A Systematic Study About The Crop Yield Prediction With Machine Learning Techniques - Review

Shevanthe Sekar[1*], Sathiyamoorthy.E[2]

Abstract

Nowadays Food security and Agriculture were already becoming incredibly prominent issues on a worldwide scale. In addition, the significance of food production increases along with the size of the population. New ways of monitoring and managing agriculture must be introduced to satisfy the Future accuracy of the intended training data from the precious experience of previous generations of common types of problems, including technology. However, Machine Learning in Agriculture helps to enhance Crop Productivity and Quality within the Agricultural Sector. Increasing crop yields has become one of the most frequently discussed issues among farmers in modern agriculture. Due to the growing importance of crop yield prediction, this can be done accurately using a mathematical procedure to reduce repetition or organizing the data based on similarities in yield prediction across countries. An in-depth summary of broadly used components and prediction algorithms is also provided. We evaluate contemporary Machine Learning approaches and compare relevant studies. The strength and weaknesses of Machine Learning algorithms supported the prediction of current and forthcoming agricultural concerns are explored. This study examines yield using Machine Learning and associated techniques. A machine learning-supported agricultural yield prediction architecture was presented based on current studies. This challenges researchers to create an accurate crop yield prediction model with minimum computation.

Keywords: Machine Learning, Agriculture, Crop Yield Prediction, Precision Farming, Artificial intelligence.

[1*,2]School of Information Technology and Engineering, VIT, Vellore 632014, India, Email:
[1](shevanthe.s@vit.ac.in) [2](esathiyamoorthy@vit.ac.in)

***Corresponding Author:** Shevanthe Sekar
*School of Information Technology and Engineering, VIT, Vellore 632014, India, Email: shevanthe.s@vit.ac.in

INTRODUCTION

Realizing that agriculture produces a significant percentage of the global food supply, agriculture is one of the major societal concerns. Many countries still experience hunger due to a lack of food or a shortage of food due to population growth (Elavarasan & Durairaj Vincent, 2020). The effects of a growing population, variable weather patterns, land degradation, and climatic changes require measures to guarantee timely and dependable agricultural development and yield, some techniques are required so that we may meet the stipulation of a burgeoning population. In addition, it's essential to coordinate with improving the deep-routed viability of agricultural food production (Huang et al., 2019). These factors imply that land appraisal, crop management, and predicting crop yields are becoming increasingly important to the production of food on a global scale(Li et al., 2013). The nation's legislators must reckon on an error-free crop production projection to acquire practical transfer evaluations to boost worldwide food security. The procedure of estimating crop output based on characteristics including weather, soil, and seed quality is known as crop yield prediction. Planning their crops for the best yields is made easier for farmers. A crop yield prediction model's accuracy is based on how well it anticipates future rainfall patterns in the area where it was created. The algorithm predicts future rainfall patterns using historical information collected in previous years and then determines which crops will be best for that region at that time. It also considers additional elements that may have an impact on a crop's productivity over its lifetime, such as soil nutrients, land slope, etc. Farmers used to forecast crops based on their past experiences and well-grounded historical data and then implement strategic cultivation decisions on their yield prediction. Meanwhile, over the last few years, the development of fresh technologies, such as crop model simulation(CSM) and machine learning(ML), as well as the ability to estimate a tremendous quantity of data using revved-up computing, have proven to predict yield more precisely. Currently, several studies show that using machine learning algorithms has a larger potential than using standard statistics. In the artificial intelligence, sector referred to as machine learning, computers can be learned without specific coding. By assuring a significant prediction capability, such strategies beat agricultural systems, which can be neither non-linear nor linear. The techniques are obtained from the Machine Learning(ML) agricultural system's learning methodology. This flow requires performing a precise job while being given training material. After finishing the training phase, the model recommends conducting tests to evaluate the acquired information (Rashid et al., 2021). Consequently, the adherent was a synopsis of this paper's main contributions:

1. Describing Machine Learning in agriculture sectors.
2. Outlining the output of crop yield in India at the moment.
3. Outlining the key components of the procedure for predicting crop yield.
4. Critical evaluation of the feature sets that have been used, comparative analysis of similar research, and a detailed critical review of CYP based on ML techniques.
5. An exhaustive analysis of the qualities and failings of using ML methods to predict crop yield with their characteristics.

MACHINE LEARNING(ML)

ML is an area that assists in understanding the structure of the computational process and styles that use data to boost performance on sets of tasks. Specifically, ML is a field that helps to understand how neural networks work.ML is also part of Artificial Intelligence(AI). The main grail of ML is to classify data based on models which have been developed and to make predictions for future issues. Algorithms used in machine learning construct a model using examples or "training data" from which the machine learns to make inferences or take actions deliberately specific to indoctrinate to do so. Machine Learning(ML) algorithms were utilized in many contexts, including those where it would be impractical or inappropriate to create custom algorithms for tackling difficult problems, such as in the pharmaceutical, dispatch filtering, speech recognition, and computer vision sectors. While some machine learning techniques are statistical learning, others are more closely tied to machine learning and data mining, which also emphasizes the use of computers to make predictions. The study of delicate optimization contributes to machine learning with new methodologies, concepts, and application domains. Exploratory data analysis using unsupervised learning is at the heart of data mining, another closely related field of research. Some applications of machine learning take data and neural networks and model their behavior after how the human brain processes information. Machine learning is also applied to predictive analytics in its application across marketing riddles. Machine learning algorithms (ML) are divided into training and testing segments.

- **Training** is about collecting random training data, also which learns a model of correlation between average physical characteristics (identity, shape, size, etc) and their quality.
- **Testing** is about measuring test data characteristics and may internally use rules similar to those manually written earlier in training data.

In the past few years, machine learning has become more popular, and there are now a lot of algorithms to choose from. Even though there are many different models, they can all be broken down into three basic parts. Every machine learning algorithm has three components(Fig 1) in it. Representation, evaluation, and optimization are directly connected to supervised and unsupervised learning.

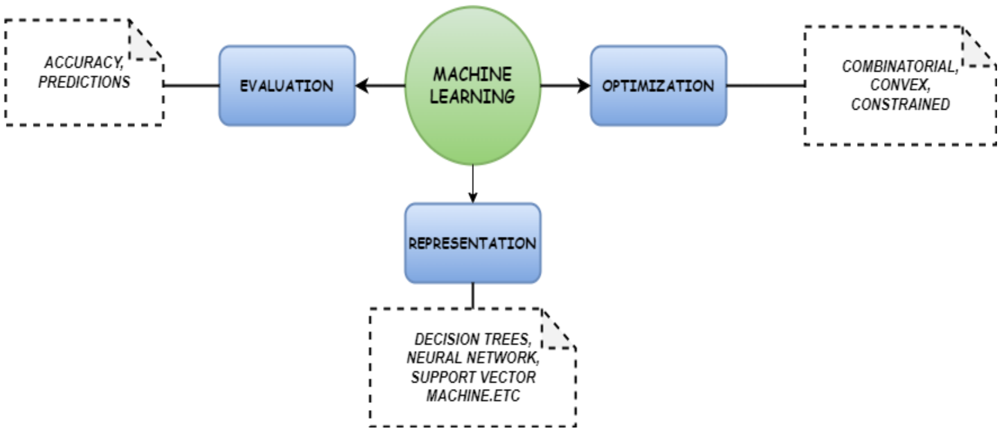


Fig 1: components *of machine learning*

Representation -This indicates how you should examine data. When using k-nearest neighbors or a graph, might wish to think about your data as a collection of specific individuals at times (like in Bayesian networks).

Evaluation - This must assess the learner's performance for supervised learning purposes for it to advance. Utilizing an evaluation function, this evaluation is performed (also known as an objective function or scoring function). Accuracy and squared error are two examples.

Optimization -The learner with the highest score from this evaluation function must be located utilizing the evaluation function from above using a variety of optimization techniques. Examples include gradient descent and greedy searches.

TYPES OF MACHINE LEARNING

A computer will process input and output data to generate an application. To illustrate, suppose that a machine learning method represents a seed, that data represents nutrients, that the garden represents programming, and that the plant represents a program. There are distinct kinds of machine learning which have been specified in Fig2 Namely,

1. *Supervised learning*
2. *Unsupervised learning*
3. *Semi-supervised learning*
4. *Reinforcement*

Supervised Learning focuses on specific tasks. It was educated on labeled data. The model is trained by making predictions and being given information about how well it did. Models are trained until they "converge" on the training data. The "Logistic Regression Algorithm" and the "Back Propagation Neural Network" are both algorithms that can be used in the Regression and Classification challenges. Unsupervised learning refers to the process of learning from data that has not been labeled. In order to get a model ready to go, it's necessary to infer underlying results are generalized standards. Mathematical procedures that look for patterns in data and eliminate duplicates are useful. The "Apriori algorithm" and the "K-Means clustering method" are examples of solutions to the problems of clustering, dimensionality reduction, and learning association rules. For semi-supervised learning, both labeled and unlabeled data are used. Classification and cluster analysis are two types of problems that need to be solved. Modeling unlabeled data with the help of examples requires flexible methodologies, such as example algorithms. Generate adversarial networks" and "self-trained naive Bayes classifier" are two examples of learning algorithms. Reinforcement learning is learning based on rewards. It employs feedback. Doing what has to be done to maximize gain. With reinforcement learning, rational choice is emphasized. It is the output of the current input that determines the value of the subsequent input. Examples include Q-learning and MDP.

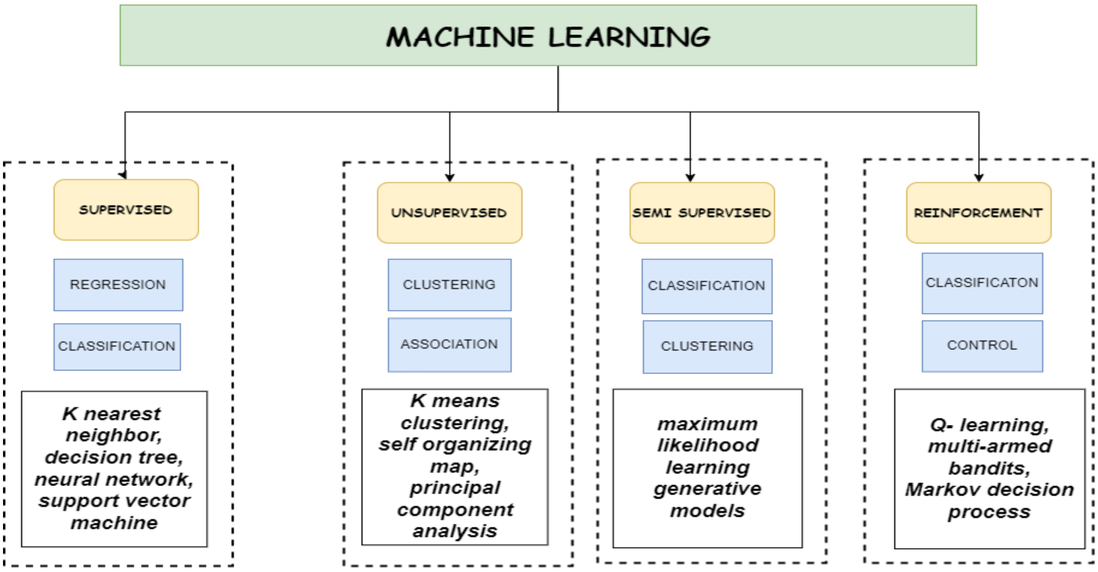


***Fig 2:** Different aromas of ML*

MACHINE LEARNING IN AGRICULTURE

Nowadays, the agricultural industry benefits from the valuable experience of previous-generation farmers, achieving modern science and technology. As a result, agricultural supply and quality are boosted. Retailers were the primary users of machine learning in agriculture, but AI also helped increase crop yields and find pest control specialists. Moreover, it seeks to improve field management by coordinating agricultural practices with crop needs (fertilizer inputs), lowering environmental hazards and the agricultural footprint (by limiting nitrogen leaching), and increasing competitiveness via more efficient procedures. Machine learning has many practical applications in agriculture. It helps farmers keep track of their crops, make more informed choices, promote greater traceability, boost product visibility and sales, strengthen relationships with landlords and tenants, and enhance the quality of their crops themselves. The advantages of modern agriculture include sustainability, maximum productivity, and a secure environment. Moreover, there are likely to be some obstacles

a) Global Population Explosion

b) *Increasing Food Demand*

c) Climate Changes

d) Natural Resource Depletion

e) Alteration Of Dietary Choices

f) *Safety And Health Concerns*

AGRICULTURAL APPLICATION OF MACHINE LEARNING

Key uses of ML in agriculture include crop, water, soil, and animal management. Crop management encompasses the study of a wide range of topics, including yield prediction, disease detection, weed identification, crop recognition, and crop quality. Water management is concerned with the most efficient use of water supplies. Soil management is the thread that ties together soil preservation and management practices. Cattle management includes both the care of animals and the raising of livestock for human consumption. Soil and crop management are the most complex uses. Crop management is a broad concept that incorporates a wide range of features that have arisen through the combination of arms to exert influence over the crop environment's biological, chemical, and physical aspects and meet both quantity and quality objectives. Predicting crop yields is a challenging issue in agriculture. With the help of a reliable model, farmers may make well-informed management decisions about crop rotation and plants for maximum profit in the current market. Several factors are involved, including the setting, management strategies, genetic and phenotypic characteristics of the crop, and the interplay between these factors. need robust ML models and lots of information. Disease detection poses a threat to the agricultural production system. in terms of quality and quantity, at the stages of manufacturing, storage, and transit. cause a major problem for the food supply. Insect and disease detection in a timely manner is crucial. Field scouting, traditionally used by specialist agronomists for disease identification, is time-consuming and solely relied on visual inspection. Due to their ability to produce large numbers of seeds and survive for extended periods of time, weeds can quickly take over large parts of a field, choking out desirable crops and threatening the sustainability of the entire agricultural ecosystem. Most crop identification efforts focus on analyzing leaf characteristics such as color, shape, and texture. Agricultural classification using remote sensing is now commonplace due to the widespread usage of satellites and aviation for sensing crop properties. The automatic recognition and classification of crops is the outcome of the combination of recent developments in computer software and image processing technologies with machine learning. It matters so much that our crops are of high quality for the marketplace. It delivers a premium service that helps farmers earn more money. Better management decisions lead to higher quality crop yields, hence it's in the farmers' best interest to develop a decision support system. Sophisticated machine learning methods can be applied as "selected harvesting" to considerably boost the quality form of "selected harvesting" to boost quality considerably through managed processes.

CROP YIELD PREDICTION

A crop's yield is a standard measurement of agricultural production per unit of land required to grow it. Cereal, grain, and legume yields are usually quantified in bushels, tons, or pounds per acre. It measures grain production and land efficiency. Farm automation, crop genetics, fertilizers, and pesticides have improved crop yields and farm efficiency.(van Klompenburg et al., 2020) National and regional decision-makers need crop yield projections for quick decisions. Before assessing a crop's yield, farmers count how much has been harvested from a certain region. After that, the harvested crop is weighed, and the field yield is extrapolated from that sample. Pre-harvest agricultural yield estimation is crucial since crop yields affect global business, food supplies,

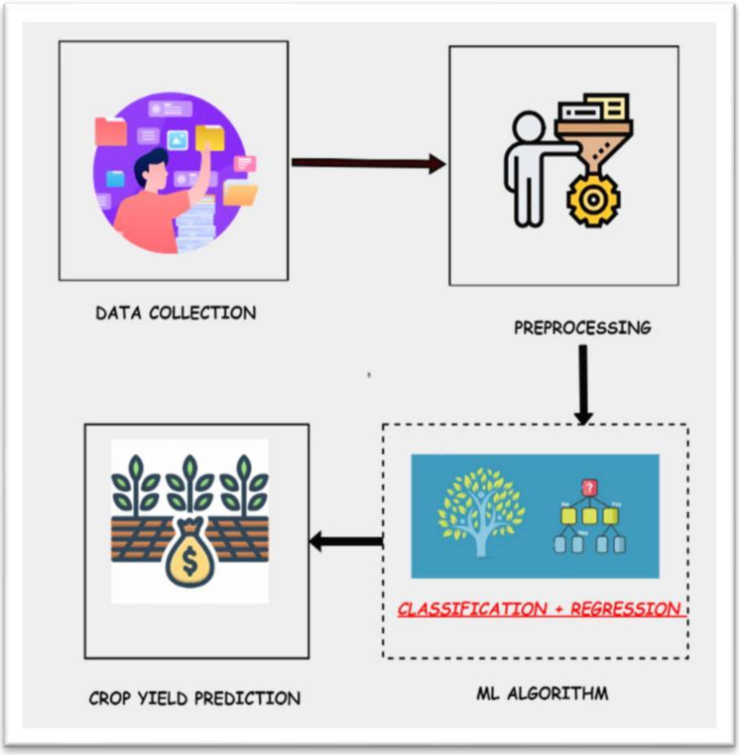


Fig 3: *Basic architecture of crop yield prediction*

and market pricing. In addition, early prediction of yield is valuable information for policymakers. Efficient land use and economic policy planning require good forecasts of crop productivity. Recently, there has been an increase in crop productivity forecasting at the field level. The most important factor affecting crop productivity is weather conditions. More accurate weather forecasts can warn farmers in time, reduce major damage, and help grow the economy. The forecast also helps farmers make decisions such as choosing alternative crops or discarding crops early in crises. Additionally, predicting yields gives farmers a better idea of how to grow and plan their seasonal crops. Therefore, it is necessary to simulate and predict yields before cultivation to achieve efficient crop management and expected results. Machine learning techniques may be efficient for yield prediction because of the non-linear relationship between crop yield and factors affecting harvest. Fig 3 describes the basic architecture for Crop yield prediction.

ANALYTICAL SURVEY ON PROJECTING CROP YIELD PREDICTION

ML is crucial to the agricultural sector. We have searched multiple reputable databases, including IEEE Explore, Science Direct, SpringerLink, and Taylor & Francis, to prepare a comprehensive literature evaluation of the issues included in Table 1. We've been perusing a selection of available online sites in our quest for reliable information. The scope of this assessment was determined by the terms "machine learning" and "crop yield prediction," both of which will be defined in greater detail below. Therefore, we have conducted the vast majority of our searches using these two phrases in conjunction with others, such as "yield prediction" OR "yield estimation," etc. Documents were collected from a wide variety of sources, as shown in Fig 4. Documents such as articles, journals, reviews, books, conference papers, presentation notes, and reports have been amassed. The results of a study tracking the publication of various article kinds in different libraries pertaining to a narrow portion of the canon are presented in Table 1. We've included a graph that shows the annual growth in the number of articles added to digital archives. This bar chart can be found beneath the table. These studies' results lead us to conclude that machine learning has been undergoing significant development in the agricultural sector in recent years.

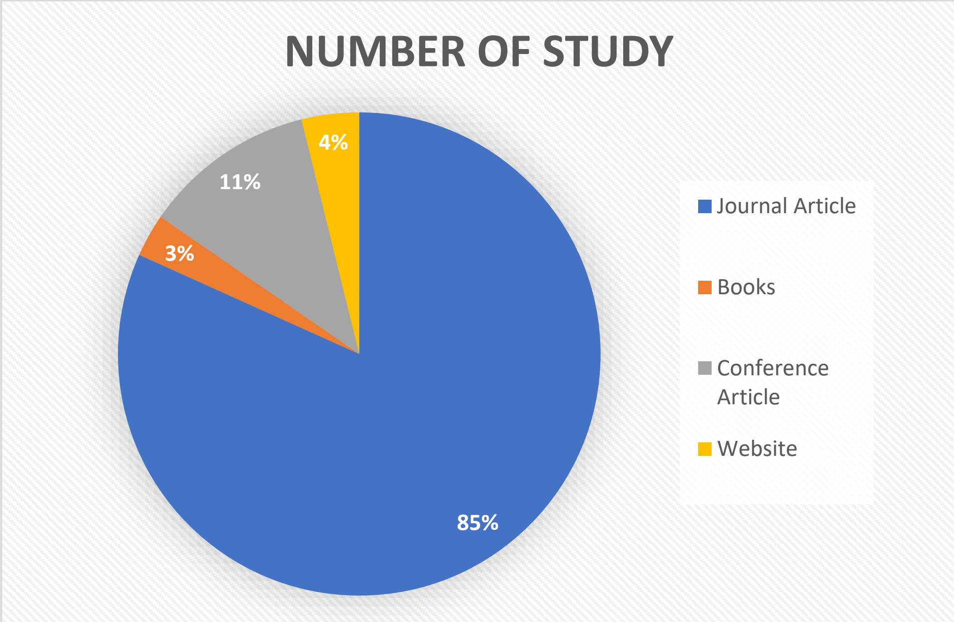


Fig 4: *Percentage of Various Publications (including Research Works)*

Table1: Tracking the outcomes of string searches performed in various digital libraries

S.NO	Digital library explored	URL	Track	search string
1	ScienceDirect (Elsevier)	https://www.sciencedirect.com	6134	"Crop yield prediction"
2	Taylor & Francis	https://www.tandfonline.com	1043	and "machine learning"
3	Springer	https://www.springer.com	1510	OR "yield prediction"
4	IEEE	https://ieeexplore.ieee.org	941	OR "yield estimation"
5	Nature publication	https://www.nature.com	1110	

Table 2: Track Article Types Across Multiple Libraries Using a Selected String.

Publications Type	ScienceDirect (Elsevier)	Taylor& Francis	Springer	IEEE	Nature Publication
Research Articles	4749	869	1258	15	1110
Conferences	40	63	42	492	NA
Book Chapter	414	3	58	140	NA
Review Articles	682	45	152	312	NA

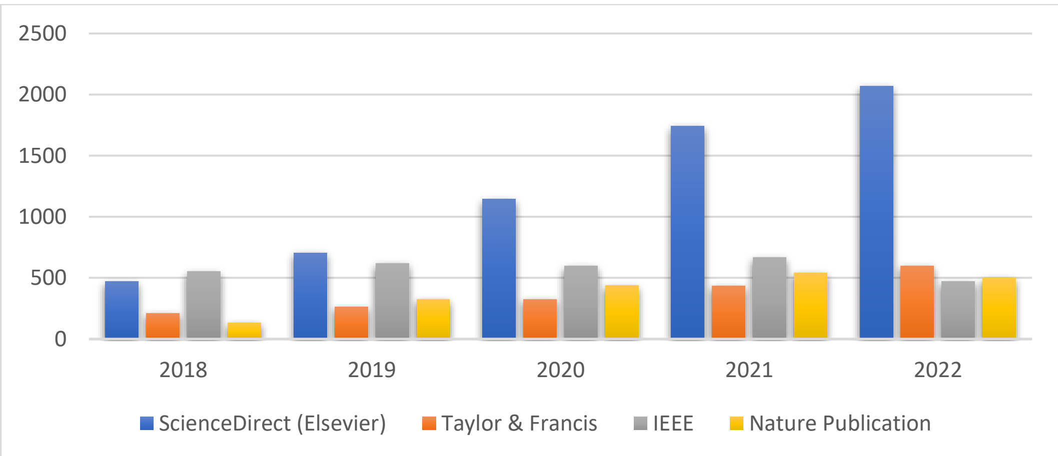


Fig 5: *Analysis by year of the articles published in digital libraries*

FUNDAMENTAL ASPECTS OF CYP

The steps used in the process of predicting agricultural yield using machine learning are shown in Fig 3. These steps include gathering data, cleaning it, dividing it, and analyzing it. A wide variety of information, including soil data (Shahhosseini et al., 2021), weather data (Deines et al., 2021), yield data (Cho et al., 2021), cropland data (Paudel et al., 2022), etc., can be collected and analyzed during the data collection(Babaie Sarijaloo et al., 2021) stages. The second step is eliminating the null values and standardizing the missing information. The data analysis step, when algorithms are used to predict crop yield, may deal with values after they have been segmented.

POPULAR FEATURES

Crop yield and yield cultivation uncertainty are influenced by an assortment of external factors. For accurate yield forecasting, the attributes are crucial. Soil factors, meteorological data, historical yield data, phenology data, cropland data, fertilization information, vegetation indices, and satellite data have played a significant role in agricultural yield prediction in recent years. In Table 3, we saw a breakdown of the most widely-used features for estimating crop yields, along with the information they provide. Increases in crop yield are essential, thus any method that can better predict it is worth pursuing. Organic matter, pH values, CN ratios, soil types, and the availability of potassium, zinc, nitrogen, and manganese are among the most often used "Soil Parameters" and "Weather Information" datasets (Biswas & Naher, 2018), as are precipitation, humidity, temperature, and wind speed (Varinderpal-Singh et al., 2022). Nutrients might be either those that occur naturally in the soil or those that are supplied to it. The degree to which saturation occurs depends on these features. Nitrogen, magnesium, potassium, sulfur, zinc, boron, calcium, phosphorus, and manganese are the nutrients that were assessed. Cropland management allows for the making of land-related decisions. Farmers that have to make changes to their fields utilized categorized These systems are in charge of watering crops and plants. in addition to fertilization, hence "Cropland Management" could also mean "Nutrient Management" (Gautron et al., 2022). The data on the sun is comprehensive and includes a wealth of facts. refers to heat or light. The "Historical Yield Data" is a compilation of information on crop production over the years, including yields, the permeability of rock layers(Felipe Maldaner et al., 2021), and the availability of various micronutrients. Dates of seed sowing(Alabi et al., 2022), transplanting, tillering, boosting, turning green, heading, and flowering is all part of the "Phenology data" feature category. Features such as wind speed and pressure data and images fall under the category of those that are utilized less frequently. The features that have been computed have included MODIS Enhanced Vegetation Index (MODIS-EVI), Normalized Vegetation Index (NDVI), and Enhanced Vegetation Index(EVI) (Paudel et al., 2022).

Table 3: list of prominent features with data used in cyp

Features	**Data**
Soil data	Soil organic matter, PH value, CN ratio, manganese, copper, potassium, zinc, available nitrogen, soil type (Ansarifar et al., 2021)(Shahhosseini et al., 2021)(Elavarasan & Durairaj Vincent, 2020) (Tesfaye et al., 2021) (Jha et al., 2022)(Biswas & Naher, 2018)(Diaz-Gonzalez et al., 2022) (Choudhury et al., 2021) (Jiru & Wegari, 2022)(Dubois et al., 2021; Khanal et al., 2021)
Weather information	Rainfall, humidity, season type, temperature, wind (Ansarifar et al., 2021) (Shahhosseini et al., 2021) (Bali & Singla, 2021) (Deines et al., 2021) (Beyene et al., 2021)(Elavarasan & Durairaj Vincent, 2020) (Ramu & Sri, 2021) (Seireg et al., 2022) (Jha et al., 2022) (Meroni et al., 2021) (Al-Gunaid et al., 2021) (W. Zhou et al., 2022) (Deines et al., 2021)(Amaratunga et al., 2020)
Historical yield data	Old data, number of micronutrients, transmissivity, rock layer permeability (Ansarifar et al., 2021) (Shahhosseini et al., 2021)(Bali & Singla, 2021)(Deines et al., 2021)(Beyene et al., 2021) (Elavarasan & Durairaj Vincent, 2020) (Ramu & Sri, 2021) (Seireg et al., 2022)(Jha et al., 2022) (Rashid et al., 2021) (W. Zhou et al., 2022) (Jiru & Wegari, 2022) (Felipe Maldaner et al., 2021) (Khanal et al., 2021) (Juventia et al., 2021)

Phenology data	Dates of sowing, transplanting, tiller, boosting, turning green, heading, flowering (Shahhosseini et al., 2021) (Beyene et al., 2021) (Alabi et al., 2022) (Tufail et al., 2021)
Cropland data	Plantation area, cropland census, texture condition (Shahhosseini et al., 2021)(Bali & Singla, 2021) (Khaki et al., 2021) (Deines et al., 2021)(Beyene et al., 2021) (Elavarasan & Durairaj Vincent, 2020) (Ramu & Sri, 2021) (Jafarbiglu & Pourreza, 2022) (Qiao et al., 2021) (Huang et al., 2019) (Meshram et al., 2021) (Gautron et al., 2022) (Tufail et al., 2021) (Juventia et al., 2021)
Fertilizer information	Potassium, nitrogen, phosphorus, sulfur, micronutrients, macronutrients, consumption of agricultural pesticide, consumption of chemical fertilizer (Elavarasan & Durairaj Vincent, 2020) (Tesfaye et al., 2021) (Seireg et al., 2022) (Meshram et al., 2021)
Vegetation indices	EVI, EVI2, GCI, GNDVI, NDVI, SAVI, WDRVI, MODIS, RDVI (Paudel et al., 2022) (Alebele et al., 2021) (Jafarbiglu & Pourreza, 2022) (Myers et al., 2021) (Alabi et al., 2022) (Debalke & Abebe, 2022) (Ji et al., 2021) (Choudhury et al., 2021) (Priya & Ramesh, 2020)
Satellite data	visible, infrared, multispectral, hyperspectral, SENTINEL1A, SENTINAL 1B, LANDSAT (Khaki et al., 2021) (Deines et al., 2021) (Paudel et al., 2022) (Tesfaye et al., 2021) (Myers et al., 2021) (Qiao et al., 2021) (Geipel et al., 2021) (W. Zhou et al., 2022) (Debalke & Abebe, 2022) (X. Zhou et al., 2021) (Choudhury et al., 2021) (Khanal et al., 2021) (Priya & Ramesh, 2020)

AN ALGORITHM FOR PREDICTION

To better predict agricultural output, many different prediction algorithms based on classification and regression have been implemented. Methods like Linear Regression (LR) and Multiple Linear Regression(MLR), Multivariate Adaptive Regression Splines (MARS),Nearest Neighbors(NN), Inverse Distance Weighting (IDW),K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) and Support Vector Regression(SVR), Decision Tree (DT) and Random Forest(RF), Artificial Neural Networks (ANN) and Deep Neural Networks (DNN), LASSO, Light GBM, and XG boost have all been used to predict crop yields. Long-Short Term Memory (LSTM), Principal Component Analysis (PCA), Generalized Linear Regression (GLR), and Convolutional Neural Networks (CNNs).

ML ALGORITHM

Nowadays the main use of Machine Learning algorithms are advantageous because they prompt consideration of how input data and the model-building procedure affect the outcome. This paves the way for picking the approach that works best for solving a certain problem, leading to a more reliable result. When examining data to model possible business decisions, one typically employs both supervised and unsupervised learning methodologies. In areas like picture classification, where there are large datasets but few cases with labels, the use of semi-supervised learning approaches is receiving a lot of interest. Regression is a statistical method used to model the connections between various variables. The high accuracy of the model's forecasts is monitored and refined in an iterative process. Regression methods are widely used in statistical arithmetic, and statistical machine learning has taken these methods and run with them. Because "regression" can mean both a type of problem and a type of method, it has the potential to confuse if used interchangeably. Linear Regression (LR) (Shahhosseini et al., 2021) (Meroni et al., 2021) (Khanal et al., 2021) (Paudel et al., 2021) Logistic Regression (Shahhosseini et al., 2021) (Meroni et al., 2021) (Khanal et al., 2021) (Paudel et al., 2021), Splines based on multivariate adaptive regression (MARS) (Paudel et al., 2021), and Smoothing of the Scatterplot Based on Local Estimates (Paudel et al., 2021)are the most widely used regression methods (LOESS). An instance-based learning model solves a problem by focusing on the training data instances or examples that the model deems most important or relevant. In order to make a prediction, these techniques typically begin with gathering a set of sample data, which is then compared to the incoming data using a similarity measure to determine the best match.

Thus, it is not surprising that instance-based approaches are also referred to as memory-based learning approaches and first-to-complete approaches. Instance storage representation and similarity measures for comparison are the primary areas of interest. The most popular instance-based algorithms are (SVM)Support Vector Machines (Alabi et al., 2022) (Qiao et al., 2021) (W. Zhou et al., 2022) (Kok et al., 2021) (Khanal et al., 2021) (Tufail et al., 2021), K-Nearest Neighbour (KNN)(Shahhosseini et al., 2021) (Khaki et al., 2021) (Deines et al., 2021) (Ramu & Sri, 2021) (Tesfaye et al., 2021) (Meroni et al., 2021) (W. Zhou et al., 2022) (X. Zhou et al., 2021) (Khanal et al., 2021)and [RF] Random Forest. Using the data's actual attribute values, decision tree methods construct a model of potential actions to take. Some of the most common (Qiao et al., 2021) decision tree techniques include Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0 (variations on a strong method), and (CHAID) Chi-squared Automatic Interaction Detection.

DL ALGORITHM

Deep learning techniques are an up-to-date version of Artificial Neural Networks (ANN)(X. Zhou et al., 2021) (Khanal et al., 2021) (Tiwari & Shukla, 2020) (Amaratunga et al., 2020)that take advantage of plentiful and economical computation. They are focused on producing neural networks that are considerably larger and more complicated than those previously constructed. In addition, as was mentioned early, Many of these strategies focus on massive unstructured or semi-structured collections of analogy signals (e.g., images, texts, audio, and video). A selection of the most widely used deep learning algorithms is provided below(Khaki et al., 2021). Long Short-Term Memory Networks (LSTMs), Convolutional Neural Networks(CNN), and Recurrent Neural Networks(RNN)(Cho et al., 2021)(Cho et al., 2021) (Qiao et al., 2021).

OTHER MISCELLANEOUS ALGORITHM

Other classification and regression algorithms, such as XGBoost (Shahhosseini et al., 2021), MARS, Gradient boosting, (Alabi et al., 2022)ridge regression (RIDGE), SNN, DRL, and a large range of others, have also been used in agricultural production prediction with a great deal of success. These algorithms include a wide number of different techniques. A few examples of these algorithms are hybrid approaches like CNN-RNN (Qiao et al., 2021), CNN-LSTM, RNN-LSTM (Bali & Singla, 2021)and MLR-ANN have been used in the process of agricultural yield prediction.

Table 4: list of popular algorithms used for crop yield prediction

Algorithms	Key features	Advantages	Disadvantages
ANN	The human brain inspired both its name and its structure, which replicates the way organic neurons communicate with one another.	highlight associations and structures in the data that are not visible;	Big computation burden,
SVM	The SVM algorithm classifies data points by finding an N-dimensional hyperplane.	Extremely precise	In the context of both training and testing machine algorithms, there is a greater demand for fast processing and compact storage..
DT	It has a node, branches, and leaf nodes. In decision analysis, a decision tree represents decisions visually and unambiguously.	Can lead to an interpretable model	Not working well on small training datasets, overfitting problem.

REGRESSION	Regression examines the link between independent factors and a dependent variable. It is used in machine learning to predict continuous outcomes using an algorithm.	Works on any size of the dataset and nonlinear problems	Choosing right for good bias/ variance trade-off.
RF	In order to produce a forest of trees whose collective forecast is more accurate than that of any individual tree, it employs bagging and feature randomization.	Powerful, accurate, great performance on many nonlinear problems	No interpretability
DNN	Due to the flexibility of the network's input layer, DNNs can simply include query and item features to capture a user's interests and improve recommendations.	Self-learning capabilities	Massive Data Requirement, High Processing Power, Black Box Problem
CNN	CNN is used for image and speech recognition. Its convolutional layer decreases image dimension without losing data.	Reduces the need for human power	Hard time classifying images with different positions
LSTM	Long-term dependency-learning RNN. LSTMs avoid long-term reliance. Long-term memory is not something they struggle to learn.	longer-lasting special units	LSTMs demand more memory and are sensitive to random weight initializations. longer
LASSO	LASSO improves OLS and Ridge regression. LASSO shrinks and selects variables to improve prediction and model interpretation.	It creates easier-to-understand models with fewer predictors.	LASSO saturates after selecting n variables. LASSO cannot choose groups.
ENSEMBLE	Using alternative modeling algorithms or training data sets, numerous models are developed to predict an outcome.	More precise prediction improves stability	Need more resources and time, non-interpretable

EFFICIENCY

Measures of performance were applied to evaluate the accuracy of various learning models' predictions of crop yields. (Khaki et al., 2021) The effectiveness of the regression model is measured in terms of the mean absolute error (MAE), mean squared error (MSE), root means square error (RMSE), (Ramu & Sri, 2021) (Alabi et al., 2022) determination coefficient (R-squared), and mean absolute percentage error (MAPE). To calculate the average importance of errors, the mathematical mean of the absolute difference between the predicted and actual observations is used (Bali & Singla, 2021). The MSE evaluates how close the regressor line is to the points in the dataset. Estimating residuals with RMSE (Shahhosseini et al., 2021) (Babaie Sarijaloo et al., 2021)(Tesfaye et al., 2021) measures how well data is concentrated on the best possible t-line. Using the determination coefficient, researchers can evaluate how much more effective the new framework is in comparison to the original. The Mean Absolute Percentage Error(MAPE) (Qiao et al., 2021)quantifies how off-base a model's predictions are. Machine learning-based agricultural yield prediction algorithms are evaluated on a variety of metrics, including their accuracy (Jha et al., 2022), precision, recall, sensitivity, specificity, and F1 Score. The success and popularity of a metric are often measured by how well it can classify data.

EVALUATION AND CRITIQUE OF RECENT STUDIES INVOLVED IN CYP RICE YIELD PREDICTION

For fighting world poverty, rice was the most useful staple food. It's one of the most important crops because it's used to nourish the diets of half the world's population. To feed the world's expanding population, rice cultivation has expanded dramatically over the globe. It is difficult to exaggerate the significance of rice to the global community. More than half of the world's population relies on this crop for mere survival. Many people in Asia, Latin America, Africa, and the Caribbean rely heavily on rice as a staple food because it is expected to provide more than a fifth of the calories that humans consume worldwide. Rice is used as a primary food source in these regions. Predicting rice crop yield (Elavarasan & Durairaj Vincent, 2020) in a timely manner prior to harvest is important for a number of reasons, including, but not limited to, deciding on rice, developing rice, predicting grain market prices, and ensuring food security. (Alebele et al., 2021)Helping farmers make educated management decisions, accurate and precise rice yield mapping can also reveal areas linked to low grain output. By collecting data throughout the growing season, an experienced agronomic can calculate an estimated yield using agrometeorological models. Even though it takes time and effort, estimations based on agrometeorological models can be quite difficult to reconcile with field measurements.(Alebele et al., 2021)Since SAR imaging is readily available in all weather situations and may supplement optical data in various climates, it provides a fantastic opportunity for remote rice yield prediction, Satellite data rich in structural detail can be combined with hyperspectral data on the required produced target to achieve high precision and efficiency. From 2015 to 2020, India's rice output was depicted in Fig 6. FAOSTAT is a global data repository from which we gathered this information it’s listed the production of rice in India.

Table 5: Some of the articles analyzed for the analytical survey.

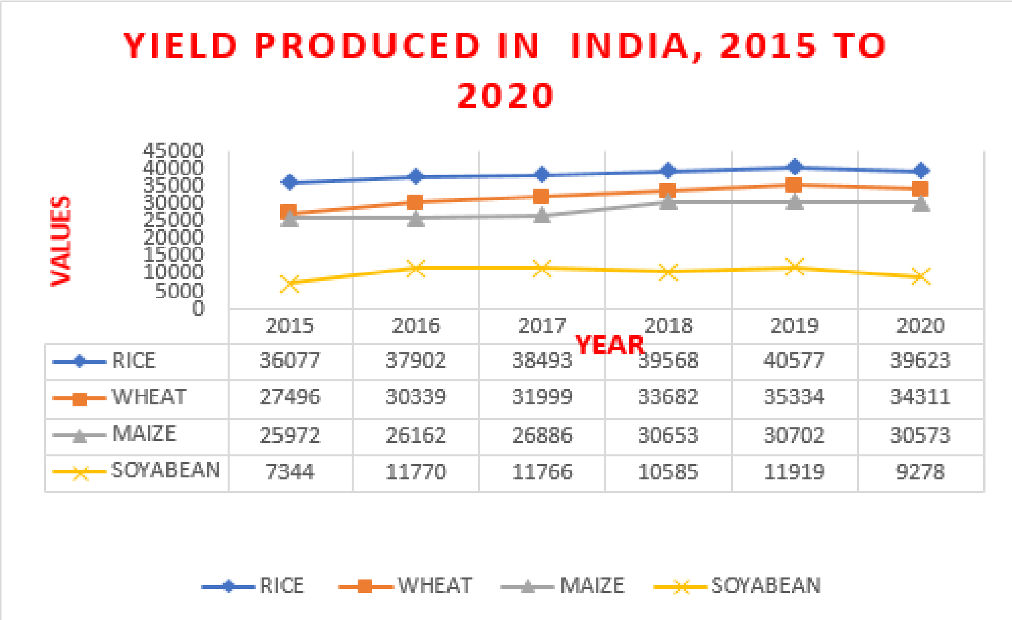
References	Objectives	Features	Algorithms	Performance
(Ansarifar et al., 2021)	Corn and soybean yield prediction	Soil parameters, weather parameters, historical data.	Multiple linear regression model	Relative root means square error (RRMSE)-8%
(Shahhosseini et al., 2021)	Maize yield prediction	Soil properties, Metrological data, management data, phenology data	Linear regression, LASSO, light GBM, random forest, XG boost.	Estimates of error: RMSE – (7 – 20%),RRMSE,MBE, Variance explained by models: R2(Coefficient of determination)
(Bali & Singla, 2021)	Wheat yield prediction	Climatic factors, historical data, cropland features.	RNN, LSTM, ANN, RF, MLR.	RMSE and MAE. RNN-LSTM: RMSE-147.12, MAE-60.50, ANN: RMSE-730.14, MAE-623.13, RF: RMSE-540.88, MAE-449.36, MVR: RMSE-915.64, MAE-796.07.
(Khaki et al., 2021)	Soyabean and corn yield prediction	Historical data, Yield efficiency, satellite images, agricultural production data.	Random forest, Deep feed-forward neural network (DFNN),3Dimentional-convolutional neural network(3D-CNN), regression tree, LASSO, ridge	Mean absolute error (MAE) – 8.74% & 8.70%

(Deines et al., 2021)	Maize and corn yield mapping and yield gap analysis	Yield monitor data, satellite data, weather, and crop data.	Random forest	R2, RMSE, MAE
(Cho et al., 2021)	Corn and grain yield monitoring	Historical data, yield monitor data	Spatial estimation: Nearest neighbors (NN), Inverse distance weighting (IDW), Kriging with exponential	Normalized root means square error (NRMSE)
(Beyene et al., 2021)	Wheat yield prediction	Phenological data, soil weather & crop data, and leaf area index data.	[ensemble Kalman filter (ENKF), world food studies (WOFOST) – assessed by linear regression], leaf area index (LAI)	R2, RMSE. **WOFOST:** R2-0.80, RMSE-413 **LAI:** R2-0.58, RMSE-592
(Elavarasan & Durairaj Vincent, 2020)	Rice yield prediction	Data on fertilization, groundwater properties, crop management, croplands, harvests, irrigation, soil and climate, and historical data	Deep reinforcement learning	Accuracy – 93.7%
(Paudel et al., 2022)	Multicrop yield prediction (including (but not limited to) spelt, spring barley, sunflower, grain corn, sugar beets, and potatoes	Wofost crop model data, meteo data, remote sensing, crop area, irrigated area, soil data, cropland management	KNN, SVR, GBDT Ensemble Learning Method	NRMSE - 0.95
(Alebele et al., 2021)	Rice yield estimation	Yield data, SAR imagery data	Gaussian kernel regression, Bayesian linear inference, RDVI	R2 - 0.81 ,RMSE - 0.55
(Ramu & Sri, 2021)	Wheat yield prediction	Temperature data, wheat production data	Artificial Neural Network, Random Forest	R2 - ANN: 0.9999, RF:11.1782
(Babaie Sarijaloo et al., 2021)	Hybrid corn yield estimation	Yield data, environmental data.	Random Forest, XG boost, ADA boost, Gradient Boosting Model, Decision Tree, Neural Network	RMSE – 0.0524
(Jafarbiglu & Pourreza, 2022)	Nuts crops yield management	Sensor parameters, lighting, geometric angles, and environmental factors	NDVI and Canopy Response Salinity Index (CRSI), principal component analysis (PCA), regression	Accuracy
(Rasti et al., 2022)	Cereal crop growth monitoring	crop canopy cover, above-ground biomass, leaf area	Artificial neural networks (ANN)	Residual value (r v) = 7.3% LCC (R2 = 0.86) and CCC (R2 =0.90)

		index , chlorophyll content, and growth stage		
(Tesfaye et al., 2021)	Wheat yield prediction	Satellite data, fertilizer data, crop phenotype data, crop rainfed data, crop gap fill data	GLR and RF	RMSE : 0.001 t/ha and 0.136 t/ha
(Myers et al., 2021)	Maize yield correlation accuracy	Satellite Image data, hyperspectral image data, Yield data, Soil data	Vegetation indices: VI , EVI,EVI2,GCI,GNDVI,NDVI,SAVI,WDRVI,MODIS (GNDVI) Green normalized difference vegetation index	If the R-squared value is greater than 0.95 and the last day of the time series is 65 days after green-up or later, then the Flex fit is appropriate. An R-squared value of 0.92 indicates a close fit of the shape model.
(Seireg et al., 2022)	Wild blueberry yield prediction	Weather data, yield data, computer simulation programmed data	stacking regression and cascading regression with a combination of MLA(VIF+ SFFS+ SBEFS),	SR: $R2$ = 0.984 and RMSE =179.898 CR: $R2$ =0.938 and RMSE =343.026
(Alabi et al., 2022)	Soybean yield prediction	Phenology data, multispectral images, vegetation indices, texture features	Regression models, Cubist, Extreme Gradient Boosting (xgboost), GBM, SVM, and RF.	R2 = 0.89
(Jha et al., 2022)	Paddy and wheat yield gap estimation	Yield data, fertilizer information, climate data, cropland information, irrigation information, seed data	Stochastic Frontier	Accuracy
(Qiao et al., 2021)	Wheat yield prediction	Satellite images, Location data, crop management information	SVM,RF,DT,2-D CNN,LSTM, 3D-CNN, Multikernal Gaussian Process.	RMSE, R2, MAPE

WHEAT YIELD PREDICTION

After maize, wheat is the most widely produced cereal grain in the world. It is also the most widely traded. Wheat output around the world reached 760 million tonnes in 2020. About 41% of global wheat production comes from the top three producers China, India, and Russia. Wheat yield prediction (Bali & Singla, 2021)guides farmers to be more precise about their production. For the prediction (Tesfaye et al., 2021), some techniques of ML, DL, IoT, etc. which have been used for their accuracy. Experienced researchers achieve the prediction by using hyperspectral remote sensing and extreme machine learning techniques by hyperspectral image and spectral clustering which are the calibration of ML techniques. (Qiao et al., 2021) estimated the yield prediction by First, a robust 3-D CNN built to maximize spatial-spectral picture characteristics. Using multi-kernel learning(MKL), we combine deep spatial-spectral features from within an image with those from across samples to improve classification accuracy. In the MKL framework, we assign nonlinear kernels to each feature to suit features from distinct domains. A kernel-based technique determines the prediction probability distribution. The production of wheat in India is listed in Fig 6 from 2015 to 2020.



	2015	2016	2017	2018	2019	2020
RICE	36077	37902	38493	39568	40577	39623
WHEAT	27496	30339	31999	33682	35334	34311
MAIZE	25972	26162	26886	30653	30702	30573
SOYABEAN	7344	11770	11766	10585	11919	9278

Fig 6: *Year-wise yield production of rice , white,maize,and soyabean from india*

MAIZE YIELD PREDICTION

Corn's widespread use and great yields help it lead to modern agriculture. Higher energy than wheat, barley, and oats. All species of animals and fowl can eat corn kernel. With growing energy prices, corn kernels are becoming a more popular bio-ethanol raw source. In 2019, international growers planted 192 million hectares of this crop, 3 million more than in 2018. According to the USDA, the US, China, and Brazil produce 48% of the world's corn. Last year, the US yielded 10.5 t/ha, putting it first. On 33,1 million acres, American farmers harvested 347 million metric tonnes of grain or 33 percent of the world's total. Using a decision tree and other machine learning techniques, the research team led by (Babaie Sarijaloo et al., 2021) predicts the yield of corn hybrids. Other methods used by participants in the 2020 Syngenta Crop Challenge include gradient boosting machines, random forests, adaptive boosting, XGBoost, and neural networks. It was investigated by (Myers et al., 2021) how the end date of the time series and the frequency of satellite imaging impacted the validity of the correlation between VI and yield. Daily multispectral photos with a resolution of 3 meters were taken over. Fig 6 described the maize produced by India from 2015 to 2020. This data was collected from FAOSTAT where worldwide data are available.

SOYBEAN YIELD PREDICTION

Humans have been growing soybeans practically constantly since the advent of agriculture in Asia, and the crop remains crucial to the global food supply. Depending on how they are refined, these legumes can be categorized as either legumes, oil seeds, vegetables, or even fuel sources. Unlike many other plant-based foods, soybeans are a "complete protein" because their protein structures contain all nine essential amino acids.(Ansarifar et al., 2021) discovered around a dozen settings via management interactions for soybean production. Some of these are in line with the current state of agronomic understanding, while others require additional study or experimental confirmation. Using this method, agronomists may determine the relative importance of weather, soil, management, and their interactions on a given site's crop output under a given weather and management situation. Physiology is used as a foundation for many crop models such as APSIM, DSSAT, RZWQM, and SWAP/ WOFOST2. Use nonlinear equations to forecast harvest success and plant characteristics. Theodore R. Alabi et al. To expedite soybean yield estimation, researchers in Nigeria mounted a Sequoiacamera on a senseFly eBee X UAV in 2020 and collected multispectral photographs of five different variety trials. Information gathered by UAVs in the form of spectral bands, canopy height, vegetation indices (VI), and texture was included in five different machine learning (ML) regression models (Cubist, XGBoost, GBM, SVM, and Random Forest) to produce an estimate of agricultural grain yield. By far outperforming predictors based on vegetation indices (VI), grey-level co-occurrence matrix (GLCM) texture information showed promise as a replacement for VI in agricultural production estimation. Fig 6 depicts the volume of soybeans harvested in India from 2015 to 2020.

DISCUSSION

The agriculture sector desperately needs to adopt cutting-edge technology to keep up with the expansion of the global population. In addition, agriculturists need accurate, timely instructions to anticipate crop yields, which is essential for developing strategies that would maximize yields. ML frameworks reveal the method by which information is gathered and analyzed by examining the vast volumes of data that are produced. These methods are used to build models that explain the interrelationships between the various elements and various actions that are taken. Additionally, ML models can be used to anticipate how an individual would react in the future. This research shows that many different selection criteria were used to choose the included publications, with data accessibility and study scope coming out on top. Most of the cited works examine techniques for yield prediction utilizing machine learning algorithms. Nonetheless, the main distinction between the two is the incorporation of a wide variety of elements. It was also observed that there was wide diversity in the trials themselves due to differences in the crop, location, and intensity. When deciding which features to include in the dataset, researchers think about both the study's goals and the ease of access the dataset provides. The wide range of existing literature also suggests that the most precise yield estimation wouldn't be achieved by implementing a model with a comprehensive collection of features. Although it is now impossible to decide which approach is more beneficial based on available evidence, it is crucial to obtain a comprehensive picture due to the widespread use of various machine learning algorithms and their promising performance. The most promising instances of classical machine learning are the LR, RF, and NN architectures. Some DL models, such as DNN, CNN, and LSTM, are also used in agricultural yield estimation in addition to these methods. For conclusive results on the model with the best performance, it is necessary to use the data in conjunction with multiple feature selection methods. Examining why some pre-existing models do not meet the research goal is advised.

REFERENCES

1. Alabi, T. R., Abebe, A. T., Chigeza, G., & Fowobaje, K. R. (2022). Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. *Remote Sensing Applications: Society and Environment*, *27*. https://doi.org/10.1016/j.rsase.2022.100782
2. Alebele, Y., Wang, W., Yu, W., Zhang, X., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2021). Estimation of Crop Yield from Combined Optical and SAR Imagery Using Gaussian Kernel Regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 10520–10534. https://doi.org/10.1109/JSTARS.2021.3118707
3. Al-Gunaid, M. A., Salygina, I. I., Shcherbakov, M. v., Trubitsin, V. N., & Groumpos, P. P. (2021). Forecasting potential yields under uncertainty using fuzzy cognitive maps. *Agriculture and Food Security*, *10*(1). https://doi.org/10.1186/s40066-021-00314-9

4. Amaratunga, V., Wickramasinghe, L., Perera, A., Jayasinghe, J., Rathnayake, U., & Zhou, J. G. (2020). Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Mathematical Problems in Engineering*, *2020*. https://doi.org/10.1155/2020/8627824
5. Ansarifar, J., Wang, L., & Archontoulis, S. v. (2021). An interaction regression model for crop yield prediction. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-97221-7
6. Babaie Sarijaloo, F., Porta, M., Taslimi, B., & Pardalos, P. M. (2021). Yield performance estimation of corn hybrids using machine learning algorithms. *Artificial Intelligence in Agriculture*, *5*, 82–89. https://doi.org/10.1016/j.aiia.2021.05.001
7. Bali, N., & Singla, A. (2021). Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India. *Applied Artificial Intelligence*. https://doi.org/10.1080/08839514.2021.1976091
8. Beyene, A. N., Zeng, H., Wu, B., Zhu, L., Gebremicael, T. G., Zhang, M., & Bezabh, T. (2021). Coupling remote sensing and crop growth model to estimate national wheat yield in Ethiopia. *Big Earth Data*. https://doi.org/10.1080/20964471.2020.1837529
9. Biswas, J. C., & Naher, U. A. (2018). Soil nutrient stress and rice production in Bangladesh. In *Advances in Rice Research for Abiotic Stress Tolerance* (pp. 431–445). Elsevier. https://doi.org/10.1016/B978-0-12-814332-2.00021-6
10. Cho, J. B., Guinness, J., Kharel, T. P., Sunoj, S., Kharel, D., Oware, E. K., van Aardt, J., & Ketterings, Q. M. (2021). Spatial estimation methods for mapping corn silage and grain yield monitor data. *Precision Agriculture*, *22*(5), 1501–1520. https://doi.org/10.1007/s11119-021-09793-z
11. Choudhury, M. R., Das, S., Christopher, J., Apan, A., Chapman, S., Menzies, N. W., & Dang, Y. P. (2021). Improving biomass and grain yield prediction of wheat genotypes on sodic soil using integrated high-resolution multispectral, hyperspectral, 3d point cloud, and machine learning techniques. *Remote Sensing*, *13*(17). https://doi.org/10.3390/rs13173482
12. Debalke, D. B., & Abebe, J. T. (2022). Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia. *Environmental Systems Research*, *11*(1). https://doi.org/10.1186/s40068-022-00249-5
13. Deines, J. M., Patel, R., Liang, S. Z., Dado, W., & Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, *253*. https://doi.org/10.1016/j.rse.2020.112174
14. Diaz-Gonzalez, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., & Patino, D. (2022). Machine learning and remote sensing techniques applied to estimate soil indicators – Review. In *Ecological Indicators* (Vol. 135). Elsevier B.V. https://doi.org/10.1016/j.ecolind.2021.108517
15. Dubois, A., Teytaud, F., & Verel, S. (2021). Short term soil moisture forecasts for potato crop farming: A machine learning approach. *Computers and Electronics in Agriculture*, *180*. https://doi.org/10.1016/j.compag.2020.105902
16. Elavarasan, D., & Durairaj Vincent, P. M. (2020). Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. *IEEE Access*, *8*, 86886–86901. https://doi.org/10.1109/ACCESS.2020.2992480
17. Felipe Maldaner, L., de Paula Corrêdo, L., Fernanda Canata, T., & Paulo Molin, J. (2021). Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Computers and Electronics in Agriculture*, *181*. https://doi.org/10.1016/j.compag.2020.105945
18. Gautron, R., Maillard, O. A., Preux, P., Corbeels, M., & Sabbadin, R. (2022). Reinforcement learning for crop

management support: Review, prospects and challenges. In *Computers and Electronics in Agriculture* (Vol. 200). Elsevier B.V. https://doi.org/10.1016/j.compag.2022.107182

19. Geipel, J., Bakken, A. K., Jørgensen, M., & Korsaeth, A. (2021). Forage yield and quality estimation by means of UAV and hyperspectral imaging. *Precision Agriculture*, *22*(5), 1437–1463. https://doi.org/10.1007/s11119-021-09790-2
20. Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J. H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, *276–277*, 107609. https://doi.org/10.1016/J.AGRFORMET.2019.06.008
21. Jafarbiglu, H., & Pourreza, A. (2022). A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. In *Computers and Electronics in Agriculture* (Vol. 197). Elsevier B.V. https://doi.org/10.1016/j.compag.2022.106844
22. Jha, G. K., Palanisamy, V., Sen, B., & Kumar, A. (2022). Explaining Rice and Wheat Yield Gaps in Eastern Indian States: Insights from Stochastic Frontier Analysis. *Agricultural Research*. https://doi.org/10.1007/s40003-021-00599-z
23. Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensors (Switzerland)*, *21*(4), 1–17. https://doi.org/10.3390/s21041406
24. Jiru, E. B., & Wegari, H. T. (2022). Soil and water conservation practice effects on soil physicochemical properties and crop yield in Ethiopia: review and synthesis. In *Ecological Processes* (Vol. 11, Issue 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s13717-022-00364-2
25. Juventia, S. D., Rossing, W. A. H., Ditzler, L., & van Apeldoorn, D. F. (2021). Spatial and genetic crop diversity support ecosystem service delivery: A case of yield and biocontrol in Dutch organic cabbage production. *Field Crops Research*, *261*. https://doi.org/10.1016/j.fcr.2020.108015
26. Khaki, S., Pham, H., & Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-89779-z
27. Khanal, S., Klopfenstein, A., KC, K., Ramarao, V., Fulton, J., Douridas, N., & Shearer, S. A. (2021). Assessing the impact of agricultural field traffic on corn grain yield using remote sensing and machine learning. *Soil and Tillage Research*, *208*. https://doi.org/10.1016/j.still.2020.104880
28. Kok, Z. H., Mohamed Shariff, A. R., Alfatni, M. S. M., & Khairunniza-Bejo, S. (2021). Support Vector Machine in Precision Agriculture: A review. In *Computers and Electronics in Agriculture* (Vol. 191). Elsevier B.V. https://doi.org/10.1016/j.compag.2021.106546
29. Li, S., Peng, S., Chen, W., & Lu, X. (2013). INCOME: Practical land monitoring in precision agriculture with sensor networks. *Computer Communications*, *36*(4), 459–467. https://doi.org/10.1016/j.comcom.2012.10.011
30. Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., & Rembold, F. (2021). Yield forecasting with machine learning and small data: What gains for grains? *Agricultural and Forest Meteorology*, *308–309*. https://doi.org/10.1016/j.agrformet.2021.108555
31. Meshram, V., Patil, K., Meshram, V., Hanchate, D., & Ramkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences*, *1*, 100010. https://doi.org/10.1016/j.ailsci.2021.100010
32. Myers, E., Kerekes, J., Daughtry, C., & Russ, A. (2021). Effects of Satellite Revisit Rate and Time-Series Smoothing Method

on Throughout-Season Maize Yield Correlation Accuracy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 12007–12021. https://doi.org/10.1109/JSTARS.2021.3129148

33. Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*. https://doi.org/10.1016/j.agsy.2020.103016
34. Paudel, D., Boogaard, H., de Wit, A., van der Velde, M., Claverie, M., Nisini, L., Janssen, S., Osinga, S., & Athanasiadis, I. N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, *276*. https://doi.org/10.1016/j.fcr.2021.108377
35. Priya, R., & Ramesh, D. (2020). ML based sustainable precision agriculture: A future generation perspective. *Sustainable Computing: Informatics and Systems*, *28*. https://doi.org/10.1016/j.suscom.2020.100439
36. Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Tian, Z., & Guo, H. (2021). Exploiting Hierarchical Features for Crop Yield Prediction Based on 3-D Convolutional Neural Networks and Multikernel Gaussian Process. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 4476–4489. https://doi.org/10.1109/JSTARS.2021.3073149
37. Ramu, M., & Sri, J. T. (2021). Wheat yield prediction using Artificial Intelligence models and its comparative analysis for better prediction. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, 363–367. https://doi.org/10.1109/ICACITE51222.2021.9404707
38. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. In *IEEE Access* (Vol. 9, pp. 63406–63439). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2021.3075159
39. Rasti, S., Bleakley, C. J., Holden, N. M., Whetton, R., Langton, D., & O'Hare, G. (2022). A survey of high resolution image processing techniques for cereal crop growth monitoring. In *Information Processing in Agriculture* (Vol. 9, Issue 2, pp. 300–315). China Agricultural University. https://doi.org/10.1016/j.inpa.2021.02.005
40. Seireg, H. R., Omar, Y. M. K., El-Samie, F. E. A., El-Fishawy, A. S., & Elmahalawy, A. (2022). Ensemble Machine Learning Techniques Using Computer Simulation Data for Wild Blueberry Yield Prediction. *IEEE Access*, *10*, 64671–64687. https://doi.org/10.1109/ACCESS.2022.3181970
41. Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-020-80820-1
42. Tesfaye, A. A., Osgood, D., & Aweke, B. G. (2021). Combining machine learning, space-time cloud restoration and phenology for farm-level wheat yield prediction. *Artificial Intelligence in Agriculture*, *5*, 208–222. https://doi.org/10.1016/j.aiia.2021.10.002
43. Tiwari, P., & Shukla, P. (2020). Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors. In *Advances in Intelligent Systems and Computing* (Vol. 933, pp. 585–594). Springer Verlag. https://doi.org/10.1007/978-981-13-7166-0_58
44. Tufail, M., Iqbal, J., Tiwana, M. I., Alam, M. S., Khan, Z. A., & Khan, M. T. (2021). Identification of Tobacco Crop Based on Machine Learning for a Precision Agricultural Sprayer. *IEEE Access*, *9*, 23814–23825. https://doi.org/10.1109/ACCESS.2021.3056577
45. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction

using machine learning: A systematic literature review. In *Computers and Electronics in Agriculture* (Vol. 177). Elsevier B.V. https://doi.org/10.1016/j.compag.2020.105709

46. Varinderpal-Singh, Kunal, Kaur, R., Mehtab-Singh, Mohkam-Singh, Harpreet-Singh, & Bijay-Singh. (2022). Prediction of grain yield and nitrogen uptake by basmati rice through in-season proximal sensing with a canopy reflectance sensor. *Precision Agriculture*, *23*(3), 733–747. https://doi.org/10.1007/s11119-021-09857-0
47. Zhou, W., Liu, Y., Ata-Ul-Karim, S. T., Ge, Q., Li, X., & Xiao, J. (2022). Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods. In *International Journal of Applied Earth Observation and Geoinformation* (Vol. 111). Elsevier B.V. https://doi.org/10.1016/j.jag.2022.102861
48. Zhou, X., Kono, Y., Win, A., Matsui, T., & Tanaka, T. S. T. (2021). Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Production Science*, *24*(2), 137–151. https://doi.org/10.1080/1343943X.2020.1819165