Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 23 | No. 4 | pp. 2071-2086 | 2024

Review Research Paper
https://doi.org/10.46488/NEPT.2024.v23i04.014
Open Access Journal

Advancements in Machine Learning and Deep Learning Techniques for Crop Yield Prediction: A Comprehensive Review

V. Ramesh and P. Kumaresan†
School of Computer Science Engineering and Information Systems (SCORE), Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India
†Corresponding author: P. Kumaresan; pkumaresan@vit.ac.in

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 04-03-2024
Revised: 19-04-2024
Accepted: 01-05-2024

Key Words:
Crop yield prediction
Machine learning
Deep learning
Artificial neural networks
Optimization algorithms

ABSTRACT

Agriculture is the crucial pillar and basic building block of our nation. Agriculture plays a key role as the major source of revenue for our nation. Farming is the primary financial source of India. Abrupt environmental changes affect crop yield prediction. Unpredictable climate changes, lack of water resources, deficiency of nutrients, depletion of soil fertility, unbalanced irrigation systems, and conventional farming techniques are the major causes of crop yield prediction. Today, AI, the use of machine learning, and deep learning techniques provide an achievable solution to improve crop yields. The key intent of the survey is to accurately predict and improve crop yield by combining agricultural statistics with machine learning and deep learning models. To accomplish this, we have surveyed the optimization algorithms implemented in conjunction with the Random Forest and Cat Boost models. A survey made across multiple databases to determine the effectiveness of crop yield prediction and analysis was performed on the included articles. The survey results show that a hybrid CNN DNN and RNN model with optimization algorithms outperforms the other existing traditional models.

INTRODUCTION

Indian economy is contingent on agriculture because it is crucial for the survival of both humans and animals in the country (Durai & Shamili 2022). From 2009 to 2030, the global population is projected to grow from 1 billion to 5 billion, leading to a significant increase in the need for agricultural commodities. As a result, there will be a greater demand for agricultural products among people, necessitating the efficient utilization of farmland and an increase in agricultural yields. Harmful climatic conditions caused by global warming frequently lead to spoiled harvests (Tseng et al. 2019). When a crop fails due to insufficient soil fertility, climate change, groundwater scarcity, flooding, or other similar circumstances, it directly affects farmers. Depending on geographical conditions and environmental factors, society in other countries recommends that farmers increase the production of certain crops (Reddy & Kumar 2021). Assessing and tracking crop productivity is crucial because of the population's faster growth (Alagurajan & Vijayakumaran 2020). Consequently, to select crops more effectively based on seasonal variation, it is critical to develop a suitable model that takes into account the relevant factors (Kumari et al. 2020).

Machine learning, a significant field of AI that focuses on the method of learning, can significantly enhance the accuracy of yield prediction. It incorporates several features to achieve this. Machine learning can extract information from datasets and identify patterns and correlations. Datasets should be used to train models, and experience should be used to describe the findings. To create a predictive model, several features are combined, and its parameters are derived using historical data collected during the training stage. Performance evaluation is conducted using a subset of historical data that is separate from the data used for training (Klompenburg et al. 2020).

Machine learning, deep learning, and hybrid models with optimization techniques are being widely used worldwide due to their efficiency in different sectors, including predicting weed detection and disease. These algorithms also aid in improving crop yield prediction in unfavorable conditions. Regardless of the distracting environment, ML, DL, Hybrid, and Optimization algorithms are used in predicting crop yields and minimizing losses.

Fig. 1 shows the proposed RF-CatBooster crop yield prediction, indicating that yield increases under different circumstances when ML, DL, Hybrid, and Optimization Techniques are utilized. The machine learning approach has been adopted as the foundation for accurate predictions. Crop prediction leverages a classification model, while yield prediction utilizes regression models to learn insights

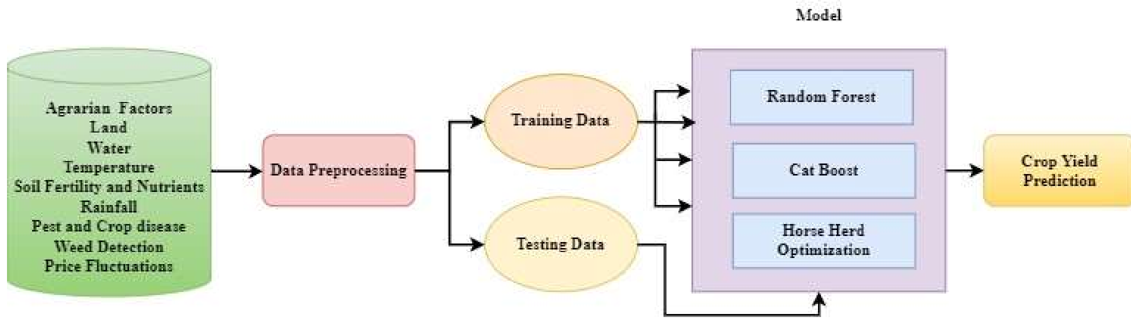

Fig. 1: Proposed RF-CatBoost architecture for crop yield prediction.

from the data. Multiple machine learning models have been surveyed based on performance metrics using a publicly accessible dataset spanning from 1999 to 2020. We propose a Random Forest and CatBoost model for predicting crop yield using preprocessed data. The dataset incorporates numerical attributes such as Crop_Name, Area, Crop_Year, Annual_Rainfall, Fertilizer, and Yield_Data. Preprocessing steps encompass encoding categorical variables like Area and Crop_Name in preparing data for the proposed model. The standardization of attributes Crop_Year, Annual_Rainfall, Fertilizer, and Yield_Data enhances consistency through the utilization of the Standard Scaler. A model is trained on 80% of the training dataset with the remaining 20% utilized to test the trained model. Prior pre-processing steps include handling missing data points, noisy entries, outlier removal, and duplication removal. MinMax Scaler resizes data proportionally within a stipulated range of 0 to 1, transforming features while retaining their original distribution shape by rescaling value to a specific range without modifying original distribution shapes. The proposed Hybrid model RF-CatBooster effectively handles non-linear data, manages categorical data, and reduces overfitting. The Horse Herd optimization algorithm is used to determine the most effective link weights in the classifier through error value computation and storage of the superior weight along with its position. Performance was measured using metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE). The results demonstrated that RF-CatBoost Regressor outperformed other models, with lower MAE and MSE scores (Moussaid et al. 2022).

The structure of this article is as follows. Section 2 presents a survey methodology for crop yield prediction. Section 3 discusses how environmental factors influence crop yield prediction. Section 4 analyzes the features and related datasets for crop yield prediction. Section 5 presents an overview of the existing machine learning, deep learning, and hybrid models for crop yield prediction. Section 6 presents the findings and discussion, and Section 7 concludes the article.

MATERIALS AND METHODS

Review Methodology

The Systematic Literature Review incorporates and provides research studies with the research questions presented in this literature, in addition to gathering them from conferences, journals, and other electronic sources.

Research Questions

The following questions served as the foundation for the review paper's analysis and exploration of each study's various facets. The following is a list of the research questions:

- RQ1: How are the features used to classify in predicting crop yield?
- RQ2: How are the data sources utilized in the crop yield prediction process?
- RQ3: How can machine learning methods be used to identify various crops in yield prediction?
- RQ4: How have crop yield predictions been implemented using machine learning models?
- RQ5: Determine the methodologies utilized to evaluate the efficacy of machine learning algorithms.

The article search was meticulously designed around the central theme of the systematic literature review and its guiding research questions. To avoid the pitfalls of generic keywords, the search strategy went beyond simply encompassing "machine learning." Instead, it focused on the intersection of "crop yield prediction" and "machine learning." This targeted approach was initially executed across nine databases (Elsevier, Springer, IEEE Explorer, MDPI, Taylor & Francis, Tech Science Press, Frontiers in Plant Science, Earth System Science Data, and Journal of Ecological Engineering), ensuring relevant results. A total of 81 documents were found during this search for evaluation and analysis.

The article identifies the records retrieved after searching through 326 papers. The duplicated records were removed and the remaining records were screened in 317 papers. 180 records not related to our research were excluded. Out of the remaining 137 papers with full-text records, 56 papers were excluded for various agri-model predictions. Finally, 81 papers were assessed for crop yield and included in the review.

Impact of Environmental Factors on Crop Yield Prediction

Developing an accurate and understandable prediction model for yield is a critical and arduous task. This is of fact that crop production is influenced by various crop-specific metrics, environmental factors, and management choices. Traditionally, crop yield predictions have been made using a combination of crop growth models, field surveys, and statistical models. A different facet of crop production prediction is addressed by each of these methods. Surveys in the field aim to obtain the truth on the ground. According to agronomic concepts of plant, environment, and management interactions, crop growth models simulate the growth and development of crops. To determine linear correlations between the predictors and crop yield, statistical models are used to predict meteorological variables and the results of the three preceding techniques as Field observations, Crop growth models, and Statistical models (Paudel et al. 2020).

Abiotic and biotic factors influencing crop yield: Abiotic factors and Biotic factors include soil, sunlight, temperature, wind, atmosphere, pH, and water, pests, and diseases, which affect the entire crop production. For example, droughts, wind, and heavy rainfall affect the crops sometimes, destroying the entire crops.

Crop yield is primarily influenced by four major agrarian factors: availability of water, soil productivity, illnesses, climate, and pests. These issues can put farmers in danger if they are not sufficiently assessed and handled. To maximize crop output while reducing risk, it is critical to examine the elements that influence crop productivity and the risks involved (Elavarasan et al. 2018).

Land, rainfall and temperature variation: Annual crop inventory maps that display both agricultural and non-agricultural land usage within Canada's agricultural area are published by Agricultural and Agri-Food Canada (AAFC). The maps are space-based operational and sensing tools that can boost agricultural productivity (Cravero et al. 2022). In a similar vein, (Amani et al. 2020) have produced high-resolution reference maps of South Asian cropland, discussing the necessity of raising output in light of regional food shortages. Too little rainfall causes crops to shrivel and die, while too much rain leads to flooding, which wastes water, fertilizer, labor, and energy, while excessive rainfall harms crop growth (Ndehedehe et al. 2018, Kalaivanan & Vellingiri 2022). Crop growth properties, such as cell division, water transport, survival, photosynthesis, growth, and yield, are impacted by low temperatures (Su et al. 2017). Additionally, advanced technologies like remote sensing, weather satellites, and weather stations help monitor and analyze rainfall patterns more accurately. By understanding how rainfall impacts crops, agricultural experts can make decisions on irrigation, selection of crops, and planting season to optimize yield and mitigate potential risks caused by rainfall variability (Khosla et al. 2019).

Soil fertility and nutrients: Many factors influence agricultural productivity, but the most important ones are soil fertility, climate, availability of water, plant diseases, and pets. Farmers can be put at significant risk if these issues are not properly managed. A crop's ability to develop as healthily as possible depends on the soil's fertility. Healthy crops require 18 essential nutrients, which are divided into macronutrients and micronutrients. Macronutrients (N, P, K, Ca, S, Mg) are needed in larger amounts, while micronutrients (Fe, Zn, Cu, B, Mn, Mo, Cl, Si) are needed in smaller amounts (Raut et al. 2020).

Pest, crop disease, and weed detection: Pest and disease activity is another important factor that affects crop output (Ip et al. 2018). The most common plant diseases are spot, blight, canker, and rust (Amudha & Brindha. 2022). Pests and diseases come in various sizes and shapes, and they pose different risks to crops. Some insects, such as plant parasites, can harm crops both directly and indirectly (Serra & Tagliaferri. 2018**,** Roldán-Serrato et al. 2018). Weed control is a major problem for farmers during the growing season. For example, a single weed can grow approximately 10 million weed seeds, which, if not quickly eliminated, can drastically lower agricultural production or cause difficulties for years (Kumar Nagothu et al. 2023**).**

Price fluctuations: Advanced data analytics, satellite imagery, and machine learning models are commonly used to forecast crop yields. The forecasts can provide valuable information to farmers, merchants, and policymakers, helping them make effective decisions regarding harvesting, planting, and marketing strategies. Even though these methods can be accurate, they are not infallible, and unexpected events can still impact prices (Sun et al. 2023).

Features and Related Dataset for Crop Yield Prediction

A study question (RQ1) was addressed by analyzing and presenting the features used with the machine learning

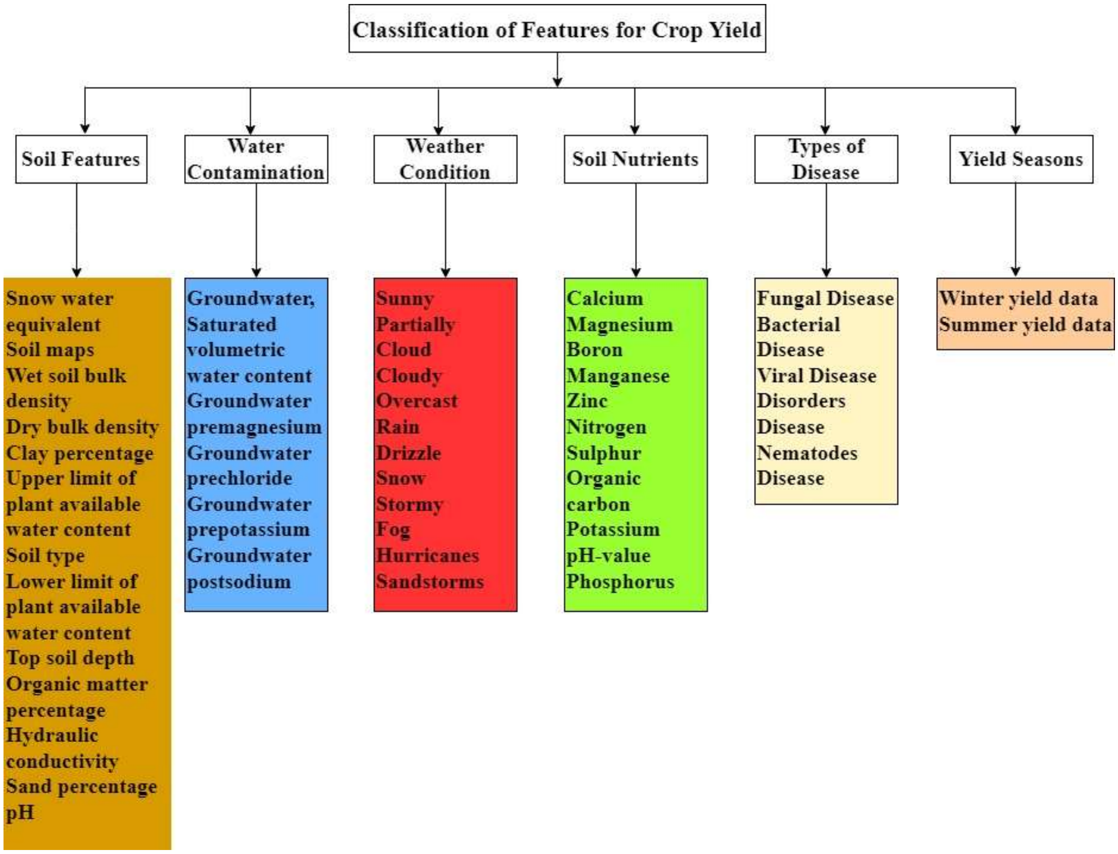

Fig. 2: Features classification for crop yield prediction.

techniques used for yield prediction. The vast quantity of data features used in yield estimation improves feature comprehension. Six categories were created from the features: soil features, water contamination, weather conditions, soil nutrients, types of disease, and yield seasons. The categorization of crop yield prediction features is seen in Fig. 2.

For instance, all satellite and aerial data features were combined, and data features of groundwater magnesium, groundwater sodium, groundwater chloride, and groundwater potassium were grouped with water content. Together with soil features, soil characteristics such as pH, type, wet soil density, dry bulk density, organic matter percentage, soil maps, snow water equivalent, clay percentage, and the upper and lower limits of plant-available water content were combined into one category. Weather data was combined with weather-related features, such as vapor pressure, daily minimum and maximum air temperatures, daily solar radiation, wind speed, temperature, rain, and precipitation. Data on diseases and yield seasons were grouped together with the names of other participants. Table 1 shows dataset features utilized in the survey article to forecast crop yields, specifically in relation to the Research Question (RQ2).

Crop selection for yield prediction using machine learning methods: A wide range of crop yields are estimated using machine learning techniques. Research Question (RQ3) was addressed through an analysis and presentation of the crops used in the machine learning techniques. Various crops such as Rice, Potato, Soybean, Cotton, Ragi, Barley, Apple, Coffee, Wheat, and Mango were examined in the articles that were reviewed. Fig. 3 shows the distribution of the various crop kinds employed in the examined articles. A machine learning algorithm has been used to predict crop yield including rice, soybeans, and wheat. The most prevalent crop whose output can be broadly predicted with machine learning approaches is rice.

In the reviewed articles, 19 papers were used to predict rice yield, 12 for maize yield, 10 for wheat yield, 3 for groundnut, cotton, and banana yield, and 2 for ragi, jowar, bajra, apple yield, and other crops.

Overview of the Existing Machine Learning, Deep Learning, and Hybrid Models for Crop Yield Prediction

Machine learning approaches: In supervised machine learning, the machines are trained using labeled datasets,

Table 1: Overview of features of Machine Learning and Dataset Description.

Authors	Dataset	Soil Info.	Water Info.	Weather Data	Nutrients	Yield Data	Synthetic Image
Goldstein et al. (2018)	MANAGE, Meteorological data	√		√			
Aghighi et al. (2018)	Silage maize dataset, the United States Geological Survey data center.					√	√
Zhong et al. (2018)	Syngenta dataset	√		√			
Kouadio et al. (2018)	Food and Agriculture Organization data	√		√		√	
Taherei Ghazvinei et al. (2018)	Daily Basis dataset	√		√			
Deepa and Ganesan (2019)	Agriculture sites of Tamilandu(Tiruvannamalai)	√	√	√			
Bondre and Mahagaonkar (2019)	The past five years' data	√			√	√	
Khosla et al. (2019)	Rainfall Data and Crop Related Data collected from data.gov.in		√	√			
Leroux et al. (2019)	MODIS - MOD13Q1, MOD11A2, NDVI, LST	√		√		√	√
Maya Gopal and Bhargavi (2019 a)	Statistical Department of Tamilnadu,		√	√	√	√	
Filippi et al. (2019)	spatial and temporal data collected on-farm	√		√			√
Cai et al. (2019)	MODIS MOD13C1 EVI, Spatial Production Allocation Model (SPAM), Australian Bureau of Statistics (ABS) from 2000 to 2014 at the SD level (unit: t/ha).					√	√
Shiu and Chung (2019)	Spot-7 Multispectral Satellite Image						√
Elavarasan et al. (2020)	Meteorological Department of India, Agricultural Department of Tamilnadu	√	√	√		√	
Kamir et al. (2020)	MODIS dataset - MOD13Q1	√		√		√	√
Guo et al. (2021)	Chinese Meteorological Administration			√			
Nyeki et al. (2021)	spatiotemporal database	√		√			
Pant et al. (2021)	Food and Agriculture Organization data			√	√	√	
Joshua et al. (2021)	data.gov.in and indiastat.org data	√		√	√	√	
Batool et al. (2022)	Data collected from NTHRI.	√		√		√	√
Burdett and Wellen (2022)	Professional agronomists(Last year)	√				√	
Jeevaganesh et al. (2022)	Over the past two decades, agricultural data from across India has been analyzed.	√		√		√	
Cedric et al. (2022)	World Bank's knowledge portal CCKP,			√	√	√	
Croci et al. (2022)	Crop and Yield Data, Meteorological, Soil, Satellite	√		√		√	√
Gopi and Karthikeyan (2022)	Kaggle - Crop Recommendation and Crop Yield Prediction			√	√	√	
Rahman and Aktar (2022)	Food and Agriculture Organization data	√		√		√	
Joshua et al. (2022)	Agricultural Department of Tamilnadu, Regional Meteorological Centre - Chennai, Tata-Cornell Institute for Agriculture and Nutrition, Statistical Department of Tamilnadu.			√	√	√	
Iniyan et al. (2023)	Agriculture csv dataset for Crop Yield Prediction	√		√		√	
Ed-Daoudi et al. (2023)	The Regional Agricultural Development Office Ouarzazate Morocco.	√		√		√	
Sathya and Gnanasekaran (2023)	Joint Director of Agriculture Office, Kattuthottam, Thanjavur and Indian Meteorological Department, Chennai.	√	√	√	√		

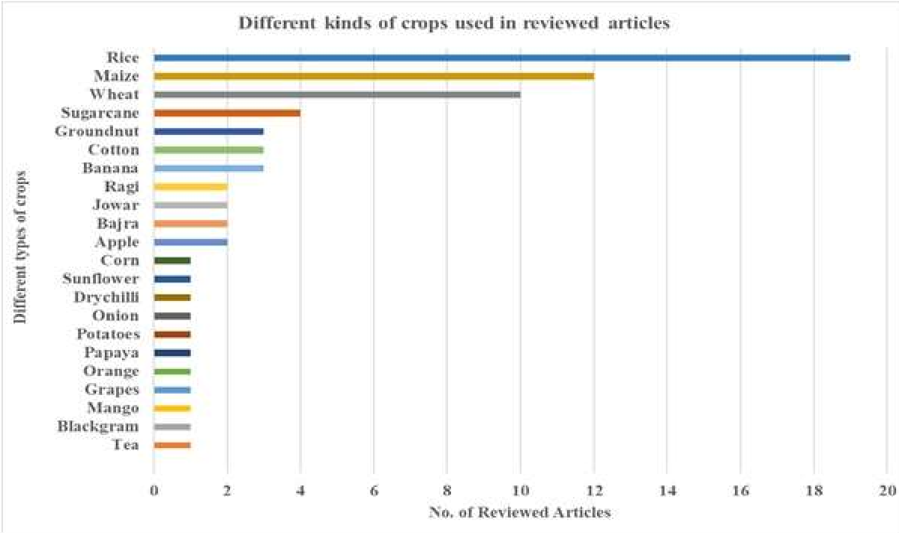

Fig. 3: Various crops distribution.

allowing them to predict outputs based on the training. The labeled data shows that certain inputs are already linked to the output. The process involves training the machine by providing it with input data and corresponding output data. After this training phase, the machine is then tested for predicting the output using separate test data. Unsupervised machine learning is trained using an unlabeled dataset to make predictions without any supervision. Semi-supervised lies between supervised learning and unsupervised learning algorithms. During the training process, it utilizes the merging of labeled and unlabeled datasets. Reinforcement learning operates through a feedback-based process. AI software agent explores their surroundings by taking action and learning from their training. Performance is improved, and the agent automatically adapts and becomes more proficient. Fig. 4 shows the general classification of Machine Learning Models.

Machine learning algorithms are crucial in predicting crop yields. The review paper examined various machine learning approaches, including Linear Regression, Logistic Regression, K-Nearest Neighbour, Gradient Boosting, Decision Tree, Random Forest, Cat Boost, XGBoost, CNN, DNN, RNN, Long Short-Term Memory, Artificial Neural Network and Hybrid Networks. The various algorithms were listed to address the Research Question (RQ4) to highlight their benefits.

Linear, logistic, and gradient-boosting regression: The most straightforward machine learning method is linear regression. Liang et al. (2023) predicted crop yield using the multiple linear regression method while considering socioeconomic and natural factors. The example phrase is:

$$y = b1x1 + bnxi + b0 \qquad \ldots(1)$$

When y is the crop yield, x1... xi are natural and socioeconomic factors, b0 is a constant term, and b1... bn is the regression coefficient. Logistic regression, a statistical model with a logistic function, represents a binary dependent variable in its simplest form. Islam et al. (2022) proposed a model called the attention-based dilated CNN logistic regression for the detection of tomato leaf disease with the highest accuracy rate. Gradient boosting, an ensemble approach, is a powerful technique in practical machine learning. Nihar et al. (2022) predicted the regional wise sugarcane crop yields from the Uttar Pradesh agriculture dataset using satellite images. Verma (2022) predicted the crop yield from weather and soil conditions. The five ML models were used, such as KNN, SVC, RF, DT, and Gradient Boosting. Gradient Boosting has achieved the highest accuracy.

Random forest and cat boost models: Random Forest is widely recognized for its high accuracy, robustness, versatility, scalability, and importance in determining features in classification and regression tasks. Random Forest reduces overfitting by incorporating the average predictions of multiple decision trees. This averaging process enhances its robustness against noise and outliers present in the data. It offers a way to determine the importance of features, which can help select features and interpret data. Random Forest architecture is shown in Fig. 5. Choudhary et al. (2022) developed Sentinel-2 data that is suitable for predicting rice yield and conducting advanced classification with high yield prediction accuracy. Jain & Choudhary (2022) have developed a Soil-Based Machine Learning Comparative

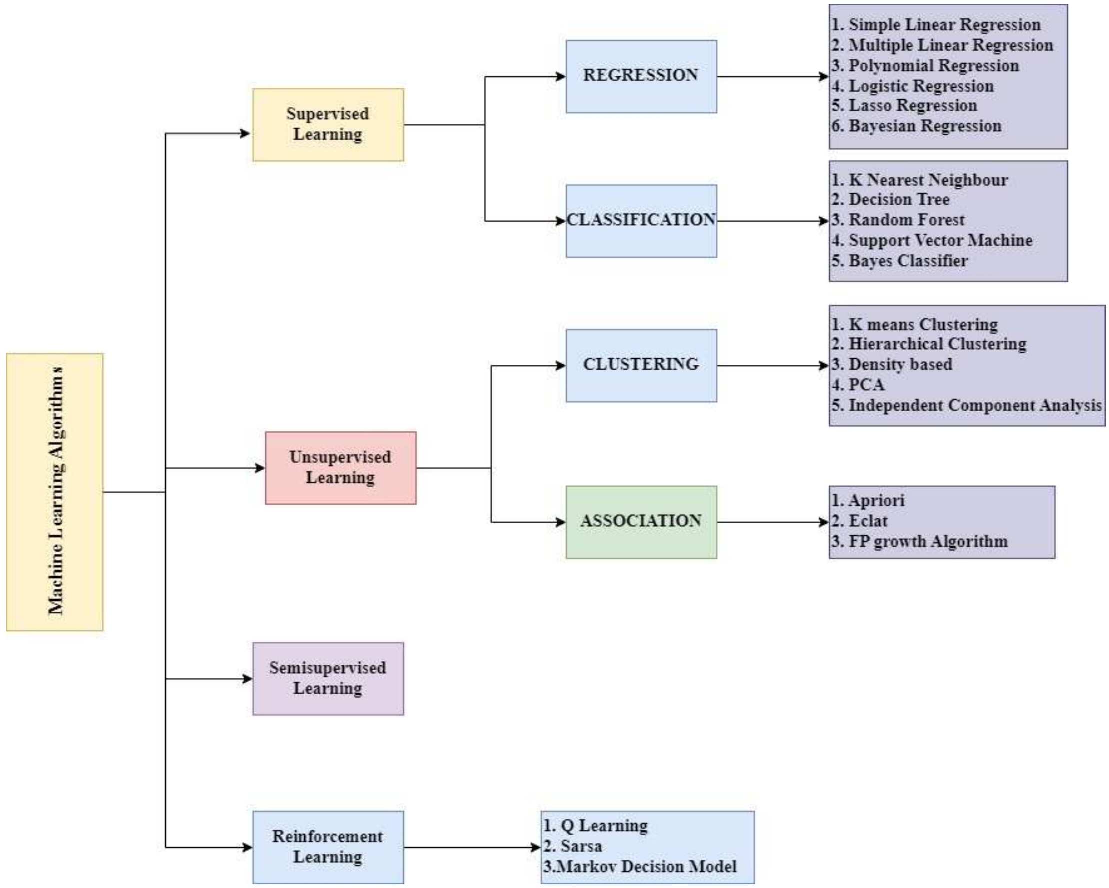

Fig. 4: Classification of Machine Learning Models.

Analytical Framework to forecast crop yield production. The SMLF utilizes soil features and climate factors to create a feature vector. The performance of SMLF is higher than other methods in the yield prediction. Croci et al. (2022) developed a machine-learning framework to predict maize yield. They incorporated different data sources, namely vegetation indices, soil, and meteorological data. The framework is used to identify the machine learning configuration that performs the best and the optimal lead time.

Cat Boost is an algorithm for gradient boosting that is specifically designed to handle datasets containing numerous categorical variables. It employs gradient descent to optimize the decision tree parameters, thereby enhancing the model's performance. The study evaluated various algorithms, including CatBoost Regressor, for predicting tree crop yield.

XGBoost: The XGBoost algorithm, short for eXtreme Gradient Boosting, is a combination of gradient-boosted regression trees. It is an improved gradient-boosting machine developed specifically for enhancing the output prediction speed and performance. Hazra et al. (2023) used seven machine-learning algorithms for predicting crop yields, such as KNN, RF, XGBoost, LightGBM, ANN, SVM, and MLP. Among these, the XGBoost model gave the highest accuracy. Panigrahi et al. (2022) conducted a study in the Telangana region of India from 2016 to 2018, where they developed a crop yield prediction model using XGBOOST. The model was specifically designed for maize, groundnut, and Bengal gram. The researchers obtained the data from the Open Data Source of the State of Telangana, India, which was provided by the Department of Agriculture and Cooperation, Government of Telangana. Their findings showed that the accuracy of the XGBoost model surpassed that of other models. Kaur Dhaliwal et al. (2022) predicted historical cotton lint yield in the southeastern United States using six different ML techniques. Among these six techniques, the XGBoost showed the highest accuracy.

Deep Learning Approaches

A deep learning algorithm is a component of a machine learning algorithm that is utilized to execute complex calculations on a large amount of data in sophisticated manners (Muruganantham et al. 2022). The input, hidden,

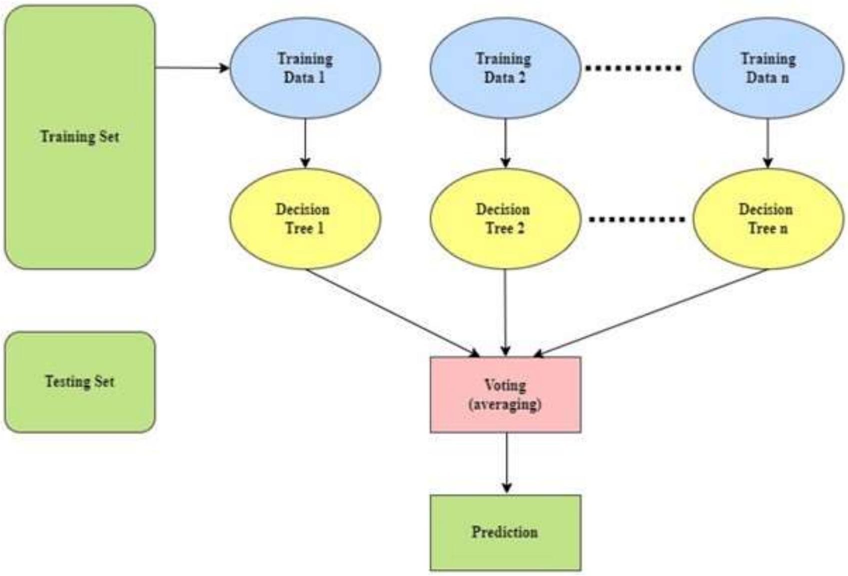

Fig. 5: Architecture of Random Forest.

and output layers of the neural network are all made up of nodes or artificial neurons. Each node receives data inputs and combines them with random weights. Finally, activation functions determine which neuron should be activated. By automatically detecting hidden patterns in the data, deep learning methods can develop more effectual decision rules. Deep learning algorithms generally outperform conventional machine learning algorithms in terms of prediction accuracy (Elavarasan & Vincent 2021).

Deep learning techniques can capture the spatio-temporal relationships in datasets (Tian et al. 2021a). Several deep learning approaches, including MLP, RNN, DNN, CNN, LSTM, and autoencoders, have been used in numerous studies to predict crop yields. Deep learning algorithms can automatically extract salient features from the data, eliminating the need for manual data preparation. An LSTM (Long Short-Term Memory) network can effectively mitigate the vanishing gradient problem that can occur with deep recurrent neural networks (RNNs) (Liu et al. 2022).

Neural networks for crop yield prediction: Several artificial neurons stacked on top of one another make up a CNN. The layers that make up a CNN consist of the pooling layer, the convolution layer, and the fully connected layer. The CNN Layers are used to process the dataset and extract its features (Wang et al. 2020). To identify crops and weeds, assess biomass, and forecast the production of wheat and barley crops with multispectral data, Nevavuori et al. (2019) developed a model using Deep CNN. The technique of convolutional neural networks produces outstanding results in problems related to object detection and image classification. The results show that using RGB images improves the accuracy of yield estimates made by CNN algorithms. A deep neural network is a form of feed-forward neural network containing multiple fully connected hidden layers. Ang et al. (2022) predicted the yield for oil palm at the block level from multi-source data using Multiple Linear Regression, XGBoost, SVR, RF, and DNN approaches. Kalaiarasi & Anbarasi. (2022) introduced the concept of MDNN (Multi-parametric Deep Neural Network) for predicting crop yield by incorporating various factors like climate and soil conditions. The RNN can process the arbitrary length of an input sequence, extract the features from the input sequence, and store them in its hidden state (Fig. 6).

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) known for its ability to learn long-term dependencies and retain sequential data, including relevant information from previous inputs (Tian et al. 2021 b). The LSTM was developed to overcome the challenges faced by traditional RNNs, specifically the issues of exploding and vanishing gradients. It is particularly valuable in time series forecasting due to its ability to remember past inputs. The LSTM unit consists of a cell (Fig. 7), which includes an input gate, an output gate, and a forget gate. These gates regulate the flow of data within the cell and retain values for an unlimited duration. Elavarasan and Vincent (2020a) have developed a framework that uses a deep recurrent Q-learning network with 38 features to accurately estimate agricultural yield. This algorithm, known as Q-learning, offers improved accuracy and reliability in crop yield forecasts when compared to other models.

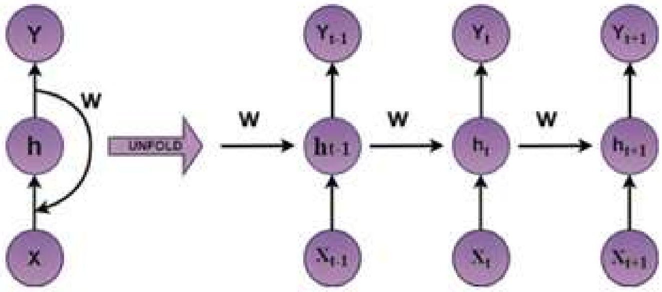

Fig. 6: Architecture of RNN.

Fig. 8 shows the standard architecture of an Artificial Neural Network (ANN) consisting of three layers: the input layer, the hidden layer, and the output layer. Each layer is comprised of multiple neurons or nodes. The input is initially received through the input layer and then forwarded to the hidden layer, commonly referred to as the core of the ANN. There may be one or more hidden layers, which are responsible for processing the information and uncovering hidden patterns or features. The output layer, which takes input from the last hidden layer, is the final layer responsible for providing the output. An ANN model was utilized by Anurag Satpathi et al. (2023) to predict rice yield.

Hybrid and Optimization Algorithms for Crop Yield Prediction

The purpose of optimization is to decrease the loss function, leading to improved prediction accuracy. The horse Herd optimization algorithm derives the optimal weights of the classifier links. Krishna et al. (2023) utilized A-BiLSTM-MFA to predict crop yields in India, focusing on specific features. The resulting predictions were both accurate and fast.

RESULTS AND DISCUSSION

The selected articles are analyzed and summarized in the review. Fig. 9. shows the number of articles published from 2016 and 2023.

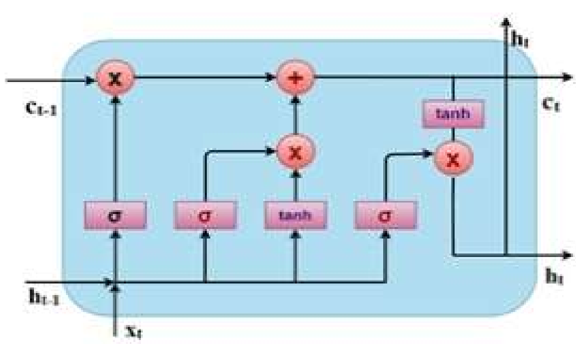

Fig. 7: Architecture of LSTM.

Selection of Crops for Yield Prediction Based on ML Approaches

In the articles reviewed, authors have utilized approximately 37 different crops for yield prediction. Among these, 7 crops are the most commonly used and are illustrated in Fig. 10. The remaining 30 crops, which are either used solely for comparison or to support yield prediction, have been grouped under others.

Performance Evaluation Metrics

Different evaluation metrics are used to measure the performance of machine learning in predicting crop yield. A total of 18 metrics are considered, including RMSE, Coefficient of determination (R^2), MAE, MSE, Precision, Accuracy, F1-Score, Recall, MAPE, Correlation Coefficient(R), NMSE, RRMSE, Sensitivity, Specificity, ME, CV, RMAE, MCC, Index of Agreement were investigated in the reviewed papers. Fig. 11 shows performance evaluation metrics for crop yield prediction Algorithm.

Mean Absolute Error (MAE)

The mean absolute error refers to the absolute difference between the predicted value and the actual value (Aghi et al. 2018). Eq. 2. Gives the mathematical expression of MAE.

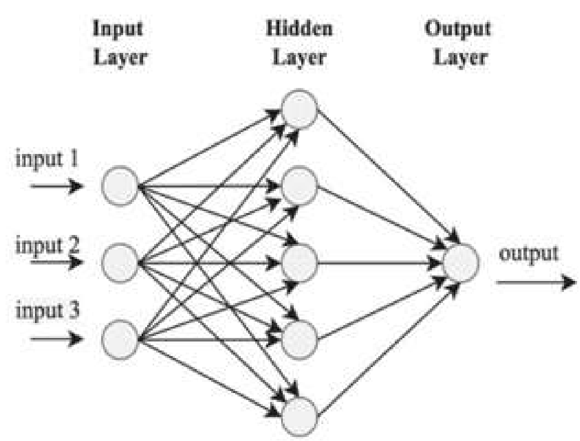

Fig. 8: Architecture of Artificial Neural Networks.

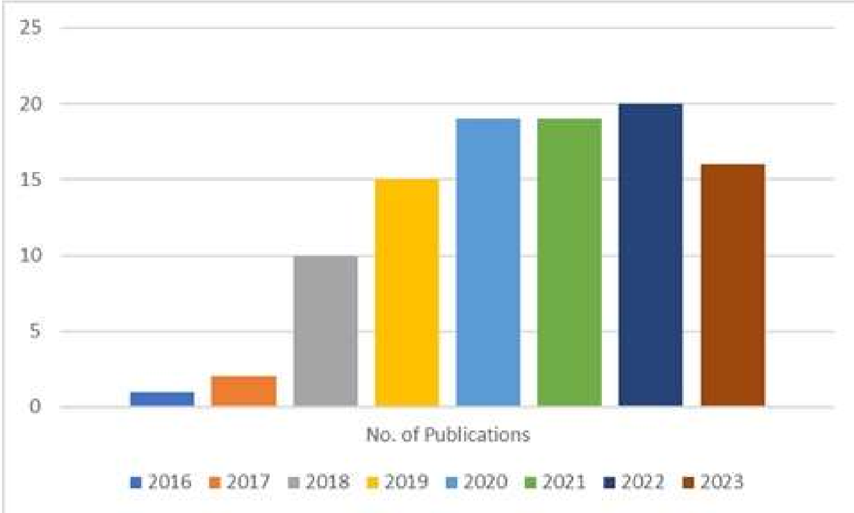

Fig. 9: The number of reviewed articles with respect to the publication year.

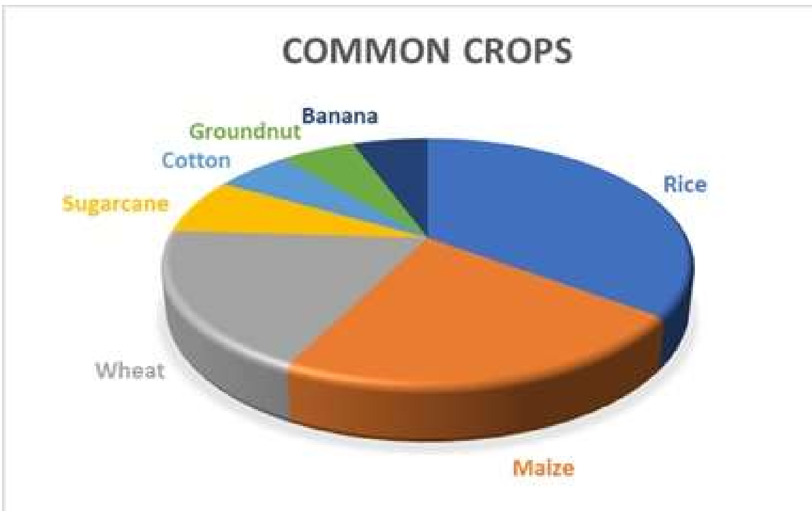

Fig. 10: Analysis of Common Crops.

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|Y_i-\hat{Y}_i\right| \quad \ldots(2)$$

Mean Square Error (MSE)

The mean square error is obtained by average squared errors. This value represents the average of the squared differences between the predicted and actual values (Elavarasan et al. 2020 b). Eq. 3. Gives the mathematical expression of MSE.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i}-y_i)^2 \quad \ldots(3)$$

Precision

Precision is determined by calculating the sum of accurately predicted crop yield ratings (True Predicted) divided by the total number of crop yield predictions (True Predicted, False Predicted) (Deepa and Ganesan. 2019). Eq. 4. Gives the mathematical expression of Precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \ldots(4)$$

Accuracy

Accuracy refers to the proportion of accurate crop yield predictions out of the total number of crop yield predictions made (Nyeki et al. 2021). Eq. 5. Gives the mathematical expression of Accuracy.

$$\text{Acc} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad \ldots(5)$$

F1 Score

The F1 Score is a measurement of accuracy for a test. It considers both precision and recall when determining the score. The F1 score is calculated by taking the harmonic mean of precision and recall (Bondre & Mahagaonkar 2019). Eq. 6. Gives the mathematical expression of F1 Score.

$$F1 = \frac{(2TP)}{(2TP+FP+FN)} \quad \ldots(6)$$

The F1 score is calculated as the weighted average of precision and recall. A value of 1 represents the best performance, while a value of 0 indicates the worst.

Recall

Recall is defined as the sum of the crop yield ratings that were accurately predicted, divided by the total number of attempts made to predict crop yield (both successful and unsuccessful predictions) (Singh Boori et al. 2023). Eq. 7. Gives the mathematical expression of Recall.

$$\text{Recall}= \frac{TP}{TP+FN} \quad \ldots(7)$$

Correlation Coefficient (R)

The correlation coefficient (R) is included to quantify the strength of the linear relationship between the predictions of the regression model and the actual values (Aghighi et al. 2018). Eq. 8. Gives the mathematical expression of correlation coefficient (R).

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}} \quad \ldots(8)$$

Specificity

Specificity is determined by dividing the number of individuals who test negative by the total number of individuals without the disorder (Gopi & Karthikeyan. 2023). Eq. 9. Gives the mathematical expression of Specificity.

$$\text{Specificity}= \frac{TN}{TN+FP} \quad \ldots(9)$$

Square Root of Mean Absolute Error (RMAE)

RMAE stands for the square root of the Mean Absolute Error (MAE). Its naming convention is similar to that of Root Mean Square Error (RMSE), which is the square root of Mean Square Error (MSE) (Kouadio et al. 2018). Eq. 10. Gives the mathematical expression of RMAE.

$$RMAE = 100*\frac{MAE}{\hat{Y}m} \quad \ldots(10)$$

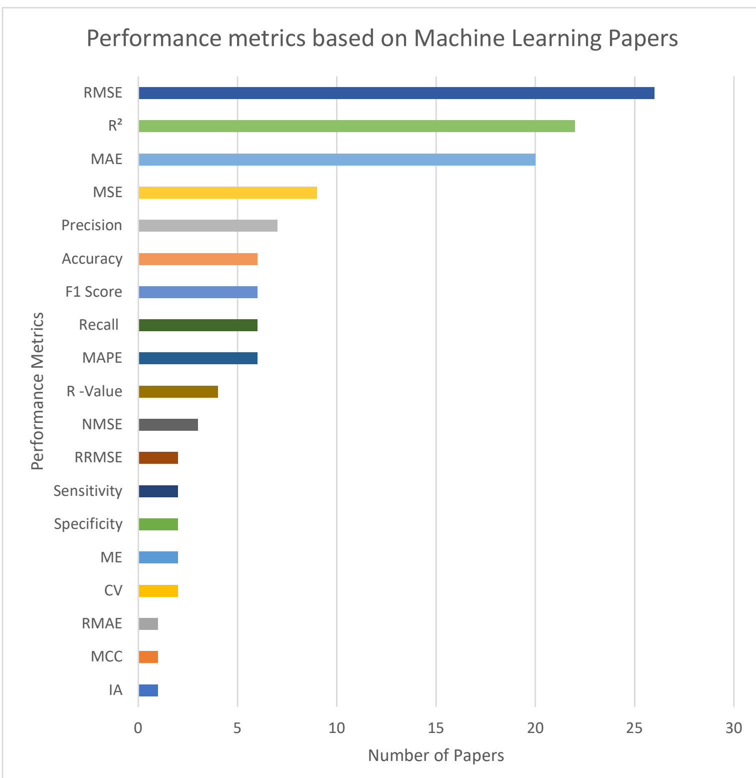

Fig. 11: Performance evaluation metrics for crop yield prediction Algorithm.

Table 2: Overview of different machine learning algorithms, crop utilized, and evaluation metrics used in crop yield prediction.

Authors	Machine Learning Algorithms	Crops	Performance Evaluation Metrics
Goldstein et al. (2018)	LR, GBRT, BTC	Alfalfa, Barley, Corn	RMSE, ME
Aghighi et al. (2018)	BRT, RFR, SVR, GPR	Maize	R, RMSE, MAE
Zhong et al. (2018)	RF, Stochastic decision model	Soyabean	RMSE, MAE
Kouadio et al. (2018)	ELM, RF, MLR	Coffee	RMSE, MAE, RRMSE, RMAE
Taherei Ghazvinei et al. (2018)	ELM, ANN, GP	Sugarcane	RMSE, r, R^2
Deepa and Ganesan (2019)	SVM, NB, J48	Rice, Groundnut, Sugarcane, Cumbu, Ragi	Accuracy, Precision, Sensitivity, Specificity
Bondre and Mahagaonkar (2019)	SVM, RF	Rice, Jowar, Wheat, Soybean, Sunflower, Cotton, Sugarcane, Tobacco, Onion, Drychilli	Precision, Recall, F1-Score
Khosla et al. (2019)	MANN, SVR	Bajra, Maize, Rice, Ragi	Not Available
Leroux et al. (2019)	MLR, RF	Maize	R^2, RMSE, RRMSE, MAE
Maya Gopal and Bhargavi (2019 a, b)	SVR, KNN, RF, MLR, ANN, Hybrid MLR-ANN	Rice	RMSE, MAE, R
Filippi et al. (2019)	RF	Wheat, Barley, Canola	MSE, RMSE
Khaki and Wang (2019)	LASSO, SNN, RT, DNN	Maize	RMSE
Cai et al. (2019)	SVM, RF, NN, LASSO	Wheat	R^2
Shiu and Chung (2019)	OLS, SVR, GWR	Rice	R^2
Elavarasan et al. (2020)	RF, DT, Gradient Boosting	Rice	MSE, MAE, RMSE, MAPE, R^2
Kamir et al. (2020)	CUB, XB, RF, SVMl, MARS, GP, SVMr, KNN, MLP	Wheat	RMSE,R^2
Guo et al. (2021)	MLR, BP, SVM, RF	Rice	R^2, RMSE, MAE
Nyeki et al. (2021)	CP-ANN, XY-Fs, SKNs, SVM, XGBoost, ReLu	Maize	ROC, Sensitivity, Accuracy
Pant et al. (2021)	GBR, RFR, SVR, DTR	Maize, Potatoes, Rice, Wheat	R^2
Singh Boori et al. (2022)	LR, DT, RF	Wheat	R^2, RMSE
Prasad et al. (2021)	RF	Cotton	R^2, RMSE, MAPE, IA
Joshua et al. (2021)	SVM, RBFNN, GRNN, BPNN	Rice	R^2, RMSE, MAE, MSE, MAPE, CV, NSME
Batool et al. (2022)	SVR, AdaBoost Regressor, ARDR, DTR, MLPR, MLR, RANSACR, SLR, XGBoost, SVMR	Tea	MAE, MSE, RMSE
Burdett and Wellen (2022)	MLR, ANN, DT, RF	Corn, Soyabean	R^2, MAE, RMSE
Jeevaganesh et al. (2022)	AdaBoost, RF	Rice, Maize, Blackgram, Lentil, Banana, Mango, Grapes, Apple, Orange, Papaya	Precision, Recall, F1-score
Choudhary et al. (2022)	DT, LR, RF	Rice	RMSE, R^2
Jain and Choudhary (2022)	SVM, RF, NB, LR	Wheat, Maize	Precision, Recall, F1-score, Accuracy
Cedric et al. (2022)	DT, MLR, KNN	Rice, Maize, Cassava, Seed Cotton, Yams, Banana	R^2, MAE
Croci et al. (2022)	KNN, RF,GPR,SVMr, SVMI, NNET, CUB	Maize	R^2, MAE, RMSE, MAPE, Nrmse

Table Cont....

Authors	Machine Learning Algorithms	Crops	Performance Evaluation Metrics
Gopi and Karthikeyan (2022)	MMML-CRYP	Groundnut, Maize, Moong, Rice, Urad	Accuracy, Precision, Recall, Specificity, PR-Score, ROC-Score, F1-Score, MCC
Rahman and Aktar (2022)	LR, PR, SVR	Rice	MSE, MAE, MedAE, R^2
Joshua et al. (2022)	BPNN, SVM, GRNN	Rice	R^2, MAE, RMSE, MAPE, NMSE, CV, ME
Torsoni et al. (2023)	MLR, MLP, SVM, RF, XGBOOSTING, GradBOOSTING	Soyabean	R^2, RMSE, MSE, MAE, MAPE
Wu et al. (2023)	RF, XGBOOST, LSTM	Rice	R^2, RMSE
Iniyan et al. (2023)	MLR, DT, GBR, ENet, Lasso Regression, Ridge Regression, PLS Regression, LSTM	Bajra, Wheat, Jowar	R^2, MAE, RMSE
Islam et al. (2023)	RF, XGBOOST, LightGBM, GradientBoost, LR	Rice	RMSE, MAE
Ed-Daoudi et al. (2023)	DT, RF, NN	Wheat, Apples, Dates, Almonds, Olives	MSE, R^2
Sathya and Gnanasekaran (2023)	MLR-LSTM(Hybrid), LSTM, SVM, RF	Rice	R^2, RMSE, MAE, MSE, F1-Score, Recall, Precision, Accuracy
Wang et al. (2023)	LR, DT, SVM, EL, GPR	Wheat	RMSE, MAE, MSE, R^2

The articles were analyzed to answer Performance evaluation metrics (RQ5), presenting the ML algorithms and various evaluation approaches used in Table 2.

CONCLUSIONS

The objective of this systematic literature review is to identify areas where further research is needed in machine learning, deep learning, hybrid, and optimization for crop yield prediction. It also offers valuable insights into how vegetation indices and environmental factors impact the prediction of crop yield. The studies were conducted on various crops with different temperatures, rainfall levels, and other factors. Ultimately, the machine learning deep learning and hybrid approaches outperform in predicting crop yield with the highest accuracy. The surveyed models are all equally capable of predicting crop yields based on the model's parameters and factors. However, the most effective approaches for crop yield prediction are RF, CatBoost, CNN, and LSTM, with Optimization. As per the survey, the RF algorithm can provide better results and insights depending on the specific parameters and problem at hand. CNN has the capability to identify significant features that could impact the prediction of crop yield. In addition, LSTM identifies the variation pattern of interdependence time-series data. RMSE is the evaluation metric most commonly used in the reviewed articles. It is followed by MAE, R2, MAPE, and MSE. Future investigation will be carried out on implementing crop yield prediction using the proposed RF-CatBoost hybrid model with Horse Herd optimization.

REFERENCES

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H.S. and Radiom, S., 2018. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12), pp.4563-4577.

Alagurajan, M. and Vijayakumaran, C., 2020. ML methods for crop yield prediction and estimation: An exploration. *International Journal of Engineering and Advanced Technology*, 9(3), pp.3506-3508.

Amani, M., Kakooei, M., Moghimi, A., Ghorbanian, A., Ranjgar, B., Mahdavi, S., Davidson, A., Fisette, T., Rollin, P., Brisco, B. and Mohammadzadeh, A., 2020. Application of Google Earth Engine cloud computing platform, Sentinel imagery, and neural networks for crop mapping in Canada. *Remote Sensing*, 12(21), pp.1-18.

Amudha, M. and Brindha, K., 2022. Multi techniques for agricultural image disease classification and detection: A review. *Nature Environment and Pollution Technology*, 21(5), pp.2165-2175.

Ang, Y., Shafri, H.Z.M., Lee, Y.P., Bakar, S.A., Abidin, H., Mohd Junaidi, M.U.U., Hashim, S.J., Che'Ya, N.N., Hassan, M.R., Lim, H.S., Abdullah, R., Yusup, Y., Muhammad, S.A., Teh, S.Y. and Samad, M.N., 2022. Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. *Earth Science Informatics*, 15(4), pp.2349-2367.

Batool, D., Shahbaz, M., Shahzad Asif, H., Shaukat, K., Alam, T.M., Hameed, I.A. and Luo, S., 2022. A hybrid approach to tea crop yield prediction using simulation models and machine learning. *Plants*, 11(15), p.1925.

Bondre, D.A. and Mahagaonkar, S., 2019. Prediction of crop yield and fertilizer recommendation using machine learning algorithms. *International Journal of Engineering Applied Sciences and Technology*, 4(5), pp.371-376.

Burdett, H. and Wellen, C., 2022. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precision Agriculture*, 23(5), pp.1553-1574.

Cai, Y.P., Guan, K.Y., Lobell, D., Potgieter, A.B., Wang, S.W., Peng, J., Xu, T.F., Asseng, S., Zhang, Y.G., You, L.Z. and Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*, 274, pp.144-159.

Cedric, L.S., Adoni, W.Y.H., Aworka, R., Zoueu, J.T., Mutombo, F.K., Krichen, M. and Kimpolo, C.L.M., 2022. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, 2, p.100049.

Choudhary, K., Shi, W., Dong, Y. and Paringer, R., 2022. Random forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. *Advances in Space Research*, 70(8), pp.2443-2457.

Cravero, A., Pardo, S., Sepúlveda, S. and Muñoz, L., 2022. Challenges to use machine learning in agricultural big data: A systematic literature review. *Agronomy*, 12(3), pp. 114-124.

Croci, M., Impollonia, G., Meroni, M. and Amaducci, S., 2022. Dynamic maize yields predictions using machine learning on multi-source data. *Remote Sensing*, 15(1), p. 178.

Deepa, N. and Ganesan, K., 2019. Hybrid rough fuzzy soft classifier based multi-class classification model for agriculture crop selection. *Soft Computing*, 23(21), pp.10793-10809.

Durai, S.K.S. and Shamili, M.D., 2022. Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, 3, p.100041.

Ed-Daoudi, R., Alaoui, A., Ettaki, B. and Zerouaoui, J., 2023. Improving crop yield predictions in Morocco using machine learning algorithms. *Journal of Ecological Engineering*, 24(6), pp.392-400.

Elavarasan, D. and Vincent, P.D., 2020a. Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE Access*, 8, pp.86886-86901.

Elavarasan, D. and Vincent, D.R., 2020b. Reinforced XGBoost machine learning model for sustainable, intelligent agrarian applications. *Journal of Intelligent and Fuzzy Systems*, 39(5), pp.7605-7620.

Elavarasan, D. and Vincent, P.M.D.R., 2021. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *Journal of Ambient Intelligence and Humanized Computing*, 12(11), pp.10009-10022.

Elavarasan, D., Durai Raj Vincent, P.M., Srinivasan, K. and Chang, C.Y., 2020. A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling. *Agriculture (Switzerland)*, 10(9), pp.1-27.

Elavarasan, D., Vincent, P.M.D.R., Sharma, V., Zomaya, A.Y. and Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, 155(August), pp.257-282.

Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D.S.N., Pozza, L.E., Ugbaje, S.U., Jephcott, T.G., Paterson, S.E., Whelan, B.M. and Bishop, T.F.A., 2019. An approach to forecasting grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, 20(5), pp.1015-1029.

Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O. and Ravid, G., 2018. Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. *Precision Agriculture*, 19(3), pp.421-444.

Gopi, P.S.S. and Karthikeyan, M., 2023. Multimodal machine learning-based crop recommendation and yield prediction model. *Intelligent Automation and Soft Computing*, 36(1), pp.313-326.

Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Bryant, C.R. and Senthilnath, J., 2021. Integrated phenology and climate in rice yield prediction using machine learning methods. *Ecological Indicators*, 120, p.106935.

Hazra, S., Karforma, S., Bandyopadhyay, A., Chakraborty, S. and Chakraborty, D., 2023. Prediction of crop yield using machine learning approaches for agricultural data. *TechRxiv. Preprint*, pp.0-10.

Iniyan, S., Varma, V.A. and Naidu, C.T., 2023. Crop yield prediction using machine learning techniques. *Advances in Engineering Software*, 175(August 2022), p.103326.

Ip, R.H.L., Ang, L.M., Seng, K.P., Broster, J.C. and Pratley, J.E., 2018. Big data and machine learning for crop protection. *Computers and Electronics in Agriculture*, 151(November 2017), pp.376-383.

Islam, M.D., Di, L., Qamer, F.M., Shrestha, S., Guo, L., Lin, L., Mayer, T.J. and Phalke, A.R., 2023. Rapid rice yield estimation using integrated remote sensing meteorological data and machine learning. *Remote Sensing*, 15(9).

Jain, K. and Choudhary, N., 2022. Comparative analysis of machine learning techniques for predicting production capability of crop yield. *International Journal of System Assurance Engineering and Management*, 13(s1), pp.583-593.

Jeevaganesh, R., Harish, D. and Priya, B., 2022. A machine learning-based approach for crop yield prediction and fertilizer recommendation. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp.1330-1334.

Joshua, S.V., Priyadharson, A.S.M., Kannadasan, R., Khan, A.A., Lawanont, W., Khan, F.A., Rehman, A.U. and Ali, M.J., 2022. Crop yield prediction using machine learning approaches on a wide spectrum. *Computers, Materials and Continua*, 72(3), pp.5663-5679.

Joshua, V., Priyadharson, S.M. and Kannadasan, R., 2021. Exploration of machine learning approaches for paddy yield prediction in the eastern part of Tamil Nadu. *Agronomy*, 11(10), pp.1-19.

Kalaiarasi, E. and Anbarasi, A., 2022. Multi-parametric multiple kernel deep neural networks for crop yield prediction. *Materials Today: Proceedings*, 62, pp.4635-4642.

Kalaivanan, K. and Vellingiri, J., 2022. Survival study on different water quality prediction methods using machine learning. *Nature Environment and Pollution Technology*, 21(3), pp.1259-1267.

Kamir, E., Waldner, F. and Hochman, Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160(9), pp.124-135.

Kaur Dhaliwal, J., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S. and Mengistu, A., 2022. Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture*, 199(6), p.107107.

Khaki, S. and Archontoulis, S.V., 2020. A CNN-RNN framework for crop yield prediction. *Frontiers in Plant Science*, 10, p.1750.

Khaki, S. and Wang, L., 2019. Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10(6), pp.1-10.

Khosla, E., Dharavath, R. and Priya, R., 2019. Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. *Environment, Development and Sustainability*, 22(6), pp.5687-5708.

Klompenburg, T., Kassahun, A. and Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177(01), p.105709.

Kouadio, L., Deo, R.C., Byrareddy, V., Adamowski, J.F., Mushtaq, S. and Phuong Nguyen, V., 2018. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. *Computers and Electronics in Agriculture*, 155(8), pp.324-338.

Krishna, M.V., Swaroopa, K., Swarna Latha, G. and Yasaswani, V., 2023. Crop yield prediction in India based on mayfly optimization empowered attention-bi-directional long short-term memory (LSTM). *Multimedia Tools and Applications*, 13, p.789.

Kumar Nagothu, S., Anitha, G., Siranthini, B., Anandi, V. and Siva Prasad, P., 2023. Weed detection in agriculture crop using unmanned aerial vehicle and machine learning. *Materials Today: Proceedings*, 16, pp.412-418.

Kumari, P., Rathore, S., Kalamkar, A., Kumari, P., Rathore, S., Kalamkar, P.A. and Kambale, T., 2020. Prediction of crop yield using SVM approach with the facility of E-MART system. *Easychair,* 17, p.318-339.

Leroux, L., Castets, M., Baron, C., Escorihuela, M.-J., Bégué, A. and Lo Seen, D., 2019. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. *European Journal of Agronomy*, 108, pp.11-26.

Liang, J., Li, H., Li, N., Yang, Q. and Li, L., 2023. Analysis and prediction of the impact of socio-economic and meteorological factors on rapeseed yield based on machine learning. *Agronomy*, 13(7), pp.1-17.

Liu, Y., Wang, S., Wang, X., Chen, B., Chen, J., Wang, J. and Zhu, K., 2022. Exploring the superiority of solar-induced chlorophyll fluorescence data in predicting wheat yield using machine learning and deep learning methods. *Computers and Electronics in Agriculture*, 192, p.106612.

Maya Gopal, P.S. and Bhargavi, R., 2019a. A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*, 165(6), p.104968.

Maya Gopal, P.S. and Bhargavi, R., 2019b. Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. *Applied Artificial Intelligence*, 33(7), pp.621-642.

Moussaid, A., El Fkihi, S., Zennayi, Y., Lahlou, O., Kassou, I., Bourzeix, F., El Mansouri, L. and Imani, Y., 2022. Machine learning applied to tree crop yield prediction using field data and satellite imagery: A case study in a citrus orchard. *Informatics*, 9(4), pp.111-121.

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N.H. and Islam, N., 2022. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, 14(9), pp.17-30.

Ndehedehe, C.E., Agutu, N.O. and Okwuashi, O., 2018. Is terrestrial water storage a useful indicator in assessing the impacts of climate variability on crop yield in semi-arid ecosystems? *Ecological Indicators*, 88(January), pp.51-62.

Nevavuori, P., Narra, N. and Lipping, T., 2019. Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*, 163, p.104859.

Nihar, A., Patel, N.R. and Danodia, A., 2022. Machine-learning-based regional yield forecasting for sugarcane crop in Uttar Pradesh, India. *Journal of the Indian Society of Remote Sensing*, 50(8), pp.1519-1530.

Nyéki, A., Kerepesi, C., Daróczy, B., Benczúr, A., Milics, G., Nagy, J., Harsányi, E., Kovács, A.J. and Neményi, M., 2021. Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods. *Precision Agriculture*, 22(5), pp.1397-1415.

Panigrahi, B., Kathala, K.C.R. and Sujatha, M., 2022. A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. *Procedia Computer Science*, 218, pp.2684-2693.

Pant, J., Pant, R.P., Kumar Singh, M., Pratap Singh, D. and Pant, H., 2021. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, 46, pp.10922-10926.

Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C. and Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187(October), p.103016.

Prasad, N.R., Patel, N.R. and Danodia, A., 2021. Crop yield prediction in cotton for regional level using random forest approach. *Spatial Information Research*, 29(2), pp.195-206.

Rahman, T. and Aktar, S., 2022. Machine learning approaches to predict rice yield of Bangladesh. *2022 International Conference on Innovations in Science, Engineering and Technology*, ICISET 2022, February, pp.329-333.

Raut, J. and Mittal, S., 2020. Soil fertility and crop recommendation using machine learning and deep learning techniques: A review. *Turkish Journal of Computer and Mathematics Education*, 11(02), pp.1119-1127.

Reddy, D.J. and Kumar, M.R., 2021. Crop yield prediction using machine learning algorithm. *Proceedings - 5th International Conference on Intelligent Computing and Control Systems*, ICICCS 2021, pp.1466-1470.

Sathya, P. and Gnanasekaran, P., 2023. Paddy yield prediction in Tamilnadu Delta region using MLR-LSTM model. *Applied Artificial Intelligence*, 37(1).

Satpathi, A., Setiya, P., Das, B., Nain, A.S., Jha, P.K., Singh, S. and Singh, S., 2023. Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. *Sustainability*, 15(3), p.2786.

Serra, A. and Tagliaferri, R., 2018. Unsupervised learning: Clustering. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3, pp.350-357.

Shiu, Y.S. and Chuang, Y.C., 2019. Yield estimation of paddy rice based on satellite imagery: Comparison of global and local regression models. *Remote Sensing*, 11(2), pp.1-18.

Singh Boori, M., Choudhary, K., Paringer, R. and Kupriyanov, A., 2023. Machine learning for yield prediction in Fergana Valley, Central Asia. *Journal of the Saudi Society of Agricultural Sciences*, 22(2), pp.107-120.

Su, Y.X., Xu, H. and Yan, L.J., 2017. Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi Journal of Biological Sciences*, 24(3), pp.537-547.

Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H. and Liu, P., 2023. Agricultural product price forecasting methods: A review. *Agriculture*, 13(9), p.1671.

Taherei Ghazvinei, P., Darvishi, H.H., Mosavi, A., Bin Wan Yusof, K., Alizamir, M., Shamshirband, S. and Chau, K.W., 2018. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Engineering Applications of Computational Fluid Mechanics*, 12(1), pp.738-749.

Tian, H., Wang, P., Tansey, K., Han, D., Zhang, J., Zhang, S. and Li, H., 2021a. A deep learning framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, PR China. *International Journal of Applied Earth Observation and Geoinformation*, 102, p.102375.

Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S. and Li, H., 2021b. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agricultural and Forest Meteorology*, 310, p.108629.

Torsoni, G.B., de Oliveira Aparecido, L.E., dos Santos, G.M., Chiquitto, A.G., da Silva Cabral Moraes, J.R. and de Souza Rolim, G., 2023. Soybean yield prediction by machine learning and climate. *Theoretical and Applied Climatology*, 151(3–4), pp.1709-1725.

Tseng, F.H., Cho, H.H. and Wu, H.T., 2019. Applying big data for intelligent agriculture-based crop selection analysis. *IEEE Access*, 7, pp.116965-116974.

Verma, R., 2022. Crop analysis and prediction. *2022 5th Multimedia, Signal Processing and Communication Technologies (IMPACT)*, 44, pp.1-5.

Wang, J., Tian, G., Tao, Y. and Lu, C., 2023. Prediction of Chongqing's grain output based on support vector machine. *Frontiers in Sustainable Food Systems*, 7.

Wang, X., Huang, J., Feng, Q. and Yin, D., 2020. Winter wheat yield prediction at the county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. *Remote Sensing*, 12(11), p.1744.

Wang, Y., Shi, W. and Wen, T., 2023. Prediction of winter wheat yield and dry matter in North China Plain using machine learning algorithms for optimal water and nitrogen application. *Agricultural Water Management*, 277(January), p.108140.

Wu, H., Zhang, J., Zhang, Z., Han, J., Cao, J., Zhang, L., Luo, Y., Mei, Q., Xu, J. and Tao, F., 2023. AsiaRiceYield4km: Seasonal rice yield in Asia from 1995 to 2015. *Earth System Science Data*, 15(2), pp.791-808.

Zhong, H., Li, X., Lobell, D., Ermon, S. and Brandeau, M.L., 2018. Hierarchical modeling of seed variety yields and decision making for future planting plans. *Environment Systems and Decisions*, 38(4), pp.458-470.

ORCID DETAILS OF THE AUTHORS

P. Kumaresan: https://orcid.org/0000-0001-5563-8325