Systematic Review

Artificial Intelligence for Sustainable Agriculture: A Comprehensive Review of AI-Driven Technologies in Crop Production

Zulfiqar Ali [1], Asif Muhammad [2], Nangkyeong Lee [3], Muhammad Waqar [4] and Seung Won Lee [3,5,6,7,*]

[1] Department of Computer Science, COMSATS University, Islamabad 44000, Pakistan; zulfiqarali@comsats.edu.pk
[2] FAST School of Computing, National University of Computer & Emerging Sciences, A. K. Brohi Road, Sector H-11/4, Islamabad 44000, Pakistan
[3] Department of Precision Medicine, Sungkyunkwan University School of Medicine, Suwon 16419, Republic of Korea
[4] School of Technology, Business and Arts, University of Suffolk, Ipswich IP4 2QJ, UK
[5] Department of Artificial Intelligence, Sungkyunkwan University, Suwon 16419, Republic of Korea
[6] Department of Metabiohealth, Sungkyunkwan University, Suwon 16419, Republic of Korea
[7] Personalized Cancer Immunotherapy Research Center, Sungkyunkwan University School of Medicine, Suwon 16419, Republic of Korea
* Correspondence: swleemd@g.skku.edu

Abstract: Smart farming leverages Artificial Intelligence (AI) to address modern agricultural sustainability challenges. This study investigates the application of machine learning (ML), deep learning (DL), and time series analysis in agriculture through a systematic literature review following the PRISMA methodology. The review highlights the critical roles of ML and DL techniques in optimizing agricultural processes, such as crop selection, yield prediction, soil compatibility classification, and water management. ML algorithms facilitate tasks like crop selection and soil fertility classification, while DL techniques contribute to forecasting crop production and commodity prices. Additionally, time series analysis is employed for demand forecasting of crops, commodity price prediction, and forecasting crop yield production. The focus of this article is to provide a comprehensive overview of ML and DL techniques within the farming industry. Utilizing crop datasets, ML algorithms are instrumental in classifying soil fertility, crop selection, and various other aspects. DL algorithms, when applied to farming data, enable effective time series analysis and crop selection. By synthesizing the integration of these technologies, this review underscores their potential to enhance decision-making in agriculture and mitigate food scarcity challenges in the future.

Keywords: artificial intelligence; Internet of Things; machine learning; deep learning; smart farming; smart agriculture; time series analysis; PRISMA

1. Introduction

Despite facing previously unheard-of difficulties, including resource scarcity, population expansion, and climate change, agriculture continues to be a vital component of the world's prosperity and social stability. To tackle these problems, cutting-edge technologies like deep learning (DL) and machine learning (ML) are being embraced more and more to transform farming methods globally. Even though nations like Pakistan have particular difficulties because of their lack of resources and technological advancements, the concepts presented here apply to a variety of agricultural environments. Pakistan's agriculture sector is experiencing a persistent decline due to its robust economic foundation and crucial role in several economic schemes. The country's economy is highly dependent on crops, serving as

the cornerstone of its farming heritage. Various tasks are included in agricultural activities, such as crop-relevant decisions, irrigation control, soil health assessment, and fertilizer selection. Neural networks and random forests are two ML algorithms that are particularly effective at managing vast amounts of multidimensional data (Niazian and Niedbała [1]). These techniques improve in vitro breeding optimization, yield prediction, and genotype categorization. Precision phenotyping is made possible by combining machine learning and imagery, which advances precision farming and plant breeding studies in the future.

Smart farming relies heavily on ML. ML is applied to crop management and selection. For example, various crops grow well in different types of soil. Farmers must play their role in selecting the best possible acreage for their crops. ML classification approaches can be used to determine if the land is suitable for a given yield. ML regression approaches can be used to know the water resources required. The manuscript focuses mainly on several supervised and unsupervised ML techniques for yield selection and management. Several ML techniques, including Decision Trees (DTs), Random Forest (RF), k-Nearest Neighbour (KNN), and Support Vector Machines (SVM) [2], have been examined. One of the tedious tasks in intelligent farming is to determine important parameters for the selection of crops and their management. ML classification algorithms such as SVM [3], DTs [4], KNN, RF [5], and several others are used to recommend crops based on soil attributes, ground composition, and ecological conditions. ML must be used to forecast agricultural yield [6] so that appropriate measures can be taken to raise crop productivity. One of the crucial steps in ML is the selection of features, which is assessed using a variety of techniques [7]. Using statistical approaches to forecast the climate and predict the frequency of rainfall will increase crop productivity. Research on the choice of crops and recommendation prediction utilizing a variety of soil and meteorological data was conducted using a Recurrent Neural Network (RNN) and ML [8]. Figure 1 depicts a visual representation of our study, which aims to aid farmers with limited literacy in their native language throughout each of these phases.

Figure 1. Benefits of Smart IoT applications in the farming industry.

This paper examined various deep learning methods that can be used for the selection of crops, with a particular emphasis on the Radial Basis Function Network (RBFN) and Extreme Learning Machine (ELM). Theoretically, the ELM algorithm is best suited to learning the pattern in a short time duration [9]. There is no need to perform training iteratively, as

all the parameters are updated once. Compared with traditional learning algorithms, ELM often achieves the smallest weight norm and the least amount of training error. Another approach called RBFN learns more efficiently and quickly than other conventional neural networks [10]. RBFN can be used in a more efficient way for crop selection. This article addresses the possibility of implementing smart farming practices and offers an assessment of RBFN. Several time series approaches have been reviewed, including RNN [11] and Autoregressive Integrated Moving Average (ARIMA), as covered by Asy'ari et al. [12]. The article provides novel perspectives and avenues for inquiry into creative approaches to creating new models that can aid in improved crop management decision-making. Time series models can benefit from using multi-head attention enabled by transformers [13], which can handle long-term dependencies better than existing methods. Time series analysis plays a crucial role, especially considering the growing need for precise crop yield forecasting. The authors also explore the benefits of using RNN and ARIMA for time series prediction of agricultural data.

Predictive analytics and real-time decision-making made possible by the integration of ML, DL, and time series analysis (TSA) approaches in smart farming have the potential to completely transform agricultural processes. However, putting these cutting-edge technologies into practice presents several difficulties, including integrating disparate data sources, guaranteeing robustness to environmental unpredictability, and optimizing resource allocation. In the context of smart farming, this manuscript attempts to explore how well ML, DL, and TSA may improve crop output prediction, disease detection, and irrigation management while considering the unique challenges and limitations of agricultural contexts.

The inspiration for this article came from the difficulties farmers were facing at the time of implementing smart farming. Some of the problems were rigorously analyzed and solved by utilizing deep learning and sophisticated machine learning techniques. Various technologies have been showcased, including DL, ML, and TSA. The goal of this evaluation is to strengthen algorithms in terms of performance and learning rate. The benefits of the reviewed algorithm over conventional methods are covered in the following sections. Several regions of Pakistan are lagging in agricultural output because of farmer ignorance, delayed retrieval of vital data, and proactive decision formulation [14].

In these areas, Information and Communication Technologies (ICTs), AI, ML, and DL play a significant role in making access to data available to farmers [9]. Food insecurity is a result of poor agricultural performance brought on by climate change and a lack of agricultural resources [15]. This enables farmers to degrade soil with stronger pesticides, which negatively impacts agricultural methods [16]. These include disease-related reduced production, unpredictably changing environment, and loss of soil fertility [17].

Some of the key contributions of this research manuscript are as follows:

- An extensive overview of ML and DL algorithms in the context of smart farming, featuring a comprehensive classification of the current data within this field.
- An exploration of applications in smart farming and an overview of the ML and DL approaches, including Linear Regression, Support Vector Machines (SVMs), and artificial neural networks (ANNs).
- A rigorous overview of time series models using multi-head attention enabled by transformers.

2. Related Work

The conventional agricultural era 1.0 was characterized by agricultural practices that were focused on the creation of food in sophisticated fields for human endurance and the breeding of animals [18]. Among the fundamental farming tools utilized were sickles and shovels. Productivity remained low since manual labor accounted for most of the work.

The farming era 3.0 was introduced because of the 20th century's tremendous growth of computing and automation. Agriculture has been improved by robotic techniques, agricultural machines with programming, and other technologies. With proper distribution, accurate irrigation, compact chemical use, site-specific fertilizer delivery, effective pest control technologies, etc., the shortcomings in the smart farming 2.0 era are mitigated, and several principles were reviewed for the smart farming era 3.0. Modern technology is harnessed in smart farming to bolster precision agriculture, granting farmers the capacity to remotely oversee their crops. The utilization of sensors and automated equipment in smart farming has resulted in enhanced productivity for farming personnel, positively impacting harvesting and crop yields [19].

A technological revolution in agriculture has been driven by technology that automates conventional farming practices. The IoT has reinvented long-standing practices and changed how farming is currently conducted because of technology [20]. Agriculture is pivotal in global efforts to combat climate change and enhance sustainability. This article explores the emissions and removals of greenhouse gases (GHGs) in agriculture, examining sources such as rice cultivation, livestock enteric fermentation, and synthetic fertilizers. SaberiKamarposhti et al. [21] highlighted challenges in reducing emissions and investigated innovative solutions, including AI-powered monitoring systems and carbon capture technologies. Advanced techniques like precision agriculture and renewable energy integration can reduce emissions while boosting productivity. The paper thoroughly assesses agriculture's role in climate change mitigation, offering insights and future research directions to enhance understanding and practical solutions for sustainability. The authors further explore the integration of hydrogen energy and AI within smart infrastructure, aiming to revolutionize the global energy sector. It discusses the progress, challenges, and potential breakthroughs in using AI technologies, such as deep learning and machine learning, to optimize energy generation, distribution, and utilization. Key benefits include predictive maintenance, real-time decision-making, and efficient demand-side management, which enhance energy system resilience and sustainability. Significant challenges are highlighted, including data privacy and security, interoperability, and the technical limitations of AI in grid management. The study advocates for standardizing communication protocols and further research to address these issues. It emphasizes the role of AI in autonomous energy management, improving flexibility, proactive maintenance, and decentralized energy generation and storage, which supports rapid decision-making and enhances grid durability.

SaberiKamarposhti et al. [22] comprehensively examine agriculture's role in climate change, focusing on GHG emissions and mitigation strategies. It analyzes emissions from various agricultural sources, including carbon dioxide, methane, and nitrous oxide, and explores potential carbon sequestration methods like soil carbon storage, afforestation, and reforestation. The study emphasizes both the positive and negative impacts of emissions reduction policies and identifies sustainable agricultural practices, improved livestock management, and precision agriculture as key mitigation strategies. It also addresses the challenges of implementing these strategies, including socioeconomic and regulatory obstacles, and stresses the importance of equitable solutions for smallholder farmers. The research highlights advancements in measurement, climate-smart technologies, and the need for cross-sectoral collaboration. It underscores agriculture's potential to reduce emissions, enhance sustainability, and ensure food security, advocating for a transformative approach to achieve a sustainable and resilient agricultural future.

The benefits of smart agriculture encompass real-time crop data collection, precise assessments of crops and soil, remote monitoring capabilities for farmers, sustainable management of water and other natural resources, and increased agricultural and livestock

output. In essence, smart agriculture represents the evolution of precision farming through modernization and intelligent approaches to collect data across various farm operations, which are subsequently monitored remotely and supported with relevant real-time maintenance solutions. The evolution of conventional farming towards smart farming is depicted in Figure 2.
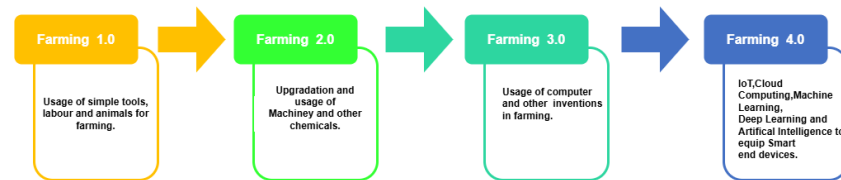

Figure 2. Evolution of conventional farming towards smart farming.

The integration of DL and ML technology has brought about a radical shift in precision agriculture, also known as smart farming, in recent years. With an emphasis on the use of ML and DL approaches, this literature review examines the state of research, developments, and applications in the field of smart farming.

2.1. Overview of Machine Learning Algorithms in the Context of Precision Farming

By analyzing historical and current conditions, it is possible to forecast future crop yields. This predictive insight extends its benefits to farmers, the food industry, several agencies regarding food, and individuals relevant to food security [23]. The methods for predicting yields vary depending on the crop, the types of input data used, and the specific prediction model. Researchers have been actively engaged in predicting crop yields using available data. Researchers concentrated on refining their predictions through the utilization of the Random Forest algorithm, using a sugar cane dataset that included macroclimatic data. Existing research employed both ANN and Multiple Linear Regression (MLR) techniques to forecast maize yields in response to climate variations, incorporating climatic, crop, soil, and fertilizer data as inputs.

The focus of yield prediction research has been on using the data to produce precise projections. For example, ref. [24] used RF to apply feature selection and model fine-tuning to a sugarcane dataset that includes macroclimatic data, emphasizing the significance of these processes. To project maize yield in response to climatic fluctuations, existing research uses both ANN and MLR approaches, combining crop, soil, and fertilizer data as input variables to predict soybean yield, Deep Gaussian Process modeling, and remote sensing data.

ANN and SVM are mostly used to estimate the quantity of berries per cluster in grapes. Several factors are included in growth analysis, and one of the most basic measurements is plant height. Plant breadth and the quantity of leaves per plant are other frequently used measurements. The goal of farming is yield, and a plentiful harvest can only be obtained if the crop is allowed to grow under the right conditions. A Knowledge and Data-Driven Model (KDDM) was presented as a descriptive tool for plants. It includes data on climate, growth media (e.g., fertilizer usage), and plant development metrics (e.g., biomass output and fruit setting).

An extended version of RF was proposed by Geng et al. [25] to predict food safety. Feng et al. [26] used time series data and an SVM to detect crop types. 9 crop kinds were classified using the SVM model with an accuracy of more than 86%. Authors in [2] illustrate crop and resource management automation using a range of classifiers, such as KNN, SVM, and Back Propagation Neural Network (BPNN). Further in [27], Loresco et al. showed how to use KNN for image segmentation to determine the plant's growth stages. Abbas et al. [28] extracted relevant data for agricultural yield using a variety of ML algorithms, including LR and SVM. In the study in [29], Hamza et al. use ML algorithms to

compare different agricultural yield parameters, such as temperature, humidity, and so forth. Table 1 outlines several applications of yield expectation and expansion analysis.

Abbas et al. [28] extracted relevant data for agricultural yield using a variety of ML algorithms, including LR and SVM. In another article [30], authors use ML algorithms to compare different agricultural yield parameters, such as temperature, humidity, and so forth. The Advanced Decision Tree (ADT) classification technique is described by [31] as a means of designing mobile applications and turning on the global positioning system. Babu et al. [32] describe feature selection using PSO-SVM followed by fuzzy-based decision tree classification. Using RF analysis, authors in [33] examined agricultural yield prediction, assisting farmers in implementing best management strategies. A tree algorithm is used in [34] to show how agricultural land might be transformed into urban areas and adjusted accordingly. To boost agricultural productivity, Nagasubramanian et al. [35] demonstrated the use of the SVM algorithm to identify the disease at early stages. ML techniques are a major component of modern farming data processing. This section explores some of the most used ML models in this field. The analysis of data from smart farming has led to a broad use of these machine-learning techniques. The ML methods employed for this purpose can be broadly categorized into three groups: ANN-based models, which include deep learning models; tree or kernel-based models; and conventional regression. Prediction, detection, and optimization are the three basic purposes of smart farming applications. Although there are many other ML approaches, we have chosen three of the most popular ones—linear regression, SVMs, and ANNs—for a closer look. Most farm management activities, including controlling fertilizer use [36] and maximizing water use [37], have been the focus of linear regression models. SVMs have generally been used in quality maintenance applications, as demonstrated by studies [38,39]. The widely used ANNs and their variants have been applied in many fields, such as quality maintenance [40].

Table 1. Precision, recall, and specificity scores for different algorithms.

Research Area	Data Type	ML Algorithm	Modeling	Records in Dataset	Precision	Recall
Instantaneous Yield Formation	Macro Climate, Soil (fertilizer), Crop Data	ANN and MLR	Classification	152 Images	0.8705	0.9572
Instantaneous Yield Formation	Remote Sensing Data	CNN with Deep Gaussian Process	Proposed the crop model	NA	NA	NA
Crop Yield Gap Analysis	Remote Sensing Data (Satellite Data)	Linear Regression	LSTM	NA	NA	NA
Prediction of Oil Extraction	Macro Climate, Soil, Crop Data, Other (Oil yield)	ANN	Statistical approach	NA	NA	NA
Instantaneous Yield Formation	Macro Climate, Soil, Crop Data	Random Forest	DNN	10,196 Images	AUC weed = 0.782	NA
Interpreting Yield Predictions with Factors	detection of diseases in plants	Deep CNN	NA	NA	NA	NA

ANN models are now used in most detective and predictive applications. Smart agricultural data, particularly time series farming data, now has forecasts that are noticeably more accurate because of developments in ANN and deep learning. These new advancements have made previously impossible tasks, such as diagnosing and forecasting diseases, insects, and the best times to harvest, more doable. This has proven to be crucial in resolving the difficulties brought forth by complex data. On the other hand, LR has

shown usefulness in applications related to both detection and optimization. Even though it is a fundamental method in predictive modeling, its use for prediction in smart farming applications has not been as common. This is because most linear regression models are less effective in real-world scenarios because they have difficulty capturing complex and non-linear correlations among many parameters relevant to smart farming. SVMs have been useful in detection tasks and have shown a high degree of accuracy in anomaly identification, particularly in situations where there are several non-linearly connected predictor features. This is mostly because the quantity of input predictor characteristics has no bearing on the SVM parameters. Therefore, regardless of the magnitude and complexity of the input features, SVMs are highly suited for identifying weeds or damages (which are comparable to anomalies) in the farming setting. Additionally, they function well even when there are non-linearities in the data. When it comes to smart farming regression forecasting models, time series-based regression approaches are the most used method for predicting crop yield. This technique predicts future yield values by using historical yield data at predetermined periods of time, or "lag values".

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

The relationship between the independent and dependent variables is the fundamental component of a regression model. The coefficient β_1 reflects the linear relationship between X_i and Y_i [41]. Let us take an example where we want to estimate the yield for a given farm, and we know that the environmental elements of the farm are represented by X_i, and the yield values are designated by Y_i. Since the yield values we are forecasting depend on the values of the environmental factors X_i, which are the independent variables, the yield values themselves become the dependent variable. The linear relationship between X_i and Y_i is represented by the coefficient β_1. For both classification and regression problems, SVM analysis is a popular machine learning approach. Significant non-linear correlations between dependent and independent variables make it impossible for a linear model to adequately address some regression problems. In these situations, the Lagrange dual formulation allows the previously outlined technique to be extended to handle non-linear functions. Since non-linear kernel functions aid in the more precise resolution of non-linear regression problems, which in turn results in accurate anomaly detection, this is very helpful for fruit grading.

Furthermore, SVMs were used to identify fruit that was damaged and to recognize fruit and work in tandem with a harvesting robot. Equations (2) and (3) provide a positive result; the SVM technique predicts the presence of the positive class; conversely, it produces a negative result, as it predicts the presence of the negative class [41,42].

$$3w^T x + b = 1 \tag{2}$$

$$4w^T x + b = -1 \tag{3}$$

where w is a weight vector, x is input vector, and b is bias.

The SVM technique tries to solve the dual problem by using the Lagrangian dual equation [41].

$$\max : f(c_1, \ldots, c_n) = \sum_{i=1}^{n} C_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i c_i (x_i \cdot x_j) y_j c_j \tag{4}$$

$$\text{subject to } \sum_{i=1}^{n} c_i y_i = 0$$

where $C_i \geq 0$ and $C_i \leq \frac{1}{2nL}$ for all i.

The Lagrangian function uses a new 'slack variable' denoted by C_i. In the field of yield forecasting, the SVM kernel functions can help solve non-linear regression problems more successfully with accurate forecasting performance. However, there are times when the SVM regression approach matches the training set too closely, which is referred to as the "over-fitting problem". Under these conditions, the forecasting performance of testing data is lower than that of training data. Another issue with SVMs is that, when compared with more interpretable methods like linear or logistic regression, SVM regression is significantly more challenging for farming domain specialists to understand. However, due to their improved performance, deep learning techniques are being used more and more in recent research on smart farming. Because CNNs are more effective instruments for image analysis and recognition, they have been used for these kinds of analyses.

2.2. Overview of Deep Learning Algorithms in the Context of Precision Farming

ANNs, a deep learning technique, have become a popular technology in yield forecasting. Deep learning techniques are superior to classical regression and SVM regression because of their complex computing capabilities and non-linear features, which enable them to handle large datasets. DL has been used in recent research to forecast yield and achieve state-of-the-art predictive performance. To understand the link between dependent and independent variables, these techniques iteratively use training data over numerous epochs. They are built utilizing basic neural network components, such as neurons and activation functions. The handling of big datasets and non-linearity are two notable issues that deep learning successfully tackles [42]. However, in many cases, learning could take longer than using more straightforward methods like SVMs.

One subclass of deep neural networks called CNNs is primarily used for the analysis of image or video data in the context of smart farming. CNNs have a wide range of applications in yield prediction. They are made up of three fundamental components: completely connected layers, pooling, and convolution. The convolutional layer reduces the number of parameters by dividing the weight parameters between two dimensions of neurons (or inputs) in the previous layer (also known as 2D convolution). After the input image (x) with a linear filter is passed through the pooling layer, a bias term (b) is added, and then a non-linear function—typically a Rectified Linear Unit (ReLU)—is applied. This is the computation of the output O. The weights W and bias b determine the filters of the pooling function, represented by the symbol p($\cdot$).

$$0 = p\left(\int (W^* x + b) \right) \tag{5}$$

Deep CNNs were used for yield prediction by [38], while CNNs were used for fruit detection by [43]. In addition to CNNs, Xiong et al. [44] included spatial and temporal attention, which improved the interpretation of how farm factors affect tomato productivity. As an illustration, ref. [45] may provide yield projections that are more accurate since they make use of remote sensing data along with CNNs and Gaussian processes. CNNs were used to maximize workforce utilization. Although CNNs are frequently employed for feature extraction in detection tasks, further research into their accuracy and performance is necessary. Furthermore, to estimate maize yield in response to climate changes, De Alwis et al. [46] used a non-linear ANN model with a single hidden layer.

$$y = \sum_j w_j \tanh\left(\sum_i w_{ji} x_i + b_j \right) + \tilde{b} \tag{6}$$

In the above Equation (6), the function tanh acts to move the input into a non-linear space, confirming the degree of complexity needed from the ANN model and ultimately producing forecasts that are more accurate. ANN models were used for weed detection by [44] and yield prediction by [47]. A lot of DL methods can overfit, and finding a way to combine model generalization with training accuracy is still a problem in this area of study. An overview of all ML and DL approaches and their applications is given in Table 2, which produces several important findings.

Table 2. Machine learning models and big data applications in smart farming.

Ref.	ML Algorithm	Data Category	Findings	Limitations
[24]	Random forest	Prediction yield	A yield potential prediction model was created and assessed. Estimated soil parameters and remote sensing vegetation indices were the primary inputs used to determine crop potential.	SKN model can be used to predict wheat yield and to classify field area into different yield potential zones more efficiently.
[48]	ANN and MLR	Prediction yield	The manuscript accurately examines photos captured with an inexpensive camera. A novel attribute for identifying berries was created. Strong correlations in determining the overall quantity of grapes.	Assessment of the berry set of plants, which is impossible with these approaches.
[49]	ANN	Prediction yield	WSN networks are used to construct a control system between node sensors in a crop field for the purpose of optimizing agriculture through crop irrigation.	Accuracy in terms of precision, recall, and AUC is not explored.
[50]	ANN and SVM	Prediction yield	Authors investigated how wavelet texture features might be used to distinguish between weeds and crops. Extract 52 texture features from wavelet multi-resolution photos.	The main limitation is a notable variation in the textural frequencies of the plant canopies.
[38]	ANN and CNN	Detection Weed	The goal of this article is to give an overview of the updated technologies utilized in agricultural systems.	Overall, the mathematical figures, like accuracy and time analysis, is not presented.

Deng et al. [51] introduced ELM; compared with feedforward networks, ELM offers higher performance and speed. Weights are determined analytically in ELM, and concealed nodes are selected at random. A deep understanding of ELM and its benefits and drawbacks is covered in this essay. ELM applications in clustering, classification, and regression tasks were investigated by [52]. Amirian et al. [53] built a Radial Basis Function (RBF) on top of convolution networks. Combined with convolution networks, RBF has the benefit of providing comprehensive insights into the decision-making process. After reviewing RBFN networks, ref. [54] concluded that RBFN can be applied in a variety of fields and provides quick training. Rani et al. [55] presented an ML-based crop selection model that considers both soil factors and weather conditions combined. LSTM RNN is used for weather analysis, whereas RF classifier is used for crop selection. When it comes to weather prediction, this model performs better than ANN. The RMSE for the Min. Temperature forecast, Max. Temperature prediction, and Rainfall prediction using LSTM RNN are 5.023%, 7.28%, and 8.24%, respectively. In the subsequent stage, the Random Forest Classifier demonstrated 97.235% accuracy in crop selection, 96.437% accuracy in resource dependency prediction, and 97.647 accuracy in determining the crop's ideal planting timing. Numerous noise-suppression techniques based on statistics and machine learning have been put forth in the literature. Compared with Kalman and moving average filters, which are other common filters, the suggested LSTM filter in [56] performs better in suppressing noise. Our

threshold-based aquaponics automation system incorporates an LSTM filter to optimize sustainable food production while minimizing expenses.

Cordeiro et al. [57] discuss many deep neural network topologies for constructing soil moisture prediction models. We also address the issue of missing values for the features in the dataset. We employ KNN data imputation for this purpose, which calls for replacing the values of uncertain (or missing) characteristics with values that guarantee the required level of trustworthiness. To assess the prediction models' performance in terms of CPU and RAM utilization, we also implanted the models on a tiny single-board computer, which is frequently utilized as a fog node. In [58], a novel intrusion detection system for agricultural IoT networks is proposed. The NSL KDD dataset is used to assess the proposed technique, which starts with a number of pre-processing stages on the original feature set. Recursive feature elimination is used to identify important features, which are subsequently translated into square color pictures. Currently, input images can be learned by many CNN architectures, including the Xception, Inception, and VGG16 models. CNN models' performance is compared with traditional machine learning techniques and assessed using measures like accuracy, recall, F1 score, and precision.

Offering a novel ML strategy for a heterogeneous data environment with IoT-sensed data about the environment, agricultural circumstances, plant traits, requests, etc., is the main goal of this research [59]. The dataset of the following five crops—rice, ragi, gram, potato, and onion—was sourced from the Kaggle repository for Andhra Pradesh. We used a variety of deep learning and machine learning algorithms in this work, including expectation–maximization (EM), decision trees, SVM, K-Means, Naive Bayes, and AI techniques (LSTM, RNN). With a training accuracy of 99.27% for agricultural yield prediction, the Random Forest method appears to perform better than other machine learning algorithms. However, when it comes to crop yield prediction, sigmoid outperforms ReLU and tanh activation, achieving 99.71 percent accuracy with four hidden layers.

2.3. Overview of Time Series Analytical Approaches in the Context of Precision Farming

Time series analytical techniques are essential for offering insights into many facets of agricultural operations in the context of precision farming. Time series models that consider variables like weather, irrigation, and fertilization can be used to analyze historical data on agricultural yields. Future crop yields can be predicted with the use of predictive modeling approaches such as Autoregressive Integrated Moving Average (ARIMA) and machine learning algorithms. To forecast future weather conditions, time series analysis is used to examine historical weather data. With this information, farmers can organize their operations and lessen the effect of unfavorable weather events on crop production.

Time series methods help track the characteristics that determine the health of the soil over time, including moisture content, pH, and nutrient levels. Finding trends in soil health data aids in scheduling and fertilization optimization. Through the identification of patterns and trends in resource consumption over time, time series analysis aids in the optimal use of resources, such as water, fertilizers, and pesticides (Table 3).

Table 3. Information and articles on crop selection.

Title	Findings	Methods	Parameters
Future crop projection [60]	Taking climate variability forecast into account and crop productivity in the future	ARIMA	Atmosphere factors like rainfall, temperature
FCrop forecast [61]	Crop productivity is also greatly influenced by environmental factors. Utilising various machine learning algorithms, forecasting crop choice	Recursive feature elimination with bagging results best accuracy algorithm. When compared with other algorithms	N, K, pH, Sulfur, Zinc, etc.

Table 3. *Cont.*

Title	Findings	Methods	Parameters
Crop yield projection [62]	Examination of several machine learning methods for predicting crop projection.	ANN, CNN	temperature, rainfall, and soil type
Crop type classification [26]	Identifying crop types with machine learning methods	Random forest, SVM	Sentinel-2A images
Internet of Things and organizing data [63]	Examine the role that Internet of Things and big data play in smart farming.	Big data and intelligent cropping	Investigating IoT importance
Intelligent farming with Internet of Things [63]	Implementing smart farming with big data without compromising computer performance.	Big data and IoT	Nitrogen, temperature, soil Settings
Smart farming with Internet of Things [64]	Using IoT for smart farming also leads to the most efficient use of energy.	SVM, logistic regression, and random forest classifiers	water capacity, period, soil potency, water quality, temperature

A mixed model utilizing SVM and ARIMA was created by [65]. SVM designed the non-linear portion, and ARIMA is utilized to control time series for a linear model. Results from the mixed model outperformed those from the single model. To forecast the agricultural production of different crops, Yunli et al. [66] used the ARIMA model, considering several variables such as changes in the environment and the use of fertilizers. Further in [60], authors forecasted agricultural output by taking climate variables into account and utilizing ARIMA. The RNN model was shown by [11] to be a competitive model for time series data forecasting.

3. Principles and Methods

The well-known PRISMA standards, which outline how to gather and evaluate data from existing research, are followed in this study. A useful tool for assisting researchers in the compilation of reviews and meta-analyses is the PRISMA statement, a systematic framework consisting of 27 items in the form of a checklist. Every stage of the systematic review process is covered in detail in this part, which is in perfect harmony with the four core stages of the PRISMA approach. The detail of the PRISMA framework is explained below in the form of subsections: Identification Phase, Screening Phase, Eligibility Phase, Inclusion Phase, and PRISMA Overview, which gives a comprehensive grasp of the methodical methodology used in this investigation.

3.1. Identification Phase

The identification phase is the first step in identifying records from a variety of sources. A search string was created to make this procedure easier in Table 4. "Agricultur*" OR "Farm*" AND "Machine Learning" AND "application" OR "implementation" OR "case study" OR "experimental" OR "practical" is the last string.

Table 4. Keywords utilized for extracting records from repositories.

Group 1	Group 2	Group 3
"agricultur*"	"machine learning"	"application"
"farm*"	"deep learning"	"implementation"
"crop"	"time series"	"case study"
		"precision"

The search query was modified to fit the syntax of each digital repository, and it included records indexed in the designated repositories (Web of Science (WoS) and Scopus)

through 15 December 2024. To retrieve all possible variations of a given keyword, the wildcard symbol ("*") was added to the end of some of the terms in the search string. Keywords from separate groups in the search string were connected using the boolean "AND" and keywords from the same group were connected using the boolean "OR".

3.2. Screening Phase

A methodical procedure is used in the screening phase to compare the found records to the preset inclusion and exclusion criteria in Table 5. The choice of well-regarded articles shows a dedication to guaranteeing the best possible quality of research.

Table 5. Criteria for screening of an article.

Criteria	Description
Duplicates	Duplicate records and redundant articles were removed
Article Screening	Must include the title, year, abstract, and DOI
Manuscript Type	The manuscript is selected that has some original contribution.

3.3. Eligibility Phase

Following the initial screening, a thorough evaluation of the complete texts of the remaining records is used to identify the eligible studies. The research is carefully assessed in this step in relation to the predetermined inclusion and exclusion criteria in Table 6.

Table 6. Inclusion criteria for eligibility phase of an article.

Criteria	Description
Research Content	Publications specifically related to smart farming
	Articles with ML-based contents
	Articles with DL-based contents.
Full Text	Must be available.

3.4. Inclusion Phase

The studies that satisfy all inclusion requirements and are incorporated into the SLR are represented by the inclusion phase. The same author manually conducted this screening.

3.5. PRISMA Overview

In accordance with PRISMA criteria (Figure 3), searching the repositories WoS and Scopus, 2732 research articles in total matched the search phrase from Table 4. Justified by the exponential growth of ML adoption in agriculture post-2012, driven by advances in computational power, sensor technologies, and open-source frameworks (e.g., TensorFlow, PyTorch). This period aligns with key milestones, such as the rise of precision agriculture and the proliferation of IoT devices in agriculture. A final selection of 87 articles was found for an in-depth study after a detailed screening and eligibility process explained in Tables 5 and 6, respectively.

The Web of Science (WoS) database was used to gather published articles on the subject. The bibliometric analysis was conducted using the WoS database, specifically the database's primary collection. More than 68 million records from 1900 to the present day are included in this database. By adding the field tags "agricultur*", "farm*", "crop", "machine learning", "deep learning", "time series", "application", "implementation", "case study", and "precision" to the title (TIT) and abstract (ABS), we employed an advanced search. On 15 December 2024, the search was carried out. It contained the Emerging Sources Citation Index (ESCI), the Social Science Citation Index (SSCI), and the Science Citation Index—Expanded (SCI-E).
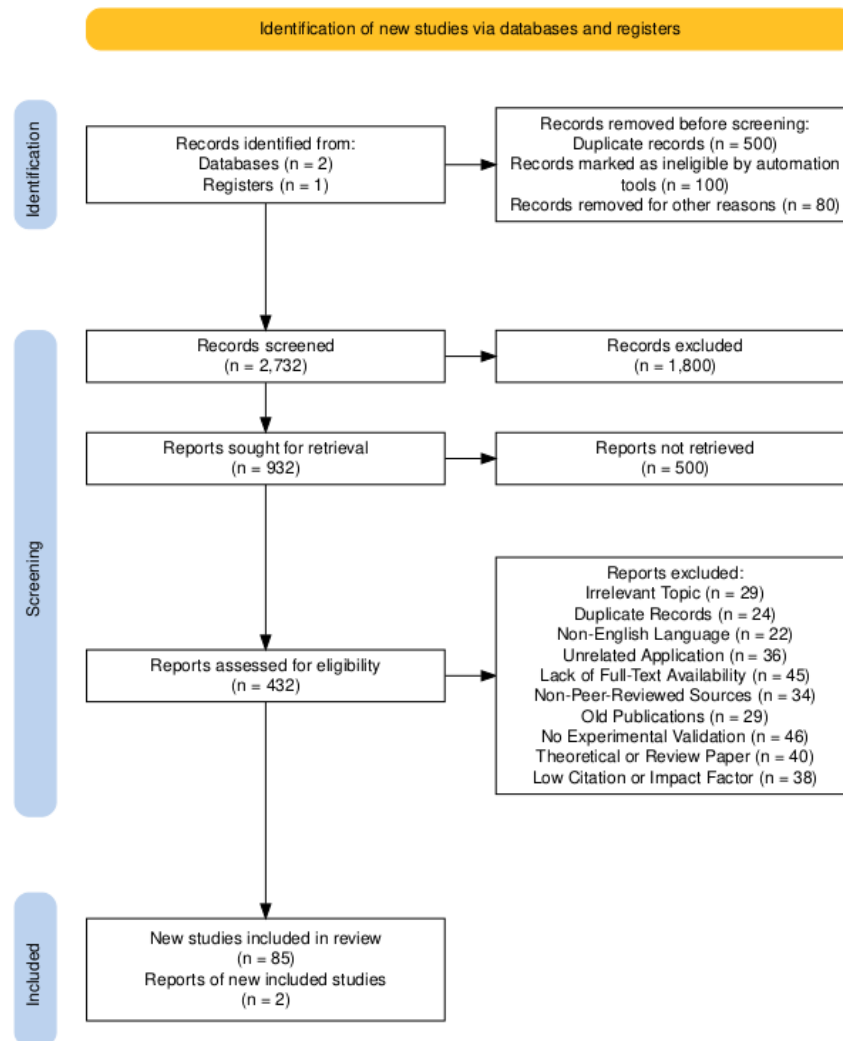
Figure 3. PRISMA flowchart.

4. Impact of Machine Learning Algorithms on Crop Choice and Oversight

ML is used to impart knowledge to machines. By classifying data like training and testing, ML entails transferring knowledge into machines. After using training instances throughout the training phase, the program is used to obtain consistent results for fresh data. Testing cases are used to validate the model after training. Supervised and unsupervised learning are the two primary classifications of ML techniques. In supervised learning, program participants are under the supervision of a supervisor. Various supervised learning techniques are used, including DTs, SVM, KNN, hidden Markov models, Bayesian networks, identification distributions, and others.

Using an approach called unsupervised machine learning, a computer is fed a lot of data to look for patterns in it. Unsupervised methods aid in revealing hidden patterns within the data. Computer science and statistics are combined in machine learning to enhance prediction abilities. KNN, self-organizing maps, hierarchical clustering, partial-based clustering, and K-Means clustering are a few examples. Numerous features are included in historical data, such as pH, temperature, humidity, precipitation, phosphorus, potassium levels, wind speed, zinc, and organic carbon. The features may be in the form of binary or numerical in terms of category. ML is used by the irrigation system in several areas, including soil management, crop management on demand, plant disease detection, and crop quality management.

Figure 4 depicts the ML approaches used in smart farming. A variety of techniques, including deep learning, machine learning, and time series analysis, are included in smart farming. The implementation of smart farming methods might require a dataset. This dataset includes crop factors such as temperature (T), humidity (H), nitrogen (N), potassium (K), and others. It also highlights several machine learning techniques, which are divided into supervised and unsupervised categories according to input and output. This article also offers insights into several ML algorithms. To improve accuracy and speed up learning, crop data will be subjected to DL. Time series analysis will be used to project crop yield and commodity prices in the future.
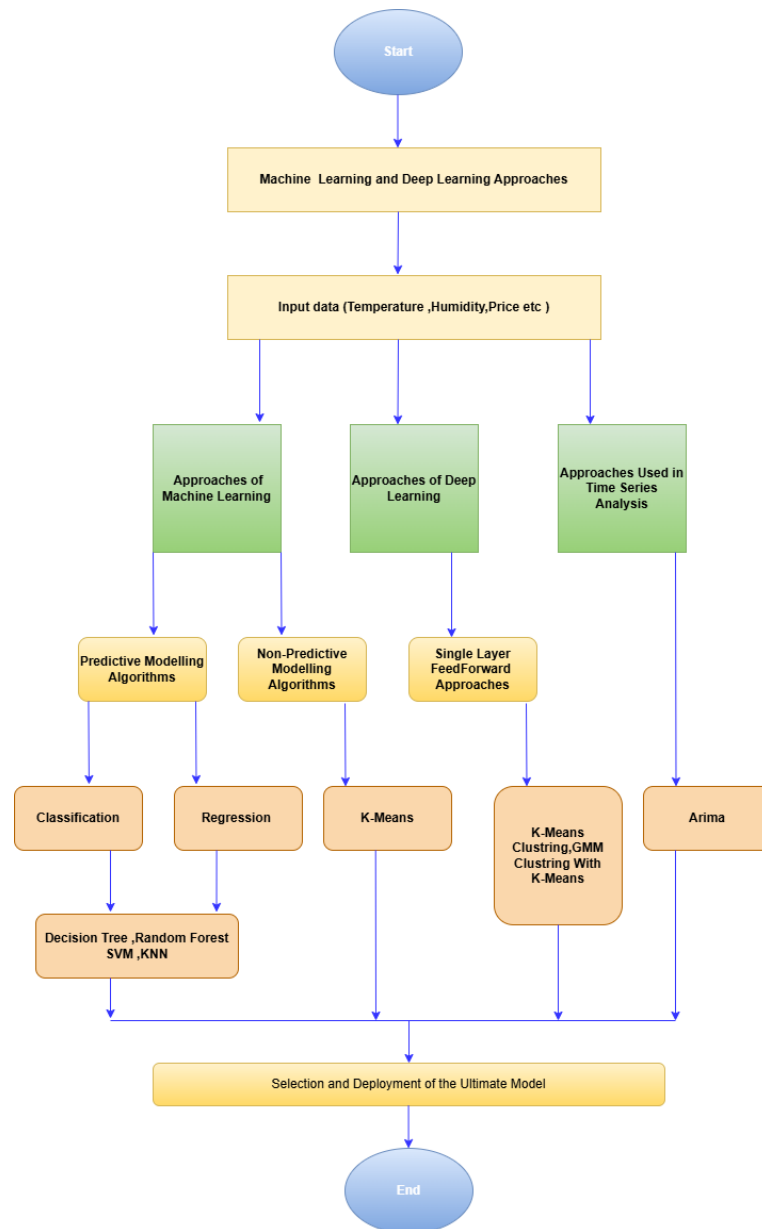

Figure 4. Artificial intelligence methods in precision agriculture.

Gosai et al. [67] developed a crop recommendation system to maximize crop yield using ML. An important factor in Pakistan's employment and GDP is agriculture. It also entails obtaining access to elements like soil via a dataset from a soil testing facility and obtaining crop data from experts in agriculture. Farmers now have a solution thanks to precision agriculture. We will feed the recommendation algorithm with soil data. To provide the farmer with extremely precise and economical crop recommendations, the

system will collect the data and then use SVM and ANN to execute an ensemble model with majority voting.

4.1. Generating a Comprehensible Decision Tree

To process agricultural data in the cloud, an Advanced Decision Tree (ADT) categorization model is created and presented in [31]. Technology development has led to notable advancements in the creation of agricultural software applications that deliver information more quickly. Nonetheless, a lot of farmers still employ traditional farming methods, which have led to comparatively low creation. The importance of fertile soil in maintaining crop growth and increasing yield cannot be overstated. By assisting farmers in identifying soil inadequacies related to nutrient content, soil type, pH value, EC (Electrical Conductivity) value, and soil texture, the soil fertility levels enable them to select the best crops to increase productivity. Initially, this approach uses the Virudhunagar District to forecast the degree of soil fertility. C5.0: The ADT classification approach provides information on soil, planting advice, and crop options.

A decision tree can be easily understood by people if it has an easy-to-understand structure and, more specifically, simple expressions. However, using simple decision trees will cause issues with accuracy. Multivariate Understandable Statistical Tree (MUST), an oblique decision tree split approach developed by researchers in [68], creates decision trees with fewer variables that appear in decision rules.

4.2. Contribution Regarding Support Vector Machine

Using Landsat NDVI data, researchers tested whether support vector machines could recognize crop types that were watered over time [69]. Most agro-environmental assessments require a specific crop variety specific to the site. SVMs were employed by researchers in the Phoenix Active Management Area to attempt to differentiate between different crops in a complex cropping system. Stratified random selection and data from the Landsat NDVI time series dataset were used to choose datasets for SVM training. Local knowledge and stratified random selection were employed. Nine major crop kinds were correctly classified with an accuracy of nearly better than 86% for both training datasets. Intelligent-gent selection reduced training set sizes and improved overall classification accuracy when compared with stratified random approaches. According to a comparative study, standard support vector machines will only consider significant samples and disregard the value of features. Characteristic Weighting is the process of giving each characteristic in a data collection a particular weight based on a set of standards. Ref. [70] made the Relief algorithm's proposal. Weights will be assigned according to the correlation between the qualities using the Relief-F technique. Some qualities that do not meet the predetermined criterion are eliminated. The SVM classifier will perform better if the Relief-F method is applied and uncorrelated or weakly correlated SVM features are eliminated. Crop selection or soil fertility classification may use the Relief-F technique with SVM for improved outcomes.

4.3. KNN Algorithm Based on Multi-Dimension Tree

Data mining techniques were employed by [71] to examine several agricultural yield forecast methodologies. The agricultural system handles a large volume of data produced by numerous components, making it incredibly complex. Crop yield prediction has drawn the interest of consultants, producers, and organizations involved in the agricultural industry. This study examines how data mining methods are applied in the agriculture sector. Relatively modern data mining techniques in agriculture use a variety of techniques, including SVM, K-Means, and KNN. These days, AI is expanding quickly, and data mining has advanced significantly as well. A recent development in agricultural crop yield analysis is data mining, which is the process of identifying patterns hidden in vast amounts of

data. With the available data, yield prediction is a crucial agricultural subject that has not yet been resolved. One difficult problem that can be solved with data mining approaches is yield prediction. Figure 4 highlights the prominent ML and DL approaches for smart farming or agriculture.

KNN is a supervised ML approach that searches for the target among the kNN in the training dataset. As such, it will take a long period to create KNN. A multidimensional binary tree is utilized by the KD-tree (K Dimension tree) to efficiently describe training data. As a result, a groundbreaking method called [10] was introduced to effectively decrease time complexity and describe training data. In a KD-tree, the hyperplane of each non-leaf node can be split into two subspaces, and each subspace can be further divided recursively using the same technique. The left subspace and the right subspace—often referred to as the upper subspace and lower subspace—are the two divisions that apply to all subspaces. Making a KD-tree from a piece of data in K dimensions indicates a split of the K-dimensional space made up of the K-dimensional dataset. In K dimensions, each node in the tree represents a hyper-rectangle area.

5. Impact of Deep Learning Algorithms on Crop Choice and Oversight

In the research arena, there is a strong correlation between AI, ML, and AL. Within ML, which itself is a subset of AI, lies deep learning. Deep learning learns the data by utilizing hidden layers. ML techniques are appropriate for simple data problems. When dealing with complex or disorganized data, deep learning approaches are essential. In a DL architecture, the typical layers are input, hidden, and output. Examples of DL architectures include RNN, CNN, and others. The next subsections look at several DL algorithms that different researchers have suggested for crop management and selection.

5.1. Single-Layer Feedforward Approaches

A novel learning method for single-hidden layer feedforward neural networks (SLFNs) called ELM [72], SLFN is a basic learning technique for single-hidden layer feedforward neural networks. The ELM might, therefore, function well in terms of generality.

As demonstrated by [72], a single-hidden layer feedforward neural network can accurately learn N different examples. Nearly any non-linear activation function and a maximum of N hidden nodes can be included in this neural network. Although conventional gradient-based learning algorithms such as Backpropagation (BP) and its variant, the Levenberg–Marquardt (LM) method, have been used to train multilayer feedforward neural networks, these learning algorithms remain relatively slow and are prone to becoming stuck in local minimums. SVMs are widely employed in algorithm learning and are renowned for their strong generalization capabilities. However, fine-tuning the SVM kernel parameters is very labor-intensive. The unified learning framework for "generalized" SLFN, which includes, but is not limited to, sigmoid networks, RBF networks, threshold networks, trigonometric networks, fuzzy inference systems, etc., was the focus of ELM research from 2001 to 2010. The theory-based universal approximation and classification powers of ELM were proved during those years, according to [73] and other investigations.

Feature learning methods that were often employed in ELM research between 2010 and 2015 were Principal Component Analysis (PCA) and Non-negative Matrix Factorization (NMF). It is shown that instead of using the BlackBox kernel that SVM uses, ELM may provide the WhiteBox kernel mapping, which is accomplished via ELM random feature mapping, and that SVM provides worse solutions than ELM. Examples of unique situations in which ELM uses linear hidden nodes are PCA and NMF. Figure 5 displays the ELM architecture for crop selection. Many aspects are considered when choosing crops, including temperature, humidity, potassium (K), and nitrogen (N).
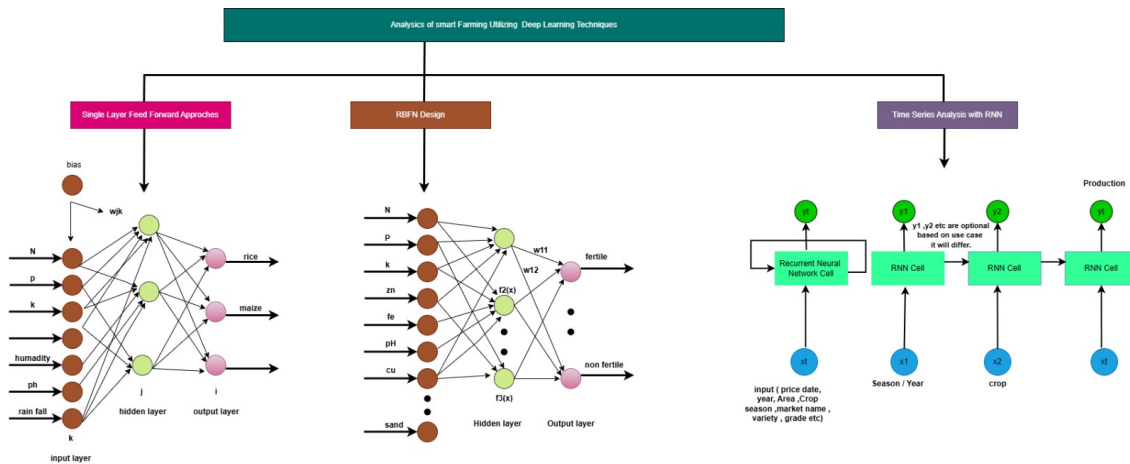
Figure 5. Deep learning techniques for crop selection [9].

The ELM algorithm's fundamental steps are listed below:

- In the first step, take the bias factor and random weight matrices.
- The weight matrix and bias sizes are (j × k) and (1 × k), where j means how many hidden nodes there are and k means how many input nodes.
- Ascertain the hidden layer's output matrix. The output matrix of the first hidden layer is achieved by taking the transpose of the weight matrix and multiplying X, which stands for the training data.
- Decide which feature to activate. It is allowed to use any activation function, such as RELU or SoftMax, and others.
- Find the Moore–Penrose pseudoinverse. There are several methods available for computing the generalized inverse of H by Moore–Penrose. These methods could include, but are not limited to, orthogonal projection, iterative methods, orthogonal value decomposition, and orthogonalization (SVD).
- Find the Moore–Penrose pseudoinverse and the output weight matrix beta, which is a function of the output.

5.2. Clustering Approach with Radial Basis Function

The radial basis function network was created by Broomhead and Lowe in 1988. Even though RBFs only have one hidden layer, they are universal approximators that fasten the convergence process. There are various applications of RBF networks, like function approximation, interpolation, classification, and time series prediction. These applications help achieve a variety of industrial goals, such as forecasting soil fertility and agricultural productivity. Figure 5 displays the RBFN [74] architecture for estimating soil fertility. The weights connecting the input vector to hidden neurons in an RBF architecture represent the center of the relevant neuron. The values of the weights connecting the hidden neurons to the output neurons are chosen to train the network. These weights are preset such that there is space in the receptive fields of the buried neurons.

Since vectors that are adjacent to one another in Euclidean space should fall into the same neuron's receptive field, K-Means clustering is used to locate the centers of hidden neurons. To set the number of cluster centers, select "K". Select K randomly chosen locations from the dataset to act as the K centroids of the data. Determine the centroid in the dataset that is closest to each given position. For each centroid, find the average of all the points that are closest to it. Convert each centroid value to the corresponding average. The range of receptive fields is chosen so that the input vector's domain is entirely enclosed by the neurons' receptive fields. The value of sigma is found using the biggest "d" distance between two buried neurons. The K-Means clustering method is used to

determine the locations of RBF centers. The rate of convergence of RBFs with a single hidden layer is significantly faster than that of multilayer perceptrons (MLPs). In low-dimensional data, RBF networks are generally preferred over MLP when deep feature extraction is not required and the results are directly related to the input vector component. RBFs are resilient learning models in contrast to most machine learning models. They are also universal approximators.

6. Overview of Time Series Analysis on Crop Choice and Oversight

The application of ARIMA and SVM mixed models in agricultural management within the scope of intellectual agriculture [75]. A sophisticated form of contemporary agriculture known as "wisdom agriculture" makes use of numerous scientific and technological advancements, including automated control systems, telecommunications, and an understanding of farm management. Additionally, it is a primary focus of national Internet strategy. This paper employed the ARIMA model to model the agricultural management time series [76] in a linear fashion. The SVM model was then used to describe the time series' non-linear component. At last, the comprehensive forecasts of the two models came to pass. Compared with any model used alone, the combined model is more accurate, precise, and able to provide more information on agricultural production. The ARIMA model for predicting agricultural productivity was studied by authors in [77]. Figure 5 analyzes different deep learning techniques: the single-layer feedforward approach, RBFN, and time series with RNN in smart farming.

Predicting any issue, event, or variable requires a thorough grasp of the elements influencing it; estimations of agricultural crop yield are no exception. India's agricultural output is greatly influenced by several elements, such as enough rainfall, timely use of pesticides and fertilizers, a nice climate and environment, and farmer subsidies.

Because agriculture is a source of income for many people, it is vital and time-consuming to predict agricultural crop yield. Numerous univariate and multivariate time series techniques can be used to forecast these variables. This article forecasted the annual production of a particular agricultural commodity using the ARIMA model. Further, in [78], authors proposed an ARIMA model to predict maize output and cultivated area. An estimated 9952.72 tonnes of maize will be produced, with 6479.8 thousand tonnes as the maximum and 13,425.64 thousand tonnes as the lowest estimate. This forecast is important because it helps formulate sensible policies about the nation's relative production, price structure, and consumption of maize. The results show that the total cultivated area can be increased in the future with the implementation of conservation and land reclamation techniques.

Every state, as well as every input and output, is independent in conventional neural networks. Unlike typical neural networks, RNNs [79] produce the current state by feeding the output of the previous state into it. When comparing different time series analyses, RNN will perform better than the ARIMA model. Using short-term time series data, RNN may generate precise prediction results.

To predict output for the upcoming years, smart farming may employ RNN networks. Projected crop production will decrease food sufficiency in future years. Additionally, producers can utilize RNN networks to forecast agricultural prices and determine the profit or loss of a particular crop in the years to come.

Difficulties Encountered in the Analysis of Time Series Data

In agriculture, time series analysis plays a critical role in estimating future crop yield based on demand. One of the main issues with time series analysis is the overfitting of the data. Overfitting of time series models is a common occurrence; hence, managing it is

an important effort. Managing missing data: Incorrect results will arise if time series data contains any missing values. Thus, the primary goal of data preparation is to prepare data that is devoid of missing values. Short-term forecasts can benefit from using the ARIMA model. Any long-term projections could have inaccurate outcomes. RNN can be applied to any long-term prediction. When utilizing RNN for time series, missing values do not really matter. Compared with ARIMA, the calculation cost for agriculture time series prediction utilizing RNN will be higher.

7. Discussion

The integration of AI into agricultural practices represents a paradigm shift in addressing global challenges such as food security, climate change, and resource optimization. This review highlights the transformative potential of AI-driven technologies, particularly ML, DL, and time series analysis, in advancing sustainable crop production. The findings underscore how these tools enhance decision-making, improve crop yield predictions, and enable precision agriculture. However, their adoption also raises critical questions about scalability, ethical considerations, and the balance between technological innovation and traditional farming practices.

One major obstacle to AI-driven solutions is still their scalability. Smallholder farmers, who make up a sizable share of the world's agricultural workforce, frequently do not have access to high-speed internet, sophisticated equipment, or technical know-how. For instance, in distant areas with unstable electricity, DL models that require GPU clusters are not feasible. Prioritizing mobile-based platforms and lightweight models (such as edge AI and federated learning) is necessary to democratize AI.

Ethical issues are also quite important. When third-party AI companies aggregate farm-specific data, data privacy concerns surface, and farmers may be subject to abuse. Furthermore, traditional or organic farming methods may be marginalized by algorithmic bias, such as models that were mostly trained on data from industrialized farms. Frameworks must be established by policymakers to provide fair access to AI tools while preserving farmer autonomy.

7.1. Current Challenges

Even though ML and DL have significantly improved smart farming, a number of obstacles prevent its widespread use the following:

High-quality, labeled datasets are essential for ML and DL models. However, because of limited historical records, weather fluctuations, and sensor failures, data collection in smart farming is frequently uneven.

- A lot of ML/DL models need a lot of processing power and cloud-based resources, which might be expensive and out of reach for small-scale farmers in isolated places.
- Farmers frequently lack the technical know-how to comprehend intricate ML/DL models, which raises questions about adoption and confidence. Making decisions is made more difficult by the fact that DL models are black boxes.
- Because of differences in climate, soil, and crop kinds, ML/DL models are frequently less effective in diverse agricultural settings than they are in controlled ones.
- Cybersecurity risks, such as data breaches and sensor manipulation, are increased when IoT devices and cloud-based ML/DL systems are integrated into agriculture.
- Standardized rules governing AI-driven smart farming are lacking, especially when it comes to data ownership, bias in AI models, and moral issues with automated decision-making.

7.2. Future Development Directions

Future Development Directions To overcome these challenges and enhance the impact of ML and DL in smart farming, future research and development efforts should focus on the following:

- Developing better interpretable ML/DL models that offer farmers clear and useful insights would boost use and trust in Explainable AI (XAI).
- Improving Federated Learning and Edge Using edge computing with federated learning to implement on-device processing would lessen reliance on cloud computing while maintaining privacy and enabling real-time decision-making.
- In a variety of farming scenarios, combining data from satellites, drones, Internet of Things sensors, and historical records can increase the accuracy and resilience of ML/DL forecasts.
- DL architectures that are optimized for low-power devices will allow for wider deployment, particularly in areas with limited resources.
- Strong encryption, blockchain technology, and anomaly detection techniques can be used to improve data security and stop illegal access to agricultural systems.
- Responsible adoption of AI-driven smart farming depends on the establishment of international regulatory standards that guarantee sustainability, equity, and moral AI use.

8. Conclusions

The agricultural sector has undergone significant transformation through technological advancement, with machine learning (ML) emerging as a critical enabler for optimizing crop management and selection. This systematic review evaluates ML methodologies applied in agricultural research between 2010 and 2023, highlighting prevalent techniques such as SVM, KNN, fuzzy neural networks, Autoregressive Integrated Moving Average (ARIMA), decision trees, ensemble learning, and random forests. Each approach exhibits distinct advantages and limitations, prompting researchers to increasingly adopt hybrid frameworks that integrate multiple ML or deep learning architectures to enhance predictive accuracy and operational efficiency.

To ensure methodological rigor, this review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework, employing a structured protocol for the identification, screening, and inclusion of relevant studies. Additionally, the analysis explored methodologies in Natural Language Processing (NLP) for agricultural data interoperability and evaluated Python-based web development frameworks (V 3.10) for deploying scalable model interfaces.

By synthesizing these insights, this review serves as a valuable resource for researchers seeking to advance precision agriculture. It provides a foundation for developing integrated ML architectures tailored to crop optimization, refining multilingual data translation systems, and designing user-centric web applications for real-time agricultural decision support. The findings underscore the potential for interdisciplinary innovation, encouraging further exploration of synergistic models to address evolving challenges in sustainable crop management.

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The completed PRISMA checklist is provided in the Supplementary Materials (Supplementary File S1). Available online: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097 (accessed on 30 December 2024).

References

1. Niazian, M.; Niedbała, G. Machine learning for plant breeding and biotechnology. *Agriculture* **2020**, *10*, 436. [CrossRef]
2. Anami, B.S.; Malvade, N.N.; Palaiah, S. Classification of yield affecting biotic and abiotic paddy crop stresses using field images. *Inf. Process. Agric.* **2020**, *7*, 272–285. [CrossRef]
3. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [CrossRef]
4. Júnior, I.d.S.T.; Torres, C.M.M.E.; Leite, H.G.; de Castro, N.L.M.; Soares, C.P.B.; Castro, R.V.O.; Farias, A.A. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. *Ecol. Indic.* **2020**, *117*, 106685. [CrossRef]
5. Tarek, Z.; Elshewey, A.M.; Shohieb, S.M.; Elhady, A.M.; El-Attar, N.E.; Elseuofi, S.; Shams, M.Y. Soil erosion status prediction using a novel random forest model optimized by random search method. *Sustainability* **2023**, *15*, 7114. [CrossRef]
6. Dang, C.; Liu, Y.; Yue, H.; Qian, J.; Zhu, R. Autumn crop yield prediction using data-driven approaches:-support vector machines, random forest, and deep neural network methods. *Can. J. Remote Sens.* **2021**, *47*, 162–181. [CrossRef]
7. Rostami, M.A.; Balmaki, B.; Dyer, L.A.; Allen, J.M.; Sallam, M.F.; Frontalini, F. Efficient pollen grain classification using pre-trained Convolutional Neural Networks: A comprehensive study. *J. Big Data* **2023**, *10*, 151. [CrossRef]
8. Jain, S.; Ramesh, D. Machine Learning convergence for weather based crop selection. In Proceedings of the 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 22–23 February 2020; pp. 1–6.
9. Chen, J.; Zeng, Y.; Li, Y.; Huang, G.B. Unsupervised feature selection based extreme learning machine for clustering. *Neurocomputing* **2020**, *386*, 198–207. [CrossRef]
10. Akkem, Y.; Biswas, S.K.; Varanasi, A. Smart farming using artificial intelligence: A review. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105899. [CrossRef]
11. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [CrossRef]
12. Asy'ari, M.Z.; Aten, J.F.C.; Prasetyo, D. Growth predictions of lettuce in hydroponic farm using autoregressive integrated moving average model. *Bull. Electr. Eng. Inform.* **2023**, *12*, 3562–3570. [CrossRef]
13. Dhanya, V.; Subeesh, A.; Kushwaha, N.; Vishwakarma, D.K.; Kumar, T.N.; Ritika, G.; Singh, A. Deep learning based computer vision approaches for smart agricultural applications. *Artif. Intell. Agric.* **2022**, *6*, 211–229. [CrossRef]
14. Saqib, S.E.; Arifullah, A.; Yaseen, M. Managing farm-centric risks in agricultural production at the flood-prone locations of Khyber Pakhtunkhwa, Pakistan. *Nat. Hazards* **2021**, *107*, 853–871. [CrossRef]
15. Agesa, B.L.; Onyango, C.M.; Kathumo, V.; Onwonga, R.; Karuku, G. Climate change effects on crop production in Yatta sub-county: Farmer perceptions and adaptation strategies. *Afr. J. Food Agric. Nutr. Dev.* **2019**, *19*, 14010–14042.
16. Subeesh, A.; Mehta, C. Automation and digitization of agriculture using artificial intelligence and internet of things. *Artif. Intell. Agric.* **2021**, *5*, 278–291. [CrossRef]
17. Khan, N.A.; Gao, Q.; Iqbal, M.A.; Abid, M. Modeling food growers' perceptions and behavior towards environmental changes and its induced risks: Evidence from Pakistan. *Environ. Sci. Pollut. Res.* **2020**, *27*, 20292–20308. [CrossRef] [PubMed]

18. El Aissi, M.E.M.; Benjelloun, S.; Lakhrissi, Y.; Ali, S. A Scalable Smart Farming Big Data Platform for Real-Time and Batch Processing Based on Lambda Architecture. *J. Syst. Manag. Sci.* **2023**, *13*, 17–30.

19. Quy, V.K.; Hau, N.V.; Anh, D.V.; Quy, N.M.; Ban, N.T.; Lanza, S.; Randazzo, G.; Muzirafuti, A. IoT-enabled smart agriculture: Architecture, applications, and challenges. *Appl. Sci.* **2022**, *12*, 3396. [CrossRef]

20. Goel, R.K.; Yadav, C.S.; Vishnoi, S.; Rastogi, R. Smart agriculture—Urgent need of the day in developing countries. *Sustain. Comput. Inform. Syst.* **2021**, *30*, 100512. [CrossRef]

21. SaberiKamarposhti, M.; Why, N.K.; Yadollahi, M.; Kamyab, H.; Cheng, J.; Khorami, M. Cultivating a sustainable future in the artificial intelligence era: A comprehensive assessment of greenhouse gas emissions and removals in agriculture. *Environ. Res.* **2024**, *250*, 118528. [CrossRef]

22. SaberiKamarposhti, M.; Kamyab, H.; Krishnan, S.; Yusuf, M.; Rezania, S.; Chelliapan, S.; Khorami, M. A comprehensive review of AI-enhanced smart grid integration for hydrogen energy: Advances, challenges, and future prospects. *Int. J. Hydrogen Energy* **2024**, *67*, 1009–1025. [CrossRef]

23. Akbar, A.; Kuanar, A.; Patnaik, J.; Mishra, A.; Nayak, S. Application of artificial neural network modeling for optimization and prediction of essential oil yield in turmeric (*Curcuma longa* L.). *Comput. Electron. Agric.* **2018**, *148*, 160–178. [CrossRef]

24. Banerjee, D.; Kukreja, V.; Rana, D.S.; Garg, A. Sugarcane Severity Assessment with Deep Learning and Random Forest: A Multi-class Approach. In Proceedings of the 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 21–22 March 2024; pp. 34–39.

25. Geng, Z.; Duan, X.; Li, J.; Chu, C.; Han, Y. Risk prediction model for food safety based on improved random forest integrating virtual sample. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105352. [CrossRef]

26. Feng, S.; Zhao, J.; Liu, T.; Zhang, H.; Zhang, Z.; Guo, X. Crop type identification and mapping using machine learning algorithms and sentinel-2 time series data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3295–3306. [CrossRef]

27. Loresco, P.J.M.; Tan, G.P.; Dadios, E.P. Hybrid Tree-Fuzzy-Rough Set Decision Support for Determining Plant Growth Using Vision-Based Descriptors. In Proceedings of the 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 12–13 December 2019; pp. 460–465.

28. Abbas, F.; Afzaal, H.; Farooque, A.A.; Tang, S. Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* **2020**, *10*, 1046. [CrossRef]

29. Hamza, K.; Rafia, M.; Noor, A.; Ayesha, H.; Sadaf, M.; Faisal, S.; Sheraz, A.; Malik, M.I.; Dengel, A. Bacterial prediction using internet of things (IoT) and machine learning. *Environ. Monit. Assess.* **2022**, *194*, 133.

30. Mahajan, J.; Banal, K.; Mahajan, S. Estimation of crop production using machine learning techniques: A case study of J&K. *Int. J. Inf. Technol.* **2021**, *13*, 1441–1448.

31. Rajeswari, S.; Suthendran, K. C5. 0: Advanced Decision Tree (ADT) classification model for agricultural data analysis on cloud. *Comput. Electron. Agric.* **2019**, *156*, 530–539. [CrossRef]

32. Babu, C.M.; Nikhila, R.; Purnachand, K.; Shaik, A.; Tara, S. Crop prediction based on soil types using machine learning. *AIP Conf. Proc.* **2023**, *2477*, 030063.

33. Basha, S.M.; Rajput, D.S.; Janet, J.; Somula, R.S.; Ram, S. Principles and practices of making agriculture sustainable: Crop yield prediction using Random Forest. *Scalable Comput. Pract. Exp.* **2020**, *21*, 591–599. [CrossRef]

34. Koc, I.; Atay, Y.; Babaoglu, I. Discrete tree seed algorithm for urban land readjustment. *Eng. Appl. Artif. Intell.* **2022**, *112*, 104783. [CrossRef]

35. Nagasubramanian, K.; Jones, S.; Singh, A.K.; Sarkar, S.; Singh, A.; Ganapathysubramanian, B. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* **2019**, *15*, 98. [CrossRef]

36. Sun, X.; Guo, M.; Ma, M.; Mankin, R.W. Identification and classification of damaged corn kernels with impact acoustics multi-domain patterns. *Comput. Electron. Agric.* **2018**, *150*, 152–161. [CrossRef]

37. de Alwis, S.; Eichhorn, A.; Held, A.; Pawlowski, J.M.; Schiffer, M.; Versteegen, F. Asymptotic safety, string theory and the weak gravity conjecture. *Phys. Lett. B* **2019**, *798*, 134991. [CrossRef]

38. Fawzy, M.; Mahmoud, A.E.D.; Hosny, M.; El Maghrabi, N.; Abdelfatah, A.M. Artificial Intelligence for Sustainable Water Management and Treatment. In *Artificial Intelligence and Modeling for Water Sustainability*; CRC Press: Boca Raton, FL, USA, 2023; pp. 37–52.

39. Maraveas, C.; Karavas, C.S.; Loukatos, D.; Bartzanas, T.; Arvanitis, K.G.; Symeonaki, E. Agricultural greenhouses: Resource management technologies and perspectives for zero greenhouse gas emissions. *Agriculture* **2023**, *13*, 1464. [CrossRef]

40. Reddy, N.K.; Deepa, K.; Lekshmi, S.; Özer, T. Image Processing Based Pest Detection and Location Sharing to Farmers. In Proceedings of the 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 26–28 May 2023; pp. 1–7.

41. De Alwis, S.; Hou, Z.; Zhang, Y.; Na, M.H.; Ofoghi, B.; Sajjanhar, A. A survey on smart farming data, applications and techniques. *Comput. Ind.* **2022**, *138*, 103624. [CrossRef]

42. Guan, S.; Wang, Z.; Cao, Y. A novel blockchain-based model for agricultural product traceability system. *IEEE Commun. Mag.* **2023**, *61*, 124–129. [CrossRef]

43. Mekonnen, Y.; Namuduri, S.; Burton, L.; Sarwat, A.; Bhansali, S. Machine learning techniques in wireless sensor network based precision agriculture. *J. Electrochem. Soc.* **2019**, *167*, 037522. [CrossRef]

44. Xiong, J.; Lipsitz, O.; Nasri, F.; Lui, L.M.; Gill, H.; Phan, L.; Chen-Li, D.; Iacobucci, M.; Ho, R.; Majeed, A.; et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *J. Affect. Disord.* **2020**, *277*, 55–64. [CrossRef]

45. Kang, M.; Wang, F.Y. From parallel plants to smart plants: Intelligent control and management for plant growth. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 161–166. [CrossRef]

46. De Alwis, S.; Zhang, Y.; Na, M.; Li, G. Duo attention with deep learning on tomato yield prediction and factor interpretation. In Proceedings of the PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, 26–30 August 2019; Proceedings, Part III 16; Springer: Cham, Switzerland, 2019; pp. 704–715.

47. Tandekar, D.; Dongre, S. A Review on Various Plant Disease Detection Using Image Processing. In Proceedings of the 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 19–20 June 2023; pp. 552–558.

48. Aquino, A.; Diago, M.P.; Millán, B.; Tardáguila, J. A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosyst. Eng.* **2017**, *156*, 80–95. [CrossRef]

49. Muangprathub, J.; Boonnam, N.; Kajornkasirat, S.; Lekbangpong, N.; Wanichsombat, A.; Nillaor, P. IoT and agriculture data analysis for smart farm. *Comput. Electron. Agric.* **2019**, *156*, 467–474. [CrossRef]

50. Bakhshipour, A.; Jafari, A.; Nassiri, S.M.; Zare, D. Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* **2017**, *157*, 1–12. . [CrossRef]

51. Deng, B.; Zhang, X.; Gong, W.; Shang, D. An overview of extreme learning machine. In Proceedings of the 2019 4th International Conference on Control, Robotics and Cybernetics (CRC), Tokyo, Japan, 27–30 September 2019; pp. 189–195.

52. Wang, L.; Khishe, M.; Mohammadi, M.; Mahmoodzadeh, A. Extreme learning machine evolved by fuzzified hunger games search for energy and individual thermal comfort optimization. *J. Build. Eng.* **2022**, *60*, 105187. [CrossRef]

53. Amirian, M.; Schwenker, F. Radial basis function networks for convolutional neural networks to learn similarity distance metric and improve interpretability. *IEEE Access* **2020**, *8*, 123087–123097. [CrossRef]

54. Mishra, S.; Dash, P. Short-term prediction of wind power using a hybrid pseudo-inverse Legendre neural network and adaptive firefly algorithm. *Neural Comput. Appl.* **2019**, *31*, 2243–2268. [CrossRef]

55. Rani, S.; Mishra, A.K.; Kataria, A.; Mallik, S.; Qin, H. Machine learning-based optimal crop selection system in smart agriculture. *Sci. Rep.* **2023**, *13*, 15997. [CrossRef]

56. Kim, J.; Yu, B.; O'Hara, S. Lstm filter for smart agriculture. *Procedia Comput. Sci.* **2022**, *210*, 289–294. [CrossRef]

57. Cordeiro, M.; Markert, C.; Araújo, S.S.; Campos, N.G.; Gondim, R.S.; Da Silva, T.L.C.; da Rocha, A.R. Towards Smart Farming: Fog-enabled intelligent irrigation system using deep neural networks. *Future Gener. Comput. Syst.* **2022**, *129*, 115–124. [CrossRef]

58. El-Ghamry, A.; Darwish, A.; Hassanien, A.E. An optimized CNN-based intrusion detection system for reducing risks in smart farming. *Internet Things* **2023**, *22*, 100709. [CrossRef]

59. Apat, S.K.; Mishra, J.; Raju, K.S.; Padhy, N. The robust and efficient Machine learning model for smart farming decisions and allied intelligent agriculture decisions. *J. Integr. Sci. Technol.* **2022**, *10*, 139–155.

60. Praveen, B.; Sharma, P. Climate variability and its impacts on agriculture production and future prediction using autoregressive integrated moving average method (ARIMA). *J. Public Aff.* **2020**, *20*, e2016. [CrossRef]

61. Suruliandi, A.; Mariammal, G.; Raja, S. Crop prediction based on soil and environmental characteristics using feature selection techniques. *Math. Comput. Model. Dyn. Syst.* **2021**, *27*, 117–140. [CrossRef]

62. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]

63. Ouafiq, E.M.; Saadane, R.; Chehri, A. Data management and integration of low power consumption embedded devices IoT for transforming smart agriculture into actionable knowledge. *Agriculture* **2022**, *12*, 329. [CrossRef]

64. Phasinam, K.; Kassanuk, T.; Shabaz, M. Applicability of internet of things in smart farming. *J. Food Qual.* **2022**, *2022*, 7692922. [CrossRef]

65. Wen, Q.; Wang, Y.; Zhang, H.; Li, Z. Application of ARIMA and SVM mixed model in agricultural management under the background of intellectual agriculture. *Clust. Comput.* **2019**, *22*, 14349–14358. [CrossRef]

66. Guo, Y. Integrating genetic algorithm with ARIMA and reinforced random forest models to improve agriculture economy and yield forecasting. *Soft Comput.* **2024**, *28*, 1685–1706. [CrossRef]

67. Gosai, D.; Raval, C.; Nayak, R.; Jayswal, H.; Patel, A. Crop recommendation system using machine learning. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2021**, *7*, 558–569. [CrossRef]

68. Hirsch, H.; Richardson, D.M.; Pauchard, A.; Le Roux, J.J. Genetic analyses reveal complex introduction histories for the invasive tree Acacia dealbata Link around the world. *Divers. Distrib.* **2021**, *27*, 360–376. [CrossRef]

69. Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 103–112. [CrossRef]

70. Huang, C.; Huang, X.; Fang, Y.; Xu, J.; Qu, Y.; Zhai, P.; Fan, L.; Yin, H.; Xu, Y.; Li, J. Sample imbalance disease classification model based on association rule feature selection. *Pattern Recognit. Lett.* **2020**, *133*, 280–286. [CrossRef]

71. Medar, R.; Rajpurohit, V.S. Supervised machine learning techniques for predicting sugarcane yield. *Int. J. Recent Technol. Eng.* **2019**, *8*, 5662–5668. [CrossRef]

72. Maitah, M.; Malec, K.; Ge, Y.; Gebeltová, Z.; Smutka, L.; Blažek, V.; Pánková, L.; Maitah, K.; Mach, J. Assessment and prediction of maize production considering climate change by extreme learning machine in Czechia. *Agronomy* **2021**, *11*, 2344. [CrossRef]

73. Akyol, K. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Syst. Appl.* **2020**, *140*, 112875. [CrossRef]

74. Ali, Z.; Muhammad, A.; Adnan, R.; Alkhalifah, T.; Aslam, S. Evaluating Machine Unlearning: Applications, Approaches, and Accuracy. *Eng. Rep.* **2025**, *7*, e13081. [CrossRef]

75. Varsha, M.; Poornima, B.; Pavan Kumar, M.; Basavarajappa, S. Novel hybrid ARIMA–BiLSTM model for forecasting of rice blast disease outbreaks for sustainable rice production. *Iran J. Comput. Sci.* **2023**, *6*, 147–159. [CrossRef]

76. Noorunnahar, M.; Chowdhury, A.H.; Mila, F.A. A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *PLoS ONE* **2023**, *18*, e0283452. [CrossRef]

77. Shankar, S.V.; Ajaykumar, R.; Ananthakrishnan, S.; Aravinthkumar, A.; Harishankar, K.; Sakthiselvi, T.; Navinkumar, C. Modeling and forecasting of milk production in the western zone of Tamil Nadu. *Asian J. Dairy Food Res.* **2023**, *42*, 427–432. [CrossRef]

78. Mahmoud Sayed Agbo, H. Forecasting agricultural price volatility of some export crops in Egypt using ARIMA/GARCH model. *Rev. Econ. Political Sci.* **2023**, *8*, 123–133. [CrossRef]

79. Ganatra, N.; Patel, A. Deep learning methods and applications for precision agriculture. In *Machine Learning for Predictive Analysis, Proceedings of ICTIS 2020, Ahmedabad, India, 2020*; Springer: Singapore, 2021; pp. 515–527.

80. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]