Received: 22 May 2025 | Revised: 11 July 2025 | Accepted: 26 August 2025 | Published online: 6 September 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 4(3) 323–335
DOI: 10.47852/bonviewAIA52026214

BON VIEW PUBLISHING

Smart Farming: Crop Recommendation Using Machine Learning with Challenges and Future Ideas

Devendra Dahiphale[1], Pratik Shinde[1], Koninika Patil[1] and Vijay Dahiphale[2,*]

[1]*Department of Computer Science and Electrical Engineering, University of Maryland, USA*

[2]*Department of Computer Science, Binghamton University, USA*

Abstract: This paper addresses the critical challenge of optimizing crop selection in agriculture to enhance food production sustainably. The problem is framed as a multi-class classification task where the goal is to recommend the most suitable crop based on a set of environmental and soil features. While traditional methods rely on time-consuming and labor-intensive expert knowledge, this work proposes a data-driven approach using machine learning. The novelty of our investigation lies in the comprehensive comparative analysis of seven machine learning algorithms and the development of a highly accurate neural network model. We utilize a publicly available dataset from Kaggle, which has been preprocessed to ensure data quality. We provide a detailed account of our feature engineering and hyperparameter tuning processes. Our proposed neural network model, with a specific architecture of 30–20–10 neurons, achieves a validation accuracy of 97.73%. This work also discusses the challenges of deploying such models, including real-world data variability and the need for model interpretability. We demonstrate that our approach, particularly the neural network model, provides a robust, scalable, and adaptable solution for crop recommendation, outperforming other models (in holistic view) like Random Forest which achieved a slightly higher accuracy of 99.5% on this specific dataset but with less generalization potential. The findings of this study can empower farmers to make informed decisions, ultimately leading to improved crop yields, enhanced soil fertility, and greater profitability.

Keywords: smart farming, precision agriculture, machine learning, deep learning, crop recommendation, big data, yield prediction, sustainable agriculture

1. Introduction

Machine learning [1, 2] is a field of study that gives computers the ability to learn without being explicitly programmed, a definition by Arthur Samuel (1959). Machine learning algorithms [2] are trained on large amounts of data to make predictions or decisions.

Agriculture, being a major sector worldwide, requires farmers to cultivate profitable and sustainable crops. Not choosing the right crop can have a significant impact on crop yield (in addition to other negative impacts such soil degradation), leading to decreased productivity and potential financial losses for farmers. When farmers fail to consider crucial factors such as climate suitability, soil conditions, and market demand, the chosen crops may struggle to thrive and achieve their full yield potential. Unsuitable crops may suffer from inadequate adaptation to the local climate, resulting in poor growth, increased vulnerability to pests and diseases, and reduced overall yield. Moreover, crops that do not align with market demand may face difficulties in finding buyers or fetching favorable prices, further exacerbating the economic impact on farmers. By leveraging machine learning-based crop recommendation systems, farmers can mitigate these challenges and make informed decisions to maximize crop yield and ensure long-term agricultural viability.

Machine learning and agricultural data converge to revolutionize how farmers understand and optimize their practices. With the increasing availability of data from sources such as weather stations, satellites, sensors, and farm equipment, machine learning algorithms can analyze vast amounts of information and extract valuable insights. These algorithms can uncover complex patterns, correlations, and predictive models that were previously hidden within the data. By combining machine learning techniques with agricultural data, farmers gain the ability to make data-driven decisions, ranging from crop selection and irrigation management to pest control and yield prediction. This integration empowers farmers to achieve higher efficiency, resource optimization, and sustainable practices, ultimately leading to improved productivity and profitability in the agricultural sector.

Crop recommendation systems can be used to analyze a variety of data, such as weather data, soil data, and market data. This data can be used to train machine learning models to predict which crops will likely be successful in a given location. Crop recommendation systems can also inform farmers about the best practices for growing specific crops.

The development of crop recommendation systems using machine learning has the potential to improve the productivity and sustainability of agriculture. By helping farmers to choose suitable crops to grow, crop recommendation systems can help to increase crop yields and reduce the use of resources. Furthermore, as climate change continues to alter weather patterns and increase the frequency of extreme events, the need for adaptive agricultural practices becomes paramount. Machine learning models can play a crucial role in climate change adaptation by providing recommendations that are resilient to these changing conditions [3, 4]. For instance, these systems can

***Corresponding author:** Vijay Dahiphale, Department of Computer Science, Binghamton University, USA. Email: vdahiphale@binghamton.edu

suggest drought-resistant crops in regions facing water scarcity or crops that can tolerate higher temperatures. Furthermore, machine learning can address several other challenges [5] in agriculture, for example, predicting crop yield, identifying pests and diseases, optimizing crop production, improving water efficiency, reducing the use of pesticides and fertilizers, soil management, etc.

Crops are a significant source of food and fiber for the world's population. The World Resource Institute is trying to solve the problem of how to feed ten billion people sustainably by 2050. Therefore, increasing high-quality crop yield is very important. The choice of crops to plant can significantly impact crop yields and profitability. Climate change and other environmental factors make it tough to predict which crop will succeed, given the location.

In this paper, we use machine learning to recommend crops to farmers. First, we collect the dataset and preprocess it. Then, we train and test models using features such as soil content and type, soil pH value, temperature, humidity, and rainfall. We also attempted feature engineering concepts to verify if the model performs better using a combination of different features and use it as a new feature in the same dataset. Agriculture has general challenges, and in the context of machine learning, therefore, we highlight these challenges thoroughly. Eventually, we present some exciting ideas for the readers to venture into.

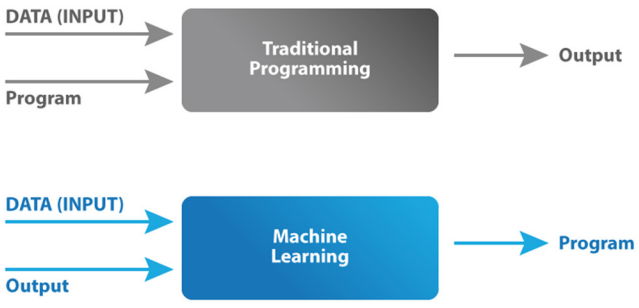
2. Background Survey

2.1. Machine learning

Machine learning gives computers the ability to learn without being explicitly programmed. In other words, machine learning is turning things or data into numbers and finding patterns in those numbers. The identified patterns help in predicting output for new data points. The fundamental difference between traditional programming and machine learning is shown in Figure 1. Traditional programming and machine learning are two different approaches to solving problems. Traditional programming involves writing code that defines the steps that the software should take to solve the problem. On the other hand, machine learning involves training a model on data so that the model can learn to solve the problem on its own. Machine learning algorithms are primarily categorized into three types based on how machines learn.

1) Supervised Learning: Models are trained on labeled data in supervised machine [2] learning. This means that the data has been tagged with the correct output. The model then learns to predict the outcome for new data that has not been labeled. Several supervised machine learning algorithms exist, such as decision trees, Logistic regression, support vector machines, and neural networks.
2) Unsupervised Learning: Unsupervised learning [2] is a type of machine learning where the model is trained on a set of unlabeled data. This means that the data does not have any labels associated with it. The model then learns to find patterns in the data and to group similar data points. Some examples of unsupervised learning are k-means clustering, hierarchical clustering, a priori algorithm, principal component analysis, etc.
3) Reinforcement Learning: Reinforcement learning (RL) [2] is a type of machine learning that allows an agent to learn how to behave in an environment by trial and error. The agent receives rewards for actions that lead to desired outcomes and punishments for actions that lead to undesired results. Over time, the agent learns to take actions that maximize its rewards. The algorithms such as q-learning, policy gradients, and actor critic fall into the category of reinforcement learning.

Figure 1
Traditional programming vs. machine learning



2.2. Machine learning algorithms used

Although many machine learning algorithms are commonly used, we are only highlighting the following algorithms in this survey because they are the ones that we used in our study. The choice of these seven algorithms was based on their popularity, diverse approaches to classification (linear models, tree-based ensembles, instance-based learning, probabilistic models, and neural networks), and their previous application in agricultural research. This allowed for a broad comparison of different modeling techniques for the crop recommendation task.

1) Logistic Regression: Logistic regression [2, 6] is a statistical method that predicts the probability of an event occurring. It is a type of regression analysis that is used to model the relationship between one or more independent variables and a categorical dependent variable.
2) Decision Tree: A decision tree [2, 7] is a supervised learning algorithm that uses a tree-like structure to represent the relationship between the input and output data. A decision tree is made up of nodes and branches. The nodes represent decisions, and the branches represent the possible outcomes of those decisions.
3) Random Forest: A random forest [2] is an ensemble learning algorithm comprising a collection of decision trees. Random forests are created by training many decision trees on different subsets of the training data. Each decision tree is trained using a random subset of the features. To make predictions, each decision tree in the random forest makes a prediction. The final prediction is made by taking the majority vote on the predictions from the individual decision trees. Random forests are often used for classification and regression tasks.
4) K-Nearest Neighbors: The K-nearest neighbors (KNN) [2] algorithm is a supervised learning algorithm. KNN works by finding the k most similar neighbors in training set to a new input instance and then predicting the label of the new input instance based on the labels of the K nearest neighbors. This algorithm can be used for both classification and regression problems.
5) Naive Bayes: The naive Bayes algorithm [2, 8] is a supervised learning algorithm that uses Bayes' theorem. It is a simple and versatile algorithm that can be used for various tasks, such as spam filtering, text classification, and medical diagnosis. The naive bayes algorithm works by assuming that the presence of a particular feature in a class is unrelated to the presence of any other feature. This assumption is not always true. Therefore, it is called "naive". However, it is a good approximation in many cases, making the algorithm very simple to train and interpret. To classify an object, the naive bayes algorithm first calculates the probability of each class. It then calculates the probability of each feature given to each class. The class with the highest probability is the class the object is assigned to.
6) SVM: A support vector machine (SVM) [2] is a supervised learning algorithm. SVMs are based on finding a hyperplane that separates

the data into two classes. The hyperplane is chosen to maximize the distance between the hyperplane and the closest data points on either side. This algorithm can be used for both classification and regression tasks.

7) Neural Network: The human brain inspires a neural network [2, 9, 10]. It is a network of interconnected (also called edges or connections) neurons called nodes. Neural networks are made up of multiple layers of nodes. The first layer of neurons is called the input layer, whereas the last layer is called the output layer. The layers in between are referred to as hidden layers. Each neuron in a neural network has a number of inputs and a single output. The inputs to a neuron are the outputs of the neurons in the previous layer. The result of a neuron is calculated using a function called an activation function [2]. The activation function is a non-linear function that transforms the input to the neuron into an output. The most common activation function is the sigmoid [2] function. However, we can provide a custom activation function based on our requirements.

2.3. Machine learning vs. big data processing frameworks

Machine learning and big data processing frameworks like MapReduce [11] are powerful tools that can be used to analyze large datasets. However, they have different strengths and weaknesses.

In some cases, machine learning, and big data processing frameworks can be used together. For example, machine learning models can be created using machine learning and then used by big data processing frameworks to generate final results.

Machine learning models are trained on large datasets and can then be used to make predictions or decisions without human intervention. Big data processing frameworks, on the other hand, are designed to process large datasets quickly and efficiently. They are often used to process data for tasks such as data mining, data warehousing, large graph processing, and analytics [12].

We need a tool that is accurate, fast, scalable, and easy to use. Therefore, we decided to use machine learning. Machine learning can be used directly to make predictions, while big data processing frameworks require additional tools to make predictions on top of the data processing framework.

2.4. Existing research in crop recommendation

In the last few years, there have been slight increases [13] in research in the field of crop recommendation. For example, Priyadharshini et al. [14] present "Intelligent Crop Recommendation System", Benos et al. [15] present "Machine Learning in Agriculture", Doshi et al. [16] present a system called AgroConsultant, PS Kiran et al. [17] in their paper titled as "System for Crop Recommendation", Pande, et al. in their paper [18] proposes a viable and user-friendly yield prediction system for farmers, Rajak et al. in their paper [19] proposes a model with a majority voting technique using a support vector machine (SVM) and ANN as learners to recommend a crop, Reddy and Kumar in their paper [20] present a survey of the existing techniques for crop recommendation, Ghadge et al. in their paper [21] present a theory on the crop recommendation, Kulkarni et al. in their research paper [22] showcase the work on improving crop productivity through a crop recommendation system using ensembling technique; and Géron in his paper [2] (most cited on IEEE Xplore) present a similar approach using machine learning on data collected from a district in Tamil Nadu, India; however, the paper does not talk about models' accuracy or have not described data used. There is some other agricultural-related literature that is indirectly related to crop recommendation, for example; Ayaz et al. in their work [23] mainly talks about the Internet of Things and sensors for collecting agricultural data. Recent works have also focused on intelligent water management practices [24] and remote sensing technologies like PlanetScope nanosatellites for land use classification [25], which are complementary to crop recommendation systems. Furthermore, advanced deep learning models like Convolutional LSTM [26] and techniques like Cost-Sensitive Learning and Ensemble Methods [27] are being explored for crop yield forecasting and handling imbalanced datasets in agriculture, respectively. Our survey suggests little research on crop recommendation; much of the above-referred literature is from the last four to five years. We believe this could be because of inherent challenges in the field of the agriculture sector (which are presented in Section 8), in addition to the difficulties related explicitly to machine learning in agriculture.

Compared to these prior methods, the approach proposed in this paper demonstrates higher accuracy, potential for reasoning capabilities and broader crop coverage. It is based on a multi-class neural network trained on publicly available Kaggle data encompassing 22 crop types and 7 environmental features. The model is validated through stratified five-fold cross-validation and achieves a validation accuracy of 97.73%. Furthermore, it is architected to support modular integration with sensor data and large language models, which can aid in generating interpretable justifications for the crop recommendation output. This extensibility makes it well suited for deployment in intelligent and dynamic agricultural ecosystems.

3. Our Contribution

Although the existing literature, primarily covered in Section 2.4, provides a good foundation for the research topic on crop recommendation models, they have some limitations. For example, many authors do not comprehensively overview their research process. This includes not mentioning their dataset sources, the accuracy of their models, or how their models were trained and tested. Additionally, much of the research lacks implementation details and does not specify the features used. Finally, many manuscripts only present surveys or theoretical work on crop recommendation topics.

Our paper addresses these limitations by developing comprehensive crop recommendation models. We describe each step of our process in detail, including our data collection, feature engineering, model training, and evaluation. We show that our system has the highest accuracy of any crop recommendation model in the literature. We achieved this by conducting feature engineering, which transforms the data to make it more useful for machine learning algorithms. We also analyzed data using seven different machine learning algorithms and with different configurations to achieve the highest accuracy, keeping these models' performance in mind.

Our key contributions include:

1) Development and rigorous evaluation of crop recommendation models using seven supervised machine learning algorithms.
2) Implementation of a 4-layer neural network yielding 97.73% validation accuracy, with a detailed architectural description.
3) Systematic feature engineering and hyperparameter tuning to maximize model performance, with specific details provided.
4) An integration-ready architecture for real-time data from sensors and natural language interfaces using Large Language Models (LLMs).
5) A comparative evaluation against existing methods demonstrating a substantial performance improvement on the used dataset.

All in all, first, we preprocess the data. Further, we apply several machine learning algorithms for recommending a crop. We train models using the following algorithms and find and compare the accuracy of each of the models for the recommendation system. Moreover, we try

various configurations for each model to achieve better performance and accuracy. Our proposed model is designed with modularity for integration with real-time sensor systems and Large Language Models (LLMs). This extensibility allows for real-time data ingestion and explainable AI outputs, enabling future deployment in intelligent decision support systems.

Second, we address the limitations of the papers highlighted in Section 2.4. We present a comprehensive crop recommendation system with all the details.

Third, we highlight the challenges in the agriculture sector, both in general and in the context of applying machine learning techniques to agricultural data.

Finally, we present multiple good ideas as future work for our work. The list of future work items is stated in Section 9 at the end of the manuscript.

We believe that our work significantly contributes to the field of crop recommendation. Our comprehensive approach to the problem, high accuracy, and attempts at feature engineering are all novel contributions. We believe that our work will be helpful to farmers, agricultural researchers, and other stakeholders in the farming sector.

4. Data Description, Methodology, and Experimentation

This section presents an overview of the methodology pictorially in Figure 2 that we have used to train various models. First, we iterated all the following steps with all the selected machine learning algorithms listed in Section 3.

1) Input Data: Because the quality and quantity of the data significantly impact a model's accuracy, we ensured the data was clean and well-labeled. As shown in Figure 2, the input to the system is a combination of soil and environmental characteristics. Table 1 and 2 shows a sample of raw data we used to train and test our models.
2) Preprocessing: The dataset was preprocessed to handle missing values and duplicates. We removed all null and duplicate records. The features were segregated from the label column. We also experimented with feature engineering by creating new features from existing ones. For instance, we created a 'N–P–K ratio' feature to capture the nutrient balance, although this did not lead to a significant improvement in model performance and was therefore not included in the final models. All data was then described and plotted to identify and handle any outliers. The data was split into a 70% training set and a 30% testing set. This split was chosen to provide a sufficiently large training set for the models to learn the underlying patterns while leaving a substantial subset for unbiased evaluation. While cross-validation is a robust technique for model evaluation, the 70–30 split was used for initial model development and comparison for computational efficiency.
3) Choose a Machine Learning Algorithm: In each iteration, we chose one of the seven algorithms we had decided to use. For every selected algorithm, we iterated steps from preprocessing to testing or validating the model to tune the model.
4) Model Configurations: To achieve higher test and cross-validation accuracy, we performed hyperparameter tuning for each model. The search space and tuning strategies were as follows:
   Decision Tree: We tuned the 'criterion' ('gini' or 'entropy') and 'max_depth' (from 5 to 20).
   Random Forest: We tuned 'n_estimators' (from 50 to 200) and 'max_depth'.
   KNearest Neighbors: We tested different values for 'n_neighbors' (from 3 to 15).
   SVM: We experimented with different kernels ('linear', 'rbf') and the 'C' parameter.
   Neural Network: We tuned the number of hidden layers, the number of neurons per layer, the activation functions ('relu', 'sigmoid', 'softmax'), the number of epochs (from 50 to 1000), and the learning rate for the Adam optimizer. The optimal configurations are presented in Section 6. We used a grid search approach for hyperparameter tuning where feasible.
5) Training Models: This is where the machine learning algorithm learns from the data prepared in the "Preprocessing" step.
6) Testing Accuracy of the Model: We evaluate the accuracy of the created model against the test data. In addition, we measured the cross-validation accuracy of the model. If the accuracy is

Table 1
Main features

Index	Feature name	Feature description
1	Nitrogen (N)	Nitrogen is largely responsible for the growth of leaves on the plant.
2	Phosphorus (P)	Phosphorus is largely responsible for root growth and flower and fruit development.
3	Potassium (K)	Potassium is a nutrient that helps the overall functions of the plant perform correctly.
4	Temperature	Temperature in degree Celsius
5	Humidity	Relative humidity in %
6	pH	pH value of the soil
7	Rainfall	Rainfall in mm

Figure 2
Methodology

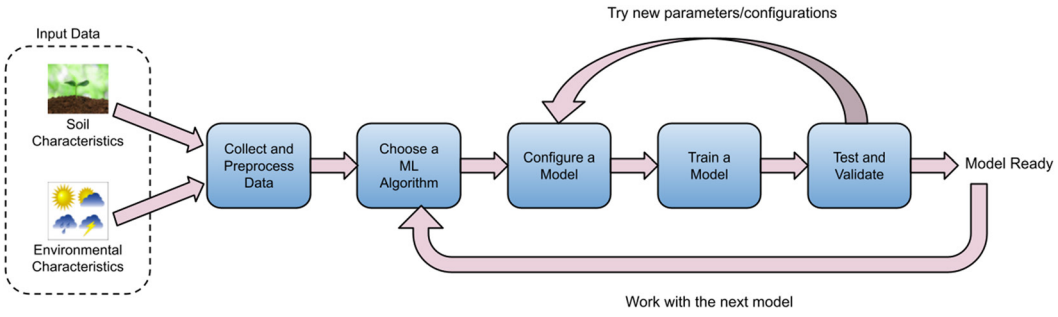


Table 2
First 5 rows of data

Index	N	P	K	Temperature	Humidity	Ph	Rainfall	Label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	Rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	Rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	Rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	Rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	Rice

unsatisfactory, we iterate the process by returning to the "Model Configuration" step. In some instances, we experimented with the feature engineering approach. Suppose the model's accuracy and performance are good at this step. In that case, we return to choosing a new algorithm step to repeat the same procedure with another advanced machine learning algorithm.

4.1. Experimentation

1) Model Using Multi-Class Neural Network: A multi-class neural network can be used to classify data into multiple classes. This is in contrast to a single-class neural network, which can only classify data into one category. We created a four-layered neural network using the TensorFlow [28, 29] framework; Figure 3 shows one of the examples. The first layer (input layer) contains thirty neurons, the second twenty neurons, the third includes ten neurons, and the fourth layer (output layer) includes twenty-two neurons. The second and third layers are called hidden layers. The rationale behind the 30–20–10 neuron architecture was to create a funnel-like structure that progressively extracts more abstract features from the input data. This architecture provided a good balance between model complexity and performance, avoiding overfitting on the relatively small dataset. We experimented with different combinations of "relu", "softmax", and "sigmoid" activation functions to tune the model for better accuracy and performance. The best performance was achieved with 'relu' in the hidden layers and 'softmax' in the output layer. Finally, we experimented with the network with multiple epochs [2] values until we found the optimal ones. Increasing epoch value decreases the performance of the neural network.

Figure 3
Neural network

```
import tensorflow as tf

model = tf.keras.models.Sequential([
    tf.keras.layers.Dense(30, activation='relu', input_shape=(7,)),
    tf.keras.layers.Dense(20, activation='relu'),
    tf.keras.layers.Dense(10, activation='relu'),
    tf.keras.layers.Dense(labels_count, activation='softmax')
])

model.compile(
    loss=tf.keras.losses.CategoricalCrossentropy(),
    optimizer=tf.keras.optimizers.Adam(),
    metrics=['accuracy']
)

model.fit(
    x_train,
    y_train,
    epochs=60,
    validation_data=(x_test, y_test),
    batch_size=32
)
```

Furthermore, we used CategoricalCrossentropy as a loss function [2, 9], called categorical crossentropy. It is used for training multi-class classification models. It measures the distance between the predicted probabilities and the actual labels. The lower the categorical crossentropy, the better the model is performing.

Finally, we use an optimizer called Adams optimizer [30]. Adam is a popular optimization algorithm for training deep learning models. It is an extension of the AdaGrad [31] and RMSProp [32] algorithms, and it is effective for a wide range of problems.

2) Rest of the Models: All models except those using neural networks were created, trained, tested, and validated similarly. We used the classifiers listed in Table 3 to build models with different algorithms. Some notable differences are presented in this section.

For the decision tree algorithm, we used Gini and entropy [2], two impurity measures used in decision trees. We also used max depth for the decision tree as another parameter.

For the K-nearest neighbors algorithm, we tried a configuration called neighbors, which determines the total number of nearest outputs that should be considered.

For SVM, we used kernel configuration. Kernel machines are a class of algorithms for pattern analysis, and their best-known member is the support-vector machine (SVM).

Furthermore, we evaluated and cross-validated models with the functionality from Figure 4.

5. Proposed Methodology

In this section, we present the proposed and recommended method for crop recommendation, which is based on a supervised learning technique using a multi-class neural network. Among all the evaluated models in our study including Decision Trees, Random Forest, SVM, Naive Bayes, and Logistic Regression the neural network emerged as the most effective, scalable, and adaptable, particularly for handling non-linear relationships inherent in agricultural data. Our experimental results (refer to Table 4) demonstrated that this model achieved a validation accuracy of 97.73%, coupled with precision and

Table 3
Primary classifiers used

Index	Model name
1	Logistic Regression (...)
2	Decision Tree Classifier (...)
3	Random Forest Classifier (...)
4	K neighbors Classifier (...)
5	Gaussian NB (...)
6	svm.SVC (...)

Figure 4
Model evaluation

```python
from sklearn.model_selection import KFold, cross_val_score
from sklearn.metrics import accuracy_score

# Initialize KFold for 5-fold cross-validation
kfold = KFold(n_splits=5, shuffle=True, random_state=42)

def evaluate(my_model, x_test_data, y_test_data):
    """
    Evaluates the model on test data and returns the accuracy percentage.
    """
    predictions = my_model.predict(x_test_data)
    accuracy = accuracy_score(y_test_data, predictions)
    return round(accuracy * 100, 3)

def perform_cross_val(my_model, features, labels):
    """
    Performs k-fold cross-validation and returns the mean accuracy score.
    """
    scores = cross_val_score(my_model, features, labels, cv=kfold)
    mean_score = round(scores.mean() * 100, 3)
    return mean_score
```

recall values of 0.99, establishing it as a highly reliable method for practical deployment in smart farming systems, especially considering the recent advances in artificial intelligence.

5.1. Rationale for choosing neural network

Neural networks are computational models inspired by the human brain and are especially adept at identifying complex patterns in data. Unlike traditional algorithms that rely on explicit if-then rules or linear separability, neural networks can automatically learn relationships among features such as soil nutrients, weather patterns, and moisture levels factors that interact in highly non-linear and multidimensional ways in agricultural ecosystems. This capability makes them particularly suitable for crop recommendation systems where input variables are interdependent and noisy. The model outperforms simpler classifiers in capturing non-linear dependencies and is resilient to noise in real-world agricultural datasets.

Additionally, the recent advances in artificial intelligence, particularly in deep learning and neural architecture design, have made neural networks more interpretable, scalable, and deployable on edge devices. These advances enable better model performance as well as new opportunities for building systems that can reason about their recommendations, rather than just providing black-box outputs. This is increasingly important in agriculture, where trust and transparency are key to adoption by farmers and agribusiness stakeholders.

Neural networks also serve as a foundational layer for future integration with Large Language Models (LLMs) or other generative AI tools, which could later be used to explain the decision-making process in natural language. For instance, a future system might not only recommend "maize" as the ideal crop but also provide a rationale such as: "Maize is recommended because your region has moderate nitrogen, low potassium, and rainfall levels below 100 mm, which match known optimal growth conditions for maize."

5.2. Architecture and configuration

The proposed model uses a feedforward neural network architecture implemented with the TensorFlow deep learning library. The architecture is structured as follows:

Input Layer: 7 neurons corresponding to input features: Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Humidity, pH level, and Rainfall.

Hidden Layers: Three hidden layers with 30, 20, and 10 neurons respectively, all utilizing ReLU activation.

Output Layer: 22 neurons (one for each crop class) with Softmax activation for multi-class classification.

The model is compiled using the Categorical Crossentropy loss function, which is suitable for multi-class classification problems, and optimized using the Adam optimizer. The training was performed over 1000 epochs with a batch size of 32, based on hyperparameter tuning.

5.3. Data flow and model training

Each sample from the dataset is transformed into a 7-dimensional input vector. For example:

[N = 85, P = 58, K = 41, Temp = 21.77 °C, Humidity = 80.31%, pH = 7.03, Rainfall = 226.65 mm]

This input is propagated through the network layers. The output layer generates a probability distribution across 22 crop categories, with the crop having the highest probability selected as the recommended choice. For example:

[rice: 0.91, maize: 0.04, banana: 0.01, . . .] →
Predicted Crop: rice

Table 4
Model accuracy (with standard deviation from 5-fold cross-validation and F1-Score)

Index	Model name	Accuracy%	Validation accuracy% ± Std Dev	Configurations	Precision/Recall/ F1-Score
1	Logistic Regression	94.545	95.955 ± 0.8	–	0.95/0.95/0.95
2	Decision Tree	99.091	98.682 ± 0.5	with Gini and Max Depth = 12	0.99/0.98/0.98
3	Decision Tree	99.091	98.409 ± 0.6	with Entropy and Max Depth = 10	0.99/0.99/0.99
4	Random Forest	99.545	99.5 ± 0.3	with n_estimators = 100	0.99/0.99/0.99
5	K-Nearest Neighbors	98.636	98.045 ± 0.7	n_neighbors = 5	0.98/0.98/0.98
6	Naive Bayes	99.545	99.5 ± 0.3	–	0.99/0.99/0.99
7	SVM	97.727	97.682 ± 0.9	Kernel = rbf	0.98/0.98/0.98
8	SVM	99.242	98.682 ± 0.5	Kernel = linear	0.99/0.99/0.99
9	Neural Network (relu and softmax)	–	95.00 ± 1.2	epoch = 60, (relu and softmax activation)	0.99/0.99/0.99
10	Neural Network (sigmoid)	–	97.73 ± 0.8	epoch = 1000, (sigmoid activation)	0.99/0.99/0.99

During training, the model uses backpropagation and gradient descent to minimize prediction error between predicted and actual crop labels.

5.4. Model evaluation and generalization

The dataset is divided into 70% training and 30% testing subsets. To validate the model, we used 5-fold cross-validation, which confirmed that the neural network consistently achieved high accuracy across different subsets, demonstrating its generalization capabilities.

Moreover, the neural model is data-agnostic, meaning it can generalize across varying soil and climate conditions when retrained with region-specific data. This makes it suitable for deployment in diverse agro-climatic zones globally.

5.5. Addressing key challenges

The proposed method effectively addresses several real-world and technical challenges:

1) Non-Linearity: Captures complex interactions between features without manual rule creation.
2) Scalability: Easily retrained or extended with more features, including real-time sensor data.
3) AI Reasoning and Interpretability: Recent advancements allow integration with explainable AI and LLMs, enabling transparent, rationale-driven crop suggestions.
4) Precision Agriculture: Provides localized, condition specific crop recommendations to optimize yield.
5) Data-Driven Decision Support: Empowers farmers, policymakers, and agribusinesses with real-time, trustworthy recommendations.

This neural-network-driven method lays a robust foundation for future AI-based agriculture systems and ensures adaptability to both current needs and future technological integrations.

6. Results and Evaluation

We used an existing Kaggle dataset for our model. The dataset comprised approximately 2200 instances extracted from an original pool of over 100,000 agricultural records. The dataset includes seven environmental and soil-based features across 22 crops. While this dataset provides a good starting point, we acknowledge its limitations. It is a static dataset that does not capture temporal variations, and being from Kaggle, it may contain inherent biases. For example, the data is specific to Indian agro-climatic zones and may not be generalizable to other regions without retraining. Table 1 shows the main features of the data, and the first few rows of the data are shown in Table 2. Figure 5 represents the distribution of crop features, while Figure 6 shows the pairplot for the features used. A pairplot is a type of statistical graph that shows the relationships between multiple features in the form of a matrix in a dataset, with each row and column representing a different variable. The plots in the matrix's diagonal show each variable's distribution, while the plots in the off-diagonal show the relationships between pairs of variables. Figures 7 and 8 show the pictorial representation of the features and their count. This dataset was created by augmenting datasets of rainfall, climate, and fertilizer data available for India.

We used a total of 22 unique labels for the data used are listed in Table 5.

These labels are extracted from a database of around 100k records; because there is one good crop for a given setting, the records were reduced to approximately 2.2k. Table 4 shows the results of our experiments. We tried different splits for train and test data and eventually settled down for 70% training data and 30% testing data. For all the models, we achieved an accuracy of at least 95% when the proper configurations were used. We experimented with various configurations for each model. The configurations in Table 4 are optimal in terms of performance and accuracy.

The high accuracy achieved by most models, particularly Naive Bayes and Random Forest (99.5%), is noteworthy. The Naive Bayes model performed surprisingly well, possibly because the features in the dataset have a high degree of independence, which is the core assumption of this algorithm. The Random Forest model's high accuracy is expected, given its robustness and ability to handle complex interactions between features. Our proposed Neural Network, while achieving a slightly lower accuracy of 97.73%, offers better scalability and adaptability for more complex, real-world scenarios. The discrepancy in accuracy between the NN and RF can be attributed to the relatively small size of the dataset (2.2k rows), which may not be

Figure 5
Crop features distribution (N (kg/ha), P (kg/ha), K (kg/ha), temperature (°C), humidity (%), rainfall (mm))

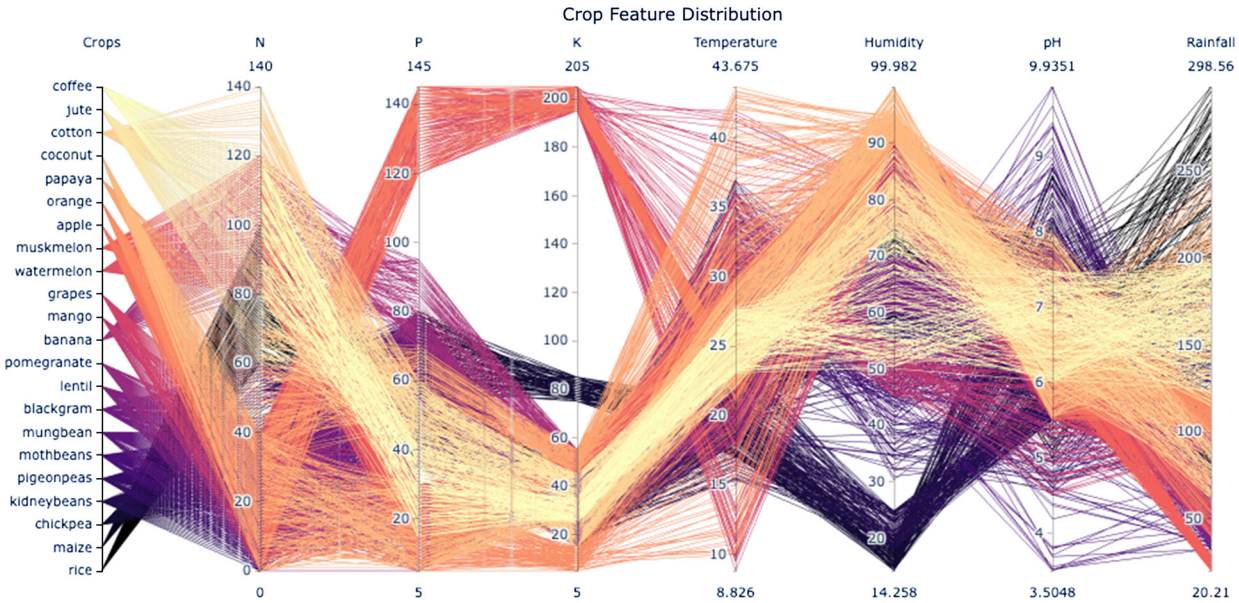
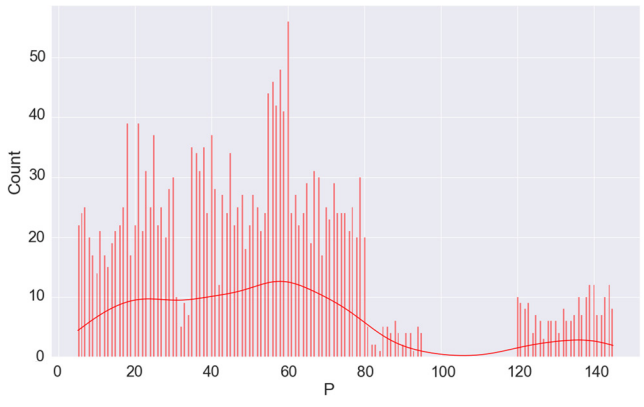


Figure 6
Pair plotting of all data

Figure 7
Feature graph for phosphorus (P) in kg/ha



large enough for the NN to fully leverage its learning capacity without overfitting. The computational cost of training the neural network was higher than the other models, especially with a large number of epochs, but its prediction time is fast, making it suitable for real-time applications.

For example, increasing the depth value of the decision tree increases the accuracy but also increases the training and prediction time. We achieved an accuracy of 99.5% with a random forest algorithm with 100 estimators.

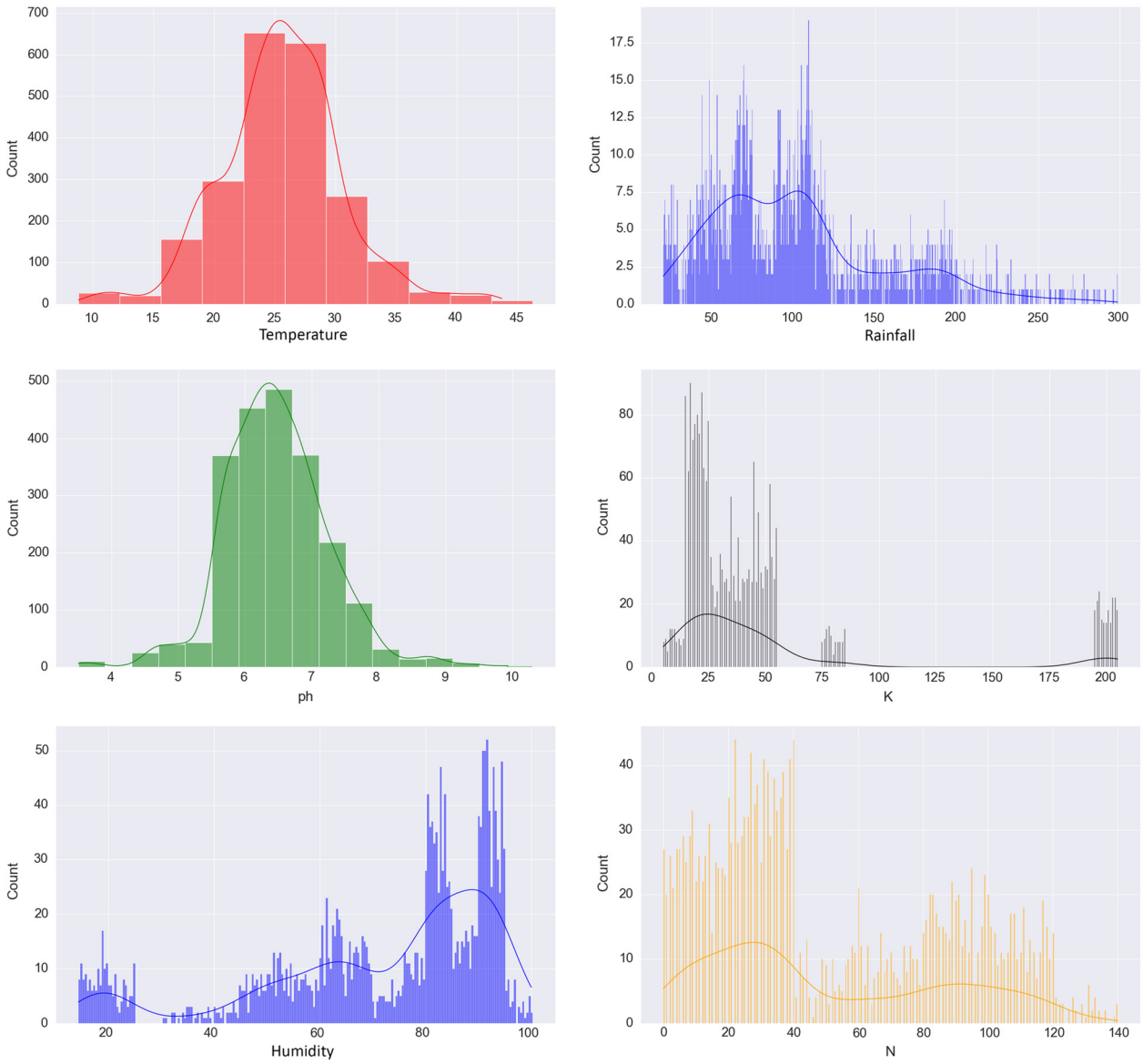
For the neural network model, we observed that the number of epochs plays a significant role in the accuracy and performance. The experiments clearly show that the performance is inversely proportional to the number of epochs, and the accuracy is directly proportional to the number of epochs. For example, we achieved an accuracy of 97.73% with 100 epochs. In machine learning, precision and recall [2] are two metrics used to evaluate the performance of a model. Precision measures the fraction of positive predictions that are positive, while recall measures the fraction of positive instances that are correctly identified. The last column in Table 4 shows precision and recall for each model. The formulas to calculate precision and recall are: Precision = TP / (TP + FP) Recall = TP / (TP + FN) where, TP = True Positives, FP = False Positives, and FN = False Negatives.

Based on our experimentation, the model using the Naive Bayes algorithm showed the highest accuracy. However, we believe that general neural networks would perform much better when the dataset size is much larger; also, it would be data agnostic.

We believe this work will help other developers or researchers understand the impact of different configurations on the accuracy and performance of machine learning models.

1) Average Conditions for Each Crop: Table 6 shows the average weather and soil characteristics values for each crop. Given the specific conditions in their region, this table can help farmers and other agricultural stakeholders make informed decisions about which crops to grow.

Figure 8
Feature graphs (temperature (°C), humidity (%), rainfall (mm), N (kg/ha), K (kg/ha))



7. Limitations

Despite strong model performance, several limitations exist:

Data Dependency: The model relies on structured, labeled data, which may be unavailable or inconsistent in real-world farming scenarios, particularly in developing regions.

No Real-Time Inputs: The current system does not integrate sensor or live weather data, limiting its adaptability to changing field conditions.

Economic Factors Ignored: Crop recommendations are based solely on environmental features, without considering market demand, pricing, or profitability.

Limited Regional Applicability: The model is trained on Indian conditions and may not generalize well to other agro-climatic zones without retraining.

Low Interpretability: The neural network offers high accuracy but lacks transparent explanations, which may hinder user trust and adoption.

Static Dataset: The model does not account for temporal trends such as seasonal shifts, soil degradation, or evolving climate patterns.

These constraints highlight areas for future work, particularly in enhancing data diversity, economic modeling, and system explainability.

8. Challenges in Agriculture

Agriculture faces many challenges today, both in general and in the context of machine learning. Some of the most concerning challenges are listed and explained below.

1) Climate change: Climate change [3] is already significantly impacting agriculture, and the effects are expected to worsen. Climate change is causing more extreme weather events, such as droughts and floods, which can damage crops and livestock. Climate change is also causing changes in temperature and

Table 5
All unique labels

Index	Label name
1	Rice
2	Maize
3	Chickpea
4	Kidney beans
5	Pigeon peas
6	Moth beans
7	Mung bean
8	Black gram
9	Lentil
10	Pomegranate
11	Banana
12	Mango
13	Grapes
14	Watermelon
15	Muskmelon
16	Apple
17	Orange
18	Papaya
19	Coconut
20	Cotton
21	Jute
22	Coffee

precipitation patterns, making it difficult for farmers to grow the necessary crops.

2) Increasing population: The world's population is expected to reach 9.7 billion by 2050. This means that we will need to produce more food to feed everyone. The increasing population puts a strain on agricultural resources, such as land, water, and fertilizer.
3) Water scarcity: Water scarcity [33] is a significant challenge in many parts of the world. Agriculture is a primary water user, and growing crops will become more challenging as water resources become scarce.
4) Soil degradation: Soil degradation [34] is a significant problem in many parts of the world. Various factors, such as overgrazing, deforestation, and poor agricultural practices, can cause soil degradation. Soil degradation makes it difficult to grow crops and can lead to erosion.
5) Pests and diseases: Pests and diseases [35] can damage crops and livestock, leading to significant losses for farmers. Pests and diseases are becoming more resistant to pesticides, making it more difficult to control them.
6) Labor shortages: There is a labor shortage in many parts of the world, including the agricultural sector. This is due to several factors, such as an aging population, migration, and low wages. Labor shortages make it difficult for farmers to harvest crops and care for livestock.
7) Economic challenges: Farmers face several economic challenges, such as low crop prices, high input costs, and competition from imported food. These challenges can make it difficult for farmers to make a living.
8) Data availability: Machine learning algorithms require large amounts of data to train. This data can be difficult and expensive, especially in developing countries.
9) Data quality: The data used to train machine learning algorithms must be high quality. This means that the data must be accurate, complete, and consistent.
10) Model interpretability: Understanding how machine learning models make decisions is essential. This is important for farmers who need to be able to trust the decisions made by the models.
11) Awareness: Many farmers lack resources and government subsidiaries to tackle their problems. Developing countries like India are getting better at this challenge. However, some countries still lack interest access and thus related resources to educate farmers.
12) Losses, inefficiencies, and waste in the global food system: First, because global agricultural dry biomass consumed as food is 6% (energy 9.0% and protein 7.6%); and second, 44% of harvested crops dry matter lost before human consumption. This is more detailed in Reference [36].
13) Crop damage: Crop damage by wild animals [37, 38] is a serious problem affecting farmers worldwide. Wild animals can damage crops in various ways, such as eating, trampling, polluting, and transmitting diseases. As a result, crop damage by wild animals can significantly impact farmers' livelihoods. In some cases, it can even lead to financial ruin.

9. Future Work and Ideas

There are multiple ways this work can be extended. Some of the examples are listed below. The readers are encouraged to extend this work by picking any of the following ideas.

1) Given the rapid advancements in Generative AI [39, 40], particularly large language models (LLM), a compelling future research direction involves developing an LLM-based crop recommendation system. This system would not only provide recommendations but also generate natural language explanations for them, addressing the interpretability challenge. A comparative study of this LLM-based system against the models evaluated in this paper would be highly valuable. This enhanced interpretability would allow farmers to understand the reasoning behind a recommendation and make more informed decisions.
2) Conduct a survey among farmers to evaluate the economic impact of using the proposed recommendation system. This would help quantify the financial benefits in terms of cost savings and increased profitability.
3) Develop a mobile application that integrates the proposed models, providing an end-to-end solution for farmers and agribusinesses. This would make the technology more accessible and user-friendly.
4) Collect and use data from diverse geographical regions to improve the model's generalizability and create region-specific recommendation systems.
5) Utilize a larger and more comprehensive dataset, including data on soil health, water quality, and pest infestations, to build more robust and accurate models.
6) Evaluate the economic and environmental impact of the recommendations, considering factors like water usage, carbon footprint, and market prices.
7) Integrate real-time data from on-farm sensors (for soil moisture, temperature, etc.) to provide dynamic and highly contextualized

Table 6
Features' mean values for each crop

Index	Crop name	Nitrogen	Phosphorous	Potassium	Temperature	Humidity	pH	Rainfall
1	Rice	79.89	47.58	39.87	23.69	82.27	6.43	236.18
2	Maize	77.76	48.44	19.79	22.39	65.09	6.25	84.77
3	Chickpea	40.09	67.79	79.92	18.87	16.86	7.34	80.06
4	Kidney beans	20.75	67.54	20.05	20.12	21.61	5.75	105.92
5	Pigeon peas	20.73	67.73	20.29	27.74	48.06	5.79	149.46
6	Moth beans	21.44	48.01	20.23	28.19	53.16	6.83	51.20
7	Mung bean	20.99	47.28	19.87	28.53	85.50	6.72	48.40
8	Black gram	40.02	67.47	19.24	29.97	65.12	7.13	67.88
9	Lentil	18.77	68.36	19.41	24.51	64.80	6.93	45.68
10	Pomegranate	18.87	18.75	40.21	21.84	90.13	6.43	107.53
11	Banana	100.23	82.01	50.05	27.38	80.36	5.98	104.63
12	Mango	20.07	27.18	29.92	31.21	50.16	5.77	94.70
13	Grapes	23.18	132.53	200.11	23.85	81.88	6.03	69.61
14	Watermelon	99.42	17.00	50.22	25.59	85.16	6.50	50.79
15	Muskmelon	100.32	17.72	50.08	28.66	92.34	6.36	24.69
16	Apple	20.80	134.22	199.89	22.63	92.33	5.93	112.65
17	Orange	19.58	16.55	10.01	22.77	92.17	7.02	110.47
18	Papaya	49.88	59.05	50.04	33.72	92.40	6.74	142.63
19	Coconut	21.98	16.93	30.59	27.41	94.84	5.98	175.69
20	Cotton	117.77	46.24	19.56	23.99	79.84	6.91	80.40
21	Jute	78.40	46.86	39.99	24.96	79.64	6.73	174.79
22	Coffee	101.20	28.74	29.94	25.54	58.87	6.79	158.07

crop recommendations. This would also help in reducing crop losses and improving decision-making.

8) Explore the use of hybrid or ensemble models, such as stacking, to potentially achieve even higher accuracy and robustness.

10. Conclusion

In conclusion, this research paper has presented crop recommendation models to predict the best crops to grow using multiple advanced machine learning algorithms and a deep neural network. The technique is scalable and easily adapted to new data and regions or countries.

The results of this study have several positive implications for the agricultural industry. First, the technique can be used by farmers to make more informed decisions about what crops to grow. Second, the method can be used by governments to develop policies that support the agricultural sector. Third, the method can be used by businesses to create new products and services that support the agricultural industry; Fourth, it will help keep the agricultural goods prices stable.

Next, we thoroughly presented agricultural challenges and some interesting future ideas to venture into.

Overall, this research has made a significant contribution to the field of agriculture. The technique is scalable, accurate, and easy to use, making it a valuable tool for farmers, governments, and businesses.

Acknowledgement

An earlier version of this work was made available as a preprint [41].

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset/data and in TechRxiv at https://www.techrxiv.org/doi/full/10.36227/techrxiv.23504496, reference number [41].

Author Contribution Statement

Devendra Dahiphale: Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing –

original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Pratik Shinde:** Validation, Investigation, Writing – review & editing. **Koninika Patil:** Writing – review & editing. **Vijay Dahiphale:** Validation, Writing – review & editing, Visualization, Project administration.

References

[1] Saleem, M. H., Potgieter, J., & Arif, K. M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precision Agriculture*, *22*(6), 2053–2091. https://doi.org/10.1007/s11119-021-09806-x

[2] Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. USA: O'Reilly Media, Inc.

[3] Ogundeji, A. A. (2022). Adaptation to climate change and impact on smallholder farmers' food security in South Africa. *Agriculture*, *12*(5), 589. https://doi.org/10.3390/agriculture12050589

[4] Haq, M. A., & Khan, M. Y. A. (2022). Crop water requirements with changing climate in an arid region of Saudi Arabia. *Sustainability*, *14*(20), 13554. https://doi.org/10.3390/su142013554

[5] Meena, A. K., & Maikhuri, R. K. (2024). Analysis of traditional hill agriculture: Unveiling challenges and opportunities in the Urgam Valley of Central Himalayan Region. *Journal of Mountain Research*, *19*(2), 363–374. https://doi.org/10.51220/jmr.v19-i2.38

[6] McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*. USA: John Wiley & Sons.

[7] Jijo, B. T., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20–28. https://doi.org/10.38094/jastt20165

[8] Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, *192*, 105361. https://doi.org/10.1016/j.knosys.2019.105361

[9] Prince, S. J. (2023). *Understanding deep learning*. USA: MIT Press.

[10] Kneusel, R. T. (2021). *Practical deep learning: A python-based introduction*. USA: No Starch Press.

[11] Dahiphale, D., Karve, R., Vasilakos, A. V., Liu, H., Yu, Z., Chhajer, A., ..., & Wang, C. (2014). An advanced mapreduce: Cloud MapReduce, enhancements and applications. *IEEE Transactions on Network and Service Management*, *11*(1), 101–115. https://doi.org/10.1109/TNSM.2014.031714.130407

[12] Dahiphale, D. (2023). MapReduce for graphs processing: New big data algorithm for 2-edge connected components and future ideas. *IEEE Access*, *11*, 54986–55001. https://doi.org/10.1109/ACCESS.2023.3281266

[13] Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: A systematic literature review. *New Zealand Journal of Crop and Horticultural Science*, *51*(1), 1–26. https://doi.org/10.1080/01140671.2022.2032213

[14] Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. In *5th International Conference on Computing Methodologies and Communication*, 843–848. https://doi.org/10.1109/ICCMC51019.2021.9418375

[15] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758. https://doi.org/10.3390/s21113758

[16] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018). AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In *Fourth International Conference on Computing Communication Control and Automation*, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697349

[17] Kiran, P. S., Abhinaya, G., Sruti, S., & Padhy, N. (2024). A machine learning-enabled system for crop recommendation. *Engineering Proceedings*, *67*(1), 51. https://doi.org/10.3390/engproc2024067051

[18] Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., & Shaurya, K. (2021). Crop recommender system using machine learning approach. In *5th International Conference on Computing Methodologies and Communication*, 1066–1071. https://doi.org/10.1109/ICCMC51019.2021.9418351

[19] Rajak, R. K., Pawar, A., Pendke, M., Shinde, P., Rathod, S., & Devare, A. (2017). Crop recommendation system to maximize crop yield using machine learning technique. *International Research Journal of Engineering and Technology*, *4*(12), 950–953.

[20] Reddy, D. J., & Kumar, M. R. (2021). Crop yield prediction using machine learning algorithm. In *5th International Conference on Intelligent Computing and Control Systems*, 1466–1470. https://doi.org/10.1109/ICICCS51141.2021.9432236

[21] Ghadge, R., Kulkarni, J., More, P., Nene, S., & Priya, R. L. (2018). Prediction of crop yield using machine learning. *International Research Journal of Engineering and Technology*, *5*(2), 2237–2239.

[22] Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018). Improving crop productivity through a crop recommendation system using ensembling technique. In *3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions*, 114–119. https://doi.org/10.1109/CSITSS.2018.8768790

[23] Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., & Aggoune, E. H. M. (2019). Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. *IEEE Access*, *7*, 129551–129583. https://doi.org/10.1109/ACCESS.2019.2932609

[24] Haq, M. A. (2024). Intelligent sustainable agricultural water practice using multi sensor spatiotemporal evolution. *Environmental Technology*, *45*(12), 2285–2298. https://doi.org/10.1080/09593330.2021.2005151

[25] Haq, M. A. (2022). Planetscope nanosatellites image classification using machine learning. *Computer Systems Science & Engineering*, *42*(3), 1031–1046. http://dx.doi.org/10.32604/csse.2022.023221

[26] Gavahi, K., Abbaszadeh, P., & Moradkhani, H. (2021). DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert Systems with Applications*, *184*, 115511. https://doi.org/10.1016/j.eswa.2021.115511

[27] Adegbenjo, A. O., & Ngadi, M. O. (2024). Handling the imbalanced problem in agri-food data analysis. *Foods*, *13*(20), 3300. https://doi.org/10.3390/foods13203300

[28] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ..., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, 265–283.

[29] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[30] Zhang, Z. (2018). Improved adam optimizer for deep neural networks. In *IEEE/ACM International Symposium on Quality of Service*, 1–2. https://doi.org/10.1109/IWQoS.2018.8624183

[31] Lydia, A. A., & Francis, F. S. (2019). Adagrad—An optimizer for stochastic gradient descent. *International Journal of Information and Computing Science*, *6*, 566–568.

[32] Taqi, A. M., Awad, A., Al-Azzo, F., & Milanova, M. (2018). The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance. In *IEEE Conference on Multimedia Information Processing and Retrieval*, 140–145. https://doi.org/10.1109/MIPR.2018.00032

[33] Vallino, E., Ridolfi, L., & Laio, F. (2020). Measuring economic water scarcity in agriculture: A cross-country empirical investigation. *Environmental Science & Policy*, *114*, 73–85. https://doi.org/10.1016/j.envsci.2020.07.017

[34] Pérez-Rodríguez, P., & Alhaj Hamoud, Y. (2024). The restoration of degraded soils: Amendments and remediation. *Frontiers in Environmental Science*, *12*, 1390795. https://doi.org/10.3389/fenvs.2024.1390795

[35] Donatelli, M., Magarey, R. D., Bregaglio, S., Willocquet, L., Whish, J. P., & Savary, S. (2017). Modelling the impacts of pests and diseases on agricultural systems. *Agricultural Systems*, *155*, 213–224. https://doi.org/10.1016/j.agsy.2017.01.019

[36] Alexander, P., Brown, C., Arneth, A., Finnigan, J., Moran, D., & Rounsevell, M. D. (2017). Losses, inefficiencies and waste in the global food system. *Agricultural Systems*, *153*, 190–200. https://doi.org/10.1016/j.agsy.2017.01.014

[37] Pathak, A., Lamichhane, S., Dhakal, M., Karki, A., Dhakal, B. K., Chetri, M., ..., & Ferdin, A. E. (2024). Human–wildlife conflict at high altitude: A case from Gaurishankar conservation area, Nepal. *Ecology and Evolution*, *14*(7), e11685. https://doi.org/10.1002/ece3.11685

[38] Alemayehu, N., & Tekalign, W. (2022). Prevalence of crop damage and crop-raiding animals in southern Ethiopia: The resolution of the conflict with the farmers. *GeoJournal*, *87*(2), 845–859. https://doi.org/10.1007/s10708-020-10287-0

[39] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

[40] Shaikh, T. A., Rasool, T., Veningston, K., & Yaseen, S. M. (2025). The role of large language models in agriculture: Harvesting the future with LLM intelligence. *Progress in Artificial Intelligence*, *14*(2), 117–164. https://doi.org/10.1007/s13748-024-00359-4

[41] Dahiphale, D., Shinde, P., Patil, K., & Dahiphale, V. (2023). *Smart farming: Crop recommendation using machine learning with challenges and future ideas*. TechRxiv. https://doi.org/10.36227/techrxiv.23504496.v1

How to Cite: Dahiphale, D., Shinde, P., Patil, K., & Dahiphale, V. (2026). Smart Farming: Crop Recommendation Using Machine Learning with Challenges and Future Ideas. *Artificial Intelligence and Applications*, *4*(3), 323–335. https://doi.org/10.47852/bonviewAIA52026214