MDPI

Article

Crop Prediction Model Using Machine Learning Algorithms

Ersin Elbasi *[ID], Chamseddine Zaki, Ahmet E. Topcu, Wiem Abdelbaki, Aymen I. Zreikat [ID], Elda Cina [ID], Ahmed Shdefat [ID] and Louai Saker

College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; chamseddine.zaki@aum.edu.kw (C.Z.); ahmet.topcu@aum.edu.kw (A.E.T.); wiem.abdelbaki@aum.edu.kw (W.A.); aymen.zreikat@aum.edu.kw (A.I.Z.); elda.cina@aum.edu.kw (E.C.); ahmed.shdefat@aum.edu.kw (A.S.); louai.saker@aum.edu.kw (L.S.)
* Correspondence: ersin.elbasi@aum.edu.kw

Abstract: Machine learning applications are having a great impact on the global economy by transforming the data processing method and decision making. Agriculture is one of the fields where the impact is significant, considering the global crisis for food supply. This research investigates the potential benefits of integrating machine learning algorithms in modern agriculture. The main focus of these algorithms is to help optimize crop production and reduce waste through informed decisions regarding planting, watering, and harvesting crops. This paper includes a discussion on the current state of machine learning in agriculture, highlighting key challenges and opportunities, and 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, incorporating online IoT sensor data that were obtained in a real-time manner, farmers can make more informed verdicts about factors that affect crop growth. Eventually, integrating these technologies can transform modern agriculture by increasing crop yields while minimizing waste. Fifteen different algorithms have been considered to evaluate the most appropriate algorithms to use in agriculture, and a new feature combination scheme-enhanced algorithm is presented. The results show that we can achieve a classification accuracy of 99.59% using the Bayes Net algorithm and 99.46% using Naïve Bayes Classifier and Hoeffding Tree algorithms. These results will indicate an increase in production rates and reduce the effective cost for the farms, leading to more resilient infrastructure and sustainable environments. Moreover, the findings we obtained in this study can also help future farmers detect diseases early, increase crop production efficiency, and reduce prices when the world is experiencing food shortages.

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

1. Introduction

Agriculture is a vital element that has a significant role in nourishing the world's growing population. To keep pace with the increasing demand for foodstuffs, 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. Smart farming aims to increase reap output while minimizing using resources such as water, fertilizer, and energy [3].

Figure 1 illustrates IoT and machine learning-based crop analysis and prediction processes. Over the years, numerous elements and technologies have been integrated into the architecture of a smart farm, such as sensing and monitoring systems, Internet of Things (IoT) sensors, data analytics and Artificial Intelligence (AI), precision agriculture techniques, remote monitoring and control, automated systems, livestock management systems, cloud computing and big data storage, energy management, and farm management software,

to enhance farming practices and boost productivity [4–6]. In smart farming, the Internet of Things is considered one of the key contributing technologies used. IoT sensors can be utilized to monitor soil moisture, temperature, and other environmental aspects [7], and 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 receive the right amount of water and nutrients, which can improve their quality and yield [8,9].
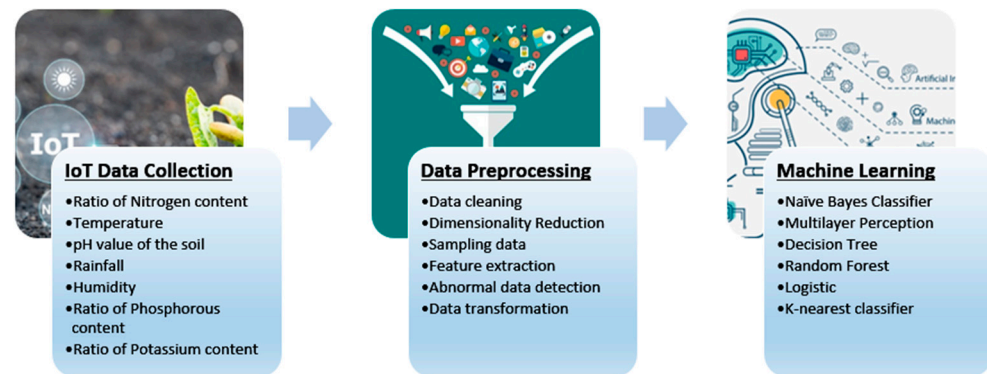
Figure 1. IoT and machine learning-based crop analysis and prediction process.

In recent years, machine learning applications have entered our lives in many areas, from health to defense industries and education to urbanization, and have 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 from the OpenAI [10], Google Bard [11], 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 technology field will increase its effectiveness as a huge multiplier, and many sectors, such as chip design [12] and traffic estimations [13], would be changed by enforcing machine learning models.

Generally, it is essential to collect and analyze accurate data using machine learning algorithms. The data collection is critical in both quality and size to obtain accurate results and make high predictions. In general, big data have size, speed, and various characteristics. Their large size helps eliminate randomness and allows the data to provide detailed results. In addition, large-scale analysis data could be more structured. Using more than one dataset from different sources in the analysis will provide a higher success rate. Many sources, such as sensors, social media, digital networks, physical devices, the stock market, and health centers, are sufficient data sources. This data can be accessed through APIs, web collection, and direct access paths. Data can be in two forms: 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.

Machine learning algorithms can analyze vast amounts of data from 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 [14]. In agriculture, especially 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 [15].

A noteworthy application of machine learning in agriculture is precision farming, which includes employing data and technology to optimize agricultural conducts such as fertilization, irrigation, and pest control to improve reap 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 reap growth, nutrient levels, and moisture content. Farmers can utilize these maps to regulate their farming conducts, such as applying fertilizer or watering specific field areas, to maximize reap output and minimize waste [16]. 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 optimal planting times and locations [17]. 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 [18,19].

There are several encounters for deploying machine learning in agriculture, such as the lack of data groundwork, high cost of sensors and other technology, and need for specialized proficiency to develop and maintain the different solutions. However, as more farms implement precision agriculture and gather data, the potential profits of deploying machine learning in agriculture will become more evident. It is worth mentioning that machine learning in agriculture is still in its early stages, and more research needs to be conducted in this area to realize this technology's potential fully. So far, the results are promising, and machine learning will likely become increasingly important [20].

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 correctly predict labels based on the input data. This study discusses several classification algorithms, such as Decision Tree (DT), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), and Random Forest (RF), which can be employed to build such models. The authors predict that adopting machine learning-focused systems and solutions will significantly enhance efficiency and productivity, resulting in massive industry changes.

This paper is structured as follows: First, the introduction and background reading are presented in Section 1 to provide a comprehensive understanding of the topic. Then, in Section 2, a relevant literature review is presented. Section 3 introduces artificial intelligence in smart farming to delve deeper into the subject. Crop analysis and prediction- benefits and challenges are given in Section 4. Section 5 outlines the methodologies used in this study, followed by the presentation of the experimental results in Section 6. Lastly, Section 7 is dedicated to the conclusion and future recommendations.

2. Literature Review

Machine learning approaches and algorithms are utilized in crop yield prediction methods to improve the quality of the crop so that the farmer's profit is maximized. The quality of the agricultural sector is improved; hence, the overall economy is enhanced. This issue has been discussed in detail in the literature [21–35]. In [21], a review of machine learning algorithms to predict palm oil yield is discussed. The authors conducted a comparative analysis of the related work, focusing on the suggested approaches' advantages, disadvantages, and limitations. 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.

The authors in [22] focused on crop prediction and yield by studying soil quality, 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 concern three machine learning algorithms: Naïve Bayes, Logistic Regression, and Random Forest.

Moreover, the authors conducted a comparison between these algorithms concerning accuracy. A crop production model proposed in [23] aims to manage the produced crop using machine learning algorithms to help farmers in developing countries who are still using traditional methods and cannot 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; Random Forest Regression algorithms; and proper real climate, weather, and soil data.

Due to the scarcity of natural resources around the world, the authors in [24] 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; therefore, the country's overall 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 [24]. The authors suggested a prediction method using the following supervised machine learning algorithms: Random Forest, Support Vector Machine, Decision Tree, K-Nearest Neighbor (K-NN), Gaussian Naïve Bayes (GNB), 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 [25], a dataset for soil prediction was collected from Tamil Nadu Agricultural University (TNAU), India, which involved 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 [26] recommended 22 types of crops in this study and proposed a three-step 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 performed 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 [27] provided a comprehensive survey on crop monitoring techniques concerning crop yield estimation and disease detection using deep learning models. Based on the results and comparison, the authors claimed that crop monitoring techniques using deep learning methods are more accurate and powerful than some developing countries' traditional methods. In [28], the authors proposed a novel technique based on Support Vector Machines 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. The farmers can use this proposed methodology to gain a better yield of the crops, and different governmental sectors can use it to improve crop productivity. However, the authors did not suggest any recommendations for fertilizer systems to improve crop management.

In [29], a machine learning and Multilayer Perception method, along with the quantity of rainfall, have been suggested by the authors to help farmers in making a proper decision regarding a harvest even before they start planting. Furthermore, the suggested method focuses on the optimal process for the marketing and storage of the crop. Based on the provided results, the authors claimed that the proposed method would be beneficial for farmers to improve agricultural yield outcomes. In [30], the main objective of this research was to predict crop productivity loss by utilizing linear regression methods based on data taken from the previous year's statistics. The models were created using real-world data, and the evaluation process was based on samples. The authors applied both Naïve Bayes and Decision Tree algorithms in this evaluation. They claimed that the proposed method

improves production and maximizes the farmer's profit. In [31], the authors developed a web-based application for crop yield prediction to be used by farmers. This tool provides farmers with a list of various crops 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 make the best decision regarding market demand and prices.

In [32], the author's survey's main objective was to evaluate the performance of different publications from 2016 to 2020 that aimed to predict fungal illnesses on the crops. The authors 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 Naïve Bayes has been widely utilized and gained the best results regarding the yearly prediction of crop diseases. In [33], using machine learning, the authors proposed a system for the early prediction of crop diseases in plants by utilizing the 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 predict plant diseases and make the best decision regarding the type of crop to be planted. The authors in [34] recognized problems facing farmers in India regarding crop yield prediction; therefore, they collected datasets published online. To facilitate analyzing and studying data, the dataset was clustered using the K-Means Clustering algorithm, and the Naïve Bayes algorithm was used to recognize the best crop to plant. The provided analysis and results show that the proposed system is beneficial for farmers, not only for the early prediction of crop yields but also for selecting the best crop to plan.

Based on the above literature review, it can be noticed that this issue has been discussed in the literature from different perspectives. Most of the work in this field utilized machine learning algorithms to help farmers with crop prediction gain better yield and improve overall production. However, many research analyses did not consider a real problem for which to perform classification and performance evaluation. Further, the authors did not give clear notifications about the obtained accuracy. Exceptionally, in [26], the authors provided analysis and performance evaluation with an accuracy of 99.45%, but still on 22 selected crops only and not based on real data. With this paper, we aim to improve the performance of machine learning usage in smart farming by:

- Presenting experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms.
- The research outcome recommends that farmers could make more informed verdicts about factors that affect crop growth by analyzing wide-ranging data collected from farms, including real-time data from IoT sensors.
- As per the provided analysis in this research work, the machine learning algorithms demonstrate a high level of classification accuracy. Notably, the Bayes Net algorithm achieved an impressive accuracy of 99.59%, while both the Naïve Bayes Classifier and Hoeffding Tree algorithms yielded a remarkable accuracy of 99.46%. These results highlight the efficacy and reliability of these algorithms in accurately classifying the given data.
- Therefore, by integrating different technologies, the achieved results can be used as an indicator for the farmers to have early and informed decisions regarding crop prediction to improve productivity, and hence the overall economy will be improved, accordingly.

3. Machine Learning in Smart Farming

Innovative farming methods have profoundly transformed agriculture using advanced technologies to increase productivity, enhance sustainability, and lessen environmental harm. Machine learning (ML), a vital component of this change, has enabled various applications that simplify farming operations and better inform decision-making processes.

ML applications are widely used in livestock, water, soil, and crop management. ML improves animal welfare for livestock and boosts production, increasing sustainability via predictive modeling and real-time health monitoring [35]. ML is exploited to optimize irrigation and water usage for water management by analyzing various parameters. The application of newly emerging ML technologies and large amounts of available weather and water data makes it easier to manage water resources. This is especially helpful because natural events can be unpredictable, and the relationships between them can be complicated [36].

In soil management, ML helps analyze soil health, predict nutrient needs, and conclude the factors affecting soil distribution controls [37]. As in crop management, ML detects diseases and weeds, evaluates crop quality, recognizes species, and predicts crop yield.

Crop yield prediction is immensely important for farmers and policymakers, governments concerned with food security, and food marketing organizations [38]. These stakeholders can use yield prediction models 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 crop production changes. However, predicting crop production is a complex task, as it is affected by many factors, such as weather conditions, the kind of fertilizer used, soil type, and the variety of seeds. Consequently, tackling this task necessitates the incorporation of diverse datasets and a range of attribute types.

Supervised learning techniques benefit crop yield prediction among the numerous ML categories. This is due to their robust predictive capabilities and ability to handle different attribute types. These methods employ labeled data to forecast outcomes based on specific inputs. For example, they project crop yields based on weather conditions and soil quality data.

Forecasting crop yield can be achieved using a broad spectrum of ML techniques, including Artificial Neural Networks (ANNs), Support Vector Machines, and Random Forests [39–42]. These algorithms' ability to process historical and up-to-date information regarding weather, soil conditions, and crop health facilitates precise crop yield predictions. The resulting insights enable farmers to make educated decisions about planting, irrigation, and fertilization, ultimately leading to optimized yields and less resource wastage.

Machine learning is already playing an important role in providing farmers with information to make agriculture more efficient and productive, hence maximizing their profits. Therefore, farmers are highly encouraged to apply ML algorithms and techniques efficiently, especially while collecting, processing, and analyzing data. ML technology can provide a solution to most challenges farmers face [43]. It can help them predict the weather more accurately, decrease waste, boost output, and increase profit margins. In this regard, farmers are motivated to use the advanced technology to collect data, such as autonomous vehicles, variable rate technology, GPS-based soil sampling, automated hardware, telematics, software, sensors, cameras, robots, drones, GPS guidance, and control systems [44].

According to [45], two-thirds of the farmers worldwide struggle to use technology, and more than 50% are unaware of the existing solutions. Teaching farmers to work with machine learning can be a transformative step in modernizing agriculture and improving productivity. While AI technology may seem daunting to some, there are ways to educate farmers about the benefits of smart farming and the usage of machine learning algorithms. Farmers need to be aware of the great benefits they may achieve if they use automation and AI on their farms. Training and informative workshops are the fastest ways to do so. This training needs to include foundation knowledge of ML, identifying use cases in agriculture, the importance of data collection and preparation for better prediction, ML algorithms and their best applications, and addressing concerns or misconceptions related to ethical and privacy considerations. They should be introduced to success stories of smart farms and encouraged to increase collaboration and share knowledge with their peers. A very important point is to introduce them to user-friendly tools and platforms that do not

require programming skills for easy adaption. Since ML is a rapidly evolving field and new techniques and tools emerge frequently, we should encourage farmers to continue learning and stay updated with the latest developments in the field.

Despite the optimistic outlook of ML in agriculture, similar to any ML issue, the quality of the results is predominantly influenced by the quality of the input data. The efficacy of crop yield prediction depends heavily on the quality and availability of data. Crop prediction requires wide-ranging data, including weather, soil, historical yield, and satellite imagery. Guaranteeing data quality through efficient collection, preprocessing, and feature selection is critical in model development. However, agricultural big data introduces several challenges, which will be examined in the following section.

4. Crop Analysis and Prediction Benefits and Challenges

As stated earlier, while machine learning is being exploited in multiple fields, it remains an active area of research and a challenging one in the agricultural domain. This section summarizes the main benefits and challenges ML faces when used in crop analysis and predictions based on recent research [46–50]. Figure 2 shows AI-based crop analysis and prediction phases.
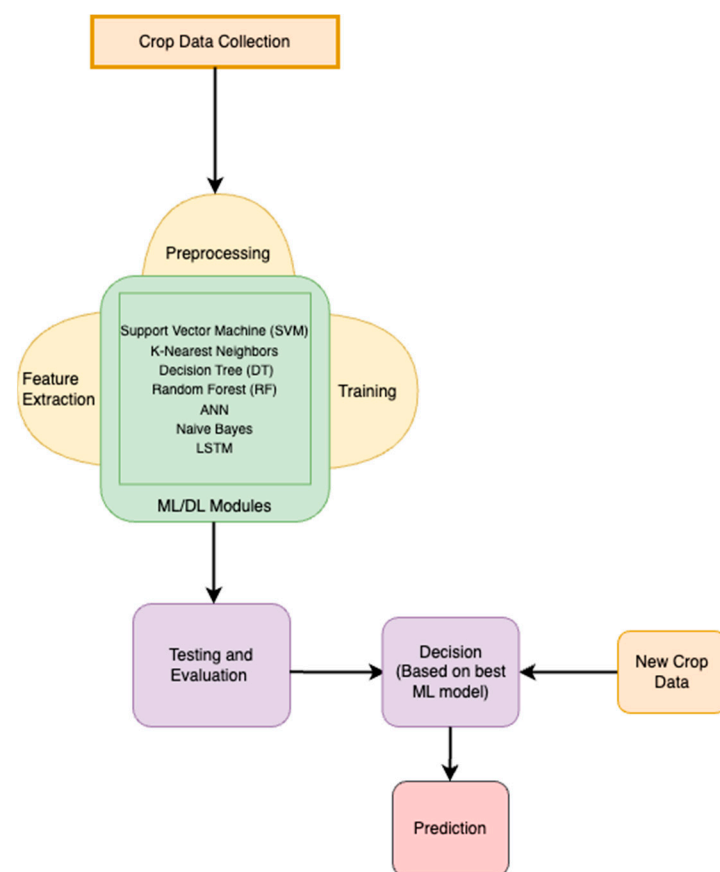
Figure 2. AI-based crop analysis and prediction.

4.1. Benefits

The following are the key advantages farmers can receive 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, such as water, fertilizer, and pesticides, by offering insights into crop development and health. This can save expenses while 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 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 humans would notice.
- Improved crop management: By offering insights into variables such as 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 promoting sustainable farming practices.

4.2. Challenges

Although crop analysis and prediction can significantly benefit from machine learning, there are also various challenges we need to consider. The main difficulties consist of the following:

- Data quality: The accuracy and dependability of machine learning 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. Ref. [51] discusses the main challenges related to fruit detection and recognition based on deep learning. They have concluded that most of the factors leading to low accuracy, slow speed, and poor robustness of fruit detection and recognition are related to the scarcity of high-quality fruit datasets, detection of small target fruits, fruit detection in occluded and dense scenarios, detection of multi-scale and multi-species fruits, and lightweight fruit detection models.
- Data volume: ML models frequently need a large amount of data for efficient training. Large data management and collection in agriculture can be complex, especially for small farms. Ref. [52] considers volume, velocity, variety, and veracity as the main challenges of big data.
- Model complexity: Because agricultural systems are intricate, it can be challenging to develop machine learning models that account entirely for all the important variables affecting crop development and output. Selecting the best model architecture for a specific crop analysis or forecast activity can be difficult and requires extensive knowledge [53]. Additionally, the most common use of ML techniques provides analysis for prediction, recommendations, situation determination, and automation. NN, RF, SVM, DT, and Naïve Bayes algorithms are the most popular techniques used in agronomy. The main challenges for these algorithms are the large volume of data, which increases the complexity of the training time and computation for SVM [53], and the need to tailor the algorithm for each specific problem in the case of RF [54]. The design of big data architecture is one of the most complex challenges, considering that it must be flexible and highly scalable [55]. Ref. [56] analyzed the factors affecting soil temperature and concluded that the relationship between variables affecting the soil temperature is quite complex and challenging, leading to the estimation of it using physically and statistically based models with a tradeoff between resolution, accuracy, and computational efficiency. According to them, the best ML technique for soil temperature retrieval generally depends on training datasets, model structure, and target level of accuracy.

Other challenges related to the usage of ML in agriculture are:

- 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, obtaining access to the hardware and software infrastructure required for developing and deploying ML models may be challenging.
- Privacy and security: These concerns exist around collecting, storing, and using 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 is possible that farmers and other interested parties need more time to be ready to 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.

5. Methodology

In our research endeavor, we deployed a comprehensive array of 15 diverse machine learning algorithms to construct models based on agricultural data, including 2200 records encompassing 22 distinct crop labels. These models provide farmers with recommendations on the most suitable crops to cultivate. Our methodology for crop analysis in Figure 3 adheres to the standard stages of data analysis [53–56]. A significant improvement is the inclusion of multiple classifiers, which are fine-tuned and evaluated to identify the most suitable ones for the input data. Moreover, our methodology incorporates feature reduction and augmentation techniques [57], essential for highlighting pertinent features that enhance crop detection and assist farmers in selecting the most appropriate features for accurate predictions.
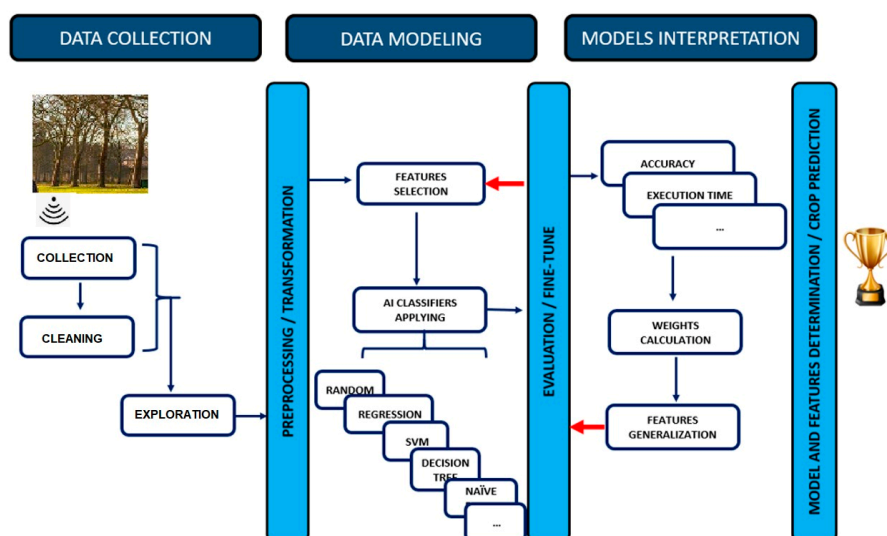
Figure 3. Methodology for crop prediction using IoT and ML.

In our study, we proposed applying various machine learning algorithms with different features for crop analysis. These algorithms were carefully chosen based on their unique capabilities and characteristics and included the Naïve Bayes Classifier, Random Forest, and Multilayer Neural Network. We thoroughly analyzed the results, emphasizing the essential features contributing to high accuracy. To ensure high-quality data, we emphasize the importance of data collection and preprocessing, including tasks such as cleaning and transformation. Figure 3 provides an overview of our crop analysis and prediction of our methodology, incorporating IoT and machine learning algorithms.

We have the following steps in our models for crop prediction.

– Data Collection: The collection of data from IoT devices on farms is vital for conducting machine learning analysis. By gathering crucial information on crop usage, crop type, water requirements, and harvest methods for agriculture, we can significantly increase productivity on smart farms.

– Data Modeling: The accuracy of data analysis was tested through experiments that involved altering labels. To group crops, we categorized them into four broad groups based on various factors, rather than predicting individual crop types. We analyzed a dataset of crop types using machine learning, utilizing seven different features to classify them. Additionally, we determined the minimum number of features necessary for precise learning and prediction.

– Model Evaluation and Interpretation: We want to achieve precise crop detection by selecting appropriate features for our machine learning algorithms. To ensure optimal results, we considered parameter properties on the relevant properties. This allowed us to obtain accurate and reliable data for our agricultural operations. During our experiments, we analyzed the effects of modifying the labels on the accuracy of our data analysis algorithm. This enabled us to understand the impact of minor label changes better and helped us optimize our approach toward achieving greater accuracy in our results. To achieve success in crop classification, it is vital to utilize broader labels. It is crucial to thoroughly investigate and ascertain the most efficient classification techniques within this domain.

Several classification algorithms, such as Naïve Bayes Classifiers, Random Forest, and Multilayer Neural Network, are used to build a model to predict the correct labels based on provided data. The constructed model was initially achieved by training using training data; then, the results were evaluated using test data to ensure that predictions were accurate and desired values.

The following algorithms were used in our experiment:

– 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. In practice, the Naïve Bayes algorithm performs well, especially when the data are sparse, and the number of features is extensive. 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}.

• For each unique class value c in y, calculate the prior probability for class c

$$P(c) = count(c)/m$$

where count (c) is the count of observations with class c.

• For each feature i in X and each unique class value c in y, calculate the conditional probability of feature i given class c as follows:

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

where count (i, c) is the count observations with the feature i and class c.

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

• 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. Random Forest creates many Decision Trees and combines their predictions to make a final prediction. More trees created yield high accuracy and robust results. Below is the pseudocode for the 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.
  ○ 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, ..., T_T.
- For each input data:

  ○ For each decision, find the prediction.
  ○ Obtain predictions of all the trees.
  ○ Calculate the final predicted class.

- Multilayer Neural Network is a machine learning algorithm consisting of multiple layers of interconnected nodes between the input and output layers.

To minimize the error between the predicted output and the actual output, the neural network involves adjusting each neuron's weight and biases by using an optimization algorithm such as backpropagation. Below is the pseudocode for the Multilayer Neural Network algorithm:

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

  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.

By integrating these methodologies, we aimed to provide farmers with efficient and effective crop recommendations, ensuring they receive tailored advice based on the most pertinent features while minimizing the required effort and time investment.

6. Experimental Results

In this work, we used 15 different machine learning algorithms to model agriculture data, which recommend to farmers the most suitable crops to produce on the farm. Tabular data are used in this work for classification of the crop data. Data are collected from the Kaggle database, which is an online platform for scientists to share their research data [58]. 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. The crop prediction dataset 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. The dataset includes several features, such as the ratio of Nitrogen content (N), temperature, pH value of the soil, rainfall, humidity, ratio of Phosphorous content (K), and the ratio of Potassium content (P) in the soil. Sixty-seven percent of the data is used in training, and the rest of the data are used for testing. Table 1 shows accuracy values and error rates for the algorithms considered. Bayes Net, Naïve Bayes Classifier, Hoeffding Tree, and Random Forest algorithms yield the best

accuracy. The DT algorithm yields 88.50%, and the rest of the classification algorithms have more than 90% accuracy. Error metrics can be formalized as follows:

$$K \text{ (Kappa)} = (P_o - P_e)/(1 - P_e) \tag{1}$$

where P_o is relative observed agreement and P_e is hypothetical probability of chance agreement.

$$MAE \text{ (Mean Absolute Error)} valueI = \frac{\sum\limits_{i=1}^{N} |y_i - x_i|}{n} \tag{2}$$

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

$$RMSE \text{ (Root Mean Square Error)} = \sqrt{\frac{\left[\sum\limits_{i=1}^{n} (x_i - x'_i)\right]}{n}} \tag{3}$$

where x_i is observed and x_i' is the predictive value.

$$RAE \text{ (Relative Absolute Error)} = \frac{\sum\limits_{i=1}^{n} |y_i - y'_i|}{\sum\limits_{i=1}^{n} |y_i - y|} \tag{4}$$

where y is the average value of the data.

$$\text{Root Relative Squared Error (RRSE)} = A = \sqrt{\frac{\sum\limits_{i=1}^{n} (P_i - T_j)^2}{\sum\limits_{i=1}^{n} (T_i - T)^2}} \tag{5}$$

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

Table 1. Accuracy and error values for each classification algorithm.

Method	Accuracy (%)	Kappa (0~1)	MAE (0~1)	RMSE (0~1)	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

Table 1. *Cont.*

Method	Accuracy (%)	Kappa (0~1)	MAE (0~1)	RMSE (0~1)	RAE (%)	RRSE (%)
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 each algorithm's build and test time. It can be noticed that the Multilayer 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 the input and output layer. The MLP algorithm runs with several iterations to find the best model for the given input and output set, which increases the build time. KSTAR and LWL algorithms have higher testing times than others. The NBC algorithm is the most efficient algorithm in classifying crop data based on accuracy and process time.

Table 2. Build and test times for classification algorithms.

Method	Build Time (Seconds)	Test Time (Seconds)
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 the Multilayer Perception algorithm with different sample sizes of training data. When 10% of the data is used in training, accuracy with MLP is 93.53%; however, when 90% of data is used, accuracy is 97.72%. In the MLP algorithm, reaching high accuracy and efficient time complexity is important. The MLP algorithm can be very efficient in numeric-based data; however, it will be slow if collected data are images or videos.

Table 4 presents the build and test times for the Multilayer Perception (MLP) algorithm under different scenarios involving modifications to the training and testing data percentage. It is worth noting that the MLP algorithm has longer build times than other algorithms [59,60]. 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%. On the other hand, test times are consistently low across all scenarios, with the highest being at 0.05 when the training set is at 10%. Experiments show that the MLP algorithm has higher build times than the other ML algorithms, but testing time is very efficient.

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

Training Set	Accuracy (%)	Kappa (0~1)	MAE (0~1)	RMSE (0~1)	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. Build and test times for Multilayer Perception algorithm (training set from 10% to 90%).

Training Set	Build Time (Seconds)	Test Time (Seconds)
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 recording 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. In this work, our primary focus is to determine the meaningfulness and correlation of the numerous features used for predicting crop outcomes. To achieve this, we utilized the Variance Inflation Factor (VIF), a statistical measure in machine learning, to assess multicollinearity [61]. By calculating the VIF for each variable, we gained valuable insights into potential correlations among predictor variables in the model. This crucial step allowed us to identify the best VIF values, indicating collinearity, and values below 10, signifying non-collinearity. As a result of this preparation, we were able to select the most suitable combination of features for our dataset. Table 5 shows the results of four scenarios where we selected three to four 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, and K, where the best accuracy obtained was 68.04% with Random Forest.

Table 5. Accuracy (%) for different feature sets.

Method	N, P, K	K, P, Rainfall	Temperature, Humidity, pH, Rainfall	N, Temperature, Humidity, pH
Bayes Net	67.64	85.69	97.05	89.70
Naïve Bayes Classifier	65.37	85.16	96.39	87.03
Logistic	66.17	74.19	85.42	76.07
Multilayer Perception	66.84	80.34	89.17	82.88
Simple Logistic	66.84	72.86	85.16	74.73
IBK	66.57	79.27	91.04	81.02
KSTAR	65.10	81.14	91.71	80.74
LWL	42.11	46.39	61.23	50.26
Ada BoostM1	6.81	6.81	6.81	6.81
Regression	65.37	84.22	95.98	86.49
Decision Table	63.77	79.27	74.73	72.19
Hoeffding Tree	65.37	85.29	96.52	86.89
J48	65.10	83.55	94.65	84.49
Random Forest	66.57	82.88	97.32	87.03
Random Tree	68.04	79.27	94.92	83.15

The results 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, and Rainfall) we identified can be used as a guide for selecting relevant features for crop determination in future agricultural data analysis.

More experiments were conducted to investigate the impact of changing the label on the accuracy of the data analysis algorithm. The results are displayed in Table 5 Instead of predicting the specific crop type, 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; 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 being cereal, whereas chickpeas and pigeon peas are classified as bushes, with the type being legumes.

This experiment aims to explore whether the algorithm could accurately predict more general categories of crops, which could be helpful in situations where general characteristics of crops are investigated, and specific crop types are not known or easily identifiable. The experiment results 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.

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
Kidney beans	Bush	Food	Legume	Drought	By Hand And Machine
Pigeon peas	Bush	Food	Legume	Drought Resistant	By Hand
Mothbeans	Bush	Fiber	Legume	Drought Resistant	Both
Mungbean	Bush	Food	Legume	Drought	Hand Picked
Black gram	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

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

7. Conclusions

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. This study identifies the challenges and opportunities associated with integrating these technologies in agriculture. It presents 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 suggest 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 in that machine learning approaches will become increasingly crucial for production predictions in agriculture in the future. In this experiment, crops were investigated according to general characteristics using different machine learning algorithms, and valuable results were obtained by making predictions in cases where certain crop types are unknown or cannot be easily identified. Our work indicated that appropriate feature selection is 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 could help optimize crop production, reduce waste, and improve food security globally.

In future work, more crop data will be evaluated using GPS-based IoT and sensor data from different geographic regions. All these results will be analyzed using a machine learning algorithm. Thus, our data evaluation pool will be established. In addition, different species of the same plant variety will be analyzed separately, and it will be possible to reveal which product is the best type of product among the same species using different machine learning algorithms.

References

1. Li, L.; Wang, B.; Feng, P.; Liu, D.L.; He, Q.; Zhang, Y.; Wang, Y.; Li, S.; Lu, X.; Yue, C.; et al. Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Comput. Electron. Agric.* **2022**, *194*, 106790. [CrossRef]
2. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]
3. Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* **2023**, *13*, 225. [CrossRef]
4. Xu, W.; Kaili, Z.; Tianlei, W. Smart Farm Based on Six-Domain Model. In Proceedings of the IEEE 4th International Conference on Electronics Technology (ICET), Chengdu, China, 7–10 May 2021; pp. 417–421.
5. Moysiadis, V.; Tsakos, K.; Sarigiannidis, P.; Petrakis, E.G.M.; Boursianis, A.D.; Goudos, S.K. A Cloud Computing web-based application for Smart Farming based on microservices architecture. In Proceedings of the 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 8–10 June 2022; pp. 1–5.
6. Ranjan, P.; Garg, R.; Rai, J.K. Artificial Intelligence Applications in Soil & Crop Management. In Proceedings of the IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 21–23 December 2022; pp. 1–5.
7. 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 Sens.* **2020**, *12*, 615. [CrossRef]
8. Gehlot, A.; Sidana, N.; Jawale, D.; Jain, N.; Singh, B.P.; Singh, B. Technical analysis of crop production prediction using Machine Learning and Deep Learning Algorithms. In Proceedings of the International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 24–25 September 2022; pp. 1–5.
9. Vashisht, S.; Kumar, P.; Trivedi, M.C. Improvised Extreme Learning Machine for Crop Yield Prediction. In Proceedings of the 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 27–29 April 2022; pp. 754–757.
10. OpenAI. New and Improved Content Moderation Tooling. OpenAI. 2022. Available online: https://openai.com/blog/new-and-improved-content-moderation-tooling/ (accessed on 1 April 2023).
11. Google. Bard Chatbox. Google. Available online: https://bard.google.com (accessed on 2 April 2023).
12. Dean, J. The deep learning revolution and its implications for computer architecture and chip design. In Proceedings of the IEEE International Solid-State Circuits Conference-(ISSCC), San Francisco, CA, USA, 16–20 February 2020.
13. Cui, Y.W.; Henrickson, K.; Ke, R.; Pu, Z.; Wang, Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4883–4894. [CrossRef]
14. Shahrin, F.; Zahin, L.; Rahman, R.; Hossain, A.J.; Kaf, A.H.; Abdul Malek Azad, A.K.M. Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning. In Proceedings of the 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; pp. 21–24.
15. Tawseef, A.S.; Tabasum, R.; Faisal, R.L. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **2022**, *198*, 107119.
16. Senthil KS, D.; Mary, D.S. Smart farming using Machine Learning and Deep Learning techniques. *Decis. Anal. J.* **2022**, *3*, 100041.
17. Senthil, K.M.; Akshaya, R.; Sreejith, K. An Internet of Things-based Efficient Solution for Smart Farming. *Procedia Comput. Sci.* **2023**, *218*, 2806–2819.
18. Vivek, S.; Ashish, K.T.; Himanshu, M. Technological revolutions in smart farming: Current trends, challenges & future directions. *Comput. Electron. Agric.* **2022**, *201*, 107217.
19. Mamatha, J.C.K. Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques. *Meas. Sens.* **2023**, *25*, 100665. [CrossRef]
20. 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. *IEEE Access* **2021**, *9*, 63406–63439. [CrossRef]

21. Babber, J.; Malik, P.; Mittal, V.; Purohit, K.C. Analyzing Supervised Learning Algorithms for Crop Prediction and Soil Quality. In Proceedings of the 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 29–31 March 2022; pp. 969–973.

22. Ishak, M.; Rahaman, M.S.; Mahmud, T. FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing. In Proceedings of the 13th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 17–20 April 2021; pp. 224–229.

23. Patel, K.; Patel, H.B. A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction. In Proceedings of the Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 15–17 September 2022; pp. 1–5.

24. Memon, R.; Memon, M.; Malioto, N.; Raza, M.O. Identification of growth stages of crops using mobile phone images and machine learning. In Proceedings of the International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, 26–27 October 2021; pp. 1–6.

25. Chandraprabha, M.; Dhanaraj, R.K. Soil Based Prediction for Crop Yield using Predictive Analytics. In Proceedings of the 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 17–18 December 2021; pp. 265–270.

26. Ray, R.K.; Das, S.K.; Chakravarty, S. Smart Crop Recommender System-A Machine Learning Approach. In Proceedings of the 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 27–28 January 2022; pp. 494–499.

27. 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. In Proceedings of the International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 10–11 March 2022; pp. 1–5.

28. Malathy, S.; Vanitha, C.N.; Kotteswari, S.; Mohankkanth, E. Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques. In Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 437–442.

29. Chowdary, V.T.; Robinson Joel, M.; Ebenezer, V.; Edwin, B.; Thanka, R.; Jeyaraj, A. A Novel Approach for Effective Crop Production Using Machine Learning. In Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 16–18 March 2022; pp. 1143–1147.

30. Yamparla, R.; Shaik, H.S.; Guntaka, N.; Marri, P.; Nallamothu, S. Crop Yield Prediction using Random Forest Algorithm. In Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 22–24 June 2022; pp. 1538–1543.

31. Apeksha, R.G.; Swati, S.S. A brief study on the prediction of crop disease using machine learning approaches. In Proceedings of the 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 18–19 June 2021; pp. 1–6.

32. Kumar, R.; Shukla, N.; Princee. Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning. In Proceedings of the International Mobile and Embedded Technology Conference (MECON), Noida, India, 10–11 March 2022; pp. 168–172.

33. Bhosale, S.V.; Thombare, R.A.; Dhemey, P.G.; Chaudhari, A.N. Crop Yield Prediction Using Data Analytics and Hybrid Approach. In Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018.

34. Alwis, S.D.; 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]

35. Lyu, Y.; Li, J.; Hou, R.; Zhang, Y.; Hang, S.; Zhu, W.; Zhu, H.; Ouyang, Z. Precision Feeding in Ecological Pig-Raising Systems with Maize Silage. *Animals* **2022**, *12*, 11. [CrossRef] [PubMed]

36. Ghobadi, F.; Kang, D. Application of Machine Learning in Water Resources Management: A Systematic Literature Review. *Water* **2023**, *15*, 4. [CrossRef]

37. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review aided by machine learning tools. *SOIL* **2020**, *6*, 35–52. [CrossRef]

38. Ramos, P.J.; Prieto, F.A.; Montoya, E.C.; Oliveros, C.E. Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.* **2017**, *137*, 9–22. [CrossRef]

39. Sengupta, S.; Lee, W.S. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. *Biosyst. Eng.* **2014**, *117*, 51–61. [CrossRef]

40. 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**, *24*, 537–547. [CrossRef]

41. Adankon, M.M.; Cheriet, M. Support Vector Machine. In *Encyclopedia of Biometrics*; Li, S.Z., Jain, A., Eds.; Springer: Boston, MA, USA, 2009; pp. 1303–1308.

42. 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**, *10*, 3254–3264. [CrossRef]

43. Jadhav, M.; Kolambe, N.; Jain, S.; Chaudhari, S. Farming Made Easy using Machine Learning. In Proceedings of the 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021; pp. 1–5. [CrossRef]

44. How Is ML Is Used in Agriculture? Available online: https://www.dtn.com/how-is-machine-learning-used-in-agriculture/ (accessed on 1 August 2023).

45. Sawhney, D. Redefining Agriculture through Artificial Intelligence: Predicting the Unpredictable. p9. GG—May 2022—M&C 19416. Available online: https://www.pwc.in/assets/pdfs/grid/agriculture/redefining-agriculture-through-artificial-intelligence.pdf (accessed on 1 August 2023).

46. Pyingkodi, M.; Thenmozhi, K.; Karthikeyan, M.; Kalpana, T.; Palarimath, S.; Kumar, G.B.A. IoT-based Soil Nutrients Analysis and Monitoring System for Smart Agriculture. In Proceedings of the 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 17–19 August 2022; pp. 489–494.

47. 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. *Inf. Process. Agric.* **2018**, *5*, 21–32. [CrossRef]

48. Patel, N.S.; Kumar, H.P.M. Soil Quality Identifying and Monitoring Approach for Sugarcane Using Machine Learning Techniques. In Proceedings of the Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 26–27 December 2022; pp. 1–5.

49. Puengsungwan, S. IoT-based Soil Moisture Sensor for Smart Farming. In Proceedings of the International Conference on Power, Energy, and Innovations (ICPEI), Chiangmai, Thailand, 14–16 October 2020; pp. 221–224.

50. 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. *IEEE Access* **2022**, *10*, 87333–87360. [CrossRef]

51. Xiao, F.; Wang, H.; Xu, Y.; Zhang, R. Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review. *Agronomy* **2023**, *13*, 1625. [CrossRef]

52. Cravero, A.; Pardo, S.; Sepúlveda, S.; Muñoz, L. Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy* **2022**, *12*, 748. [CrossRef]

53. L'heureux, A.; Grolinger, K.; Elyamany, H.; Capretz, M. Machine learning with Big Data: Challenges and approaches. *IEEE Access* **2017**, *5*, 7776–7797. [CrossRef]

54. Del Río, S.; López, V.; Benítez, J.M.; Herrera, F. On the use of Map Reduce for imbalanced Big Data using Random Forest. *Inf. Sci.* **2014**, *285*, 112–137. [CrossRef]

55. Salma, C.A.; Tekinerdogan, B.; Athanasiadis, I.N. Chapter 4—Domain-Driven Design of Big Data Systems Based on a Reference Architecture. In *Software Architecture for Big Data and the Cloud*; Morgan Kaufmann: Burlington, MA, USA, 2017; pp. 49–68.

56. Taheri, M.; Schreiner, H.K.; Mohammadian, A.; Shirkhani, H.; Payeur, P.; Imanian, H.; Cobo, J.H. A Review of Machine Learning Approaches to Soil Temperature Estimation. *Sustainability* **2023**, *15*, 7677. [CrossRef]

57. Maharana, K.; Mondal, S.; Nemade, M. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [CrossRef]

58. Available online: https://www.kaggle.com/code/theeyeschico/crop-analysis-and-prediction (accessed on 30 March 2023).

59. Elbasi, E.; Mostafa, N.; AlArnaout, Z.; Zreikat, A.I.; Cina, E.; Varghese, G.; Shdefat, A.; Topcu, A.E.; Abdelbaki, W.; Mathew, S.; et al. Artificial Intelligence Technology in the Agricultural Sector: A Systematic Literature Review. *IEEE Access* **2023**, *11*, 171–202. [CrossRef]

60. Elbasi, E.; Zreikat, A.I.; Mathew, S.; Topcu, A.E. Classification of influenza H1N1 and COVID-19 patient data using machine learning. In Proceedings of the 44th International Conference on Telecommunications and Signal Processing (TSP), Brno, Czech Republic, 26–28 July 2021; pp. 278–282.

61. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42. [CrossRef]