

Article

Crop prediction model using machine learning algorithms

Ersin Elbasi ^{1,*}, Chamseddine Zaki ¹, Ahmet E. Topcu ¹, Wiem Abdelbaki ¹, Aymen I. Zreikat ¹, Elda Cina ¹, Ahmed Shdefat ¹ and Louai Saker ¹

¹ College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait

* Correspondence: ersin.elbasi@aum.edu.kw

Abstract: This research investigates the potential benefits of integrating machine learning algorithms and IoT sensors in modern agriculture. The focus is on optimizing crop production and reducing waste through informed decisions about planting, watering, and harvesting crops. The paper discusses the current state of machine learning and IoT in agriculture, highlighting key challenges and opportunities. It also presents experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms. The findings recommend that by analyzing wide-ranging data collected from farms, including real-time data from IoT sensors, farmers can make more informed verdicts about factors that affect crop growth. Eventually, the integration of these technologies can transform modern agriculture by increasing crop yields while minimizing waste. In our studies, we achieve a classification accuracy of 99.59% using the Bayes Net algorithm and 99.46% using Naïve Bayes Classifier, and Hoeffding Tree algorithms. Our results indicate that we achieved high accuracy results in our experiments in order to increase crop growth.

Keywords: crop prediction; machine learning; feature selection; artificial intelligent; smart farming

1. Introduction

Agriculture is a vital product that has a significant role in nourishing the growing population of the world. To keep pace with the increasing demand for foodstuff, farmers need to make the best use of them to reap output while minimizing losses. Forecasting and examining reap growth is a serious part of modern agriculture, and machine learning has become a powerful tool to achieve this goal line [1,2]. Smart farming, or precision agriculture, is a modern farming conduct that utilizes recent technology to optimize reap production and minimize waste. The objective of smart farming is to increase reap output while minimizing the use of resources such as water, fertilizer, and energy [3].

In the field of smart farming, the Internet of Things (IoT) is considered one of the key contributing technologies used. IoT sensors can be utilized to monitor soil moisture, temperature, and other environmental aspects [4] the gathered data from the IoT sensors can be used to define the best time to plant, water, and harvest reaps. By using IoT sensors, farmers can guarantee that the reaps get the right amount of water and nutrients, which can improve their quality and yield [5,6]. Figure 1 illustrates IoT and machine learning based crop analysis and prediction process.

Machine learning is another technology used in smart farming. Machine learning algorithms can analyze vast amounts of data collected by IoT sensors and other sources. It is a rapidly growing field that has the potential to transform the way we predict and analyze crop growth and output. Machine learning algorithms use statistical/mathematical models and algorithms to analyze data and make predictions, enabling computer systems to learn and improve from experience without being explicitly programmed [7]. In the agriculture field, precisely in the cultivation area, machine learning algorithms can be trained on comprehensive data collected from farms, such as weather patterns, soil properties, crop growth stages, and pest and disease outbreaks. By evaluating the collected data, machine learning models can forecast reap growth, output, and quality with high accuracy [8].

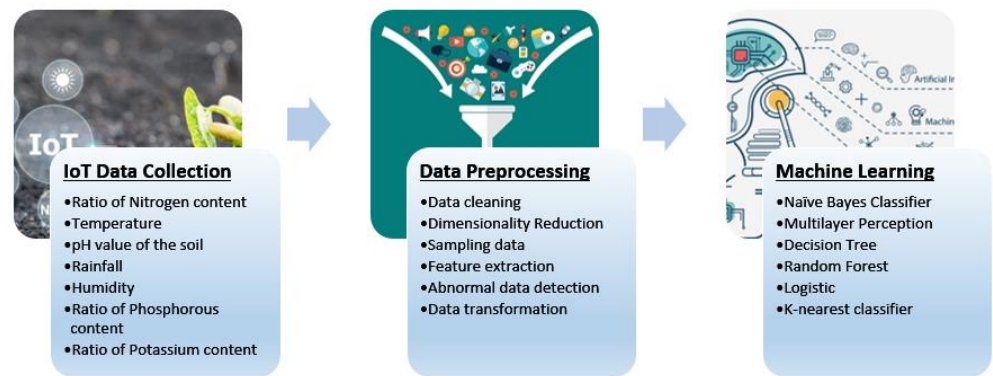


Figure 1. IoT and machine learning based crop analysis and prediction process

A new noteworthy application of machine learning in the agriculture field is precision farming, which includes the employment of data and technology to optimize agricultural conducts such as fertilization, irrigation, and pest control to improve reaped output and quality. Machine learning models can examine bulky amounts of data from several sources, such as satellite imagery, drone footage, and soil sensors, to craft comprehensive maps of reaped growth, nutrient levels, and moisture content. Farmers can utilize these maps to regulate their farming conducts, such as applying fertilizer or watering specific areas of the field, to maximize reaped output and minimize waste [9]. Machine learning can also help farmers identify the most profitable crops to plant based on market demand and environmental factors. By analyzing historical market data and weather patterns, machine learning models can predict the demand for different crops and suggest the optimal planting times and locations [10]. This can help farmers maximize their profits while minimizing the risk of crop failure. In addition to predicting crop growth and output, machine learning can also analyze the quality of the harvested crops. Machine learning models can analyze the color, texture, and shape of fruits and vegetables to determine their ripeness and quality. This information can be used to optimize the harvesting process and ensure that only high-quality produce is sold to consumers [11,12].

There are a number of encounters for deploying machine learning in the agriculture field, such as the lack of data groundwork, the high cost of sensors and other technology, and the need for specialized proficiency to develop and maintain the models [13]. However, as more farms implement precision agriculture conducts and gather data, the potential profits of deploying machine learning in agriculture will become more evident. It is worth mentioning that machine learning in the field of agriculture is still in its early stages and more research needs to be done in this area to fully realize the potential of this technology. However, the results that have been achieved so far are promising, and it is likely that machine learning will become increasingly important [14].

In this study, the authors explore the effects of machine learning on multiple industries and present an overview of the methodologies utilized in various research studies. They emphasize the significance of gathering and analyzing precise data by applying machine learning algorithms to construct models that can correctly predict labels based on the input data. The study discusses several classification algorithms, such as Decision Tree, Naïve Bayes Classifier, Support Vector Machine, and Random Forest, which can be employed to build such models. Initially, the model is developed by training it with training data, and then the results are evaluated using test data to ensure the accuracy and desired outcomes of the predictions. The authors predict that the adoption of machine learning-focused systems and solutions will significantly enhance efficiency and productivity, resulting in massive changes in various industries.

The paper is structured as follows: First, the background reading and literature review are presented in Section 1 to provide a comprehensive understanding of the topic. Then, Section 2 introduces the artificial intelligence in smart farming to delve deeper into the subject matter. Section 3 about crop analysis and prediction- benefits and challenges.

Section 4 outlines the methodologies used in this study, followed by the presentation of the experimental results in Section 5. Lastly, Section 6 is dedicated to the conclusion and future recommendations. Additionally, the subsequent section will cover the related work.

2. Literature Review

Machine learning approaches and algorithms are utilized in the crop yield prediction methods to improve the quality of the crop so that the farmer's profit is maximized and the quality of the agriculture sector is improved, and hence, the overall economy is enhanced. This issue has been discussed in detail in the literature [15-29]. In [15], a review of machine learning algorithms to predict palm oil yield prediction crop is discussed. The authors conducted a comparative analysis of the related work focusing on the advantages, disadvantages, and limitations of the suggested approaches. Furthermore, based on the discussion and evaluation of the existing studies, the authors provided a new architecture based on machine learning methods to predict palm oil yield prediction crop.

The authors in [16] focused on crop prediction and yield through the study of the quality of the soil, considering that soil properties have a major effect on crop production. The authors studied different soil properties such as; NPK (Nitrogen, Phosphorous, Potassium) levels, temperature, rainfall, moisture, PH value, and humidity. Comparative analyses are conducted concerning three machine learning algorithms; namely, Naïve Bayes, Logistic Regression, and Random Forest. Moreover, the authors conducted a comparison between these algorithms concerning accuracy. A crop production model proposed in [17] aims to manage the produced crop using machine-learning algorithms to help farmers in developing countries who are still using traditional methods and are still not able to recognize the correct market value of their products. The proposed system is based on three scenarios; firstly, choosing the best crops based on the farmer's location, secondly, providing guidance on soil preparation, and thirdly, providing the best way of crop marketing from farmer to consumer. The authors applied Support Vector Regression, Voting Regression techniques, and Random Forest Regression algorithms along with proper real data for climate, weather, and soil.

Due to the scarcity of natural resources around the world, the authors in [18] proposed to utilize supervised machine learning algorithms such as K-Nearest Neighbor Support Vector Machine, Random Forest, and Artificial Neural Network to help farmers make the proper decision regarding crop selection and production and therefore, the overall country's economic status will be improved. Observing the growth process of Chili and cotton crops using mobile phone images and machine learning techniques is the subject of the study in [19]. The authors suggested a prediction method using the following supervised machine-learning algorithms; Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor (K-NN), Gaussian Naïve Bayes (NB) and logistic regression (LR). The authors claimed that the study might help in the smart-farming process by recognizing the best machine-learning algorithm for better crop prediction and analysis, especially for Chili and cotton crops. In [20], a dataset for soil prediction is collected from Tamil Nadu Agricultural University (TNAU), India, which involves 32 districts. Based on the comparative analysis of different machine learning algorithms such as Naïve Bayes, Bayes Net, and Instance-Based Learner (IBK) algorithms, the authors claimed that the provided comparative results help farmers to make the proper decision regarding crop selection and production. The authors in [21] recommended 22 types of crops in this study and proposed a three steps framework. Firstly, data preprocessing and feature extraction, secondly, classification, and finally performance evaluation. As a result of the comparison, the authors claimed that the best classifier for this problem is Naïve Bayes with an accuracy of 99.45%. This research work would provide a better outcome if the authors did the classification and performance evaluation for a real problem.

Since crop monitoring is considered the main domain in the smart farming process and crop diseases are the main reason for yield losses, especially in developing countries,

the authors in [22] provided a comprehensive survey on crop monitoring techniques concerning crop yield estimation and disease detection using deep learning models. Based on the obtained results and comparison, the authors claimed that crop monitoring techniques using deep learning methods are more accurate and powerful compared to the traditional methods used by some developing countries. In [23], the authors proposed a novel technique based on support vector machine (SVM) for auxiliary information on real applications of the agriculture sector. The authors claimed that they obtained an accuracy of 91% compared to the existing applications. This proposed methodology can be used by the farmers to gain a better yield of the crops and can be used by different governmental sectors to improve crop productivity. However, the authors did not suggest any recommendations for fertilizer systems to improve crop management.

In [24], a machine learning and multi-perception method along with the quantity of rainfall have been suggested by the authors to help farmers in making a proper decision regarding harvest even before they start planting. Furthermore, the suggested method focuses on the optimal process for marketing and storage of the crop. Based on the provided results, the authors claimed that the proposed method would be very useful for farmers to improve agricultural yield outcomes. In [25], the main objective of this research is to predict crop productivity loss by utilizing linear regression methods based on data taken from the previous year's statistics. The models are created using real-world data and the evaluation process was based on samples. The authors applied both Naive Bayes and Decision tree algorithms in this evaluation. They claimed that the proposed method improves the production process and maximizes the farmer's profit. In [26], the authors have developed a web-based application for crop yield prediction to be used by farmers. This tool provides farmers with a list of various crops that have been planted previously, to predict and learn about the best crop to cultivate in the future. Furthermore, the tool can provide farmers with climate data and information to help them to take the best decision regarding the market demand and prices.

In [27], the author's survey's main objective is to evaluate the performance of different publications from the year 2016-2020 that aim to predict fungal illnesses on the crops. The authors have evaluated different machine learning algorithms utilized in the literature. As per the provided comparative results, the authors concluded that the best performance among all machine learning models, SVM, variations of choice trees, and Naive Bayes has been widely utilized and gained the best results regarding the yearly prediction of crop diseases. In [28], using machine learning, the authors proposed a system for the early prediction of crop diseases in plants by utilizing Convolutional Neural Network (CNN) method. The dataset that is taken from a village is trained and tested. Different diseases are collected in a database and the classifier is trained to compare the accuracy and choose the one with high accuracy. The provided model helps farmers to predict plant diseases and take the best decision regarding the type of crop to be planted. The authors in [29] recognized problems facing farmers in India regarding crop yield prediction, therefore, they have collected datasets that are published online. To facilitate analyzing and studying data, the dataset is clustered using the K-means clustering algorithm, and the Naive Bayes algorithm is used to recognize the best crop to plant. The provided analysis and results show that the proposed system is very helpful for farmers, not only for early prediction of yield crops but also for selecting the best crop to plant.

3. Artificial Intelligent in Smart Farming

Smart farming has brought about a significant transformation in agriculture by utilizing advanced technologies to maximize productivity, increase sustainability, and reduce environmental impact. A critical aspect of this revolution is machine learning, which drives various applications that enhance decision-making and streamline agricultural operations. In this section, the discussion delves into three essential areas where machine learning plays a crucial role: A) Predictive Analytics, employed for forecasting crop yields

and identifying potential threats; B) Decision Support Systems, offering actionable insights and tailored recommendations for farmers; and C) Image and Data Analysis, allowing for accurate monitoring of crop health, growth, and overall farm performance. This examination of machine learning in smart farming showcases the intersection of innovation and tradition, ultimately shaping the future of agriculture.

The predictive analysis involves gathering insights from past data by applying advanced methods such as statistical/mathematical modeling and machine learning. The goal is to make well-informed forecasts about future occurrences or actions, which can then guide strategic decisions and focused efforts. It is widely employed in smart farming for a variety of purposes, including yield prediction. Based on historical and current AgriData, it predicts future crop production. These forecasts benefit not only farmers, but also policymakers, governments concerned with food security, and food marketing organizations [30]. These entities can use the insights provided by yield predictions to make data-driven decisions and develop strategies for efficient resource allocation, food distribution, and price stabilization. This leads to a more resilient food system that can be assured by anticipating changes in crop production.

Crop yield forecasting can be accomplished using a wide range of machine learning techniques, such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests (RFs)[31, 32, 33]. These algorithms' capacity to process both historical and up-to-date information on aspects like weather, soil conditions, and crop health allows for accurate crop yield predictions. The resulting insights empower farmers to make well-informed choices about planting, watering, and fertilizing, which ultimately leads to optimized yields and less resource squandering.

SVM is a linear classifier based on the margin maximization principle. They minimize structural risk to improve the complexity of the classifier and achieve excellent generalization. SVMs construct hyperplanes that separate data into two categories, in a higher dimensional plane [34]. Because of the sophisticated classification abilities of these algorithms, they have been useful for a variety of agricultural tasks, such as automatic fruit counting and identification [31]. SVMs have been proven capable of assessing fruit maturity. In one study [32], were employed to identify unripe green citrus fruits in their natural environment. In another study, they were used to predict the growth stages [33].

Artificial neural networks are inspired by the functionality of the human brain, simulating functions such as pattern generation, learning, and decision-making. ANN can serve both as supervised models that are commonly used for regression and classification problems, as well as unsupervised models that operate without labeled data, identifying patterns and relationships within the data. Their flexibility makes them very useful for yield prediction. They are used to estimate grassland biomass in Irish farms [35] as well as to predict yield variations within fields [36].

Both SVMs and ANNs are effective in detecting pests and diseases. For example, SVMs have been used to classify parasites and detect thrips in strawberries [37] as well as to identify Bakanae disease in rice seedlings [38] and to classify and identify damaged corn kernels with images as inputs [39]. Support Vector Machines-Radial Basis Function (SVM-RBF) is applied to remote sensing and satellite imagery to process and analyze data on nutrient levels, soil moisture, and crop health. SVM-RBF classifiers are used to assess the quality of malting barley using image data and descriptors like color, shape, and texture [40]. It has also been demonstrated that ANNs can also be useful when detecting diseases and pests. For instance, it distinguishes healthy *Silybum marianum* plants from those affected by the Smut fungus [41]. In addition, it is also demonstrated to be efficacious at detecting various conditions in winter wheat canopy conditions as well [32].

A further variation of ANN is known as deep learning systems (DLs) or deep neural networks (DNNs), which are a variation of artificial neural networks [42]. They are a relatively new area of ML research. They focus on computational models composed of multiple processing layers, to learn complex data representations using multiple abstraction levels. DNNs have proven to be a significant resource for advanced farming, thanks to

their proficiency in managing intricate and diverse datasets. One major reason deep learning thrives in this area is its skill in dealing with vast quantities of data. Today's farms generate an enormous amount of information from a range of sources like satellite images, meteorological data, soil monitoring devices, and IoT gadgets.

The use of DL models makes it possible to efficiently analyze these complex datasets to extract meaningful insights and patterns that can be critical for identifying and addressing diseases and pests. Furthermore, deep learning models, specifically Convolutional Neural Networks (CNNs), demonstrate exceptional proficiency in feature extraction, a crucial aspect of disease and pest detection in smart farming [42]. These models automatically identify vital features from data. This enables the detection of plant diseases from images or crop health analysis based on spectral data. A DNN and a CNN were employed to detect and diagnose plant diseases using leaf images [42] reaching an accuracy of 99.53 %. Therefore, this automatic feature extraction not only streamlines the process of identifying potential threats but also aids in the development of targeted and timely interventions. This contributes to improved agricultural productivity and sustainability.

A second field that has been found very efficient for smart farming is Decision Support Systems (DSS). DSS has embraced various machine learning techniques to enhance numerous aspects of farm management. A prime example is evapotranspiration estimation, where Regression and Multivariate Adaptive Regression Splines (MARS) have accurately estimated monthly mean reference evapotranspiration in arid and semi-arid regions [43]. In tandem with this, ANNs have been used for daily [44] and weekly [45] evapotranspiration estimation based on meteorological data, as well as predicting daily dew point temperature using weather data inputs [46].

Another field that has shown potential in revolutionizing agriculture is computer vision and image analysis. Image and data analysis have become essential driving forces behind agricultural practices transformation. The vast amounts of data generated by remote sensing, on-field sensors, and other sources are beneficially processed by farmers using high-tech solutions and intricate analytical techniques. Adopting this data-centric strategy allows for better decision-making, which in turn leads to enhanced resource utilization, healthier crops, and a general improvement in agricultural output.

Robotics continues to gain significance, especially when it comes to the agricultural use of computer vision and pattern recognition. Among the fields where these advancements have been most significant is harvesting robot development. These robots use advanced machine learning techniques and farm data to optimize their performance. For example, a recent study [47] unveiled a tea plucking robot that used branch stiffness as input data and a Gaussian distribution mechanism. Another study designed strawberry-picking and cucumber-detecting robots using CNN and fruit color and shape as input data [48]. In addition, other studies [49,50] have integrated RGB images with SVM and deep CNN, respectively, to develop innovative harvesting robots. By utilizing computer vision and machine learning, these robots can work more efficiently and accurately, ultimately improving agriculture productivity and profitability.

In conclusion, machine learning has emerged as a transformative force in intelligent agriculture, fueling innovation via predictive analytics, decision assistance systems, and image and data examination. Utilizing these sophisticated instruments, farmers can make knowledgeable choices, streamline resource use, and attain greater sustainability. As the agricultural environment progresses, machine learning will surely take on a more critical function in defining the future of farming, guaranteeing food safety, and encouraging environmental responsibility.

4. Crop Analysis and Prediction- Benefits and Challenges

While Machine learning is being exploited in multiple fields, it remains an active area of research and a challenging one in the agricultural domain. In this section, we have summarized the main benefits and challenges ML faces when used in crop analysis and predictions based on recent research [51,52,53,54,55]. Figure 2 shows AI-based crop analysis and prediction phases.

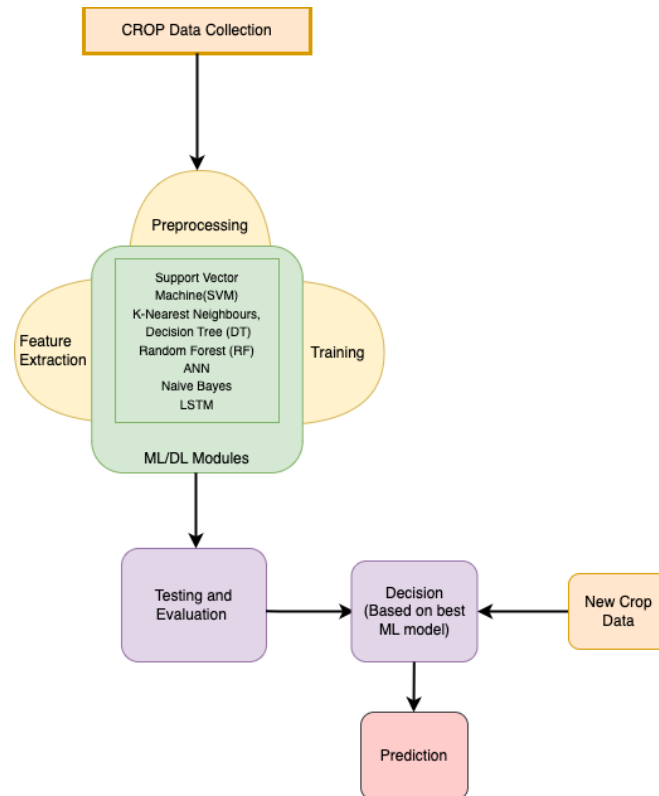


Figure 2. AI-based crop analysis and prediction

4.1. Benefits

The following are the key advantages farmers can get from utilizing machine learning on their farms:

- **More effectiveness:** This approach is more effective and accurate in identifying patterns and saving farmers' time and resources because a larger volume of data can be evaluated by machine learning in a shorter amount of time than with previous methods.
- **Increased crop yield:** Using many data sources for analysis, including weather patterns, soil quality, and historical machine learning algorithms, can help farmers make more informed decisions that increase crop yields.
- **Lower costs:** Machine learning may assist farmers in maximizing the use of resources like water, fertilizer, and pesticides by offering insights into crop development and health. This can save expenses while also lowering how much of an impact agriculture has on the environment.
- **Early disease detection:** Farmers can take preventative measures to stop the spread of illness and reduce crop loss by identifying early indicators of crop diseases with the aid of machine learning. Once the model is sufficiently trained it can detect anomalies such as discoloration on growth size in the early stages of disease, much faster than it would be noticed by humans.
- **Improved crop management:** By offering insights into variables like soil moisture, temperature, and nutrient levels, ML algorithms can assist farmers in improving their crop management tactics. This can assist farmers in making data-driven decisions regarding the best time to water, fertilize, and sow their crops

Overall, using ML in crop analysis and prediction can help farmers optimize their crop yields, reduce waste, and increase profitability while also promoting sustainable farming practices.

4.2. Challenges

Although crop analysis and prediction can greatly benefit from machine learning, there are also a number of drawbacks. The main difficulties consist of:

- **Data volume:** ML models frequently need a lot of data for efficient training. Large data management and collection in agriculture can be difficult, especially for small farms.
- **Data quality:** The accuracy and dependability of machine learning (ML) models depend on the caliber of the training data. Obtaining high-quality data in agriculture can be challenging because of changes in the soil, climate, geography, and other environmental factors. As a result, gathering and cleansing data might be difficult.
- **Model complexity:** Because agricultural systems are intricate, it can be challenging to develop machine learning (ML) models that completely account for all the important variables affecting crop development and output. Selecting the best model architecture for a certain crop analysis or forecast activity can be difficult and may call for extensive knowledge.
- **Interpretability:** Analyzing the outcomes of ML models, particularly those that use deep learning techniques, which are quite complex, can be challenging. Because of this, it may be difficult for farmers to comprehend the elements that go into making a particular crop prediction or suggestion.
- **Accessibility:** In situations with limited resources, it may be challenging to obtain access to the hardware and software infrastructure required for developing and deploying ML models.
- **Privacy and security:** These concerns exist around the collection, storage, and use of sensitive agricultural data. It can be challenging to ensure privacy and security while still allowing access to the data for ML research.
- **Human factors:** It's possible that farmers and other interested parties won't readily adopt new methods and technology, such as ML-based systems. For technology to be used more widely, it must be made accessible, user-friendly, and capable of providing real benefits

Addressing these challenges requires collaboration between data scientists, farmers, and other stakeholders to ensure that ML algorithms are effective, usable, and ethical.

4. Methodology

In recent years, machine learning applications have entered our lives in many areas from the health to the defense industry, from education to urbanization, and taken an effective way in decision-making situations. At the same time, it started to produce information and technology solutions by forming the basis of the newly emerging search engine infrastructure such as ChatGPT (Chat Generative Pretrained Transformer [56] from the OpenAI, Google Bard [58], and similar AI-based chatbots and some other tools. Many research companies reveal that new trends will grow even more in various platforms. In this respect, the effect of machine learning-oriented systems and solutions in the field of technology will increase its effectiveness as a huge multiplier, and many sectors such as the chip design [59] and traffic estimations [60] would be changed by enforcing machine learning models.

Generally, it is important to collect and analyze accurate data using machine learning algorithms. The data collection is critical in both quality and size to get accurate results and make high predictions. In general, big data has the characteristics of size, speed, and

variety. Their large size helps eliminate randomness and allows the data to provide detailed results. In addition, large-scale analysis data is unstructured. Using more than one dataset from different sources in the analysis will provide a higher success rate. Many different sources such as sensors, social media, digital networks, physical devices, the stock market, and health are used as sufficient data sources. These data can be accessed through APIs, web collection, and direct access paths. Data can be in two different forms such as static datasets or stream data. Data from different platforms are incorporated into the data processing operations. Analysis using these collected data makes data cleaning and preprocessing more critical while using machine learning algorithms.

Data preprocessing is an important step to improve data quality by using data cleaning and data transformation to construct the data to make it credible and usable [61]. Besides these operations, data reduction is also a very critical stage that helps models to have good quality and credible results. For example, missing, incomplete, and noisy data lead to incorrect results in the analysis. Even though analysts use correct and suitable machine learning algorithms, results may not be promising. Data extraction is the process of obtaining data from sources and machine learning is important for the correct creation of models. Data transformation is also the process of converting the obtained form of the data to a different format to be able to analyze it as an important step in the data preprocessing phase. By having preprocessing operations, machine learning algorithms can easily interpret data features. Figure 3 shows detail of the methodology for crop analysis and prediction using IoT and ML algorithms.

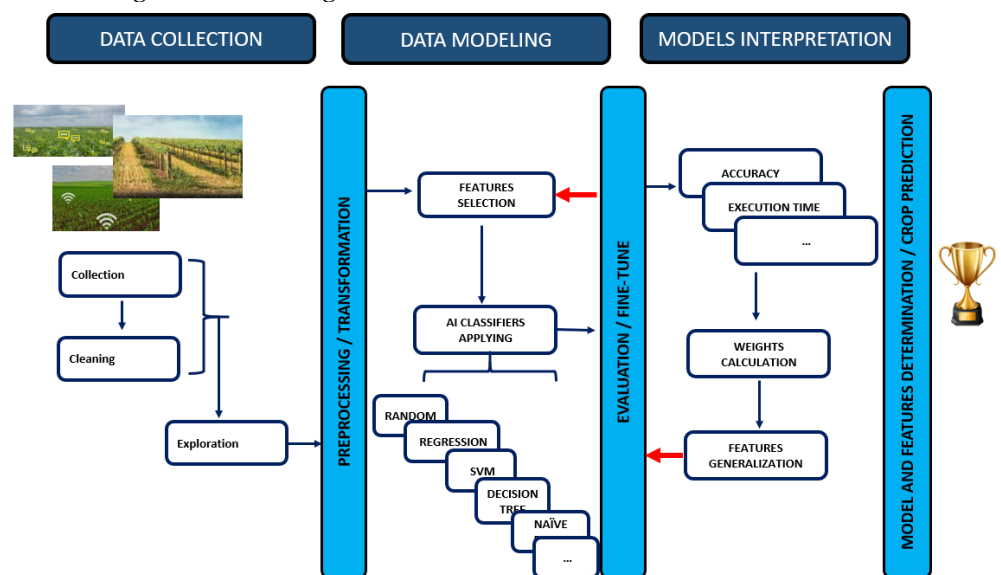


Figure 3. Methodology for crop prediction using IoT and ML

There are many classification algorithms such as Decision Tree (DT), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), and Random Forest (RF) that can be used to build a model to predict the correct labels based on provided data. The built model initially was achieved by training using training data, then the results are evaluated using test data to make sure that predictions are accurate and desired values. Authors in the research [62] focus on behavioral classification in neuroscience and another study to measure the water quality using machine learning algorithms [63], and finally, in medicine, liver disease prediction using classification methods [63] indicate that classification is one of the important supervised learning methods to do prediction.

In classification, the dataset is divided into two separate categories as training and test sets. The decisions are done based on which model would be used, and predictions are established. Later, these predictions are evaluated according to the accuracy of the results. Model complexity and performance values are revealed with these models based on the machine learning algorithms, which will be classified according to the problem area and the situation, and the model will be applied to the dataset [64].

Naïve Bayes Classifier is a supervised learning algorithm that uses Bayes' theorem to classify objects. It is used in machine learning and data mining applications for text analysis, medical diagnosis, spam filtering, and other similar tasks. Naïve Bayes assumes that features in a class are considered independent of each other. The basic idea behind Naïve Bayes is to calculate the conditional probability of each feature given the class, as well as the prior probability of each class, and then using Bayes' theorem calculate the posterior probability of each class given the features of the data point. It then chooses the class with the highest posterior probability as the predicted class for the data point. In practice, the Naïve Bayes algorithm performs well, especially when the data is sparse and the number of features is large. Below is the pseudocode for Naïve Bayes Classifier:

Input: {X: Training set; m: Number of observations; n: Number of features; Labels y for the training set}

Output: {New data point x to be classified; predicted class for x}

Step 1:

For each unique class value c in y:

- Find the count of observations with class c, let it be count(c)
- Calculate the prior probability for class c, let it be $P(c) = \text{count}(c) / m$

Step 2:

For each feature i in X and each unique class value c in y:

- Find the count of observations with the feature i and class c, let it be count (i, c)
- Find the count of observations with class c
- Calculate the conditional probability of feature i given class c.

$$P(i | c) = (\text{count}(i, c)) / (\text{count}(c))$$

Step 3:

Calculate the posterior probability for each class given the new data point x:

For each unique class value c in y:

- Initialize the posterior probability $P(c|x)$ to $P(c)$
- For each feature i in x:

Multiply $P(c|x)$ by $P(i | c)$ if $x[i]$ is observed in the training data, otherwise ignore the term.

Step4:

Choose the class with the highest posterior probability as the predicted class for x.

Random forest is a type of supervised machine learning algorithm that is used in regression and classification problems. It is an ensemble learning algorithm that uses decision trees to make a prediction. It is based on the concept of decision trees, where a tree-like model of decisions and their possible consequences is constructed. Random Forest creates a large number of decision trees and combines their predictions to make a final prediction. More trees created give high accuracy and robust results. Below is the pseudocode for Random Forest algorithm:

Inputs: {(x, y): Training data set; T: Number of trees; d: Maximum depth of each tree; f: Number of features to consider at each split}

Outputs: {Learned trees for classification}

For each tree in the random forest:

- Select a bootstrap sample from the training data set.
- Using the bootstrap sample, create a decision tree T_t with a maximum depth of d.
- Randomly select f features to consider at each split of T_t
- Use the selected features to find the best split at each node of T_t

Create the list of decision trees T_1, T_2, \dots, T_T

For each input data:

- For each decision, find the prediction. Obtain the predictions of all the trees.

- Using the prediction from all decisions trees, calculate the final predicted class.

Multilayer Neural Network is a type of machine learning algorithm known as Artificial Neural Network (ANN). It consists of multiple layers of interconnected nodes between the input layer and the output layer.

To minimize the error between the predicted output and the actual output, the neural network involves adjusting the weights and the biases of each neuron by using an optimization algorithm such as backpropagation. Multilayer neural network algorithm is used in image and speech recognition, natural language processing, and financial forecasting. It can achieve high accuracy with sufficient training data and computational resources. Below is the pseudocode for the Multilayer Neural Network algorithm:

- Define the number of layers, number of neurons per layer
- Initialize randomly the weights and biases for each neuron in the network.

For each input data:

- Forward propagate the input through the neural network to obtain the predicted output.
- Calculate the error rate between the predicted output and the actual output.
- Backward propagate the error through the neural network to adjust the weights and biases using the optimization algorithm.
- Repeat until the error converges to a satisfactory level.

5. Experimental Results

There are several application areas of machine learning in agriculture. In this work, we used 15 different machine learning algorithms to model agriculture data which recommend farmers the most suitable crops to produce in the farm. The dataset includes several features such as ratio of Nitrogen content (N), temperature, pH value of the soil, rainfall, humidity, ratio of Phosphorous content (K), and ratio of Potassium content (P) in the soil. Crop prediction data set has 2200 records which have 22 crop labels such as apple, banana, rice, coffee, cotton, black gram, watermelon, chickpea, coconut, grapes, jute, kidney beans, grape, lentil and orange. 67% of the data is used in training and the rest of the data is used for testing [65]. Table 1 shows accuracy values and error rates for the algorithms taken into consideration.: Bayes net, Naïve Bayes Classifier, Hoeffding tree and random forest algorithms give the best accuracy. Error metrics can be formalized as follows:

$$K (Kappa) = (P_o - P_e) / (1 - P_e) \quad (1)$$

Where P_o is relative observed agreement and P_e Hypothetical probability of chance agreement.

$$MAE (Mean Absolute Error) value = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (2)$$

Where n is the total number of data, x_i is true value and y_i is prediction.

$$RMSE (Root Mean Square Error) = \sqrt{\frac{\sum_{i=1}^n (x_i - x'_i)^2}{n}} \quad (3)$$

Where x_i observed and x'_i is predictive value.

$$RAE (Relative Absolute Error) = \frac{\sum_{i=1}^n |y_i - y|}{\sum_{i=1}^n |y_i - y|} \quad (4)$$

Where y is the average value of the data.

$$Root Relative Squared Error (RRSE) = A = \sqrt{\frac{\sum_{i=1}^n (P_i - T_i)^2}{\sum_{i=1}^n (T_i - T)^2}} \quad (5)$$

Where P is the predicted value and T is the target value.

Table 1. Accuracy and error values for each classification algorithms

Method	Accuracy	Kappa	MAE	RMSE	RAE	RRSE
Bayes Net	99.59	0.995	0.0010	0.018	1.14	8.64
Naïve Bayes Classifier	99.46	0.994	0.0009	0.020	1.05	9.73
Logistic	97.99	0.979	0.0020	0.038	2.30	18.24
Multilayer Perception	98.79	0.987	0.0046	0.033	5.33	16.18
Simple Logistic	98.66	0.986	0.0025	0.029	2.88	14.03
IBK	97.86	0.977	0.0032	0.043	3.69	21.05
KSTAR	97.86	0.977	0.0036	0.038	4.11	18.47
LWL	76.74	0.756	0.0752	0.188	86.59	90.26
Ada BoostM1	6.82	0.036	0.0829	0.203	95.51	97.79
Regression	98.38	0.983	0.0099	0.042	11.41	20.44
Decision Table	88.50	0.879	0.0565	0.145	65.10	69.61
Hoeffding Tree	99.46	0.994	0.0009	0.020	1.05	9.74
J48	98.79	0.987	0.0012	0.032	1.35	15.36
Random Forest	99.46	0.994	0.0032	0.024	3.63	11.75
Random Tree	98.12	0.980	0.0017	0.041	1.96	19.79

Table 2 shows build and test time for each algorithm. It can be noticed that the Multiplayer perception algorithm build time is greater than other algorithms Because this algorithm is an artificial neural network algorithm including several hidden layers in addition to input and output layer [65]. MLP algorithm runs with several iterations to find best model for given input and output set which makes build time higher. KSTAR and LWL algorithms have higher testing time than others.

Table 2. Build and test times for classification algorithms

Method	Build Time (second)	Test Time (second)
Bayes Net	0.48	0.25
Naïve Bayes Classifier	0.03	0.67
Logistic	4.83	0.06
Multilayer Perception	17.39	0.05
Simple Logistic	3.86	0.02
IBK	0.03	0.69
KSTAR	0	6.9
LWL	0	9.56

Ada BoostM1	0.04	0
Regression	2.4	0.05
Decision Table	0.75	0.01
Hoeffding Tree	0.41	0.06
J48	0.27	0.03
Random Forest	1.57	0.13
Random Tree	0.02	0

Table 3 demonstrates accuracy and error measurement in Multilayer perception algorithm with different sample size of training data. When 10% of the data is used in training, accuracy with MLP is 93.53%; on the other hand, when 90% of data is used, accuracy is 97.72%.

Table 3. Accuracy and error values for Multilayer perception algorithm (Training set from 10 % to 90%)

Training Set	Accuracy	Kappa	MAE	RMSE	RAE	RRSE
10%	93.53	0.9323	0.0166	0.0726	19.06	34.69
20%	95.39	0.9518	0.0096	0.0568	11.10	27.22
30%	95.91	0.9571	0.0082	0.0545	9.47	26.14
40%	97.87	0.9778	0.0065	0.0436	7.51	20.92
50%	97.90	0.9790	0.0057	0.039	6.50	18.87
60%	97.95	0.9786	0.0056	0.0433	6.41	20.76
70%	98.63	0.9857	0.0043	0.033	4.95	15.83
80%	98.41	0.9833	0.0038	0.0315	4.42	15.10
90%	97.72	0.9761	0.0038	0.0331	4.41	15.88

Table 4 presents the build and test times for the multilayer perceptron (MLP) algorithm under different scenarios involving modifications to the percentage of training and testing data. It is worth noting that the MLP algorithm has longer build times compared to other algorithms [65]. The results indicate that the lowest build time is observed when the training set is at 40%, while the highest build time occurs when the training set is at 10%. Test times, on the other hand, are consistently low across all scenarios, with the highest being at 0.05 when the training set is at 10%.

Table 4. Build and test times for multilayer perception algorithm (Training set from 10 % to 90%)

Training Set	Build Time	Test Time
10%	19.18	0.05
20%	17.03	0.03

30%	17.47	0.01
40%	14.26	0.02
50%	14.54	0.02
60%	14.12	0
70%	14.79	0.01
80%	13.81	0
90%	13.21	0

After performing machine learning classification on a dataset with seven features and one label representing the crop type, and recorded the accuracies of different algorithms in Table 1, further simulations were conducted to determine the minimum number of features required to achieve high accuracy in algorithm learning and prediction. Table 5 shows the results of four scenarios where we randomly selected 3 to 4 features and evaluated the accuracy of crop detection for each set. The second set of features (Temperature, Humidity, pH, Rainfall) achieved the highest accuracy, reaching 97.05% with Bayes Net and 97.32% with Random Forest. In contrast, the worst prediction outcome occurred when using the features (N, P, K) where the best accuracy obtained was 68.04% with Random Forest.

The results displayed in Table 5 highlight the importance of selecting the appropriate features for achieving high accuracy in crop detection using machine learning algorithms. The set of features (composed of Temperature, Humidity, pH, Rainfall) we identified can be used as a guide for selecting relevant features for crop determination in future agricultural data analysis.

Table 5. Accuracy for different feature sets

Method	{N, P, K}	{Temperature, Humidity, pH, Rainfall}	{K, P, Rainfall}	{N, Temperature, Humidity, pH}
Bayes Net	67.64	97.05	85.69	89.70
Naïve Bayes Classifier	65.37	96.39	85.16	87.03
Logistic	66.17	85.42	74.19	76.07
Multilayer Perception	66.84	89.17	80.34	82.88
Simple Logistic	66.84	85.16	72.86	74.73
IBK	66.57	91.04	79.27	81.02
KSTAR	65.10	91.71	81.14	80.74
LWL	42.11	61.23	46.39	50.26
Ada BoostM1	6.81	6.81	6.81	6.81
Regression	65.37	95.98	84.22	86.49

Decision Table	63.77	74.73	79.27	72.19
Hoeffding Tree	65.37	96.52	85.29	86.89
J48	65.10	94.65	83.55	84.49
Random Forest	66.57	97.32	82.88	87.03
Random Tree	68.04	94.92	79.27	83.15

Table 6. Classification of crops based on several class labels

Item	Growth Characteristics	Use (food, feed, fiber)	Type	Water requirements	Harvest method
Rice	Grass	Food	Cereals	Drought	By Hand Or Machine
Maize	Grass	Feed, Fiber	Cereals	Drought	By Hand Or Machine
Chickpea	Bush	Food	Legume	Drought	Machine
Kidneybeans	Bush	Food	Legume	Drought	By Hand And Machine
Pigeonpeas	Bush	Food	Legume	Drought Resistant	By Hand
Mothbeans	Bush	Fiber	Legume	Drought Resistant	Both
Mungbean	Bush	Food	Legume	Drought	Hand Picked
Blackgram	Bush	Food	Legume	Drought Tolerance	Both
Lentil	Bush	Food	Legume	Drought	Hands
Pomegranate	Tree	Fiber	Fruit	Drought Tolerant	Hands
Banana	Tree	Fiber	Fruit	Water Loving	Hands
Mango	Tree	Fiber	Fruit	Drought Tolerance	Hands
Grapes	Tree	Fiber	Fruit	Drought Tolerance	Hand
Watermelon	Sprawling Vines	Fiber	Fruit	Drought Tolerance	Hands

Muskmelon	Bush	Fiber	Fruit	Drought Tolerance	Hands
Apple	Tree	Fiber	Fruit	Drought Tolerance	Hand
Orange	Tree	Fiber	Fruit	Water Loving	Hand
Papaya	Tree	Fiber	Fruit	Water Loving	Hand
Coconut	Tree	Fiber	Fruit	Water Loving	Hand
Cotton	Bush	Fiber	Plant	Drought Tolerant	Machine
Jute	Shrub	Fiber	Plant	Water Loving	Hands
Coffee	Shrub	Fiber	Fruit	Drought	Hands

In order to investigate the impact of changing the label on the accuracy of the data analysis algorithm more experiments were conducted. The results are displayed in Table 7. Instead of predicting the specific type of crop, we manually grouped the crops into four more general categories based on their growth characteristics, usage, type, water requirements, and harvest method, as described in Table 6. Then, we tested the accuracy of the algorithm to predict these new general labels one at a time. Table 6 presents a classification of crops based on several class labels. The table includes information on the growth characteristics, usage (food, feed, fiber), type, water requirements, and harvest method for various crops. The crops are classified based on their growth characteristics such as grass, bush, and tree, and their water requirements such as drought tolerance, drought, and water-loving, as well as their harvest method, which can be performed by hand or machine. For example, rice and maize are classified as grasses with the type of cereal, whereas chickpeas and pigeon peas are classified as bush with the type of legumes.

The purpose of this experiment is to explore whether the algorithm could accurately predict more general categories of crops, which could be useful in situations where general characteristics of crops are investigated, and specific crop types are not known or easily identifiable. The results of the experiment are summarized in Table 7, where we compared the accuracies of various classification methods for each general category. Table 7 shows that some classification methods perform better in predicting general categories. For example, the IBK, KSTAR, and KSTAR methods achieved high accuracies for predicting all general categories compared to the type of the crop (first column in the table), while Bayes Net, Naïve Bayes Classifier, Logistic, and Multilayer Perception had low accuracies for all categories. The experiment demonstrates the potential usefulness of using more general labels in crop classification and provides insights into which classification methods may be most effective for this task.

Further research could explore the potential benefits of using a combination of classifications (multi-labeling) and investigate the potential of using more granular or specific labels that may impact the accuracy of classification algorithms.

Table 7. Comparison of the accuracies of various classification methods.

Method	Accuracy	Growth Characteristics	Usage	Type	Water Requirements	Harvest Method
Bayes Net	99.59	96.79	91.31	99.13	85.69	89.17
Naïve Bayes Classifier	99.46	79.41	85.69	90.59	65.90	76.33
Logistic	97.99	83.28	86.76	91.04	80.62	66.57
Multilayer Perception	98.79	97.99	98.12	97.41	87.16	95.72
Simple Logistic	98.66	82.08	87.71	90.91	80.08	67.51
IBK	97.86	98.53	98.66	98.72	97.99	97.99
KSTAR	97.86	99.19	99.19	98.86	97.86	97.86
LWL	76.74	83.02	88.23	67.27	57.75	70.18
Ada BoostM1	6.82	76.87	82.08	45.32	44.11	61.23
Regression	98.38	99.19	99.06	99.09	98.93	98.93
Decision Table	88.50	96.12	95.58	95.04	93.04	94.65
Hoeffding Tree	99.46	79.41	85.43	89.82	66.31	76.60
J48	98.79	98.26	97.86	98.63	98.39	99.33
Random Forest	99.46	99.33	99.73	99.45	99.73	99.59
Random Tree	98.12	98.66	99.33	98.36	97.59	98.66

6. Conclusion

Our research highlighted the significance of incorporating machine learning algorithms and IoT sensors in modern agriculture to optimize reap production and reduce waste through informed decision-making. The study identifies the challenges and opportunities associated with integrating these technologies in agriculture and presented experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms along with accuracy, error values, build, and test time for each classification algorithm. The findings suggested that analyzing wide-ranging data collected from farms, including real-time data from IoT sensors, can enable farmers to make more informed decisions about factors that affect harvest growth. Despite the challenges associated with deploying machine learning in agriculture, our results achieved so far are very promising that machine learning approaches will become increasingly important for production predictions in agriculture in the future. In this experiment, crops are investigated according to general characteristics using different machine learning algorithms, and useful results are obtained by making predictions in cases where certain crop types are not known or cannot be easily identified. Our work indicated that appropriate feature selection is very critical to achieve better accuracy in machine learning algorithms while analyzing agricultural data. Using the Temperature, Humidity, pH, and Precipitation features in the dataset, it achieved the highest accuracy, reaching 97.05% with Bayes Net and 97.32% with Random Forest. This research provided valuable insights into the potential benefits of these technologies in modern agriculture, and further research and development in this field can help optimize crop production, reduce waste, and improve food security globally.

Author Contributions: E.E., C.Z., A.E.T., A.Z, W.A., E.C., A.S., and L.K were involved in the full process of producing this paper, including conceptualization, methodology, modeling, validation, visualization, and preparing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Linchao Li et al. Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Computers and Electronics in Agriculture* **2022**, Volume 194, 106790.
2. van Klompenburg T; Kassahun A; Catal C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* **2020**, 177.
3. Kuradusenge et al., et al. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* **2023**, vol. 13(1), pp. 225.
4. Oré, G.; Alcântara, M. S.; Góes, J. A.; Oliveira, L. P.; Yepes, J.; Teruel, B.; Castro, V. Crop Growth Monitoring with Drone-Borne DInSAR. *Remote Sensing* **2020**, vol. 12(4), pp. 615.
5. A. Gehlot; N. Sidana; D. Jawale; N. Jain; B. P. Singh; B. Singh. Technical analysis of crop production prediction using Machine Learning and Deep Learning Algorithms. *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES)*, **2022**, Chennai, India, 2022, pp. 1-5.
6. S. Vashisht; P. Kumar; M. C. Trivedi. Improved Extreme Learning Machine for Crop Yield Prediction. *3rd International Conference on Intelligent Engineering and Management (ICIEM)* **2022**, London, United Kingdom, pp. 754-757.
7. F. Shahrin; L. Zahin; R. Rahman; A. J. Hossain; A. H. Kaf; A. K. M. Abdul Malek Azad. Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. *11th International Conference on Electrical and Computer Engineering (ICECE)* **2020**, Dhaka, Bangladesh, pp. 21-24.
8. Tawseef A. S.; Tabasum R.; Faisal R. L. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture* **2022**, Volume 198.
9. Senthil K. S. D.; Mary D. S. Smart farming using Machine Learning and Deep Learning techniques. *Decision Analytics Journal* **2022**, Volume 3, pp. 100041.
10. Senthil K. M.; Akshaya R; Sreejith K. An Internet of Things-based Efficient Solution for Smart Farming. *Procedia Computer Science* **2023**, Volume 218, Pages 2806-2819.
11. Vivek S.; Ashish K. T; Himanshu M. Technological revolutions in smart farming: Current trends, challenges & future directions. *Computers and Electronics in Agriculture* **2022**, Volume 201, 107217.
12. Mamatha, J.C. K. Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques. *Measurement: Sensors* **2023**, Volume 25, 100665.
13. Sandya D. A.; Ziwei H.; Yishuo Z.; Myung H. N.; Bahadorreza O.; Atul S. A survey on smart farming data, applications, and techniques. *Computers in Industry* **2022**, Volume 138, 103624.
14. Rashid M; Bari B.S.; Yusup Y.; Kamaruddin M.A.; Khan N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. in *IEEE Access* **2021**, vol. 9, pp. 63406-63439.
15. Babber J.; Malik P.; Mittal V.; Purohit K.C. Analyzing Supervised Learning Algorithms for Crop Prediction and Soil Quality. *6th International Conference on Computing Methodologies and Communication (ICCMC)* **2022**, Erode, India, pp. 969-973.
16. Ishak M; Rahaman M.S; Mahmud T. FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing. *13th International Conference on Information & Communication Technology and System (ICTS)* **2021**, Surabaya, Indonesia, pp. 224-229.
17. Patel K; Patel H.B. A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction. *Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)* **2021**, Erode, India, pp. 1-5.
18. Memon R; Memon M; Malioto N; Raza M.O. Identification of growth stages of crops using mobile phone images and machine learning. *International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)* **2021**, Quetta, Pakistan, pp. 1-6.
19. Chandraprabha M; Dhanaraj R.K. Soil Based Prediction for Crop Yield using Predictive Analytics. *3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* **2021**, Greater Noida, India, pp. 265-270.
20. Ray R. K; Das S. K; Chakravarty S. Smart Crop Recommender System-A Machine Learning Approach. *12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* **2022**, Noida, India, pp. 494-499.
21. Johnson N; Santosh M.B; Dhannia T. A survey on Deep Learning Architectures for effective Crop Data Analytics. *International Conference on Advances in Computing and Communications (ICACC)* **2021**, India, pp. 1-10.
22. Priyadharshini K; Prabavathi R; Devi V.B; Subha P; Saranya S.M; Kiruthika K. An Enhanced Approach for Crop Yield Prediction System Using Linear Support Vector Machine Model. *International Conference on Communication, Computing and Internet of Things (IC3IoT)* **2022**, Chennai, India, pp. 1-5.

23. Malathy S; Vanitha C. N; Kotteswari S; M. E. Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques. *International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* **2022**, Salem, India, pp. 437-442.
24. Chowdary V.T; Robinson Joel M; Ebenezer V; Edwin B; Thanka R; Jeyaraj A. A Novel Approach for Effective Crop Production using Machine Learning. *International Conference on Electronics and Renewable Systems (ICEARS)* **2022**, Tuticorin, India, pp. 1143-1147.
25. Yamparla R; Shaik H. S; Guntaka N; Marri P; Nallamothu S. Crop Yield Prediction using Random Forest Algorithm. *7th International Conference on Communication and Electronics Systems (ICCES)* **2022**, Coimbatore, India, pp. 1538-1543.
26. G. A. R. and S. S. S., "A brief study on the prediction of crop disease using machine learning approaches," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, pp. 1-6.
27. Kumar R; Shukla N; Princee. Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning. *International Mobile and Embedded Technology Conference (MECON)*, **2022**, Noida, India, pp. 168-172.
28. Bhosale S. V.; Thombare R. A.; Dhemey P.G.; Chaudhari A.N. Crop Yield Prediction Using Data Analytics and Hybrid Approach. *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* **2018**, Pune, India.
29. Alwis S.D.; Hou Z.; Zhang Y.; Na M.H.; Ofoghi B.; Sajjanhar A. A survey on smart farming data, applications and techniques. *Computers in Industry* **2022**, vol. 138, pp. 103624.
30. Ramos P. J.; Prieto F.A.; Montoya E.C; Oliveros C.E. Automatic fruit count on coffee branches using computer vision. *Computers and Electronics in Agriculture* **2017**, vol. 137, pp. 9-22.
31. Sengupta S.; Lee W.S. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. *Biosystems Engineering* **2014**, vol. 117, pp. 51-61, 2014.
32. Su Y.; Xu H.; Yan L. Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi J. Biol. Sci.* **2017**, vol. 24, no. 2, pp. 537-547.
33. Adankon M.M.; Cheriet M. Support Vector Machine. *Encyclopedia of Biometrics* **2009**, S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, pp. 1303-1308.
34. Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, vol. 10, 3254–3264.
35. Pantazi X.E.; Moshou D.; Alexandridis T.K.; Whetton R.L.; Mouazen A.M. Wheat yield prediction using machine learning and advanced sensing techniques. *IEEE Transactions on Instrumentation and Measurement* **2016**, vol. 65, no. 3, pp. 572-576.
36. Senthilnath J.; Dokania A.; Kandukuri M.; Ramesh K.N.; Anand G.; Omkar S.N. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosyst. Eng.* **2016**, vol. 146, pp. 16-32, 2016.
37. Ebrahimi M.A.; Khoshtaghaza M.H.; Minaei S.; Jamshidi B. Vision-based pest detection based on SVM classification method. *Comput. Electron. Agric.* **2017**, vol. 137, pp. 52-58.
38. Chung C.L.; Huang K.J.; Chen S.Y.; Lai M.H.; Chen Y.C; Kuo Y.F. Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.* **2016**, vol. 121, pp. 404-411.
39. Sun X.; Guo M.; Ma M.; Mankin R.W. Identification and classification of damaged corn kernels with impact acoustics multi-domain patterns. *Computers and Electronics in Agriculture* **2018**, vol. 150, pp. 152-161.
40. Ramirez-Paredes J.P.; Hernandez-Belmonte U.H. Visual quality assessment of malting barley using color, shape and texture descriptors. *Computers and Electronics in Agriculture* **2020**, vol. 168, p. 105110.
41. Pantazi X.E.; Tamouridou A.A.; Alexandridis T.K.; Lagopodi A.L.; Kontouris G.; Moshou D. Detection of *Silybum marianum* infection with *Microbotryum silybum* using VNIR field spectroscopy. *Comput. Electron. Agric.* **2017**, vol. 137, pp. 130-137.
42. Ferentinos K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, vol. 145, pp. 311-318.
43. Mehdizadeh S.; Behmanesh J.; Khalili K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* **2017**, vol. 139, pp. 103-114.
44. Feng Y.; Peng Y.; Cui N.; Gong D.; Zhang K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* **2017**, vol. 136, pp. 71-78.
45. Patil A.P.; Deka P.C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* **2016**, vol. 121, pp. 385-392.
46. Mohammadi K.; Shamshirband S.; Motamedi S.; Petković D.; Hashim R.; Gocic M. Extreme learning machine based prediction of daily dew point temperature. *Comput. Electron. Agric.* **2017**, vol. 117, pp. 214-225.
47. Motokura K.; Takahashi M.; Ewerton M.; Peters J. Plucking Motions for Tea Harvesting Robots Using Probabilistic Movement Primitives. in *IEEE Robotics and Automation Letters* **2020**, vol. 5, no. 2, pp. 3275-3282.
48. Mao S.; Li Y.; Ma Y.; Zhang B.; Zhou J.; Wang K. Automatic cucumber recognition algorithm for harvesting robots in the natural environment using deep learning and multi-feature fusion. *Computers and Electronics in Agriculture* **2020**, vol. 170, p. 105254.
49. Li J.; Tang Y.; Zou X.; Lin G.; Wang H. Detection of Fruit-Bearing Branches and Localization of Litchi Clusters for Vision-Based Harvesting Robots. in *IEEE Access* **2020**, vol. 8, pp. 117746-117758.
50. Ge Y.; Xiong Y.; Tenorio G. L. Fruit Localization and Environment Perception for Strawberry Harvesting Robots. *IEEE Access*, **2019**, vol. 7, pp. 147642-147652.
51. Pyngkodi M.; Thenmozhi K.; Karthikeyan M.; Kalpana T.; Palarimath S.; Kumar G.B.A. IoT based Soil Nutrients Analysis and Monitoring System for Smart Agriculture. *3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)* **2022**, Coimbatore, India, pp. 489-494.

52. Pivoto D.; Waquil P.D.; Talamini E.; Finocchio C.P.S.; Corte V.; Mores G. Scientific development of smart farming technologies and their application in Brazil, *Information Processing in Agriculture*, **2018**, Volume 5, Issue 1, 2018, Pages 21-32, ISSN 2214-3173, <https://doi.org/10.1016/j.inpa.2017.12.002>.
53. S. P. N and H. P. M. Kumar. Soil Quality Identifying and Monitoring Approach for Sugarcane Using Machine Learning Techniques. *Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, **2022**, India, pp. 1-5.
54. Puengsungwan S. IoT based Soil Moisture Sensor for Smart Farming. *International Conference on Power, Energy and Innovations (ICPEI)*, **2020**, Chiangmai, Thailand, pp. 221-224.
55. Sahu P.; Singh A.P.; Chug A.; Singh D. A Systematic Literature Review of Machine Learning Techniques Deployed in Agriculture: A Case Study of Banana Crop. in *IEEE Access*, **2022**, vol. 10, pp. 87333-87360, 2022.
56. OpenAI, "New and Improved Content Moderation Tooling,," OpenAI, 2022. [Online]. Available: <https://openai.com/blog/new-and-improved-content-moderation-tooling/>. [Accessed 01 04 2023].
57. Google, "Bard Chatbox," Google, [Online]. Available: <https://bard.google.com>. [Accessed 2 4 2023].
58. J. Dean, "1.1 the deep learning revolution and its implications for computer architecture and chip design," in *IEEE International Solid-State Circuits Conference-(ISSCC)*, **2020**.
59. Cui Y.W. et al. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems* **2019**, vol. 21, no. 11, pp. 4883-4894.
60. S. M. a. B. N. Kiran Maharana. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, **2022**, vol. 3, no. 1, pp. 91-99.
61. N. L. S. R. N. J. J. C. a. S. A. G. Goodwin. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Current opinion in neurobiology*, **2022**, vol. 73, p. 102544.
62. N. A. K. O. A. F. B. M. S. A. S. a. A. A.-S. Nasir. Water quality classification using machine learning algorithms. *Journal of Water Process Engineering* **2022**, vol. 48, p. 102920.
63. K. N. J. N. A. a. D. D. Gupta. Liver Disease Prediction using Machine Learning Classification Techniques. in *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, **2022**.
64. Elbasi E. et al. Artificial Intelligence Technology in the Agricultural Sector: A Systematic Literature Review. in *IEEE Access* **2023**, vol. 11, pp. 171-202.
65. Elbasi E.; Zreikat A.I.; Mathew S.; Topcu A.E. Classification of influenza H1N1 and COVID-19 patient data using machine learning. *44th International Conference on Telecommunications and Signal Processing (TSP)*, **2021**, Brno, Czech Republic, pp. 278-282.