields across different scales.	DNN and RF models had relatively good performance at the field level, with mean R2 values of 0.71, 0.66 and RMSE values of 1127 kg/ha and 956 kg/ha, respectively	Applicable only to winter wheat in China. Google Earth Engine data may lack key variables. Limited to four models. Short time frame analyzed.
Scopus	[119]	European Space Agency's (ESA's)	Wheat yield data and auxiliary data, Sentinel-2 and ZY-1 02D remote sensing imagery	LSTM, RF, GBDT, and SVR	Accuracy RMSE, R2	To predict winter wheat crop yields using remote sensing data and machine learning methods.	The LSTM model outperformed SVR, RF, and GBDT, with an R2 of 0.93, capturing the temporal link between satellite data and winter wheat yield.	Utilized Sentinel-2 and ZY-1 02D data exclusively, without exploring other data types. Did not account for weather and environmental factors in yield estimation.
Science Direct	[120]	Moderate Resolution Imaging Spectroradiometer (MODIS) Version 6 (Terra/MOD09A1) 8-day composite surface reflectance product with a spatial resolution of 0.5 km.	Time-series intervals (data shapes) of the SIs (S-inputs) at an 8-day temporal resolution, meteorological data (M-inputs) and four geospatial information data (G-inputs) (LSWI and the VIs (NDVI or EVI)	LSTM and one-dimensional convolutional neural network (1D-CNN)	Accuracy MSE, RMSE, R2, NSE	To propose an approach for forecasting rice yields with pixel-level precision, achieved by combining crop and deep learning models with satellite data in both South and North Korea.	The proposed model exhibited strong performance [R2 = 0.859, Nash-Sutcliffe model efficiency = 0.858, root mean squared error = 0.605 Mg ha−1], highlighting distinct spatial patterns in rice yields across South and North Korea.	Depends on satellite data, and prediction accuracy can be influenced by data quality and availability.
Science Direct	[118] (1)	GEE (Google Earth Engine) platform, Agricultural Yearbook of each county (http://www.stats.gov.cn) from 2001 to 2015 (unit: kg/ha).	Satellite vegetation indexes, meteorological indexes, and soil properties, Google Earth Engine (GEE) platform	Least Absolute Shrinkage and Selection Operator (LASSO) regression, RF, LSTM networks	Accuracy R2, RMSE	To investigate the timely prediction of rice yield across a vast area using publicly accessible multi-source data.	LSTM (with R2 ranging from 0.77 to 0.87, RMSE from 298.11 to 724 kg/ha) performed better than others.	Combining crop models, more detailed farming management data, and higher spatiotemporal resolution of input variables, such as daily weather and 10 m resolution

(*continued*)

Databases	Authors	Data Source	Features	Methods	Measure attributes	Objectives	Findings	Limitations
								data of Sentinel 2.
Scopus	[80]	Farming Community, India	soil and environmental characteristics including Avg. soil temperature, avg. air temperature, minimum and maximum air temperature, rainfall, and air humidity	RF, Bagging, K-NN, SVM, DT, NB **Feature Selection** MRFE, RFE, Boruta	accuracy (ACC), specificity (S), recall (R), precision (P), F1 score, MAE, log loss (LL), and area under the curve (AUC).	To predict crop yield by leveraging feature selection techniques and classifiers that consider the agricultural environment's attributes.	The Ensemble technique (MRFE, with RF) outperforms the current classification method in terms of prediction accuracy.	The data was collected manually and is not accessible publicly. The proposed model's scalability was not addressed or discussed.
Science Direct.	[39]	Fergana valley, Central Asia, soil grids (https:// soilgrids.org/)	Sentinel-2 generated VIs, environmental data, soil data, field data, and topographic data	LR, DT and RF regression models with scikit-learn	Accuracy R2	To create a machine learning-based yield estimation and prediction model. To evaluate and compare the efficacy of different regression techniques, such as Decision Trees (DT), Linear Regression (LR), and Random Forest (RF).	Overall accuracy of 93 %, RF has the highest yield prediction accuracy when compared to LR and DT.	More spectral features, data, and algorithms needed to improve the prediction accuracy.
Science Direct	[11]	State of Maharashtra	previous years' climate, soil, and yield, (District-wise parameters like temperature, precipitation, humidity, soil type, crop type, season, area of the field)	Multiple Linear Regression, DTR, GBR, Elastic Net, Lasso, Ridge, Partial Least Squares Regression, and feature engineering-based LSTM	Accuracy R2, MAE, RMSE	To propose a machine learning model that can make predictions based on various soil and environmental variables.	The feature-engineered LSTM model outperforms other models, achieving an accuracy of 86.3 % and demonstrating the lowest mean absolute error and root mean square error.	Higher error rate is attributed to the dynamic environmental changes in the districts Less favorable features.
PubMed	[49]	meteorological department of Punjab, statistical abstract of Punjab issued by Economic advisor to Government, Punjab	Climatic factors like minimum and maximum temperature, minimum and maximum relative humidity, rainfall, evaporation, wind direction and speed and solar radiation	RNN-LSTM, ANN, RF and Multivariate Linear Regression (MVLR)	Efficacy RMSE MAE MSE	To discover an effective deep learning method for predicting wheat crop yields.	RNN-LSTM model proving efficiency	Study conducted in a small area Using solely statistical models for crop yield prediction.

References

[1] A. Modi, P. Sharma, D. Saraswat, R. Mehta, Review of crop yield estimation using machine learning and deep learning techniques, Scalable Computing 23 (2) (2022) 59–79, https://doi.org/10.12694/scpe.v23i2.2025.

[2] N. Goel, S. Kaur, Y. Kumar, Chapter 23 - machine learning-based remote monitoring and predictive analytics system for crop and livestock, in: AI, Edge and IoT-Based Smart Agriculture, Elsevier Inc, 2022, https://doi.org/10.1016/B978-0-12-823694-9.00016-5.

[3] T. van Klompenburg, A. Kassahun, C. Catal, Crop yield prediction using machine learning: a systematic literature review, Comput. Electron. Agric. 177 (July) (2020) 105709, https://doi.org/10.1016/j.compag.2020.105709.

[4] X. Xu, P. Gao, X. Zhu, W. Guo, J. Ding, C. Li, M. Zhu, Design of an integrated climatic assessment indicator (ICAI) for wheat production : a case study in Jiangsu Province , China, Ecol. Indicat. 101 (July 2018) (2019) 943–953, https://doi.org/10.1016/j.ecolind.2019.01.059.

[5] P. Filippi, An Approach to Forecast Grain Crop Yield Using Multi - Layered , Multi - Farm Data Sets and Machine Learning, 0123456789, 2019, https://doi.org/10.1007/s11119-018-09628-4.

[6] F. Azuaje, Review of " data mining : practical machine learning tools and techniques " by witten and frank. January 2006. https://doi.org/10.1186/1475-925X-5-51, 2014.

[7] Y. Jin, H. Wang, C. Sun, Data-driven evolutionary optimization, Stud. Comput. Intell. (SCI) 975 (2021) 103–143, https://doi.org/10.1007/978-3-030-74640-7.

[8] J. Ansarifar, L. Wang, S.V. Archontoulis, An interaction regression model for crop yield prediction, Sci. Rep. 11 (1) (2021) 1–14, https://doi.org/10.1038/s41598-021-97221-7.

[9] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, I.N. Athanasiadis, Machine learning for large-scale crop yield forecasting, Agric. Syst. 187 (December 2020) (2021) 103016, https://doi.org/10.1016/j.agsy.2020.103016.

[10] L.S. Cedric, W.Y.H. Adoni, R. Aworka, J.T. Zoueu, F.K. Mutombo, M. Krichen, C.L.M. Kimpolo, Crops yield prediction based on machine learning models: case of West African countries, Smart Agricultural Technology 2 (December 2021) (2022) 100049, https://doi.org/10.1016/j.atech.2022.100049.

[11] S. Iniyan, V. Akhil Varma, C. Teja Naidu, Crop yield prediction using machine learning techniques, Adv. Eng. Software 175 (October 2022) (2023) 103326, https://doi.org/10.1016/j.advengsoft.2022.103326.

[12] A. Chlingaryan, S. Sukkarieh, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review, Comput. Electron. Agric. 151 (November 2017) (2018) 61–69, https://doi.org/10.1016/j.compag.2018.05.012.

[13] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine learning in agriculture: a review, Sensors 18 (8) (2018) 1–29, https://doi.org/10.3390/s18082674.

[14] L.J. Young, Agricultural crop forecasting for large geographical areas, Annual Review of Statistics and Its Application 6 (August 2018) (2019) 173–196, https://doi.org/10.1146/annurev-statistics-030718-105002.

[15] D. Elavarasan, D.R. Vincent, V. Sharma, A.Y. Zomaya, K. Srinivasan, Forecasting yield by integrating agrarian factors and machine learning models: a survey, Comput. Electron. Agric. 155 (October) (2018) 257–282, https://doi.org/10.1016/j.compag.2018.10.024.

[16] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: a survey, Comput. Electron. Agric. 147 (July 2017) (2018) 70–90, https://doi.org/10.1016/j.compag.2018.02.016.

[17] R. Beulah, A survey on different data mining techniques for crop yield prediction, International Journal of Computer Sciences and Engineering 7 (1) (2019) 738–744, https://doi.org/10.26438/ijcse/v7i1.738744.

[18] A. Koirala, K.B. Walsh, Z. Wang, C. McCarthy, Deep learning – method overview and review of use for fruit detection and yield estimation, Comput. Electron. Agric. 162 (January) (2019) 219–234, https://doi.org/10.1016/j.compag.2019.04.017.

[19] N. Häni, P. Roy, V. Isler, A comparative study of fruit detection and counting methods for yield mapping in apple orchards, J. Field Robot. 37 (2) (2020) 263–282, https://doi.org/10.1002/rob.21902.

[20] Q. Zhang, Y. Liu, C. Gong, Y. Chen, H. Yu, Applications of deep learning for dense scenes analysis in agriculture: a review, Sensors 20 (5) (2020) 1–33, https://doi.org/10.3390/s20051520.

[21] M.K. Dharani, R. Thamilselvan, P. Natesan, P.C.D. Kalaivaani, S. Santhoshkumar, Review on crop prediction using deep learning techniques, J. Phys. Conf. 1767 (1) (2021), https://doi.org/10.1088/1742-6596/1767/1/012026.

[22] B. Darwin, P. Dharmaraj, S. Prince, D.E. Popescu, D.J. Hemanth, Recognition of bloom/yield in crop images using deep learning models for smart agriculture: a review, Agronomy 11 (4) (2021) 1–22, https://doi.org/10.3390/agronomy11040646.

[23] P. Maheswari, P. Raja, O.E. Apolo-Apolo, M. Pérez-Ruiz, Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—a review, Front. Plant Sci. 12 (June) (2021) 1–18, https://doi.org/10.3389/fpls.2021.684328.

[24] A. Monteiro, S. Santos, P. Gonçalves, Precision agriculture for crop and livestock farming—brief review, Animals 11 (8) (2021) 1–18, https://doi.org/10.3390/ani11082345.

[25] M. Rashid, B.S. Bari, Y. Yusup, M.A. Kamaruddin, N. Khan, A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction, IEEE Access 9 (2021) 63406–63439, https://doi.org/10.1109/ACCESS.2021.3075159.

[26] L. Benos, A.C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, D. Bochtis, Machine learning in agriculture: a comprehensive updated review, Sensors 21 (11) (2021) 1–55, https://doi.org/10.3390/s21113758.

[27] A.S.M.M. Hasan, F. Sohel, D. Diepeveen, H. Laga, M.G.K. Jones, A survey of deep learning techniques for weed detection from images, Comput. Electron. Agric. 184 (December 2020) (2021) 106067, https://doi.org/10.1016/j.compag.2021.106067.

[28] P. Muruganantham, S. Wibowo, S. Grandhi, N.H. Samrat, N. Islam, A systematic literature review on crop yield prediction with deep learning and remote sensing, Rem. Sens. 14 (9) (2022), https://doi.org/10.3390/rs14091990.

[29] A. Oikonomidis, C. Catal, A. Kassahun, Deep learning for crop yield prediction: a systematic literature review, N. Z. J. Crop Hortic. Sci. 51 (1) (2023) 1–26, https://doi.org/10.1080/01140671.2022.2032213.

[30] A. Bouguettaya, H. Zarzour, A. Kechida, A.M. Taberkit, Deep learning techniques to classify agricultural crops through UAV imagery: a review, Neural Comput. Appl. 34 (12) (2022) 9511–9536, https://doi.org/10.1007/s00521-022-07104-9.

[31] M.O. Ojo, A. Zahid, Deep learning in controlled environment agriculture: a review of recent advancements, challenges and prospects, Sensors 22 (20) (2022), https://doi.org/10.3390/s22207965.

[32] D. Tranfield, D. Denyer, P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review, Br. J. Manag. 14 (3) (2003) 207–222, https://doi.org/10.1111/1467-8551.00375.

[33] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (March) (2021), https://doi.org/10.1016/j.ijsu.2021.105906.

[34] S. Sairamkumar, Design of ANN based machine learning method for crop prediction 3 (3) (2021) 223–239.

[35] K. Amankulova, N. Farmonov, U. Mukhtorov, L. Mucsi, Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology, Geocarto Int. 38 (1) (2023), https://doi.org/10.1080/10106049.2023.2197509.

[36] K.S.M. Anbananthen, S. Subbiah, D. Chelliah, P. Sivakumar, V. Somasundaram, K.H. Velshankar, M.K.A.A. Khan, An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms, F1000Research 10 (2021) 1–18, https://doi.org/10.12688/f1000research.73009.1.

[37] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, 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. 153 (January) (2018) 213–225, https://doi.org/10.1016/j.compag.2018.07.016.

[38] N.R. Prasad, Crop yield prediction in cotton for regional level using random forest approach, Spatial Information Research (2020), https://doi.org/10.1007/s41324-020-00346-6.

[39] M. Singh, K. Choudhary, R. Paringer, A. Kupriyanov, Journal of the Saudi society of agricultural sciences machine learning for yield prediction in fergana valley , central asia, Journal of the Saudi Society of Agricultural Sciences xxxx (2022), https://doi.org/10.1016/j.jssas.2022.07.006.

[40] S. Venkatesan, J. Lim, Y. Cho, A Crop Growth Prediction Model Using Energy Data Based on Machine Learning in Smart Farms 2022 (2022).

[41] S. Ju, H. Lim, J.W. Ma, S. Kim, K. Lee, S. Zhao, J. Heo, Optimal county-level crop yield prediction using MODIS-based variables and weather data: a comparative study on machine learning models, Agric. For. Meteorol. 307 (July) (2021) 108530, https://doi.org/10.1016/j.agrformet.2021.108530.

[42] P. Hara, M. Piekutowska, G. Niedbała, Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data, Land 10 (6) (2021), https://doi.org/10.3390/land10060609.

[43] S. Gupta, A. Geetha, K.S. Sankaran, A.S. Zamani, M. Ritonga, R. Raj, S. Ray, H.S. Mohammed, Machine Learning- and Feature Selection-Enabled Framework for Accurate Crop Yield Prediction 2022 (2022).

[44] S. Fu, D. Chen, H. He, S. Liu, S. Moon, K.J. Peterson, F. Shen, L. Wang, Y. Wang, A. Wen, Y. Zhao, H. Liu, M. Clinic, M. Clinic, HHS Public Access 1–41 (2021), https://doi.org/10.1016/j.jbi.2020.103526.Clinical.

[45] V.C. Waikar, S.Y. Thorat, A.A. Ghute, P.P. Rajput, M.S. Shinde, Crop prediction based on soil classification using machine learning with classifier ensembling, International Research Journal of Engineering and Technology 7 (May) (2020) 4857–4861. www.irjet.net.

[46] D. Batool, M. Shahbaz, H.S. Asif, K. Shaukat, T.M. Alam, I.A. Hameed, Z. Ramzan, A. Waheed, H. Aljuaid, S. Luo, A hybrid approach to tea crop yield prediction using simulation models and machine learning, Plants 11 (15) (2022) 1925, https://doi.org/10.3390/plants11151925.

[47] D. Elavarasan, P.M.D. Vincent, Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications 8 (2020), https://doi.org/10.1109/ACCESS.2020.2992480.

[48] T. Islam, T.A. Chisty, A. Chakrabarty, A deep neural network approach for crop selection and yield prediction in Bangladesh, in: IEEE Region 10 Humanitarian Technology Conference, 2019, https://doi.org/10.1109/R10-HTC.2018.8629828. *R10-HTC, 2018-Decem*(1).

[49] N. Bali, A. Singla, Deep learning based wheat crop yield prediction model in Punjab region of north India, Appl. Artif. Intell. 35 (15) (2021) 1304–1328, https://doi.org/10.1080/08839514.2021.1976091.

[50] S.S. Olofintuyi, E.A. Olajubu, D. Olanike, An ensemble deep learning approach for predicting cocoa yield, Heliyon 9 (4) (2023) e15245, https://doi.org/10.1016/j.heliyon.2023.e15245.

[51] K. Alibabaei, P.D. Gaspar, T.M. Lima, Crop yield estimation using deep learning based on climate big data and irrigation scheduling, Energies 14 (11) (2021) 1–21, https://doi.org/10.3390/en14113004.

[52] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, https://doi.org/10.1038/nature14539.

[53] A. Mokhtar, W. El-Ssawy, H. He, N. Al-Anasari, S.S. Sammen, Y. Gyasi-Agyei, M. Abuarab, Using machine learning models to predict hydroponically grown lettuce yield, Front. Plant Sci. 13 (March) (2022) 1–10, https://doi.org/10.3389/fpls.2022.706042.

[54] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, Z. Li, A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach, Remote Sensing of Environment 210 (April 2017) (2018) 35–47, https://doi.org/10.1016/j.rse.2018.02.045.

[55] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, L. Zhang, Deep learning in environmental remote sensing: achievements and challenges, Remote Sensing of Environment 241 (2019) 111716, https://doi.org/10.1016/j.rse.2020.111716.

[56] M. Engen, E. Sandø, B.L.O. Sjølander, S. Arenberg, R. Gupta, M. Goodwin, Farm-scale crop yield prediction from multi-temporal data using deep hybrid neural networks, Agronomy 11 (12) (2021) 1–31, https://doi.org/10.3390/agronomy11122576.

[57] S. Khaki, L. Wang, S.V. Archontoulis, A CNN-RNN Framework for Crop Yield Prediction 10 (January) (2020) 1–14, https://doi.org/10.3389/fpls.2019.01750.

[58] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, D. Traore, Deep neural networks with transfer learning in millet crop images, Comput. Ind. 108 (2019) 115–120, https://doi.org/10.1016/j.compind.2019.02.003.

[59] N. Krishnamoorthy, L.V. Narasimha Prasad, C.S. Pavan Kumar, Bharat Subedi, S.V. Haftom Baraki Abraha, Rice leaf diseases prediction using deep neural networks with transfer learning, Environ. Res. 198 (April) (2021) 111275, https://doi.org/10.1016/j.envres.2021.111275.

[60] H. Servia, S. Pareeth, C.I. Michailovsky, C. de Fraiture, P. Karimi, Operational framework to predict field level crop biomass using remote sensing and data driven models, Int. J. Appl. Earth Obs. Geoinf. 108 (2022) 102725, https://doi.org/10.1016/j.jag.2022.102725.

[61] A. Venugopal, S. Aparna, J. Mani, R. Mathew, P.V. Williams, Crop yield prediction using machine learning algorithms 9 (13) (2021) 87–91.

[62] E.H. Bouras, L. Jarlan, S. Er-Raki, R. Balaghi, A. Amazirh, B. Richard, S. Khabba, Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco, Rem. Sens. 13 (16) (2021), https://doi.org/10.3390/rs13163101.

[63] J. Cao, H. Wang, J. Li, Q. Tian, D. Niyogi, Improving the forecasting of winter wheat yields in northern China with machine learning–dynamical hybrid subseasonal-to-seasonal ensemble prediction, Rem. Sens. 14 (7) (2022), https://doi.org/10.3390/rs14071707.

[64] F. Huber, A. Yushchenko, B. Stratmann, V. Steinhage, Extreme gradient boosting for yield estimation compared with deep learning approaches, Comput. Electron. Agric. 202 (2022), https://doi.org/10.1016/j.compag.2022.107346.

[65] A. Oikonomidis, C. Catal, A. Kassahun, Hybrid deep learning-based models for crop yield prediction, Appl. Artif. Intell. 36 (1) (2022), https://doi.org/10.1080/08839514.2022.2031823.

[66] T.V. Rajinikanth, N.T. Sri, A.Y. Saikrishna, Agriculture Crop Yield Analysis and Prediction using Feature Selection based Machine Learning Techniques 8958 (2) (2022) 99–108, https://doi.org/10.35940/ijeat.B3942.1212222.

[67] V. Amaratunga, L. Wickramasinghe, A. Perera, J. Jayasinghe, U. Rathnayake, J.G. Zhou, Artificial neural network to estimate the paddy yield prediction using climatic data, Math. Probl Eng. 2020 (2020), https://doi.org/10.1155/2020/8627824.

[68] L. Kouadio, R.C. Deo, V. Byrareddy, J.F. Adamowski, Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties, Comput. Electron. Agric. 155 (October) (2018) 324–338, https://doi.org/10.1016/j.compag.2018.10.014.

[69] A. Tripathi, R.K. Tiwari, S.P. Tiwari, A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation, Int. J. Appl. Earth Obs. Geoinf. 113 (April) (2022) 102959, https://doi.org/10.1016/j.jag.2022.102959.

[70] D. Elavarasan, D.R. Vincent P M, K. Srinivasan, C.-Y. Chang, A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling, Agriculture 10 (9) (2020) 400, https://doi.org/10.3390/agriculture10090400.

[71] S.V. Joshua, A.S.M. Priyadharson, R. Kannadasan, A.A. Khan, W. Lawanont, F.A. Khan, A.U. Rehman, M.J. Ali, Crop yield prediction using machine learning approaches on a wide spectrum, Comput. Mater. Continua (CMC) 72 (3) (2022) 5663–5679, https://doi.org/10.32604/cmc.2022.027178.

[72] V. Joshua, S.M. Priyadharson, R. Kannadasan, Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu, Agronomy 11 (10) (2021), https://doi.org/10.3390/agronomy11102068.

[73] V. Talasila, Analysis and Prediction of Crop Production in Andhra Region Using Deep Convolutional Regression Network 13 (5) (2020) 1–9, https://doi.org/10.22266/ijies2020.1031.01.

[74] J. Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, J. Han, Z. Li, Identifying the contributions of multi-source data for winter wheat yield prediction in China, Rem. Sens. 12 (5) (2020) 1–22, https://doi.org/10.3390/rs12050750.

[75] M. Qiao, X. He, X. Cheng, P. Li, H. Luo, L. Zhang, Z. Tian, Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks, Int. J. Appl. Earth Obs. Geoinf. 102 (April) (2021) 102436, https://doi.org/10.1016/j.jag.2021.102436.

[76] S. Khaki, H. Pham, L. Wang, Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning, Sci. Rep. (2021) 0123456789, https://doi.org/10.1038/s41598-021-89779-z, 1–14.

[77] S. Iniyan, R. Jebakumar, Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (MSER), Wireless Pers. Commun. 126 (3) (2022) 1935–1964, https://doi.org/10.1007/s11277-021-08712-9.

[78] S.G. Kundu, A. Ghosh, A. Kundu, G. G P, A ml-ai enabled ensemble model for predicting agricultural yield, Cogent Food Agric. 8 (1) (2022), https://doi.org/10.1080/23311932.2022.2085717.

[79] A. Moussaid, S. El Fkihi, Y. Zennayi, O. Lahlou, I. Kassou, F. Bourzeix, L. El Mansouri, Y. Imani, Machine learning applied to tree crop yield prediction using field data and satellite imagery: a case study in a citrus orchard, Informatics 9 (4) (2022), https://doi.org/10.3390/informatics9040080.

[80] S.P. Raja, B. Sawicka, Z. Stamenkovic, G. Mariammal, Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers, IEEE Access 10 (March) (2022) 23625–23641, https://doi.org/10.1109/ACCESS.2022.3154350.

[81] M. Shahhosseini, G. Hu, I. Huber, S.V. Archontoulis, Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, Sci. Rep. 123456789 (2021) 1–15, https://doi.org/10.1038/s41598-020-80820-1.

[82] B. Pejak, P. Lugonja, A. Antić, M. Panić, M. Pandžić, E. Alexakis, P. Mavrepis, N. Zhou, O. Marko, V. Crnojević, Soya yield prediction on a within-field scale using machine learning models trained on sentinel-2 and soil data, Rem. Sens. 14 (9) (2022) 1–22, https://doi.org/10.3390/rs14092256.

[83] X. Chen, L. Feng, R. Yao, X. Wu, J. Sun, W. Gong, Prediction of maize yield at the city level in China using multi-source data, Rem. Sens. 13 (1) (2021) 1–17, https://doi.org/10.3390/rs13010146.

[84] S.S. Sajid, M. Shahhosseini, I. Huber, G. Hu, S.V. Archontoulis, County-scale crop yield prediction by integrating crop simulation with machine learning models, Front. Plant Sci. 13 (November) (2022) 1–16, https://doi.org/10.3389/fpls.2022.1000224.

[85] A.K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng, Winter wheat yield prediction using convolutional neural networks from environmental and phenological data, Sci. Rep. (2022) 1–14, https://doi.org/10.1038/s41598-022-06249-w.

[86] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S.S. Id, A.K. Singh, Crop yield prediction integrating genotype and weather variables using deep learning 1–19 (2021), https://doi.org/10.1371/journal.pone.0252402.

[87] S.J.J. Jui, A.A.M. Ahmed, A. Bose, N. Raj, E. Sharma, J. Soar, M.W.I. Chowdhury, Spatiotemporal hybrid random forest model for tea yield prediction using satellite-derived variables, Rem. Sens. 14 (3) (2022) 1–18, https://doi.org/10.3390/rs14030805.

[88] S. Khaki, L. Wang, Crop Yield Prediction Using Deep Neural Networks 10 (May) (2019) 1–10, https://doi.org/10.3389/fpls.2019.00621.

[89] P. Nevavuori, N. Narra, T. Lipping, Crop yield prediction with deep convolutional neural networks, Comput. Electron. Agric. 163 (June) (2019) 104859, https://doi.org/10.1016/j.compag.2019.104859.

[90] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1 (5) (2019) 206–215, https://doi.org/10.1038/s42256-019-0048-x.

[91] M. Elbadri, M.A. Abdoon, M. Berir, D.K. Almutairi, A symmetry chaotic model with fractional derivative order via two different methods, Symmetry 15 (6) (2023) 1–12, https://doi.org/10.3390/sym15061151.

[92] F.E.L. Guma, O.M. Badawy, M. Berir, M.A. Abdoon, Numerical analysis of fractional-order dynamic dengue disease epidemic in Sudan, Journal of the Nigerian Society of Physical Sciences 5 (2) (2023) 1–6, https://doi.org/10.46481/jnsps.2023.1464.

[93] M.A. Abdoon, Programming first integral method general formula for the solving linear and nonlinear equations, Appl. Math. 6 (3) (2015) 568–575, https://doi.org/10.4236/am.2015.63051.

[94] M. Ali, R.C. Deo, N.J. Downs, T. Maraseni, Multi-stage committee based extreme learning machine model incorporating the in fl uence of climate parameters and seasonality on drought forecasting, Comput. Electron. Agric. 152 (July) (2018) 149–165, https://doi.org/10.1016/j.compag.2018.07.013.

[95] A. Chakrabarty, N. Mansoor, M.I. Uddin, M.H. Al-Adaileh, N. Alsharif, F.W. Alsaade, Prediction approaches for smart cultivation: a comparative study, Complexity 2021 (2021), https://doi.org/10.1155/2021/5534379.

[96] J. Pant, R.P. Pant, M. Kumar, D. Pratap, H. Pant, Materials Today : proceedings Analysis of agricultural crop yield prediction using statistical techniques of machine learning, Mater. Today: Proc. xxxx (2021) 1–5, https://doi.org/10.1016/j.matpr.2021.01.948.

[97] A. Patil, P. Patil, P.S. Kokate, International journal of advancements in engineering and technology (IJAET) crop prediction system using machine learning algorithms 1 (1) (2020) 1–8.

[98] F. Abbas, H. Afzaal, A.A. Farooque, S. Tang, Crop yield prediction through proximal sensing and machine learning algorithms, Agronomy 10 (7) (2020), https://doi.org/10.3390/AGRONOMY10071046.

[99] Z. Zhang, Y. Jin, B. Chen, P. Brown, California almond yield prediction at the orchard level with a machine learning approach, Front. Plant Sci. 10 (July) (2019) 1–18, https://doi.org/10.3389/fpls.2019.00809.

[100] Y. Cai, K. Guan, D. Lobell, A.B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, B. Peng, Agricultural and Forest Meteorology Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches, Agric. For. Meteorol. 274 (February) (2019) 144–159, https://doi.org/10.1016/j.agrformet.2019.03.010.

[101] P. Feng, B. Wang, D.L. Liu, C. Waters, Q. Yu, Agricultural and Forest Meteorology Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia, Agric. For. Meteorol. 275 (May) (2019) 100–113, https://doi.org/10.1016/j.agrformet.2019.05.018.

[102] X. Wang, J. Huang, Q. Feng, D. Yin, Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. https://doi.org/10.3390/rs12111744, 2020.

[103] H. Gan, Q. She, Y. Ma, W. Wu, M. Meng, Generalization improvement for regularized least squares classification, Neural Comput. Appl. 31 (2019) 1045–1051, https://doi.org/10.1007/s00521-017-3090-9.

[104] Zhang, Y., Qin, Q., Ren, H., Sun, Y., & Li, M. (n.d.). Optimal Hyperspectral Characteristics Determination for Winter Wheat Yield Prediction. 1–18. https://doi.org/10.3390/rs10122015.

[105] I. Zoppis, G. Mauri, R. Dondi, Kernel methods: support vector machines, in: *Encyclopedia Of Bioinformatics And Computational Biology: ABC of Bioinformatics* (Vols. 1–3), Elsevier Ltd, 2018, https://doi.org/10.1016/B978-0-12-809633-8.20342-7.

[106] Y. Kang, M. Ozdogan, X. Zhu, Z. Ye, C. Hain, M. Anderson, Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest, Environ. Res. Lett. 15 (6) (2020), https://doi.org/10.1088/1748-9326/ab7df9.

[107] M. Yousefi D.B, A.S. Mohd Rafie, S. Abd Aziz, S. Azrad, M. Mazmira Mohd Masri, A. Shahi, O.F. Marzuki, Classification of oil palm female inflorescences anthesis stages using machine learning approaches, Information Processing in Agriculture 8 (4) (2021) 537–549, https://doi.org/10.1016/j.inpa.2020.11.007.

[108] T. Wang, J. Liang, X. Liu, Soil moisture retrieval algorithm based on TFA and CNN, IEEE Access 7 (2019) 597–604, https://doi.org/10.1109/ACCESS.2018.2885565.

[109] P.S. Maya Gopal, R. Bhargavi, Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms, Appl. Artif. Intell. 33 (7) (2019) 621–642, https://doi.org/10.1080/08839514.2019.1592343.

[110] J. Zhao, X. Yang, S. Sun, Constraints on maize yield and yield stability in the main cropping regions in China, Eur. J. Agron. 99 (August 2017) (2018) 106–115, https://doi.org/10.1016/j.eja.2018.07.003.

[111] A. Haghverdi, R.A. Washington-allen, B.G. Leib, Prediction of cotton lint yield from phenology of crop indices using arti fi cial neural networks, Comput. Electron. Agric. 152 (July) (2018) 186–197, https://doi.org/10.1016/j.compag.2018.07.021.

[112] E. Panek, D. Gozdowski, Remote Sensing Applications : society and Environment Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data, Remote Sens. Appl.: Society and Environment 17 (August 2019) (2020) 100286, https://doi.org/10.1016/j.rsase.2019.100286.

[113] C. Vallentin, K. Harfenmeister, S. Itzerott, B. Kleinschmit, C. Conrad, D. Spengler, Suitability of satellite remote sensing data for yield estimation in northeast Germany, in: Precision Agriculture, vol. 23, Springer, US, 2022, https://doi.org/10.1007/s11119-021-09827-6. Issue 1.

[114] H.T. Pham, J. Awange, Evaluation of Three Feature Dimension Reduction Techniques for Machine Learning-Based Crop Yield Prediction Models, vols. 1–18, 2022.

[115] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.

[116] M.K. Gumma, P.S. Thenkabail, P. Panjala, P. Teluguntla, T. Yamano, I. Mohammed, Multiple agricultural cropland products of South Asia developed using Landsat-8 30 m and MODIS 250 m data using machine learning on the Google Earth Engine (GEE) cloud and spectral matching techniques (SMTs) in support of food and water security, GIScience Remote Sens. 59 (1) (2022) 1048–1077, https://doi.org/10.1080/15481603.2022.2088651.

[117] G. Morales, J.W. Sheppard, P.B. Hegedus, B.D. Maxwell, Improved yield prediction of winter wheat using a novel two-dimensional deep regression neural network trained via remote sensing, Sensors 23 (1) (2023), https://doi.org/10.3390/s23010489.

[118] J. Cao, Z. Zhang, Y. Luo, L. Zhang, J. Zhang, Z. Li, F. Tao, Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine, Eur. J. Agron. 123 (December 2020) (2021) 126204, https://doi.org/10.1016/j.eja.2020.126204.

[119] E. Cheng, B. Zhang, D. Peng, L. Zhong, L. Yu, Y. Liu, C. Xiao, C. Li, X. Li, Y. Chen, H. Ye, H. Wang, R. Yu, J. Hu, S. Yang, Wheat yield estimation using remote sensing data based on machine learning approaches, Front. Plant Sci. 13 (December) (2022) 1–16, https://doi.org/10.3389/fpls.2022.1090970.

[120] S. Jeong, S. Ko, J.M. Yeom, Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea, Sci. Total Environ. 802 (2022) 149726, https://doi.org/10.1016/j.scitotenv.2021.149726.