engineering proceedings

MDPI

Proceeding Paper

Agricultural Farm Production Model for Smart Crop Yield Recommendations Using Machine Learning Techniques †

Kandasamy Vidhya [1], Sneha George [1], Palanisamy Suresh [2], Duraipandi Brindha [1,*] and Theena Jemima Jebaseeli [1]

[1] Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, India; vidhyak@karunya.edu (K.V.); snehageorge@karunya.edu (S.G.); jemima_jeba@karunya.edu (T.J.J.)
[2] Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore 641042, India; p.suresh@skct.edu.in
* Correspondence: brindhad12345@gmail.com
† Presented at the International Conference on Recent Advances on Science and Engineering, Dubai, United Arab Emirates, 4–5 October 2023.

Abstract: Smart agricultural monitoring is the use of cutting-edge technology to manage all elements impacting plants and lowering crop yield quality. The main objective of smart crop monitoring and management is to guarantee farmers optimal productivity. Additionally, the market for worldwide smart crop management is expanding continuously as a result of the rising need for smart agricultural techniques. Machine learning techniques have the potential to be utilized to provide intelligent agricultural yield suggestions that will assist farmers in increasing their crop yields and profitability. Machine learning algorithms are used to analyze massive collections containing previous yield statistics, meteorological data, soil data, and other parameters in order to discover patterns and associations that might be used to predict agricultural yields. The methodology used in this system is that the farmer must enter the details of conditions in the field. Once entered into the system, the data are analyzed. This predicts the state of environmental conditions and predicts the crop that is suitable under these situations to give a greater yield. A web application is also built here for the farmer to analyze the information regarding their crops and to generate relevant reports. To find better crops under various conditions, the k-nearest neighbor (KNN) technique is used. Finally, the farmer achieves better results based on the conditions in the field, enabling them to plant the crop that is appropriate to those conditions. The proposed system helps a huge number of farmers by using IoT (Internet of Things) devices and web applications for smart irrigation.

Keywords: smart irrigation; climate; farmers; recommendation system; crop; yield; machine learning; IoT; KNN

check for updates

Citation: Vidhya, K.; George, S.; Suresh, P.; Brindha, D.; Jebaseeli, T.J. Agricultural Farm Production Model for Smart Crop Yield Recommendations Using Machine Learning Techniques. *Eng. Proc.* **2023**, *59*, 20. https://doi.org/10.3390/engproc2023059020

Academic Editors: Nithesh Naik, Rajiv Selvam, Pavan Hiremath, Suhas Kowshik CS and Ritesh Ramakrishna Bhat

Published: 11 December 2023

Copyright: © 2023 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

In India, many farmers are still doing agriculture by following the traditional methods that do not produce improved results. Approximately 60% of India's population relies mostly on agriculture for their livelihood. India's land area is utilized for agriculture to a degree of about 50% [1]. However, the productivity of Indian agriculture is quite poor, and the level of poverty among farmers is very high. India's agricultural markets are inefficient, which results in high costs for consumers and poor prices for farmers. Farmers struggle to turn a profit as a result. Andreas et al. [2] conducted a thorough evaluation of existing scholarly efforts in the area of big data in farming to give solutions to pertinent concerns. Crop security is put at risk by unpredictable weather patterns, chaotic seed planting, unexpected locust problems, and uneven irrigation. Van Klompenburg et al. [3] conducted research to gather and synthesize the properties of machine algorithms for agricultural usage. Farmers with limited experience in traditional agriculture have difficulties in making wise crop selection decisions in today's changing environment.

There are various deep learning algorithms used in smart farming and food manufacturing, as highlighted by Ryan et al. [4]. Olakunle et al. [5] proposed an IoT ecosystem that combines advances in technology, economics, and commercial viability to enable smart farming. Ayaz et al. [6] applied swarm intelligence-based strategies to instantaneously overcome precision agriculture optimization challenges. Nandhini et al. [7] produced a comprehensive systematic study that includes GAN variants in sustainable farming as well as other applications. Lombardo et al. [8] presented a thorough study to determine the effectiveness of precise agricultural technology (PAT) relative to the farming system. The cultivation of identical crops in each climate system degrades soil health [9]. Crop yields are forecasted to drop, with the worst reductions expected in certain emerging nations, including Southeast Asia and India, according to the Climate Change Crop Yield Assumptions study [10]. Niketa Gandhi et al. [11] applied a neural network technique to estimate rice yield. The data are subsequently checked and processed using the Waikato Environment for Knowledge Analysis (WEKA) tool. However, this study implies that incorporating extra crop-related factors might improve an ANN-based model's prediction capability. Vijo et al. [12] choose to aid farmers by offering agricultural yield prediction using wireless sensor networks (WSNs). Aruul Mozhi et al. [13] presented precision agricultural technologies based on IoT and deep learning. This method analyses and collects soil characteristics from the environment using a wireless sensor network. Furthermore, these technologies advise farmers on the optimal irrigation strategies to use by forecasting the crop to be seeded for the following crop cycle [14]. The farmers will receive this data via SMS. Soil temperature, ambient temperature, and humidity are among the elements. Raghavendhar et al. [15] are unable to distinguish between permanent and non-permanent crops to cultivate numerous harvests in a single year. Regardless of the technology stated above, the purpose of this research is to give recommendations and spark conversations on the use of smart farming approaches in agriculture.

The proposed research aims to make smart technology available to farmers. The KNN algorithm predicts the crop that farmers will cultivate by analyzing thousands of data points in the dataset. When a nation opts for this smart technology, its yield will automatically increase because technology and especially user-friendly technology will surely make the farmers do better and also produce more yield. Since many farmers are knowledgeable about farming with or without technology, then with technology, the face of agriculture and Indian crop yields will also increase.

Challenges

Farmers can benefit from using the agricultural farm production model for smart crop yield suggestions to assist in choosing the best crops, planting them, and applying fertilizer. The following are factors to take into account when creating a model for agricultural farm output that would propose smart crop yields:

i. The model needs to be capable of handling a range of crops and environmental factors;
ii. The model needs to be precise and trustworthy;
iii. The model must be simple to use and understand;
iv. The model ought to be adaptable to large datasets;
v. New data should be continually included in the model.

2. Dataset

The crop recommendation dataset on Kaggle has about thirteen properties [16]. The dataset consists of 2200 data points against 8 parameters. The proposed approach uses a dataset developed by enhancing existing precipitation, temperature, and nutrient information for India.

3. Methodology

Machine learning has emerged as an essential resource in the farming arena as technology has advanced. It is a well-known decision-making method for sustainable farming.

An IoT-based agricultural system aids in making better choices and prevents bad scenarios. Smart agriculture automation systems are less costly yet more accurate than conventional farming systems. There are various types of historical data used for agricultural production analysis. The agricultural yield data reveals how much of a crop was produced in a specific year or season. The meteorological data provides weather patterns for a specific year or season, including the temperature, precipitation, and humidity. The soil data gives the physicochemical characteristics of the soil at a certain location. The farming methods reveal how the crops were cultivated, including the sort of irrigation system used and the quantity of fertilizer used. Geographical information discloses the farm's location, including its latitude, longitude, and elevation.

3.1. The Proposed Architecture

As shown in Figure 1, the user first provides the input into the system and the machine learning algorithm will match inputs and provide a recommendation.

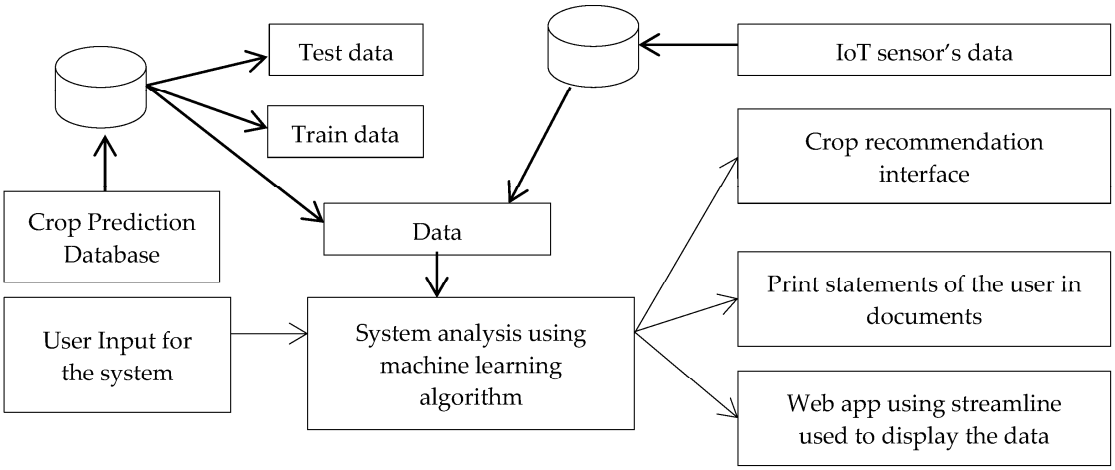


Figure 1. Block diagram of the proposed crop recommendation.

The following procedures are used to create an agricultural farm production model for machine learning-based smart crop yield recommendations:

i. Gather data on crop yield, weather patterns, soil quality, and other elements that might have an impact on crop productivity;
ii. Prepare the data for machine learning by cleaning it. This might entail normalizing the data, eliminating outliers, and entering missing numbers;
iii. Select a machine learning algorithm for crop yield prediction;
iv. Training could take some time, depending on the size of the dataset and the complexity of the machine learning technique;
v. Run the machine learning algorithm on a dataset that has been held back. This will make it feasible to avoid having the training set of data be overfit by the machine learning algorithm;
vi. Use the machine learning technique to make predictions. To develop forecasts for novel crops or novel situations, apply the machine learning algorithm;
vii. Assess the forecasts. Metrics like accuracy, precision, and recall may be used to gauge how accurate forecasts are.

The system's goal is to aid farmers in selecting crops with knowledge. The temperature of the surrounding area is measured using an LM35 temperature sensor, which is illustrated in Figure 2. The DHT22 humidity sensor measures the temperature along with the humidity ratio of the environment around it. The pH meter is used to determine the pH level. With the aid of an Arduino microcontroller, which also manages the sensors, these sensors collect data. Data received over a Wi-Fi connection are saved in an Excel spreadsheet. This project

analyses the humidity and temperature of real-world data acquired from the field using a DHT22 sensor and a history of rainfall, to produce a precise and reliable crop forecast. While recommending a crop, the location, market needs, and the crop of neighboring farmers may be taken into account.

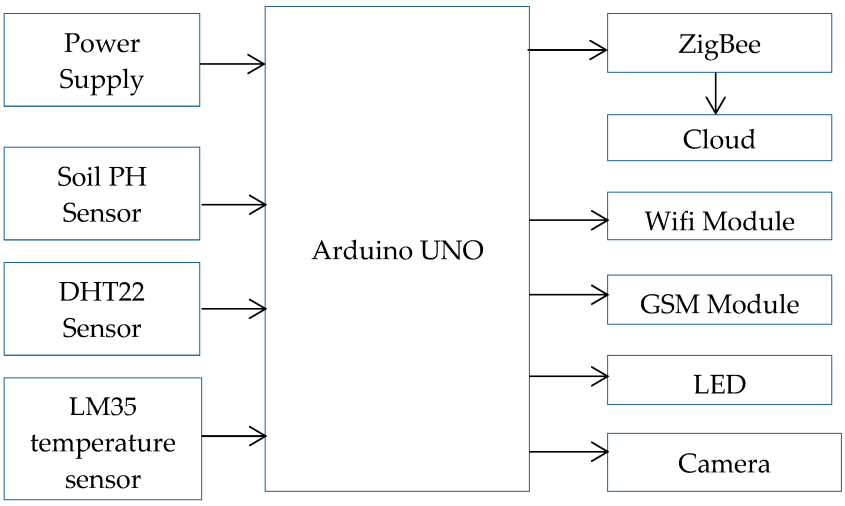


Figure 2. The proposed IoT model for the crop yield recommendation system.

Real-time data collection of a variety of variables is carried out by cameras, weather stations, GPS units, and soil moisture sensors. The gathered data are transmitted to a centralized server or cloud-based system via Bluetooth or other wireless protocols. To integrate the IoT data with the crop recommendation system, APIs or connectors are used. Hence, real-time data access is possible using these APIs. Farmers can improve crop yields, resource utilization, and decision-making while adjusting to changing environmental circumstances by integrating IoT-based real-time data into a crop recommendation system.

3.2. Pseudocode of Linear Regression Model

The linear regression model would employ a linear equation in the environment of agricultural yield prediction to forecast the yield of a crop depending on additional parameters like the weather, the type of soil, and the quantity of water and fertilizer to be applied.

```
# Load preprocessed data
data = load_preprocessed_data()
# Split the data into features (X) and target (y)
# Divide the data into two sets: training and testing.
# Initialize the linear regression model
# Train the model
# Make predictions on the test set
# Use the trained model to make predictions for a new input
    new_input = prepare_new_input()
function prepare_new_input(weather_data, soil_data, other_features):
# Combine and preprocess input data
        input_data = combine_and_preprocess_data(weather_data, soil_data, other_features)
# Ensure input data has the same format as the training data (e.g., columns, order)
        input_data = match_input_format(input_data)
        return input_data
    predicted_yield = model.predict(new_input)
# Crop selection based on predicted yield
    selected_crop = select_crop(predicted_yield)
# Display the recommended crop and predicted yield
```

print("Recommended Crop:", selected_crop)
print("Predicted Yield:", predicted_yield)

3.3. Mathematical Model of the Crop Yield Recommendation Approach

Consider the region R_n, recommended crop C_r during the season S_n of the year y_r is defined in Equation (1). Y_i is the measure of crop yield and it is the output per unit area is computed as follows:

$$Y_i = \alpha_{R_n C_r S_n} + \beta_{1 R_n C_r} \times Y_{r_i} + \beta_{2 R_n C_r} \times {Y_{r_i}}^2 + C_v \quad (1)$$

where C_v is the climate variable, 'i' is the observation of each region of the crop as per season in a year.

$$C_v = \left\{ T_i, {T_i}^2, P_i, {P_i}^2, T_i \times P_i, T_i \times P_i^2, P_i \times T_i^2 \right\} \quad (2)$$

where P is the precipitation and T is the temperature. These two parameters produce independent effects. Since the increase in temperature in the dry season may decrease yields. To estimate the climate, consider the following specifications:

$$\begin{bmatrix} T_i \\ P_i \end{bmatrix} = Y_{r_n s_n} + T_v \times \gamma \quad (3)$$

where γ is the climate variable. It is required to measure the impacts of climate variables such as temperature and precipitation and is denoted as follows:

$$\hat{T} = T_{yr_n s_n} + \left(T_i - \hat{T}_i \right) \quad (4)$$

$$\hat{P} = P_{yr_n s_n} + \left(P_i - \hat{P}_i \right) \quad (5)$$

Estimating the yield based on climate changes is given in Equation (6).

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} + \hat{C}_v \quad (6)$$

The change in yield due to climatic variables over 50 years is estimated as follows:

$$Y_{ci} = \left(\frac{Y_i - \hat{Y}_i}{Y_i} \right) \% \quad (7)$$

Based on Equation (7), the crop yield in percentage terms is predicted as per season in a year. This is based on the estimation of yields in a field based on climatic cultivation.

3.4. KNN Algorithm

The use of a KNN algorithm to differentiate crops emphasizes the importance of micronutrients and climatic parameters and has been proposed as an effective approach to crop prediction. The pseudocode representations of the main steps are as follows:

```
# Input: Data from sensors in real time (e.g., soil moisture, temperature)
# Output: Recommended crop
function recommend_crop(real_time_data, training_data, K):
# Calculate distances between real-time data and training data
distances = calculate_distances(real_time_data, training_data)
# Find the K-nearest neighbors
nearest_neighbors = find_nearest_neighbors(distances, K)
# Tally votes for each crop among the neighbors
crop_votes = count_votes(nearest_neighbors)
# Recommend the crop with the most votes
recommended_crop = get_most_voted_crop(crop_votes)
return recommended_crop
```

As shown in Figure 3, the best value is obtained when k = 2. Finding the value of K, the numbers of the nearest neighbors are taken into account while providing crop suggestions. The performance of the algorithm may be impacted by K selection. Find the difference between every vector of features in the historical data and the real-time feature vector which represents the current conditions. The K data points are considered to be the nearest neighbors. The crop class with the most support is recommended as the best choice under the current circumstances.

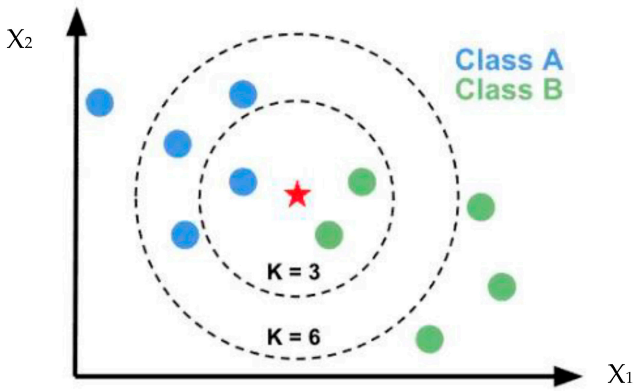


Figure 3. Optimal classification results at k = 2.

3.5. IoT Implementation

The GSM module is linked to Arduino. Sensor sensitivity readings are assessed, and alerts are given to farmers via their smartphones. As a result, farmers can receive help in deciding on crop cultivation. The data saved in the cloud is immediately accessible. ThingSpeak receives live data from NodeMCU, which may be processed and presented as graphs or conversations. It also aids in the generation of regular updates by scheduling the process to execute at a specific time. By using built-in charting features, farmers may visually grasp changes in soil and meteorological conditions. PyCharm is cross-platform software, with versions available for various operating systems. The supporting libraries of pysimple GUI are included with the print page. The web application module includes the library called Streamlit. When implementing IoT-based smart irrigation systems to serve a large number of farmers or agricultural operations, scalability considerations are crucial. Utilizing edge computing features to process data on local gateways closer to the source or IoT devices lessens the load on centralized servers. This will improve system responsiveness for a wide range of devices.

4. Results and Discussion

The recommended cultivation conditions in the location are represented in the circumstances of each crop to improve the utilization of land and air, which declines under appropriate climatic change. KNN algorithm determines the quantity of both temperature and humidity for every crop using these data. Based on the findings of the research and discussion, a conclusion comes in the form of many key factors that distinguish between internal and external variables, as well as supportive variables or elements that might boost crop yield. Several accuracy or validation methods may be used to examine the dependability of yield estimates. Some of the most prevalent measurements are described as follows: The absolute difference between projected and actual yields is averaged to calculate the mean absolute error (MAE). R-squared measures how much of the variance in actual yields can be explained by expected yields. Pearson correlation coefficient measures the linear relationship between expected and actual yields. The mean bias error (MBE) is the average of the projected and actual yield differences. As shown in Figure 4, Nitrogen (N) is essential for plant tissues. Potassium (K) is a chemical that assists in the normal functioning of plants' numerous functions. Phosphorus (P) is needed for the development

of roots in addition to blossom and fruit formation. Almost all crops require a higher Ph value when estimating soil Ph specifications per crop.

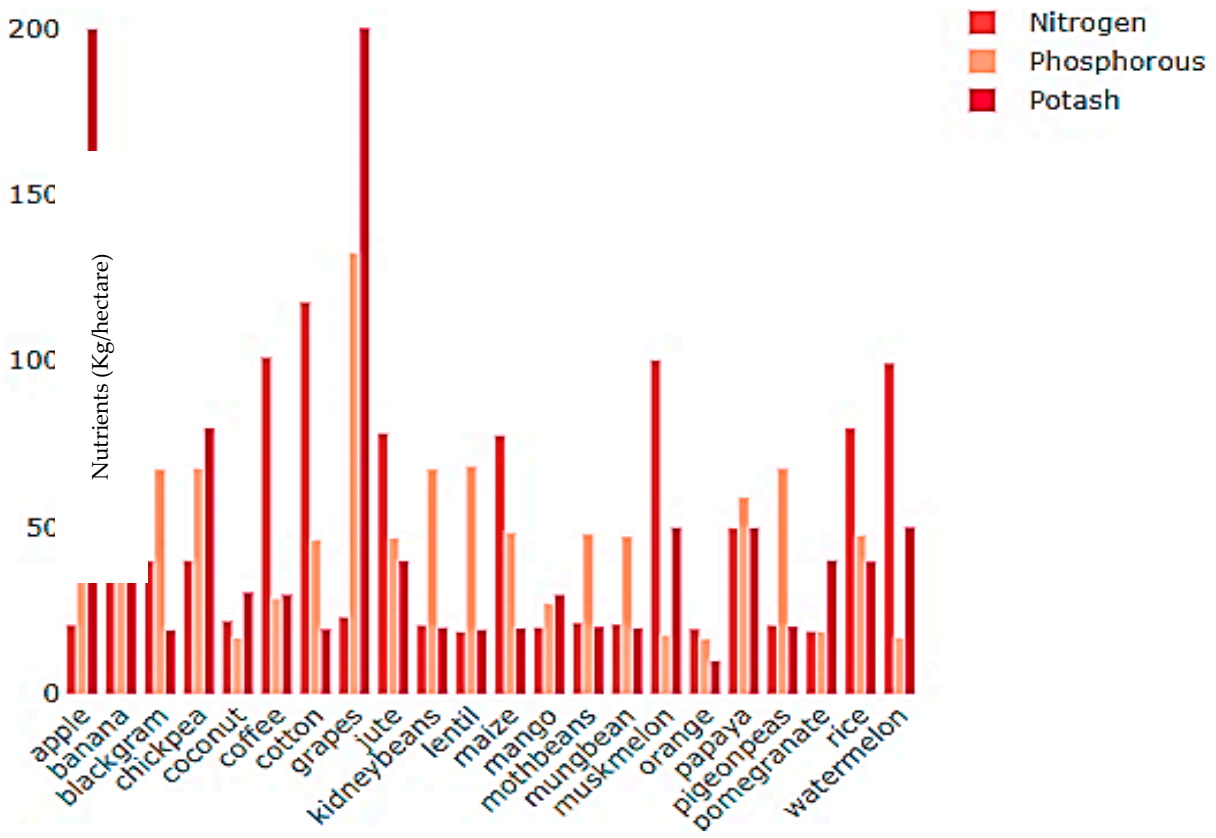


Figure 4. Required N, P, and K values for different crops.

Accuracy is used to assess the proposed method's performance.

$$Accurcay = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

The accuracy of the KNN, decision tree, and random forest algorithms is tested using performance measures based on a comparative evaluation. The KNN algorithm provides higher accuracy when analyzing performance, efficiently suggesting crops based on soil type and other characteristics. The proposed work comprises an interactive web interface that allows the user to provide the median amount of rainfall and pH value of the soil. When compared to competitive approaches, the KNN algorithm gives very accurate results, as demonstrated in Figure 5.

Smart agricultural monitoring integrates various technologies to help farmers [17]. Satellites and drones are capable of being employed to track plant health, water use, and the state of the soil using geographic technologies such as GPS and other geospatial technologies [18]. There are many advantages to implementing IoT-based smart irrigation systems at scale, including increased crop yields and water efficiency. The initial cost, infrastructure and connectivity, power supply, data security and privacy, data management, integration complexity, scalability, sensor accuracy and reliability, maintenance and calibration, and environmental factors are some of the challenges and restrictions that must be carefully taken into account [19].

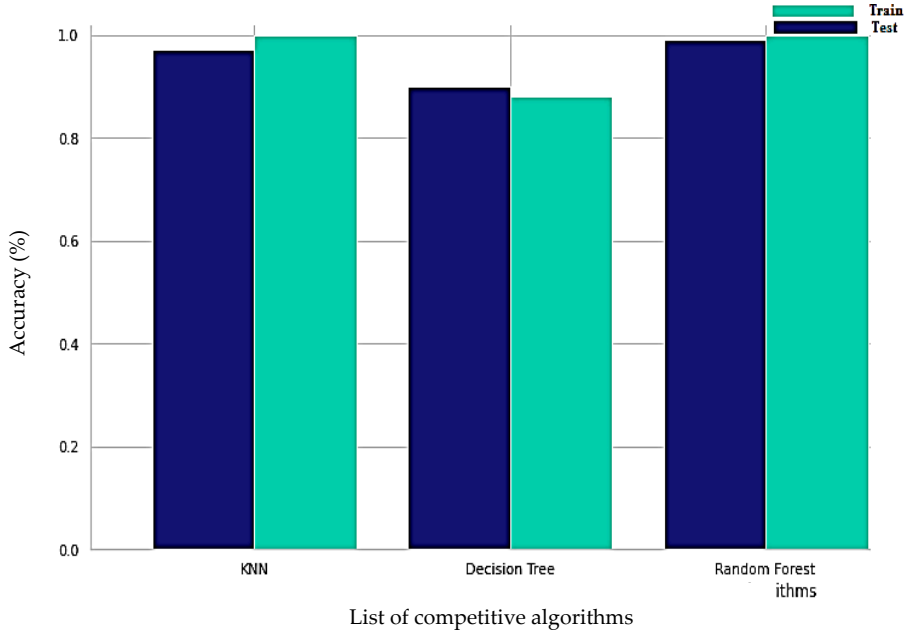


Figure 5. Testing vs. training accuracy comparison with competitive methods.

Farmers may access a plethora of information on crop production via web apps, such as weather forecasts, market pricing, and research on agriculture [20]. Farmers can then use this knowledge to make well-informed choices about when to begin planting, what to cultivate, and how to run their farms. Farmers can also use decision-support aids via web apps. These resources can assist farmers in data analysis, scenario simulation, and agricultural decision-making. Geographic coordinates, field size, topography, field faces, soil type, soil texture, organic matter, rainfall and precipitation, weather extremes, average annual temperature, and water source are the specific parameters and conditions in the field that a farmer may consider and enter into a web application for crop recommendation and decision-making.

5. Conclusions

Agriculture is the cornerstone of a developing country like India. Hence, there is a huge need to preserve crop production, which makes a major contribution to the agricultural and commercial well-being of countries throughout the world. Effective agricultural land usage is critical for a nation's achievement of food security. In this proposed system, a novel approach to smart agriculture makes use of two technological solutions. Using both live and historical data increases the precision of the outcome. Comparing several ML algorithms also improves the system's accuracy. This method will be utilized to alleviate the challenges farmers face while increasing the amount they produce and their performance of their jobs. When compared to earlier studies, the ML algorithms utilized in our proposed study provide more accuracy at a lower computational cost. The web application is being created to help farmers. The proposed system is extremely inexpensive and all of the sensors utilized are readily available and simple to use.

Author Contributions: Conceptualization, K.V.; methodology, S.G. and P.S.; formal analysis, P.S. and T.J.J.; investigation, D.B.; resources, K.V.; writing—original draft preparation, S.G. and T.J.J.; writing—review and editing, K.V., S.G., P.S. and D.B.; visualization, K.V.; supervision, P.S.; project administration, T.J.J. and D.B.; funding acquisition, K.V., S.G., P.S. and T.J.J. 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: Not applicable.

Acknowledgments: The authors would like to thank the Karunya Institute of Technology and Sciences and Sri Krishna College of Technology for all the support to complete this research.

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

References

1. How IoT-Based Smart Crop Monitoring Systems Help Maximize Farming Productivity by BIS Research. Available online: https://blog.marketresearch.com/how-iot-based-smart-crop-monitoring-systems-help-maximize-farming-productivity (accessed on 31 October 2023).
2. Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F.X. A review on the practice of big data analysis in agriculture. *Comput. Electron. Agric.* **2017**, *143*, 23–37. [CrossRef]
3. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]
4. Mark, R. Ethics of Using AI and Big Data in Agriculture: The Case of a Large Agriculture Multinational. *ORBIT J.* **2019**, *2*, 1–27. [CrossRef]
5. Elijah, O.; Orikumhi, I.; Rahman, T.A.; Babale, S.A.; Orakwue, S.I. Enabling smart agriculture in Nigeria: Application of IoT and data analytics. In Proceedings of the 3rd International Conference on Electro-Technology for National Development, Owerri, Nigeria, 7–10 November 2017; pp. 762–766.
6. Ayaz, M.; Ammad-Uddin, M.; Sharif, Z.; Mansour, A.; Aggoune, E.H. Internet-of-Things (IoT) Based Smart Agriculture: Toward Making the Fields Talk. *IEEE Access* **2019**, *7*, 129551–129583. [CrossRef]
7. Nandhini, J.; Narmadha, A.S.; Kalaiarasi, G. Intelligent Farming using Internet of Things and Cloud Computing. *Int. J. Comput. Appl.* **2021**, *183*, 22–26. [CrossRef]
8. Sarri, D.; Lombardo, S.; Pagliai, A.; Perna, C.; Lisci, R.; De Pascale, V.; Rimediotti, M.; Cencini, G.; Vieri, M. Smart Farming Introduction in Wine Farms: A Systematic Review and a New Proposal. *Sustainability* **2020**, *12*, 7191. [CrossRef]
9. Bouni, M.; Hssina, B.; Douzi, K.; Douzi, S. Towards an Efficient Recommender Systems in Smart Agriculture: A deep reinforcement learning approach. In Proceedings of the The Second International Workshop on Edge IA-IoT for Smart Agriculture, Procedia Computer Science, Niagara Falls, ON, Canada, 9–11 August 2022; Volume 203, pp. 825–830.
10. Davis, K.F.; Chhatre, A.; Rao, N.D.; Singh, D.; DeFries, R. Sensitivity of grain yields to historical climate variability in India. *Environ. Res. Lett.* **2019**, *14*, 064013. [CrossRef]
11. Gandhi, N.; Armstrong, L.J.; Petkar, O.; Tripathy, A.K. Rice crop yield prediction in India using Support Vector Machines. In Proceedings of the International Joint Conference on Computer Science and Software Engineering, Khon Kaen, Thailand, 13–15 July 2016.
12. Varghese, V.T.; Sasidhar, K.; Rekha, P. A Status Quo of WSN Systems for Agriculture. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, Kochi, India, 10–13 August 2015.
13. Varman, S.A.; Baskaran, A.R.; Aravindh, S.; Prabhu, E. Deep Learning and IoT for Smart Agriculture Using WSN. In Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, India, 14–16 December 2017; pp. 1–6.
14. Ghosh, K.; Balasubramanian, R.; Bandopadhyay, S.; Chattopadhyay, N.; Singh, K.K.; Rathore, L.S. Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal. *J. Agrometeorol.* **2014**, *16*, 1–8. [CrossRef]
15. Ikram, A.; Aslam, W.; Aziz, R.H.; Noor, F.; Mallah, G.A.; Ikram, S.; Ahmad, M.S.; Abdullah, A.M.; Ullah, I. Crop Yield Maximization Using an IoT-Based Smart Decision. *J. Sens.* **2022**, *2022*, 2022923. [CrossRef]
16. Raghavendhar, T.V.; Joshy, J.; Bhaskar, A. IOT Based Smart Agriculture System and Crop Recommendation using ML. *Int. J. Res. Electron. Comput. Eng.* **2019**, *7*, 1644–1651.
17. Crop Recommendation Dataset. Available online: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset (accessed on 31 October 2023).
18. Bharadiya, J.P.; Tzenios, N.T.; Reddy, M. Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches. *J. Eng. Res. Rep.* **2023**, *24*, 29–44.
19. Akkem, Y.; Biswas, S.K.; Varanasi, A. Smart farming using artificial intelligence: A review. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105899. [CrossRef]
20. Bakthavatchalam, K.; Karthik, B.; Thiruvengadam, V.; Muthal, S.; Jose, D.; Kotecha, K.; Varadarajan, V. IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. *Technologies* **2022**, *10*, 13. [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.