

Article

Optimizing Agricultural Data Analysis Techniques through AI-Powered Decision-Making Processes

Ersin Elbasi *, Nour Mostafa *, Chamseddine Zaki, Zakwan AlArnaout , Ahmet E. Topcu and Louai Saker

College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; chamseddine.zaki@aum.edu.kw (C.Z.); zakwan.alarnaout@aum.edu.kw (Z.A.); ahmet.topcu@aum.edu.kw (A.E.T.); louai.saker@aum.edu.kw (L.S.)

* Correspondence: ersin.elbasi@aum.edu.kw (E.E.); nour.moustafa@aum.edu.kw (N.M.)

Abstract: The agricultural sector is undergoing a transformative paradigm shift with the integration of advanced technologies, particularly artificial intelligence (AI), to enhance data analysis techniques and streamline decision-making processes. This paper delves into the integration of advanced technologies in agriculture, focusing specifically on optimizing data analysis through artificial intelligence (AI) to strengthen decision-making processes in farming. We present a novel AI-powered model that leverages historical agricultural datasets, utilizing a comprehensive array of established machine learning algorithms to enhance the prediction and classification of agricultural data. This work provides tailored algorithm recommendations, bypassing the need to deploy and fine-tune numerous algorithms. We approximate the accuracy of suitable algorithms, highlighting those with the highest precision, thus saving time by leveraging pre-trained AI models on historical agricultural data. Our method involves three phases: collecting diverse agricultural datasets, applying multiple classifiers, and documenting their accuracy. This information is stored in a CSV file, which is then used by AI classifiers to predict the accuracy of new, unseen datasets. By evaluating feature information and various data segmentations, we recommend the configuration that achieves the highest accuracy. This approach eliminates the need for exhaustive algorithm reruns, relying on pre-trained models to estimate outcomes based on dataset characteristics. Our experimentation spans various configurations, including different training–testing splits and feature sets across multiple dataset sizes, meticulously evaluated through key performance metrics such as accuracy, precision, recall, and F-measure. The experimental results underscore the efficiency of our model, with significant improvements in predictive accuracy and resource utilization, demonstrated through comparative performance analysis against traditional methods. This paper highlights the superiority of the proposed model in its ability to systematically determine the most effective algorithm for specific agricultural data types, thus optimizing computational resources and improving the scalability of smart farming solutions. The results reveal that the proposed system can accurately predict a near-optimal machine learning algorithm and data structure for crop data with an accuracy of 89.38%, 87.61%, and 84.27% for decision tree, random forest, and random tree algorithms, respectively.



Citation: Elbasi, E.; Mostafa, N.; Zaki, C.; AlArnaout, Z.; Topcu, A.E.; Saker, L. Optimizing Agricultural Data Analysis Techniques through AI-Powered Decision-Making Processes. *Appl. Sci.* **2024**, *14*, 8018. <https://doi.org/10.3390/app14178018>

Academic Editors: Andrea Prati and Alexander N. Pisarchik

Received: 14 July 2024

Revised: 5 August 2024

Accepted: 3 September 2024

Published: 7 September 2024

Keywords: machine learning; agricultural data analysis; decision making; Internet of Things; smart farming; zero hunger



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Smart farming is an emerging innovative technology to manage farm operations, mainly using the Internet of Things, cloud computing, and artificial intelligence to improve the quantity and quality of products by optimizing the use of resources and reducing environmental impact. Farming has always been a cornerstone of human societies, providing nutrition for a growing population and producing essential supplies like medicine, fiber, and fuel. Smart farming integrates digital technologies, IoT, cloud computing, robotics, sensors, GPS, and artificial intelligence into farm operations [1]. A smart farm's infrastructure

includes sensors for environmental data collection and surveillance cameras. These data are transmitted to cloud-based service platforms accessible to farmers through gateways. Like other agricultural technologies, smart farming aims to enhance food production through precise control and optimization [2]. Smart farming has expanded to encompass various domains over the past decade, including large-scale field crops, controlled-environment agriculture (greenhouses), and dairy and poultry farming. Compared to traditional farming methods, smart farming offers significant advantages due to its integration of advanced technologies, modern facilities, and data gathered throughout the farming process. Proper monitoring enables the enhancement of the quality of agricultural products. For instance, potential pest infestations or diseases can be detected early using collected data. Historical patterns can support farm workforce planning, resource optimization, and decrease excessive pesticide and fertilizer use [3].

1.1. Technologies Used in Smart Farming

Smart farming covers a variety of technologies [4], including (i) AI algorithms that perform data analysis to recognize patterns, aiding the farmers in making informed decisions about fertilization, irrigation, and pest control; (ii) Internet of Things (IoT) sensors that are used to gather data (temperature, soil moisture, humidity, and crop health) from the fields; (iii) Geographic Information System (GIS) software that is used to map and analyze spatial data, allowing farmers to optimize resource usage, field layout, and crop selection, which refers to the process of choosing the most suitable crops to cultivate, based on various data-driven insights and advanced technologies; in smart farming, this selection process involves the use of sensors, IoT devices, data analytics, and sometimes AI algorithms; and (iv) robotics and automation, which are used to perform tasks such as weeding, harvesting, and milking. Figure 1 above shows the different technologies that can be used in smart farming such as autonomous tractors, farming drones, field sensors, etc.

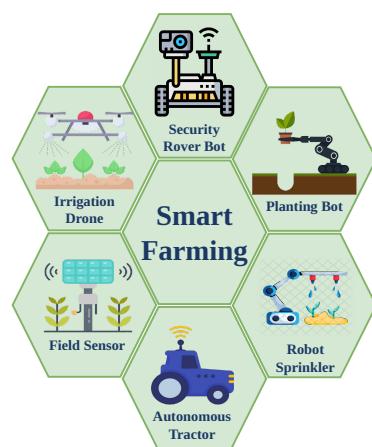


Figure 1. Technologies used in smart farming.

1.2. Benefits of Smart Farming

Smart farming has many benefits, including [5] (i) improving crop quantity and quality by optimizing resources and precise crop management, so farmers can achieve higher yields and better-quality crops; (ii) cost reduction in terms of water usage, fertilizers, and pesticides; (iii) improving productivity by using robotics and drones to efficiently reduce the time spent on tedious tasks, allowing farmers to work on other tasks; (iv) enhanced decision making that is augmented by machine learning algorithms to provide farmers with valuable insights to make better decisions about their crops and livestock; and (v) improving sustainability by reducing the environmental impact of farming on aspects such as water pollution and greenhouse gas emissions.

1.3. Challenges in Smart Farming

Despite its potential benefits, smart farming is still facing several challenges [5], including (i) high costs incurred by the investment in the technology, since the initial investment in the equipment and software is significant, making it difficult for small-scale farmers to adopt; (ii) complexity of data management, caused by the vast amounts of data generated that can be difficult to collect, store, and analyze; moreover, farmers need to have the skills to manage this data effectively; (iii) cybersecurity risks, where potential adversaries can initiate cyberattacks on vulnerable systems, causing the disruption of operations or even compromising the safety of food products; and (iv) lack of standardization, which can make it difficult for.

1.4. Applications

AI is rapidly transforming the shape of the farming industry, offering a wide range of innovative solutions to enhance crop yields, optimize resource utilization, and improve overall farm efficiency [6]. (i) Crop and soil monitoring: Analyzing sensed data to give insights into crop health, soil conditions, and nutrient deficiencies. (ii) Pest and disease detection: Using AI algorithms to identify pests, diseases, and weeds in crops using image recognition algorithms applied to drone imagery or field sensors. (iii) Precision agriculture: AI-based equipment enables farmers to apply water, fertilizer, and pesticides more precisely and efficiently. (iv) Automated machinery: AI is leading the development of autonomous farm machinery, including driverless tractors, robotic harvesters, and intelligent irrigation systems. (v) Livestock management: AI-enabled systems can monitor livestock health and movement, and optimize feeding strategies.

The contribution of this paper is to present a comprehensive approach to optimizing agricultural data analysis through the selection of appropriate machine learning algorithms. The study focuses on predicting and classifying agricultural data by experimenting with various AI models using historical datasets. The proposed approach works by preprocessing data, tuning hyperparameters, and logging the performance metrics (accuracy, precision, recall, F1-score) for each model and dataset configuration. A detailed log file is created that captures the outcomes of these experiments, which is then used to train a new predictive model. This model provides recommendations for selecting suitable algorithms for new agricultural datasets based on their features. The proposed approach aims to save time and resources by offering accurate and reliable predictions without the need for real-time algorithm execution. This tool is particularly valuable for researchers, enabling them to make informed decisions and improve the efficiency of agricultural data analysis. To the best of our knowledge, there is no previous work in the literature of smart farming that has proposed a similar optimization model.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive review of the literature, highlighting key advancements and identifying gaps in the use of AI in agriculture. This section sets the stage for the subsequent research by detailing the current state of the art. Section 3 describes the IoT and machine learning (ML) models that form the backbone of our analytical approach. We discuss the integration of IoT for data collection, emphasizing how ML can leverage this data for enhanced decision making. In Section 4, we introduce our proposed methodology, detailing the algorithms selected for optimizing agricultural data analysis and the rationale behind these choices. This section also covers the preprocessing and feature extraction techniques we employed to prepare the data for effective analysis. Section 5 discusses the experimental setup and the results obtained from implementing our AI-powered models. It includes an evaluation of the model's performance using various metrics such as accuracy, precision, recall, and F-measure, providing a critical analysis of their effectiveness in real-world scenarios. Finally, Section 6 concludes the paper with a summary of our findings and contributions. It also outlines potential future work that could further enhance the capabilities of AI in smart farming, suggesting directions for upcoming research to build on the foundation laid by this study.

2. Literature Review

The authors in [7] showed the potential of AI sensors in smart farming, specifically, using Agrobot. The study found that IoT sensors play a major role in sensing data and providing overall performance in agriculture fields. The proposed Agrobot system is designed to limit the work of farmers by performing rudimentary functions such as sowing seeds and covering them with soil. The use of wireless sensors allows for high-accuracy tracking and early detection of unwanted seeds. Cloud-based IoT agriculture provides real-time monitoring of environmental conditions to further optimize crop growth. This paper highlights the potential for AI sensors to revolutionize smart farming practices and improve agricultural productivity. While the paper outlines the Agrobot's functionality, it lacks detailed information about the methodology, such as the specifics of the AI algorithms used, sensor calibration, and data processing techniques. In addition, it lacks a comparative analysis with existing agricultural technologies, which would help in understanding the advantages or improvements this system offers over traditional methods.

Ref. [8] presented frameworks based on precision agriculture that utilized machine learning techniques to improve crop yields and reduce costs. The authors used experiments with different models to classify crops based on their attributes, and the best-performing model was used to recommend crops. The study found that machine learning significantly improved the accuracy of crop classification and recommendation compared to traditional methods. The paper emphasized the importance of data preprocessing techniques in improving the performance of machine learning models and highlighted the potential of IoT devices in real-time monitoring and decision making. The study concludes that machine learning has a significant impact on precision agriculture by providing better yields at lower costs. The paper seems to provide a broad overview of techniques rather than an in-depth analysis or novel research contributions. A deeper exploration of specific methodologies or case studies could enhance its impact. While the paper mentions various machine learning algorithms, it lacks a detailed discussion on their implementation, customization, or comparative efficacy in different agricultural scenarios.

Ref. [9] highlights the successful application of hyperspectral imaging and AI technology to detect gray mold on tomato leaves in smart agriculture. Optimal wavelengths were selected through AIC-assisted pLSA and a Bayesian network, and the predicted images using full spectral and selected wavelength models produced similar results. Clustering and directed graphs were used to aid in preserving information and human interpretation. AI technology has great potential for automating tasks like pest control patrols, and for improving crop management and disease detection in smart agriculture. Future research could explore the application of this technology to other crops and optimize the selection process for optimal wavelengths. The paper's approach to integrating AI with hyperspectral imaging for disease detection is a significant contribution, but it could benefit from comparing its methods with existing techniques.

The systematic review of 30 papers on AI in precision agriculture in [10] found that AI technologies can increase productivity and efficiency while addressing labor shortages and environmental sustainability concerns. However, it presents risks associated with the need to adapt and maintain technology, and understanding food security is crucial. The research suggests that cooperative and emerging technologies can benefit greatly from advanced technology in precision agriculture, but the impact of modern technology on farmers should be a focus of future research. Overall, AI in precision agriculture can provide multiple benefits to farmers, but its risks and social impact should be carefully considered. The paper could be strengthened by providing clear implications or recommendations for practitioners in the field of precision agriculture.

Ref. [11] discusses the integration of AI, IoT, and cloud computing in smart agriculture for sustainable and efficient food production. The use of artificial neural networks to predict crop growth and optimize water management, IoT for data collection, and cloud computing for data analysis are explored. The proposed IoT and AI technology offers high-end agricultural machinery, weather observation, and several applications for pre-

dicting and techniques for crop control. The integration of these technologies can lead to better resource management, increased production, and a more sustainable agriculture sector. The use of AI, cloud computing, and IoT in agriculture is made possible by ICT technologies. The paper provides a general overview but lacks in-depth analysis or specific case studies demonstrating the practical application and outcomes of these technologies in agriculture. In addition, more detailed information on the implementation of AI and IoT systems, including data collection, processing, and analysis methods, would enhance the paper's value.

Ref. [12] reviews the latest advancements in crop prediction using deep learning methods, including artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks with long short-term memory (RNN-LSTM), and hybrid networks in agriculture applications. The study found that CNNs outperform ANNs, with an accuracy of around 87%. RNN-LSTM and hybrid networks also showed promising results in crop yield prediction. The study highlights potential techniques to improve the accuracy and robustness of crop yield prediction models, such as data preprocessing, transfer learning, and ensemble methods. The paper also highlights the potential applications of deep learning in agriculture beyond crop yield prediction, such as plant disease detection and precision agriculture. While the paper reviews a range of techniques, it could provide a more in-depth analysis of each method, including specific strengths and limitations in different agricultural contexts.

A study was conducted in [13] in 33 districts using four machine learning algorithms and one deep learning algorithm to predict crop yields. The random forest algorithm outperformed the others with an accuracy of 91.5%. Temperature and rainfall were identified as the most important factors affecting crop yield. Machine learning algorithms were found to perform better than traditional statistical methods, providing valuable insights into the factors affecting agricultural productivity. The study provides a useful tool for policymakers to make informed decisions about agricultural planning and resource allocation. The methodology can be extended to other regions with similar climatic conditions and agricultural practices. The paper details its data preprocessing methods, which is a strength. However, it could further discuss the impact of data quality and completeness on model performance. Moreover, the study validates models using R², RMSE, and MAE, which is appropriate. However, a more in-depth discussion of these metrics in the context of agricultural data could enhance the paper.

A machine learning-enabled framework was developed in [14] to predict crop yields with an average accuracy of 92.5%. The relief algorithm selected important features, and the LDA algorithm extracted features from the input dataset, resulting in improved accuracy. PSO-SVM, KNN, and random forest algorithms were used to classify crop yields based on the selected features. The framework achieved high levels of accuracy for various crops and could predict yields up to two months before harvest time. Machine learning algorithms allowed for more efficient and accurate predictions compared to traditional methods, potentially optimizing farming practices and improving food security. Further research is needed to explore the scalability and practical applications of this framework in real-world farming scenarios. While the paper outlines the application of various machine learning methods, it could provide more detailed explanations of the algorithms and their specific configurations for the study. The paper discusses using a dataset for experiments and feature selection using the relief algorithm. More details about the dataset, such as its size, source, and characteristics, would add value.

A systematic literature review in [15] analyzed 456 studies on deep learning in crop yield prediction, selecting 44 for analysis. The review found that convolutional neural networks (CNNs) were the most used algorithm for crop yield prediction, with the best performance in terms of root mean square error (RMSE). Challenges in applying deep learning technologies to crop yield prediction include a lack of large training datasets, issues with data preprocessing and feature extraction, and the importance of using remote sensing and weather data. Ensemble learning methods were found to be effective in

improving the accuracy of crop yield prediction models. The systematic literature review approach is methodologically sound, but the paper could benefit from a more detailed discussion on the criteria for selecting and evaluating studies.

A study in [16] presents an optimized deep learning model for crop yield prediction, achieving 97% accuracy and surpassing existing models. The approach considers various factors such as weather conditions, soil qualities, water levels, and location of the farm and uses a discrete deep belief network with the VGG-Net classification method. The data were collected from India's state agriculture webpage and feature preprocessing and feature extraction using the adaptive shearlet technique. The study demonstrates the potential of the proposed method for improving global food production and helping farmers make better management and financial decisions, with millet, rice, and wheat found to be the optimal crops for a high yield. A more in-depth explanation of the chosen deep learning techniques and the rationale for using the tweak chick swarm optimization algorithm would enhance understanding. The paper claims high accuracy, but a more detailed analysis of the testing process, including datasets and validation methods, is necessary for credibility.

The study in [17] aims to predict Irish potato and maize crop yields using machine learning models. Three models, random forest, support vector regression, and polynomial regression were used, with rainfall and temperature used as climate-related predictors. The results showed that RF outperformed SVR and PR in predicting Irish potato yield, while it performed equally well as SVR in predicting maize yield. The study recommends further research on incorporating other variables, such as air humidity, soil moisture, and solar radiation, and exploring other machine learning algorithms, such as artificial neural networks and deep learning, for crop yield prediction. The findings can inform farmers, policymakers, and researchers about crop production and management. The study focuses on specific crops in a particular region, which might limit the generalizability of the results to other crops and regions. The paper could provide a more detailed justification for the selection of the specific machine learning models used.

The study in [18] presents an approach using machine learning and deep learning techniques to identify and classify crop diseases automatically. The proposed method uses image processing techniques such as image acquisition, preprocessing, segmentation, feature extraction, SVM, CNN, KNN, and GLCM to identify plant diseases. The proposed method achieved an accuracy of 98.5% in identifying plant diseases and can be used as a tool for early detection and prevention of plant diseases, which can increase crop output by detecting diseases early on and taking appropriate measures to prevent further spread. Future work can focus on improving the performance of the proposed method by incorporating more advanced machine learning and deep learning techniques. The paper lacks a detailed description of the testing process, including datasets and validation methods, to support its high-accuracy claim. The paper should discuss the real-world applicability of this model, especially considering various agricultural environments and crop types.

A deep neural network (DNN) algorithm was proposed in [19] for disease classification, pesticide recommendation, and preprocessing of leaf images from a plant village dataset. The DNN algorithm achieved an accuracy rate of 98.5% in classifying leaf diseases and recommended suitable pesticides with an accuracy rate of 95%. Preprocessing techniques and geometric manipulation of the plant leaf image dataset improved accuracy rates. The proposed DNN algorithm can detect pest infestations with an accuracy rate of 92%. The study highlights the potential for deep learning techniques to revolutionize agriculture practices in India by reducing the time and cost involved in the manual inspection and diagnosis of crop diseases. The use of transfer learning with ResNet-50 is a common approach in image recognition, so the paper could benefit from highlighting any unique adaptations or improvements made for its specific application in agriculture.

This section successfully maps out a broad spectrum of AI applications in agriculture, demonstrating a clear trajectory from basic digital applications to sophisticated machine learning (ML) and AI interventions. Key thematic areas identified include:

- Smart farming technologies: Prior studies frequently highlight the use of IoT sensors, robotics, and cloud computing. For instance, Refs. [7–9] discuss the integration of these technologies to enhance yield and operational efficiency.
- AI and machine learning models: The manuscript cites work [10–12] that employs AI algorithms for predictive analytics and operational optimization, underscoring the evolving complexity of AI applications in predicting crop yields and managing resources.
- Challenges and benefits: The paper discusses both the potential and limitations of smart farming, with particular attention to the scalability of technology and cybersecurity risks as per Refs. [13,14].
- Applications of deep learning: Several cited studies (e.g., [15–17]) delve into specific uses of deep learning and neural networks for tasks like pest detection and crop monitoring, reflecting a nuanced application of AI in agriculture.

The reviewed studies contribute significantly to the field of smart agriculture through various technological advancements, including AI sensors, machine learning frameworks, hyperspectral imaging, and IoT integration. However, these studies primarily focus on specific applications, methodologies, and performance metrics without proposing a holistic, adaptable solution for new case studies. Our proposed work distinguishes itself by introducing a generic model capable of automatically suggesting the optimal configuration for any new case study based on prior agricultural dataset analysis. Unlike previous studies that focus on specific algorithms or methodologies, our model leverages a comprehensive approach to analyze characteristics extracted from new datasets, including the number of features, records, and feature types, to recommend the best algorithm and data splitting strategy. It is an adaptable solution that can be tailored to various agricultural scenarios, improving efficiency and productivity in smart farming practices.

3. IoT ML Models

Big data has a significant impact on productivity by using ML applications. Effective results can be obtained in various fields by using the Internet of Things (IoT) and machine learning (ML) in different applications within ML models. The IoT allows ML algorithms to collect data from physical devices and sensors and share them over the internet, learn patterns from these data, and make predictions on desired topics. By using IoT and ML together, smarter and more automatic decision-making mechanisms can be obtained in agriculture. This section highlights the importance of IoT data collection in agriculture. It demonstrates how these data can be leveraged through cutting-edge machine learning models, including neural networks and deep learning.

3.1. Data Collection and IoT

In the preceding section, we reviewed several studies related to AI-driven decision-making processes in machine learning. This section delves deeper into the various methodologies used for data collection and the application of machine learning models. Collecting and analyzing sensor data from numerous IoT devices is essential for smart agriculture. The IoT enables data from sources to be communicated, organized, and primarily transmitted to cloud-based centers via the network. Currently, numerous environmental factors are reducing food production, and farmers must adopt new IT technologies on their farms to achieve the most productive crops. In smart agriculture, other relevant information, including weather, soil, and water obtained from the crops, is typically collected and analyzed using machine learning algorithms. The built models can identify solutions that utilize specific IoT sensors and sensor technology for assessing weather conditions and soil quality. Additionally, models could be improved to monitor crop development, using robots for harvesting and weeding. IoT-based data collection methods can be summarized as follows: (a) weather stations, (b) crop monitoring sensors, (c) soil sensors, (d) intelligent irrigation systems, (e) innovative agricultural machinery, (f) smart greenhouses, (g) drone technologies, and (h) data collection platforms.

From the standpoint of precision agriculture, the IoT data acquired and managed can be invaluable in enhancing efficiency across various agricultural sectors. Data-driven farms can also aid in eradicating plant diseases, analyzing farm soils, and boosting crop productivity. According to [20], work has been carried out on IoT technology for remote sensing and agriculture applications. Data collection enabled by IoT optimizes resource usage and enhances farm efficiency, contributing to precision agriculture. In another recent study [21], the authors developed a data collection model using the ZigBee wireless sensor network, covering all aspects of crops, to establish a new agricultural model.

To guarantee effective and reliable data collection and transmission, the hardware utilized in the data collection system incorporates a diverse number of sensors (including temperature and soil moisture sensors), microcontrollers (such as Arduino and Raspberry Pi), and communication modules (like Wi-Fi, LoRa, and Zigbee). This comprehensive approach enables the successful implementation of machine learning models. One of the important modules is ZigBee, which stands out as a robust wireless communication protocol with diverse applications, especially in scenarios requiring low power usage and strong network topology. It is capable of facilitating extensive mesh networks and is adaptable for various IoT applications in smart farms.

3.2. Machine Learning in Agriculture

In agriculture, machine learning algorithms are commonly employed in various ways, such as diagnosing diseases that may occur on agricultural land, deciding on the most suitable crops by conducting soil analysis, and using data obtained and collected through IoT applications to develop product prediction models. Additionally, these algorithms are used to examine and monitor the plant development using images captured by drones. Efficient and effective irrigation systems can be created using various machine learning methodologies, which are effectively utilized to classify agriculture data, significantly improving productivity and quality. The significance of machine learning methods is indispensable in intelligent systems in numerous agricultural fields.

Many studies and applications in this area aim to enhance yield and product quality in agriculture, diagnose diseases caused by insects and precipitation, and implement preventive measures. Machine learning (ML) is widely used across various domains to analyze historical data and make accurate predictions. Algorithms that can calculate the performance of suitable agricultural models for fields include decision trees, Bayes networks, random forests, Hoeffding trees, support vector machines, and artificial neural networks. These algorithms can predict crop yields, detect diseases, and optimize irrigation schedules based on diverse agricultural factors. Watering systems are crucial in regulating water consumption levels in agriculture, thereby maximizing the efficient use of resources. Identifying soil characteristics using machine learning is vital for selecting the appropriate soil and minimizing fertilizer costs [22].

Managing irrigation with machine learning is another method to enhance the production and control of water resources by economically limiting its usage for environmental protection. Drones are also increasingly important in agriculture, where they can be used for drip irrigation by employing a machine learning model trained on weather patterns [23]. Implementing this approach makes it possible to protect water resources and enhance production simultaneously. Also, drones can be used for pest control by using their cameras to capture images and analyze them with ML models to detect any diseases on the farm and during harvest and to identify defective crops in the fields. It is also crucial to assess the health of the soil in agricultural lands and determine if there are any mineral deficiencies [24].

The key to constructing the most successful models lies in analyzing and comparing ML models. This approach has proven effective and is crucial for attaining the desired outcomes for intelligent farms. In this methodology, different ML algorithms are tested, and the most successful one is identified using an optimization method, thereby achieving a high success rate in agriculture problem domains. To assess the accuracy of the mod-

els, the data were initially processed, and then, inputted into the models. Subsequently, a comparison method was employed to evaluate the performance of these models. In our model, we proposed a method to select suitable algorithms according to the characteristics of the IoT-based data that were used in our experiments. The proposed model suggests that researchers can select suitable machine learning algorithms for their agricultural dataset.

3.3. Neural Networks and Deep Learning in Agriculture

The agriculture sector is one of the most critical areas where the use of artificial intelligence can be highly beneficial. Machine learning is used to solve a variety of problems within the agricultural sector, such as agricultural productivity, resource management, and decision-making processes [25,26].

Deep learning in agriculture refers to using advanced artificial intelligence (AI) techniques, specifically deep neural networks, to analyze and interpret complex agricultural data for better decision making, farming process optimization, and addressing agricultural challenges. Deep learning is the automatic extraction of detailed patterns and representations from big and diverse datasets using deep neural networks with numerous layers.

Deep learning models are developed on neural network topologies, often deep neural networks with several layers. In agricultural applications, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory networks (LSTMs) are often used. CNNs are increasingly used in agriculture for weed detection, crop pest classification, and plant disease diagnosis [27].

Weed management is one of the major problems with crop management [28]. Since both weeds and crops have similar colors, shapes, and textures throughout the growth phase, weed detection in crops through imaging is a challenging problem. Deep learning algorithms are used to detect and classify weeds. According to [29], supervised learning algorithms are the most used in the literature, and they achieve high classification accuracy by fine-tuning pre-trained models on any plant dataset when a large amount of labeled data are available.

In the agriculture field, pest recognition is a difficult task for farmers, and it is a critical issue in yield production. Identifying and classifying all types of crop insects is the most difficult task for farmers due to their similar appearance in crop growth. In [30], a CNN is used to tackle this issue as it automatically extracts features and learns high-level features in picture classification applications.

Deep learning can be used in agriculture to monitor crop health, detect illnesses, and evaluate general field conditions by analyzing photos from remote sensing technology [31]. The authors in [32] investigate the combination of deep learning algorithms and multi-temporal images. They analyze and compare crop classification applications based on deep learning models and various time-series data to increase crop classification accuracy.

Based on estimates of crop yields, weather patterns, and other factors, deep learning is used in predictive modeling to enable farmers to make well-informed decisions. The authors in [33] use deep neural networks to predict yield corn hybrids from genotype and environment data. In [34], two deep learning models, long short-term memory and gated recurrent units, are used to analyze agricultural datasets and to predict end-of-season yields.

Because of its impact on various aspects of the industry, deep learning is important in the agricultural field. The ability of deep learning models to process and analyze complex and large datasets allows for more accurate and efficient decision making in farming practices. Deep learning plays a crucial role in crop management by enabling precise monitoring, identification of crops, and early detection of diseases or pests. Deep learning promotes sustainability by minimizing waste and reducing environmental impact.

4. Proposed Methodology

In the pursuit of optimizing agricultural data analysis techniques, the complex task of predicting and classifying agricultural data requires the efficient use of advanced machine learning and data analysis methodologies. The plethora of available algorithms necessitates

a judicious selection process, as the accuracy of predictions is profoundly linked to the choice of the model and its corresponding parameters. The task at hand is not only influenced by the nature of the data to be predicted and the dataset's features but also requires extensive experimentation with various artificial intelligence models to construct an effective predictive model.

In this work, we aim to assist researchers in the field of agriculture to select suitable machine learning algorithms for their dataset. To achieve this, we employed historical agricultural data from different dataset sources, applying the most widely used artificial intelligence algorithms to each. Records in the log file were obtained from 723 sub-tables of the crop-rec, pla-spe, and soy-dis databases to increase the efficiency of the proposed model. For every algorithm, we varied the dataset utilization in terms of the division percentage between training and testing and the number of features participating in the classification, and the dataset size. Each trial underwent a comprehensive test, and we recorded its resulting accuracy, precision, recall, and F-measure. This meticulous process resulted in creating a comprehensive log file from numerous experiments, documenting the results and accuracies of each trial.

Algorithm 1 represents the processes of building the proposed model, including a pre-analysis step to detect and overcome potential data limitations such as bias and missing data in agricultural datasets. This algorithm iterates over each dataset in the input agricultural datasets. For each dataset, it preprocesses the data and extracts features. It probes m different data splits, evaluates and experiments with various machine learning algorithms, performs cross-validation, tuning hyperparameters and training models, evaluating performance, calculating metrics, and logging the results. Indeed, the generated dataset contains the logged results, which include the performance metrics (accuracy, precision, recall, F1-score) for each dataset, algorithm, and data split combination.

Algorithm 1 AI powered agricultural modeling

```

1: Given agricultural datasets  $S$  (crop-rec,pla-spe,soy-dis), Feature set  $F$ , Selected ML
   algorithms  $A$ 
2: for each dataset in  $S$  do
3:   dataset = preprocess data(dataset) {Encode categorical variables, normalize numerical
   features}
4:    $F$  = extract meta data features(dataset)
5:   for  $i = 1$  to  $m$  { $m$  different data splits} do
6:     training set, testing set = split data(dataset,  $i$ ,  $m$ )
7:     for each algorithm in  $A$  do
8:       for  $j = 1$  to  $k$  { $k$ -fold cross-validation} do
9:         hyper parameters = tune hyper parameters(algorithm, training set,  $j$ )
10:        model = train model(algorithm, training set, hyper parameters,  $j$ )
11:        performance = evaluate model(model, testing set)
12:        metrics = calculate metrics(performance)
13:      end for
14:      average metrics = calculate average metrics(all folds performance)
15:      Log record(New log dataset, algorithm,  $F$ , metrics,  $i$ -features)
16:    end for
17:  end for
18: end for
19: Output: New Log Dataset{Use the generated dataset to train a new prediction model}

```

The parameters' descriptions are given below:

- preprocess data: Prepares the dataset for model training by encoding categorical variables into numerical values and normalizing numerical features to a standard range (e.g., 0 to 1).
- extract meta data features: This function extracts metadata features from the dataset, such as the number of records, the number and types of features, and labels.

- split data: Divides the dataset into training and testing sets based on the following training and testing ratios: (60,40), (65,35), (70,30), (75,25), (80,20), (85,15), (90,10).
- cross validation: Assesses the performance of a model by splitting the data into multiple subsets (folds), where each fold is used as a testing set exactly once.
- tune hyper parameters: Optimizes the hyperparameters of the selected algorithm using the training set.
- train model: Trains the model using the selected algorithm, training set, and tuned hyperparameters.
- evaluate model: Assesses the trained model's performance using the testing set to measure its ability to generalize to unseen data.
- calculate metrics: Computes performance metrics such as accuracy, precision, recall, and F1-score to quantify the model's success.
- calculate average metrics: Computes the average performance metrics across all folds of cross-validation for each algorithm. Aggregates the performance metrics from each fold to provide a consolidated measure of the model's effectiveness.
- log record: Logs the details of the modeling process, including the algorithm used, extracted features, performance metrics, and the split identifier.

This log file (New Log Dataset), a product of experimenting with diverse datasets, algorithms, and configurations, is valuable for further study and analysis. An illustrative example of the log file, presented in CSV format, shows the feature vector composed of attributes with the highest weight on the accuracy of the results. The log file contains the dataset name, number of features, number of records, training rate, number of classes, feature type, data structure, accuracy, and algorithm name in the feature set. Indeed, drawing from our research and an exhaustive review of the current state of the art, we identified key factors significantly influencing the accuracy of ML classifier, for instance, the number of records, the number of features, the division of data between training and testing sets, and the specific model utilized. Each record of the CSV file encapsulates the experiment result of the application of a specific ML classifier with varying configurations, and the corresponding accuracy is documented.

This CSV file subsequently served as the input for a proposed model designed to predict an estimated accuracy. This AI-trained model, in turn, plays a pivotal role in identifying the preliminary features represented that eventually optimize accuracy for a new agricultural case study. Figure 2 demonstrates the use of the proposed model on unseen agricultural data.

By approximating results based on the pre-trained model's predictions, the process avoids the real-time execution of AI algorithms, instead utilizing historical data and tested dataset features to generate recommendations. This approach optimizes time consumption while ensuring the reliability of the suggested machine learning algorithms for analyzing agricultural datasets.

Rather than deploying and fine-tuning an extensive array of algorithms, we approximate accuracy for appropriate ones, highlighting those with the highest precision. This approach significantly saves time, as suggestions are based on pre-trained AI models from previous agricultural data, providing efficient and reliable recommendations. No real-time solutions are utilized; instead, our suggestions are rooted in AI models trained on historical agricultural data.

The illustration in Figure 3 outlines our approach, which comprises three distinct phases. Each phase's output serves as the input for the subsequent phase. In phase 1, collected diverse agricultural datasets are utilized and reorganized into a new dataset where multiple classifiers are applied in phase 2. Data are preprocessed by handling missing values, encoding categorical variables, and normalizing numerical features. After training and testing operations, the generated dataset with logged results is populated. These are used to train a new prediction model in phase 2. We populate the accuracy and AI algorithm classifiers for each dataset reorganization.

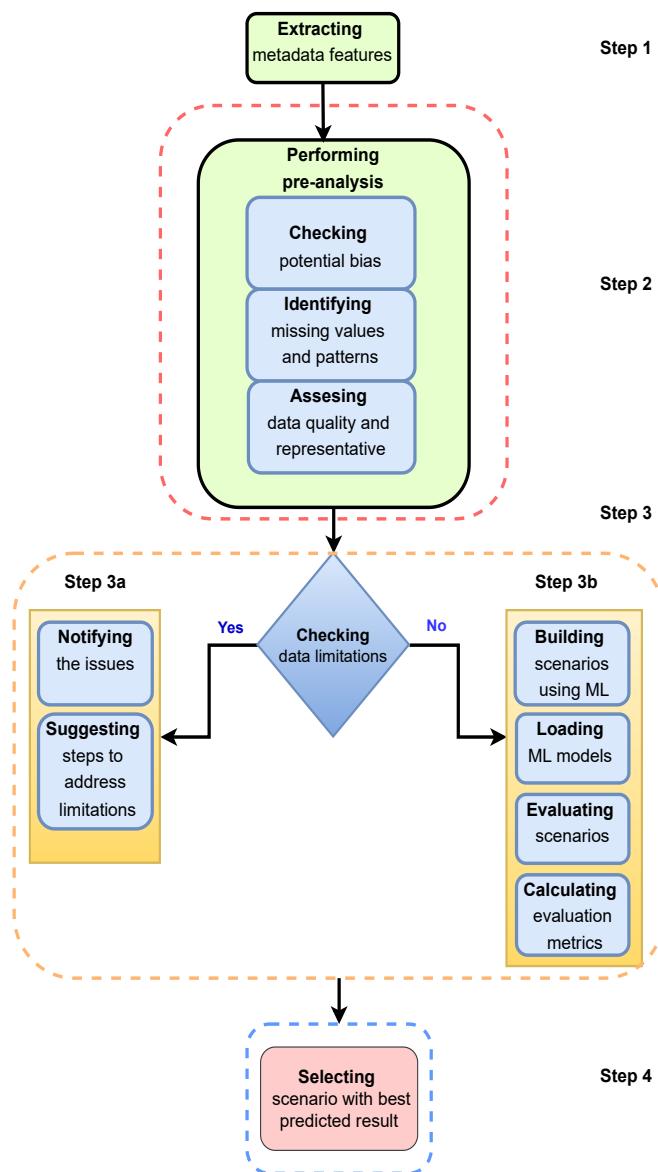


Figure 2. Use of the proposed model on unseen agricultural data.

The results from phase 2 are recorded in another file, representing the calculated outcomes for numerous simulations with datasets exhibiting different characteristics. Phase 3 is where the practical application of our approach comes into play. Here, we apply AI classifiers to create a model that can accurately predict the accuracy of a new, unseen agricultural dataset. Our approach is a crucial step in its validation. For the new dataset, we extract feature information and estimate its accuracy by applying all AI algorithms and exploring all possible data segmentations. The configuration that achieves the highest accuracy is then recommended to the user, demonstrating the practicality and relevance of our research.

During phases 2 and 3, we employed machine learning techniques such as training, data preprocessing, cleaning, testing, and validation to obtain the desired results in our experiments. In phase 3, the data generated from phase 2 served as the output of the machine learning algorithms, and we repeated the same steps as in phase 2. One key benefit of our approach is the elimination of the need for exhaustive reruns of AI algorithms. This time-saving feature allows us to identify the algorithm that yields the best results more efficiently. Our approach's predictive capability relies on a pre-trained model, which estimates outcomes based on the dataset's characteristics rather than the specific data

points within the dataset. This not only saves time but also enhances the accuracy of our predictions, providing a more reliable tool for agricultural data analysis. Finally, the best predicted results are selected based on the scenario, and we present the recommended algorithm and its performance metrics. In our study, we took a systematic approach to selecting the most effective AI-powered models for specific agricultural applications, which represents a significant advancement over the more generalized discussions found in the current literature. This was based on the following areas: (a) algorithm optimization and selection; (b) comprehensive testing and validation; and (c) practical application and scalability. This manuscript not only adds to the body of knowledge already in existence by offering a detailed examination of AI applications in agriculture, but it also makes novel contributions to the fields of algorithm selection and data analysis techniques. Its method of algorithm improvement and realistic scalability marks a major advancement in the application of AI to improve the sustainability and productivity of agriculture.

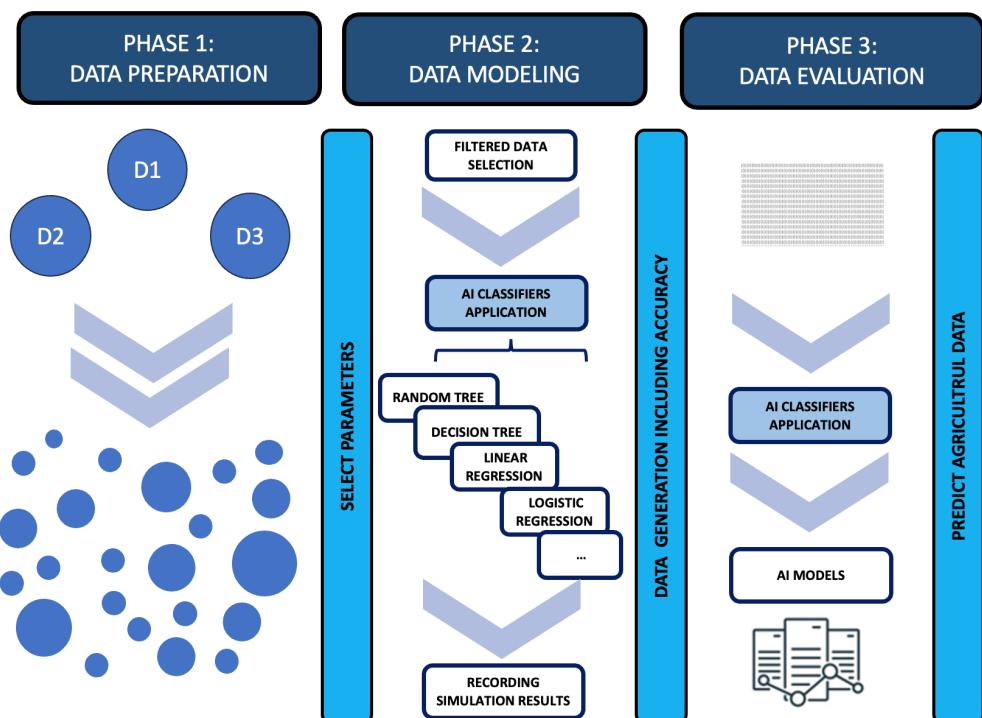


Figure 3. General structure of proposed model.

As a result, we present a tool that empowers researchers in the agricultural domain to make informed decisions when selecting the most suitable AI model based on the type and characteristics of their data.

This model is a valuable resource that minimizes time and effort by streamlining the algorithm selection process and contributing to more effective and accurate decision making in agricultural data analysis.

5. Experimental Results

This section includes data collection, description, classification, and discussion of the results of the proposed method. Predicting and classifying agricultural data involves utilizing machine learning and data analysis techniques to make forecasts or classify information. In this research work, three types of agricultural data are used in crop forecasting, disease recognition, and plant identification. Larger and more diverse datasets enable the proposed model to learn patterns and relationships more comprehensively using machine learning. Data cleaning, feature extraction, and supervised classification methods are applied to three types of agriculture data to achieve optimal prediction and classification to log their results. The agriculture datasets used in this research is given below:

- The crop-rec [35] dataset contains pH, rainfall, humidity, temperature, potassium, phosphorus, and soil nitrogen level as data features for optimum crop growth recommendations. The dataset has 22 crop yield classes: apple, banana, rice, cotton, maize, lentil, and papaya. The dataset has 2200 records and eight attributes. The feature characteristics of the crop-rec dataset are as follows:
 - The values range from 0 to 140 for nitrogen, 5 to 145 for phosphorus, and 5 to 205 for potassium.
 - Temperature ranges from about 8.83 to 43.68.
 - Humidity ranges from about 14.26% to 99.98%.
 - pH ranges from 3.50 to 9.94, indicating a wide range of soil acidity and alkalinity.
 - Rainfall ranges from about 20.21 mm to 298.56 mm.
- The pla-spe [36] dataset contains information on 100 plant species' leaves for classification. Features are extracted from images and created using an expert recommendation system. There are 1600 records with leaf information based on texture, margin, and shape.
- The soy-dis [37] dataset contains 35 features and 19 classes for large soybean diseases. The features and classes are given in the table below. Table 1 shows detailed information for the soy-dis dataset.

Table 1. Description of the soy-dis dataset.

Feature	Set
Month	April, May, June, July, August, September, October, N/A
Plant Condition	normal, lt-normal, N/A
Precip	It-normal, normal, gt-normal, N/A
Temperature	It-normal, normal, gt-normal, N/A
Hail	yes, no, N/A
Crop-history	diff 1st year, same 1st year, same 1st two years, same 1st several years, N/A
area damaged	scattered, low areas, upper areas, whole field, N/A
severity	minor, pot-severe, severe, N/A
seed-tmt	none, fungicide, other, N/A
germination	90–100%, 80–89%, 0–80%, N/A
plant-growth	normal, abnormal, N/A
leaves	normal, abnormal
spots halo	absent, yellow-halos, no-yellow-halos, N/A
spots marg	w-s-marg, no-w-s-marg, dna, N/A
spot size	lt-1/8, gt-1/8, dna, N/A
leaf shread	no, yes, N/A
leaf malf	no, yes, N/A
leaf mild	no, upper surf, lower surf, N/A
stem	normal, abnormal, N/A
lodging	yes, no, N/A
stem-cankers	no, below soil, above soil, above sec nde
canker lesion	dna, brown, dk-brown-blk, tan, N/A
fruiting bodies	no, yes, N/A
external decay	no, firm and dry, watery, N/A

Table 1. Cont.

Feature	Set
mycelium	no, yes, N/A
int discolor	none, brown, black, N/A
sclerotia	no, yes, N/A
fruit-pods	normal, diseased, few-present, dna, N/A
fruit spots	no, colored, brown-w, distort, dna, N/A
seed	normal, abnormal, N/A
mold-growth	no, yes, N/A
seed-discolor	no, yes, N/A
seed-size	normal, lt-norm, N/A
shriveling	no, yes, N/A

The soy-dis dataset has 19 classes for diseases such as charcoal rot, brown stem rot, bacterial blight, purple seed stain, and frog eye leaf spot. In this work, three agricultural datasets are used. Table 2 gives the number of samples, number of features, feature type, data distribution, and number of labels for each dataset.

Table 2. Characteristics of agriculture data.

Dataset	<i>crop-rec</i>	<i>soy-dis</i>	<i>pla-spe</i>
Number of samples	2200	1600	1985
Number of features	7	64	35
Feature type	Numeric	Numeric	Numeric
Data distribution	Right-skewed	Uniform	F
Number of labels	22	99	19

In this work, machine learning methodologies are applied to classify agriculture data to increase productivity and quality. The most frequently used classification algorithms are selected such as Bayes network (BN), naïve Bayes classifier (NBC), logistic (L), multilayer perceptron (MP), locally weighted learning (LWL), decision table (DTa), Hoeffding tree (HT), decision tree (DT), random tree (RT), and random forest (RF). The probabilistic-based algorithms BN (99.49%) and NBC (99.44%) give the highest accuracy for *crop-rec*, NBC (89.17%) for *soy-dis*, and the MP (94.29%) algorithm is the best method for the *pla-spe* dataset. The NBC and BN algorithms can achieve high accuracy because the data are high-dimensional and categorical. They perform efficiently when features are discrete and can be modeled using probability distributions. MP's performance depends on several factors, such as the complexity of the problem, the availability and size of the data, and appropriate hyperparameter tuning. MP gives the highest accuracy for the *pla-spe* data because it learns hierarchical representations of features from the data. This ability is beneficial when dealing with high-dimensional data or when the relevant features are not explicitly known. Table 3 shows an example of the obtained accuracy with some predefined configuration applied for all datasets [38].

Selecting the appropriate machine learning algorithms for agriculture datasets is an iterative process that requires understanding the problem, experimenting with various models, and fine-tuning them based on performance metrics. The key to choosing the most suitable algorithm lies in balancing model complexity, interpretability, and computational efficiency. Parameter tuning is crucial for optimizing machine learning algorithms' performance. The process involves systematically adjusting hyperparameters and evaluating their impact on performance metrics. In our work, it is essential to identify the most suitable

configuration for a given dataset and task. Table 3 demonstrates the parameter list for chosen ML algorithms.

Table 3. Accuracy, kappa, MAE, and RMSE values for all datasets.

	Accuracy			Kappa			MAE			RMSE		
	crop-rec	soy-dis	pla-spe									
BN	99.49	88.89	86.98	0.994	0.877	0.854	0.001	0.019	0.016	0.02	0.11	0.09
NBC	99.44	89.17	83.75	0.994	0.889	0.829	0.001	0.014	0.011	0.02	0.11	0.11
L	97.65	84.83	89.34	0.975	0.832	0.891	0.002	0.021	0.015	0.04	0.13	0.12
MP	99.17	86.82	94.29	0.991	0.854	0.935	0.004	0.023	0.015	0.03	0.11	0.05
LWL	79.33	57.49	83.41	0.783	0.527	0.827	0.075	0.098	0.082	0.18	0.22	0.19
DTa	87.19	75.76	90.84	0.865	0.731	0.905	0.056	0.094	0.061	0.14	0.19	0.16
HT	99.12	87.85	84.72	0.989	0.865	0.841	0.009	0.016	0.022	0.02	0.12	0.14
DT	98.62	83.69	77.69	0.985	0.821	0.774	0.001	0.025	0.041	0.03	0.13	0.15
RF	91.27	88.91	78.52	0.908	0.876	0.765	0.003	0.036	0.042	0.02	0.11	0.14
RT	98.76	85.49	77.06	0.987	0.843	0.768	0.001	0.019	0.012	0.03	0.13	0.16

The experiments show that rule-based algorithms exhibited strong performance across different datasets and parameter configurations, with the DT algorithm achieving the highest accuracy of 89.38%. The probabilistic and regression-based methods produced mixed results, with BN showing promise but needing careful parameter tuning. MP demonstrated robust performance, particularly when appropriately tuned, achieving high accuracy. Instance-based learning (LWL) had the longest training and testing times, indicating a high computational cost. Ensuring that preprocessing is sufficient for classification involves several steps and considerations to improve the quality of the input data, and thus, the performance of the classifier. In this work, the following steps are applied from raw data to training input data for agriculture data:

- Noisy and mislabeled data are removed (smoothing, normalization).
- Data augmentation techniques (rotation, shifting) are used to increase the diversity of the training set.
- The dataset is balanced to ensure an equal number of records for each label.
- Data are shuffled to ensure a random distribution, which helps reduce bias during training.
- The most relevant features are selected using correlation analysis.
- New features are created from existing data using statistical methods (min, max, range).
- Data are restructured into a consistent format.

The number of features is reduced using feature selection methods to keep only the most relevant features. The data are split into training and testing sets. Evaluation metrics such as accuracy, precision, and the F1-score show the success of the data preprocessing.

Kappa is a ratio that determines how accurate two raters are on classified items. As a case, it evaluates the reliability of multiple labels in data classification that divide the items into distinct classes. Kappa values are range between -1 and $+1$. A value close to $+1$ is the perfect result in data classification. In this work, BN (0.994) and NBC (0.994) give the best kappa values in crop-rec, NBC (0.889) in soy-dis, and MP (0.935) and DTa (0.905) in pla-spe. Another evaluation metric for the classification model is the mean absolute error (MAE). This is a metric that averages the sum of all absolute differences between predicted values and observed values. In addition to that, MAE is a measure that indicates how much a particular model is distanced from the source data on average. If the MAE value is close to 0, it indicates perfect results. In the crop-rec data classification, both the probabilistic and rule-based algorithms have an MAE rate of less than 0.03. The experiments show

that the LWL and DTa algorithms have higher error rates than the other algorithms for all three datasets. Model complexity and incomplete rules are reasons for the lower error rate in these algorithms. Another classification measure that has a lot of application in regression modeling is the Root Mean Squared Error (RMSE). It is similar to MAE except that it considers bigger disparities because it multiplies them by themselves [39–41].

$$k(\text{kappa}) = (A_o - A_e) / (1 - A_e) \quad (1)$$

where A_o is the relative observed agreement and A_e the hypothetical probability of chance agreement after classification.

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

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{n}} \quad (3)$$

where x_i is the actual value, \hat{x}_i is the predicted value, and n is the total number of observations.

$$RAE = \frac{\sum_{i=1}^N |x_i - t_i|}{\sum_{i=1}^N |t_i - \hat{t}_i|} \quad (4)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^N (x_i - t_i)^2}{\sum_{i=1}^N (t_i - \hat{t})^2}} \quad (5)$$

where x_i is predicted, t_i is target, and \hat{t}_i is the average value of the target data.

Table 3 demonstrates the TP rate, precision, recall, and F-measure values. The TP rates are more than 90% for all methods except the LWL and decision table methods. In general, the LWL method has the lowest rates for all datasets.

The true positive (TP) rate refers to the ratio of real positive cases that can be correctly identified. In many cases, it is critical, especially when correctly classifying positives, especially in medical diagnoses. The true positive rate is high in the BN and NBC algorithms for the crop-rec data, meaning that the model is effectively identifying and correctly classifying a large proportion of the positive instances in the dataset. On the other hand, LWL (79.33) on the crop-rec data and the rule-based algorithms on the soy-dis and pla-spe datasets have low TP rates, which shows a low rate of accuracy. A false positive (FP) appears when the model wrongly predicts a positive outcome and the true outcome is the opposite. There should be a balance between TPs and FPs. Our aim in this work is to increase the TP and decrease the FP rates. In medical diagnosis, TP values are very critical, they might affect the patient's health critically. In agriculture work, the expectation is a balance between TPs and FPs. In machine learning, especially in classification, precision is used as a measure of the accuracy of positive predictions from the model. High precision indicates that whenever the model predicts a positive item classification. The recall measurement is useful, especially where false negative results are expensive and undesired. A single value combining precision and recall is called the F-measure or F1-score. We can use the F-measure only in classification. It is a metric that helps us to find the optimal point in this trade-off [42–44]. The results given in Tables 3–6 are used as an attribute in the proposed data model for the farmer recommendation system. Table 4 shows the FP rate, MCC, ROC area, and PRC area values for all the algorithms.

MCC is another measure for assessing the performances of binary classification models. True positives, true negatives, false positives, and false negatives are considered, and it can be very helpful for dealing with skewed datasets. In addition, an important measure derived from the ROC curve called area under the ROC measures how well a model discriminates between positives and negatives concerning any given decision threshold.

The precision–recall curve (PRC) is another evaluation tool for binary classification models. It is useful when dealing with imbalanced datasets such as the data used in this work. The ROC area rate considers all possible classification thresholds. If the ROC value is close to +1, the classification model is reliable; otherwise it is not. In these experiments, all ML models have a value larger than 0.8, which shows that they are all appropriate models to apply to these datasets. High precision and accuracy and low error values indicate the success of the proposed models.

Table 4. Parameter list for selected ML algorithms.

Algorithm	Parameter List
BN	Maximum entropy, d separation
NBC	Alpha is 0.3, binary/Boolean features
L	Lipliner function, max iteration 240, tolerance is 1×10^{-4}
MP	3 layers, sigmoid function, rate 0.01, 182 epoch, dropout rate is 0.2
LWL	Bandwidth is 0.6, Gaussian function, Chebyshev distance, cross-validation
DTa	Threshold is 85%, rule merging
HT	Gini function, maximum depth is 6, delta is 1×10^{-7} , tau is 0.05
DT	Entropy function, random splitter, max leaf node is 8, balanced class weight
RF	Maximum depth is 12, balanced class weight, impurity is 0.03, n estimator is 74
RT	Maximum depth is 9, balanced class weight, absolute error function

Table 5. TP rate, precision, recall, and F-measure values for all datasets.

	TP Rate		Precision				Recall				F-Measure		
	crop-rec	soy-dis	pla-spe	crop-rec	soy-dis	pla-spe	crop-rec	soy-dis	pla-spe	crop-rec	soy-dis	pla-spe	
BN	99.49	88.89	86.98	0.994	0.877	0.854	0.001	0.019	0.016	0.02	0.11	0.09	
NBC	99.44	89.17	83.75	0.994	0.889	0.829	0.001	0.014	0.011	0.02	0.11	0.11	
L	97.65	84.83	89.34	0.975	0.832	0.891	0.002	0.021	0.015	0.04	0.13	0.12	
MP	99.17	86.82	94.29	0.991	0.854	0.935	0.004	0.023	0.015	0.03	0.11	0.05	
LWL	79.33	57.49	83.41	0.783	0.527	0.827	0.075	0.098	0.082	0.18	0.22	0.19	
DTa	87.19	75.76	90.84	0.865	0.731	0.905	0.056	0.094	0.061	0.14	0.19	0.16	
HT	99.12	87.85	84.72	0.989	0.865	0.841	0.009	0.016	0.022	0.02	0.12	0.14	
DT	98.62	83.69	77.69	0.985	0.821	0.774	0.001	0.025	0.041	0.03	0.13	0.15	
RF	91.27	88.91	78.52	0.908	0.876	0.765	0.003	0.036	0.042	0.02	0.11	0.14	
RT	98.76	85.49	77.06	0.987	0.843	0.768	0.001	0.019	0.012	0.03	0.13	0.16	

Figures 4–8 shows training time, testing time, root absolute error (RAE), and root relative squared error (RRSE) values for each dataset. The time required for training and testing classification algorithms can vary significantly based on multiple factors, including the complexity of the algorithm, the size and nature of the dataset, and the computational resources available. In Figures 4 and 5, our experiments show that algorithms such as L, MP, and LWL take more training time than others. LWL has the highest time for testing time; other algorithms, including MP, have a low cost by taking a shorter time on all datasets. The RAE is a way to assess the accuracy of predictions, with lower values indicating better performance. The NBC and HT algorithms on crop-res, NBC and HT on soy-dis, and NBC and HT on pla-spe give the best RAE values. HT algorithms are modeled to adapt quickly to changes in the data distribution. The simplicity of the HT model can contribute to

better generalization, and consequently, lower RAE. The highest training time is 9.17 s, for the MP algorithm on the crop-rec data. LWL has the highest testing time for crop-rec. Neural network algorithms have a high potential to solve complex problems. On the other hand, they have high complexity in training because of the number of layers, forward and backward calculations, and data sizes. This problem can be reduced with hardware acceleration, batch processing, optimization, and data cleaning. Table 7 shows that it has the highest training cost of all datasets. However, the algorithm has one of the lowest testing costs for all datasets among the applied classification algorithms. The testing time is 0.01 s for the crop-rec data, 0.001 s for the soy-dis data, and 0.02 s for the pla-spe data. Probabilistic-based algorithms such as BN and NBC have efficient training and testing times. The random forest algorithm takes the longest time of the rule-based algorithms, as shown in Figures 4 and 5 [45].

Table 6. FP rate, MCC, ROC, and PRC area rates for all datasets.

FP Rate			MCC			ROC Area			PRC Area			
crop-rec	soy-dis	pla-spe										
BN	0.001	0.014	0.011	0.994	0.877	0.885	1.000	0.983	0.992	0.999	0.914	0.923
NBC	0.001	0.015	0.009	0.993	0.890	0.898	1.000	0.992	0.952	0.998	0.952	0.961
L	0.002	0.020	0.014	0.976	0.833	0.841	1.000	0.985	0.994	0.994	0.928	0.937
MP	0.001	0.015	0.012	0.992	0.847	0.855	1.000	0.981	0.990	0.999	0.907	0.916
LWL	0.009	0.055	0.041	0.757	0.591	0.827	0.990	0.950	0.928	0.852	0.833	0.982
DTa	0.005	0.024	0.019	0.883	0.812	0.730	0.989	0.941	0.846	0.885	0.761	0.684
HT	0.001	0.018	0.009	0.994	0.876	0.788	1.000	0.994	0.894	0.999	0.966	0.869
DT	0.001	0.019	0.022	0.986	0.829	0.746	0.998	0.973	0.875	0.987	0.809	0.728
RF	0.001	0.015	0.028	0.994	0.879	0.791	1.000	0.993	0.893	1.000	0.959	0.863
RT	0.002	0.015	0.023	0.987	0.849	0.764	0.994	0.921	0.828	0.977	0.770	0.693

In this paper, a new database is created using several crop data classification results. This dataset includes a data name, number of features, number of classes, number of records, classification method, and accuracy, as shown in Table 7. Several binary features are also added into the feature vector, such as multilabel, balanced, structures, time series, and binary. The data vector consists of 13 attributes and 720 records. Table 7 shows sample data from the newly developed dataset from ML-applied crop data. All data in this table were obtained after the classification of the dataset in the Weka software version 3.6.12. Weka is a widely used machine learning tool developed by the University of Waikato in New Zealand. It includes data preprocessing, classification, clustering, and visualization.

Data vector = {data name, f1 = #of features, f2 = #of records, f3 = training rate, f4 = classification/regression, f5 = #of class, f6 = feature type, f7 = balanced/unbalanced, f8 = binary, f9 = multilabel, f10 = structured, f11 = time series, f12 = accuracy, f13 = ML method}.

The experimental results show that the best methodology and data format can be predicted using machine learning algorithms, as shown in Table 8. The results show that rule-based algorithms give promising results in predicting the best machine learning algorithm and data structure for crop data. Table 8 shows accuracy, kappa, MAE, RAE, TP rate, F-measure, and MCC values for the integrated new datasets after individual machine learning processing. The results show that the decision tree (89.38%), random forest (87.61%), and random tree (84.27%) algorithms have the highest accuracy for the proposed data selection. There are several reasons behind the rule-based algorithms' success. These algorithms can handle both categorical and numerical data. They can naturally handle a mix of data types without requiring extensive preprocessing, making them suitable for

datasets with diverse feature types. All rule-based algorithms are suitable for both binary and multiclass classification problems. They can naturally handle multiple classes without significant modifications. In addition to that, decision tree and random tree methods can handle missing values in the features without requiring imputation. Rule-based algorithms also have some limitations, such as overfitting and sensitivity to noise. On the other hand, they are less sensitive to outliers and missing values than the other classification algorithms. They can use non-linear relationships between features and the target variable. Rule-based algorithms assign low importance to irrelevant or redundant features during the training process. The TP rates are more than 90% in the MP, LWL, Dta, DT, RF, and RT algorithms. Probabilistic and regression-based methods do not perform well in this newly developed dataset for prediction and classification. Probabilistic methods are not successful when features are highly correlated, data are imbalanced, and there is a high number of outlier data in the input.

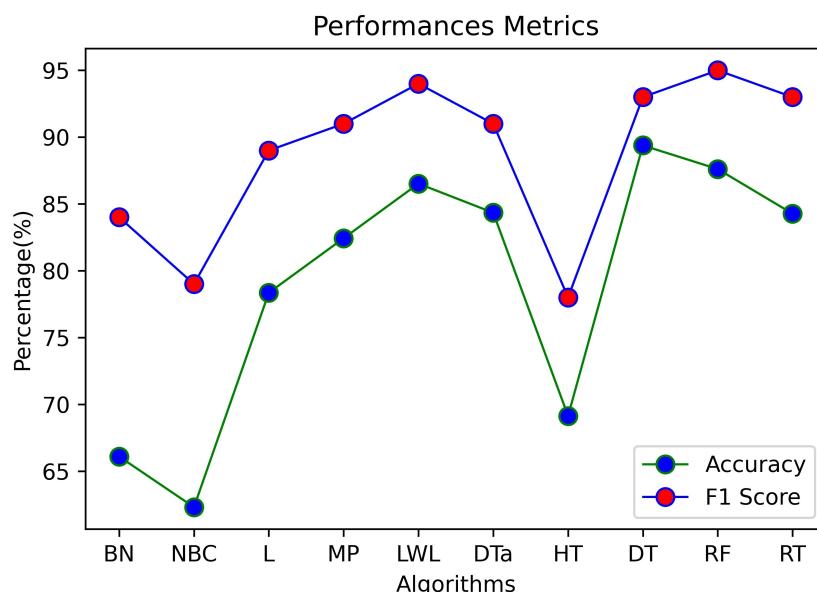


Figure 4. Performance measurement comparison.

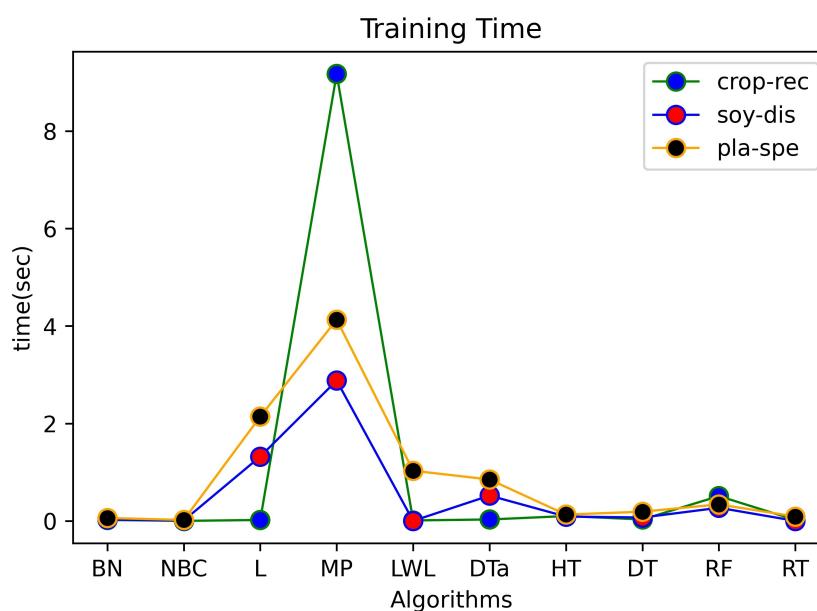


Figure 5. Training times using different algorithms.

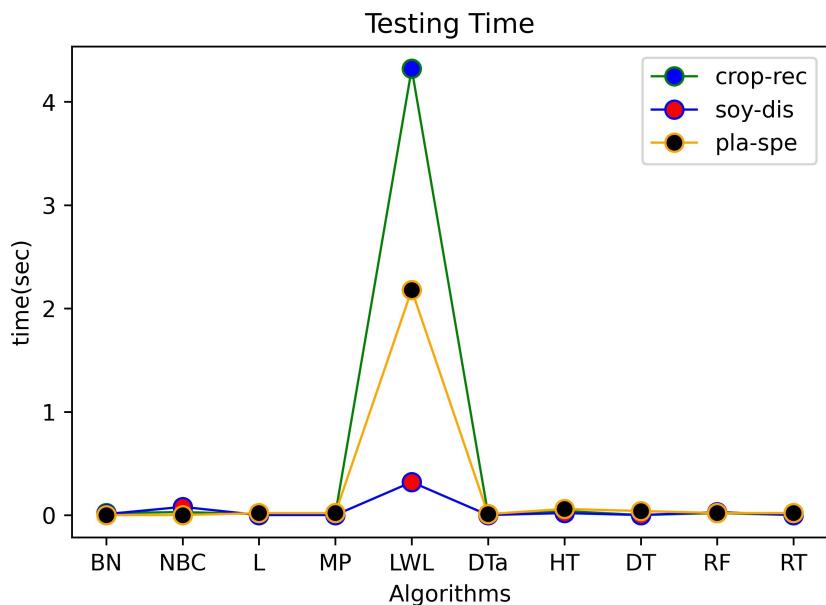


Figure 6. Testing times using different algorithms.

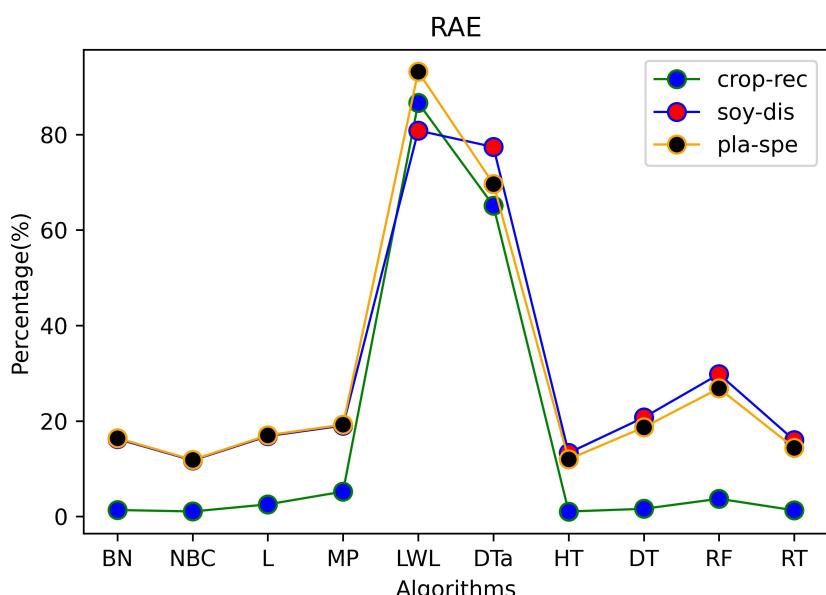


Figure 7. RAE percentage comparison.

Table 9 demonstrates the number of hidden layers, neurons per layer, learning rate, activation function, and number of epochs for each dataset. The accuracies are 99.17% on the crop-rec data, 86.82% on the pla-spe data, and 94.29% on the soy-dis data.

The experimental results in Table 8 demonstrate that rule-based algorithms, particularly decision tree, random forest, and random tree, are highly effective for predicting and classifying agriculture data. These algorithms outperform probabilistic and regression-based methods significantly in terms of accuracy, TP rate, F-measure, and MCC. The high performance of rule-based algorithms suggests that they can effectively capture the complexities and patterns in the data, making them suitable for practical applications in agriculture data recommender system. Decision trees and random trees provide clear, interpretable rules for making predictions. This is important in agriculture, where stakeholders need to understand the reasoning behind recommendations. Random forests reduce the variance of the model and improve generalization to unseen data. This is particularly useful in agriculture, where the system must perform well under varied conditions. The agriculture

data in this work have many features. Rule based algorithms can handle high-dimensional data well and are capable of selecting the most relevant features during the training process. Overall, the ability of the decision tree, random forest, and random tree algorithms to capture complex relationships, provide interpretability, handle high-dimensional and non-linear data, and maintain robustness and efficiency makes them well-suited for agricultural recommendation systems. These characteristics help in delivering accurate, reliable, and actionable recommendations to improve agricultural productivity and management. The proposed recommendation system for agriculture faces the following limitations:

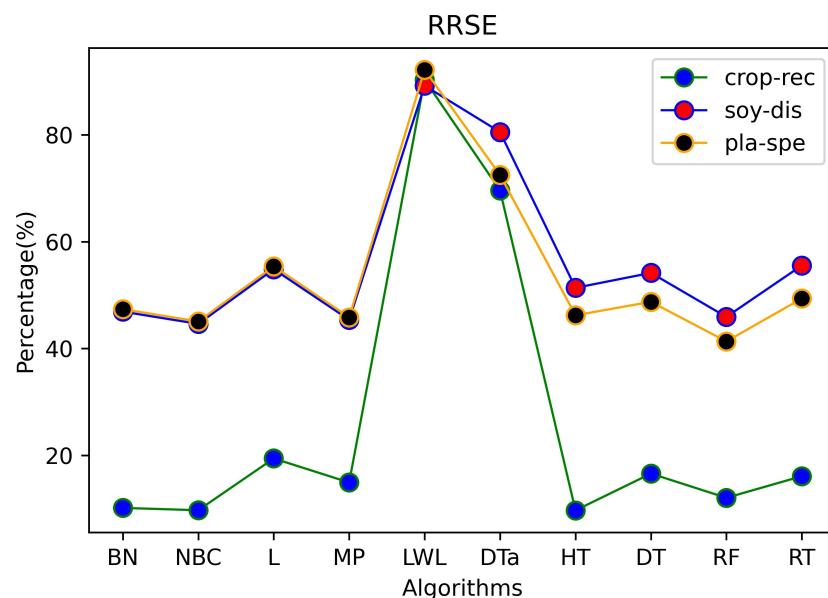


Figure 8. RRSE percentage comparison.

Table 7. Sample data from integrated ML united classification crop data.

Data	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13
crop-rec	9	2125	67	1	22	1	1	1	0	1	0	99.49	BN
crop-rec	7	1750	80	0	14	1	1	1	0	1	0	84.86	RF
crop-rec	5	500	80	0	14	1	1	1	0	1	0	81.72	RT
soy-dis	35	330	67	1	19	1	1	1	0	1	0	88.89	BN
soy-dis	35	330	67	0	19	1	1	1	0	1	0	87.85	HT
soy-dis	35	330	20	1	19	1	1	1	0	1	0	86.42	BN
soy-dis	35	330	30	1	19	1	1	1	0	1	0	86.94	MP
soy-dis	35	330	80	0	19	1	1	1	0	1	0	85.87	RF
soy-dis	35	330	80	0	19	1	1	1	0	1	0	82.57	RT
soy-dis	30	250	67	1	19	1	1	1	0	1	0	81.56	BN
soy-dis	30	250	10	1	11	1	1	1	0	1	0	84.13	BN
soy-dis	30	250	20	1	11	1	1	1	0	1	0	79.93	BN
soy-dis	30	250	30	1	11	1	1	1	0	1	0	91.73	BN
soy-dis	30	250	40	1	11	1	1	1	0	1	0	89.89	BN
soy-dis	30	250	50	1	11	1	1	1	0	1	0	90.3	NBC
soy-dis	30	250	80	1	11	1	1	1	0	1	0	71.06	DTa
soy-dis	20	200	67	1	11	1	1	1	0	1	0	75.59	L

Table 7. Cont.

Data	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13
pla-spe	64	1600	67	1	100	1	1	1	0	1	0	89.34	L
pla-spe	64	1600	10	1	100	1	1	1	0	1	0	90.39	DTa
pla-spe	64	1600	10	0	100	1	1	1	0	1	0	84.3	HT
pla-spe	52	1600	20	0	100	1	1	1	0	1	0	82.18	HT
pla-spe	52	1000	30	1	100	1	1	1	0	1	0	89.02	DTa
pla-spe	30	1000	40	0	60	1	1	1	0	1	0	75.76	DT
pla-spe	30	800	80	0	40	1	1	1	0	1	0	65.03	RT

Table 8. Accuracy and error rates for ML united classification crop data.

	Accuracy	Kappa	MAE	RAE (%)	TP Rate	F-Measure	MCC
BN	66.09	0.53	0.12	49.6	89.2	0.84	0.77
NBC	62.30	0.45	0.15	27.3	83.2	0.79	0.68
L	78.35	0.24	0.08	66.9	89.2	0.89	0.91
MP	82.43	0.13	0.07	31.4	90.1	0.91	0.87
LWL	86.52	0.71	0.06	24.3	92.4	0.94	0.89
DTa	84.34	0.77	0.11	67.1	95.3	0.91	0.86
HT	69.13	0.44	0.17	48.2	79.3	0.78	0.71
DT	89.38	0.81	0.08	45.9	91.3	0.93	0.88
RF	87.61	0.79	0.08	47.6	95.4	0.95	0.95
RT	84.27	0.75	0.08	43.9	95.4	0.93	0.89

- Insufficient and imbalanced data.
- High variability and heterogeneity of environmental conditions.
- Complexity of adapting to temporal changes.

Additionally, challenges in real applications include the need for model interpretability, technological barriers, resource constraints, and ethical concerns regarding data privacy and equity.

In this work, we tried to address one of the research problems in the agriculture field. We proposed an efficient machine learning algorithm selection recommender system for agricultural data. Training data are integrated from several crop data such as crop forecasting, disease recognition, and plant identification after applying ML algorithms. The developed system recommends efficient ML algorithms to researchers who are working on agricultural data prediction, classification, and modeling. As a result, the experimental results show that the proposed model is robust, efficient, and time-consuming for ML-based agriculture modeling.

Table 9. Model parameters and performance for agriculture datasets.

Dataset	crop-rec	pla-spe	soy-dis
Number of Parameters	8	67	35
Number of Hidden Layers	2	2	1
Number of Neurons	21; 16	43; 24	32
Learning Rate	0.01	0.01	0.001
Activation Function	Sigmoid	Tanh	Sigmoid
Number of Epochs	204	127	89

6. Conclusions

In this paper, we proposed an efficient machine learning algorithm selection recommender system for agricultural data. The system is a promising new model for helping researchers select suitable machine learning algorithms for their agricultural data. The system is based on rule-based algorithms, which can handle categorical and numerical data and are less sensitive to outliers and missing values than other classification algorithms. These algorithms work well with diverse data types and can handle missing values, making them suitable for various agricultural data challenges. After individual machine learning processing, the proposed system was evaluated on a dataset of integrated new datasets. The results show that the system can accurately predict the best machine learning algorithm and data structure for crop data. Specifically, the proposed model achieved accuracies of 89.38%, 87.61%, and 84.27% for the decision tree, random forest, and random tree algorithms, respectively. These results are significantly better than those achieved by probabilistic and regression-based methods. Probabilistic and regression-based methods perform poorly with these agricultural datasets due to factors like highly correlated features and imbalanced data.

In future work, we will explore integrating explainable AI (XAI) methods to provide insights into model predictions, to enhance the accuracy and efficiency of predictions further. This includes the potential adaptation of deep learning models that can process and analyze even larger datasets with higher dimensionalities. In addition to the scalability of the proposed models to larger, multi-regional datasets, it could help in understanding the broader applicability and limitations of the current approach. By addressing these areas, future research can not only enhance the technical capabilities of AI models in agriculture but also ensure that such innovations are accessible, practical, and beneficial at the ground level, thereby transforming agricultural practices sustainably and efficiently.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the Kaggle Dataset web link provided at reference number [35–37].

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

References

1. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [[CrossRef](#)]
2. Suebsombut, P.; Sekhari, A.; Sureepong, P.; Ueasangkomsate, P.; Bouras, A. The using of bibliometric analysis to classify trends and future directions on “smart farm”. In Proceedings of the International Conference on Digital Arts, Media and Technology (ICDAMT), Chiang Mai, Thailand, 1–4 March 2017; pp. 136–141.
3. Moon, A.; Kim, J.; Zhang, J.; Liu, H.; Woo Son, S. Understanding the impact of lossy compressions on IoT smart farm analytics. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 4602–4611.
4. Yeo, H. Smart Farming Technology Review: Keynote Address. In Proceedings of the 2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA), Las Vegas, NV, USA, 25–27 May 2022; p. 2.
5. Idoje, G.; Dagiuklas, T.; Iqbal, M. Survey for smart farming technologies: Challenges and issues. *Comput. Electr. Eng.* **2021**, *92*, 107104. [[CrossRef](#)]
6. 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](#)]
7. Ragavi, B.; Pavithra, L.; Sandhiyadevi, P.; Mohanapriya, G.; Harikirubha, S. Smart Agriculture with AI Sensor by Using Agrobot. In Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 11–13 March 2020; pp. 1–4.

8. Katarya, R.; Raturi, A.; Mehndiratta, A.; Thapper, A. Impact of Machine Learning Techniques in Precision Agriculture. In Proceedings of the 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 7–8 February 2020; pp. 1–6.
9. Torai, S.; Chiyoda, S.; Ohara, K. Application of AI Technology to Smart Agriculture: Detection of Plant Diseases. In Proceedings of the 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Chiang Mai, Thailand, 23–26 September 2020; pp. 1514–1519.
10. Harmani, V.P.; Himawan, B.M.; Alhadi, M.A.; Gunawan, A.A.S.; Anderies. Systematic Literature Review: Implementation Of Artificial Intelligence in Precision Agriculture. In Proceedings of the 2022 5th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 August 2022; pp. 479–484.
11. Baghel, S.S.; Rawat, P.; Singh, R.; Akram, S.V.; Pandey, S.; Baghel, A.S. AI, IoT and Cloud Computing Based Smart Agriculture. In Proceedings of the 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 14–16 December 2022; pp. 1658–1661.
12. Dharani, M.K.; Thamilselvan, R.; Natesan, P.; Kalaivaani, P.; Santhoshkumar, S. Review on Crop Prediction Using Deep Learning Techniques. *J. Phys. Conf. Ser.* **2021**, *1767*, 012026. [[CrossRef](#)]
13. Jhajharia, K.; Mathur, P.; Jain, S.; Nijhawan, S. Crop Yield Prediction using Machine Learning and Deep Learning Techniques. *Procedia Comput. Sci.* **2023**, *218*, 406–417. [[CrossRef](#)]
14. Gupta, S.; Geetha, A.; Sankaran, K.; Zamani, A.; Ritonga, M.; Raj, R.; Ray, S.; Sobahi, H. Machine Learning-and Feature Selection-Enabled Framework for Accurate Crop Yield Prediction. *J. Food Qual.* **2022**, *2022*, 330–338. [[CrossRef](#)]
15. Oikonomidis, A.; Catal, C.; Kassahun, A. Deep learning for crop yield prediction: A systematic literature review. *N. Z. J. Crop. Hortic. Sci.* **2023**, *51*, 1–26. [[CrossRef](#)]
16. Vignesh, K.; Askarunisa, A.; Abirami, A. Optimized Deep Learning Methods for Crop Yield Prediction. *Comput. Syst. Eng.* **2023**, *44*, 1051–1067. [[CrossRef](#)]
17. 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](#)]
18. Ahmed, I.; Habib, G.; Yadav, P.K. An Approach to Identify and Classify Agricultural Crop Diseases Using Machine Learning and Deep Learning Techniques. In Proceedings of the 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 1–3 March 2023; pp. 1–6.
19. Rajeshram, V.; Rithish, B.; Karthikeyan, S.; Prathab, S. Leaf Diseases Prediction Pest Detection and Pesticides Recommendation using Deep Learning Techniques. In Proceedings of the 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 23–25 March 2023; pp. 1633–1639.
20. Ullo, S.L.; Sinha, G.R. Advances in IoT and smart sensors for remote sensing and agriculture applications. *Remote Sens.* **2021**, *13*, 2585. [[CrossRef](#)]
21. Liu, W. Smart sensors, sensing mechanisms and platforms of sustainable smart agriculture realized through big data analysis. *Clust. Comput.* **2021**, *26*, 2503–2517. [[CrossRef](#)]
22. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access* **2020**, *9*, 4843–4873. [[CrossRef](#)]
23. Abioye, E.; Hensel, O.; Esau, T.; Elijah, O.; Abidin, M.; Ayobami, A.; Yerima, O.; Nasirahmadi, A. Precision Irrigation Management Using Machine Learning and Digital Farming Solutions. *AgriEngineering* **2022**, *4*, 70–103. [[CrossRef](#)]
24. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [[CrossRef](#)]
25. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)]
26. Albahar, M. A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities. *Agriculture* **2023**, *13*, 540. [[CrossRef](#)]
27. Hassan, S.M.; Maji, A.K.; Jasiński, M.; Leonowicz, Z.; Jasińska, E. Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. *Electronics* **2021**, *10*, 1388. [[CrossRef](#)]
28. Yu, J.; Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. Deep learning for image-based weed detection in turfgrass. *Eur. J. Agron.* **2019**, *104*, 78–84. [[CrossRef](#)]
29. Hasan, A.S.M.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [[CrossRef](#)]
30. Thenmozhi, K.; Srinivasulu Reddy, U. Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* **2019**, *164*, 104906. [[CrossRef](#)]
31. Eunice, J.; Popescu, D.E.; Chowdary, M.K.; Hemanth, J. Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. *Agronomy* **2022**, *12*, 2395. [[CrossRef](#)]
32. Li, Q.; Tian, J.; Tian, Q. Deep Learning Application for Crop Classification via Multi-Temporal Remote Sensing Images. *Agriculture* **2023**, *13*, 906. [[CrossRef](#)]
33. Khaki, S.; Wang, L. Crop Yield Prediction Using Deep Neural Networks. *Front. Plant Sci.* **2019**, *10*, 621. [[CrossRef](#)]

34. Alibabaei, K.; Gaspar, P.D.; Lima, T.M. Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling. *Energies* **2021**, *14*, 3004. [[CrossRef](#)]
35. Crop Recommendation System Using Machine Learning. 2022. Available online: <https://www.kaggle.com/code/nirmalgaud/crop-recommendation-system-using-machine-learning> (accessed on 13 December 2023).
36. Leaf Classification. Available online: <https://www.kaggle.com/c/leaf-classification> (accessed on 13 December 2023).
37. Mscse, H. Soybean Disease Dataset. 2022 Available online: <https://www.kaggle.com/datasets/shuvoalok98/soybean-disease-dataset> (accessed on 13 December 2023).
38. Şaar, F.; Topcu, A.E. Minimum spanning tree-based cluster analysis: A new algorithm for determining inconsistent edges. *Concurr. Comput.* **2022**, *34*, e6717. [[CrossRef](#)]
39. 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](#)]
40. Sharma, P.; Dadheech, P.; Aneja, N.; Aneja, S. Predicting agriculture yields based on machine learning using regression and deep learning. *IEEE Access* **2023**, *11*, 111255–111264. [[CrossRef](#)]
41. Alebele, Y.; Wang, W.; Yu, W.; Zhang, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cheng, T. Estimation of crop yield from combined optical and SAR imagery using Gaussian kernel regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10520–10534. [[CrossRef](#)]
42. Vlachopoulos, O.; Leblon, B.; Wang, J.; Haddadi, A.; LaRocque, A.; Patterson, G. Evaluation of crop health status with UAS multispectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 297–308. [[CrossRef](#)]
43. Chen, L.; Xing, M.; He, B.; Wang, J.; Shang, J.; Huang, X.; Xu, M. Estimating soil moisture over winter wheat fields during growing season using machine-learning methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3706–3718.
44. Tugrul, B.; Eryigit, R.; Ar, Y. Deep learning-based classification of image data sets containing 111 different seeds. *Adv. Theory Simul.* **2023**, *6*. [[CrossRef](#)]
45. Elbasi, E.; Zaki, C.; Topcu, A.E.; Abdelbaki, W.; Zreikat, A.I.; Cina, E.; Shdefat, A.; Saker, L. Crop prediction model using machine learning algorithms. *Appl. Sci.* **2023**, *13*, 9288. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.