IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 2, February 2026

DOI: 10.17148/IJARCCE.2026.15204

Smart Crop and Fertilizer Recommendation System with Plant Disease Identification

Vanchala Sutar[1], Sakshi Santosh Biranje[2], Sakshi Sanjay Barwade[3], Sneha Sanjay Dangare[4], Ankita Bharama Dhagate[5], Sudarshan Santaji Jadhav[6]

Assistant Professor, Department of Computer Science Engineering (Artificial Intelligence), DKTE Ichalkaranji[1]

Final Year B.Tech, Computer Science Engineering (Artificial Intelligence), DKTE Ichalkaranji[2-6]

Abstract: Agriculture is essential to the economies of countries like India, but it faces challenges such as changing climate, poor soil nutrients, and outbreaks of plant diseases. To tackle these problems, this study introduces a Smart Crop and Fertilizer Recommendation System with Plant Disease Identification that uses machine learning and deep learning techniques. The web-based system gives real-time suggestions for suitable crops and fertilizers based on data about soil nutrients, temperature, humidity, pH, and rainfall. We trained and evaluated seven machine learning models: Decision Tree, Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, XGBoost, and K-Nearest Neighbours (KNN) for crop recommendations. The Random Forest model achieved the highest accuracy and was chosen as the best option. Additionally, the system generates fertilizer recommendations to fix soil nutrient shortages and boost crop yield. The system also includes plant disease identification using a Convolutional Neural Network (CNN), which analyses leaf images to classify diseases accurately. This allows for early detection and timely response. This integrated solution helps users make informed decisions, supports sustainable farming, and improves productivity and food security.

Keywords: CNN, crop recommendation, machine learning, plant disease identification, random forest

I. INTRODUCTION

1.1 Background

Agriculture is a main source of income in developing countries like India. However, it faces several challenges. These include climate change, soil nutrient loss, poor fertilizer use, and frequent plant disease outbreaks. These issues hurt crop productivity and food security. Recent progress in machine learning and artificial intelligence (AI) allows for data-driven decisions in agriculture. Smart crop and fertilizer recommendation systems can analyze soil and environmental factors. Additionally, deep learning methods, especially Convolutional Neural Networks (CNNs), have proven accurate in identifying plant diseases from leaf images. Combining these technologies into one system can improve agricultural efficiency, sustainability, and crop yields.

1.2 Problem Statement

Farmers often depend on traditional knowledge and trial-and-error methods for choosing crops and applying fertilizers. These methods may not always fit the current soil conditions and weather factors. Choosing the wrong crops and using too much or too little fertilizer can result in lower yields, soil damage, and financial losses. Furthermore, delays in identifying plant diseases can worsen crop harm. There is a need for a complete system that can offer precise recommendations for crops and fertilizers. This system should also allow for early detection of plant diseases using modern machine learning and deep learning techniques.

1.3 Research Objectives

This study aims to:

- Develop a smart crop recommendation system using multiple machine learning algorithms based on soil and climatic parameters
- Identify the most accurate machine learning model for crop prediction through comparative analysis
- Design a fertilizer recommendation module to suggest appropriate fertilizers based on soil nutrient deficiencies
- Implement a plant disease identification system using CNN for accurate classification of leaf diseases
- Integrate all modules into a user-friendly web-based application for practical use by farmers

1.4 Significance

This research contributes to the field of intelligent agriculture by providing:

DOI: 10.17148/IJARCCE.2026.15204

- An integrated framework combining crop recommendation, fertilizer suggestion, and plant disease identification
- A comparative evaluation of multiple machine learning models for accurate crop prediction
- A deep learning–based approach for early detection of plant diseases using leaf images
- A practical and accessible decision-support system to promote sustainable farming, increase crop yield, and enhance food security

II. LITERATURE REVIEW

2.1 Machine Learning for Crop and Fertilizer Recommendation

Machine learning techniques have been widely applied in agriculture to improve crop selection and fertilizer management by analyzing soil nutrients and climatic parameters. Algorithms such as Decision Trees, Naive Bayes, SVM, KNN, Logistic Regression, and Random Forest have been used for predicting suitable crops and optimizing fertilizer usage. These methods enable data-driven decision-making but may suffer from limitations such as overfitting and reduced accuracy when used individually.

2.2 Random Forest and Ensemble Learning

Ensemble learning techniques, particularly Random Forest, have shown superior performance in agricultural prediction tasks. By combining multiple decision trees, Random Forest reduces variance and improves prediction accuracy. Studies report that Random Forest outperforms traditional classifiers when handling complex, high-dimensional agricultural datasets, making it suitable for crop and fertilizer recommendation systems.

2.3 CNN-Based Plant Disease Identification

Deep learning approaches, especially Convolutional Neural Networks (CNNs), have demonstrated high accuracy in plant disease identification using leaf images. CNNs automatically extract discriminative features and effectively classify diseases, enabling early detection and timely intervention to reduce crop losses.

2.4 Research Gap

Existing studies largely address crop recommendation, fertilizer management, and plant disease detection independently. Limited work has focused on integrating these components into a unified, web-based intelligent system with comparative evaluation of machine learning models. This research addresses this gap by combining crop and fertilizer recommendation with CNN-based plant disease identification.

III. METHODOLOGY

3.1 System Architecture

The proposed system consists of five core components designed to deliver accurate crop, fertilizer, and plant disease recommendations:

1. **Data Acquisition Module:** Collects soil and environmental parameters including Nitrogen (N), Phosphorous (P), Potassium (K), temperature, humidity, pH, and rainfall, along with plant leaf images for disease identification.
2. **Data Preprocessing Module:** Handles missing values, feature scaling, image resizing, normalization, and dataset splitting to ensure model robustness.
3. **Model Training Module:** Implements multiple machine learning algorithms for crop and fertilizer recommendation and a Convolutional Neural Network (CNN) for plant disease identification.
4. **Evaluation Module:** Assesses model performance using accuracy, precision, recall, F1-score, and confusion matrix analysis.
5. **Deployment Module:** Integrates trained models into a web-based interface for real-time prediction and farmer accessibility.

3.2 Mathematical Framework

3.2.1 Crop and Fertilizer Prediction Model

Crop and fertilizer recommendation is formulated as a supervised multi-class classification problem. Given an input feature vector

$$x = [N, P, K, T, H, pH, R]$$

where N, P, Krepresent soil nutrients, Ttemperature, Hhumidity, pHsoil pH, and Rrainfall, the classifier predicts the most suitable crop and corresponding fertilizer recommendation.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 15, Issue 2, February 2026
DOI: 10.17148/IJARCCE.2026.15204

3.2.2 CNN-Based Plant Disease Identification

The CNN model learns hierarchical feature representations from leaf images. Given an input image I, the CNN extracts spatial features through convolution and pooling layers, followed by fully connected layers to classify plant diseases. Dropout layers are incorporated to reduce overfitting.

3.3 Model Training and Optimization

3.3.1 Machine Learning Models

The following machine learning algorithms were trained and evaluated for crop recommendation:

- Decision Tree
- Gaussian Naive Bayes
- Support Vector Machine (SVM)
- Logistic Regression
- Random Forest
- XGBoost
- K-Nearest Neighbours (KNN)

Among these, Random Forest demonstrated superior accuracy and stability across diverse crop types.

3.3.2 CNN Training Parameters

- **Image size:** 128 × 128 pixels
- **Train–validation split:** 80:20
- **Loss function:** Categorical Cross-Entropy
- **Regularization:** Dropout layers

3.4 Experimental Design

3.4.1 Data Preprocessing

- Removal of missing or inconsistent values
- Feature normalization for numeric attributes
- Image resizing and pixel normalization
- Stratified 80–20 train-test split

3.4.2 Hyperparameters

- Random Forest estimators: Optimized experimentally
- CNN batch size: Tuned for stable convergence
- Dropout rate: Applied to prevent overfitting

3.4.3 Evaluation Metrics

Model performance was evaluated using:

- Accuracy: Overall correctness of predictions
- Precision: Correct positive predictions
- Recall: Disease or crop detection completeness
- F1-Score: Harmonic mean of precision and recall
- Confusion Matrix: Detailed classification analysis

3.5 Implementation Details

The system was implemented using Python with the following tools:

- **NumPy:** Numerical computations
- **Pandas:** Dataset handling
- **Scikit-learn:** Machine learning models and metrics
- **TensorFlow/Keras:** CNN implementation
- **OpenCV:** Image preprocessing
- **Streamlit:** Web-based user interface

IV. EXPERIMENTAL SETUP

4.1 Dataset Requirements

The proposed framework utilizes two primary datasets:

- **Crop Recommendation Dataset:**
- Target variable: Crop label (multi-class)
- Features: Soil nutrients (N, P, K), temperature, humidity, pH, and rainfall

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 2, February 2026
DOI: 10.17148/IJARCCE.2026.15204

- Dataset size: 2,200 instances
- Format: CSV
- **Plant Disease Identification Dataset:**
- Target variable: Disease class label
- Features: Leaf images
- Image resolution: 128 × 128 pixels
- Total classes: 38 (healthy and diseased categories)
- Format: Image directory structure

4.2 Validation Strategy
To ensure reliable and reproducible results, the following validation strategies were adopted:
- Stratified train–test split (80:20) to preserve class distribution
- Fixed random seed to ensure reproducibility of results
- Independent validation for crop recommendation and disease identification modules
- Model persistence to support deployment and inference

4.3 Comparative Analysis Protocol
For crop recommendation, multiple machine learning algorithms were evaluated using a consistent experimental protocol:
1. Train each model using identical preprocessing and training conditions
2. Evaluate performance on the held-out test dataset
3. Compute performance metrics including accuracy, precision, recall, and F1-score
4. Compare results across models to identify the optimal algorithm

V. RESULTS AND DISCUSSION

5.1 Results Overview
The experimental evaluation of the proposed system produced the following key outputs:
1. Model Performance Comparison: Accuracy comparison of seven machine learning algorithms for crop recommendation
2. Confusion Matrices: Detailed classification analysis for crop prediction models
3. Best Model Identification: Selection of the most accurate model for crop recommendation
4. CNN Training Analysis: Training and validation accuracy and loss curves for plant disease identification
5. Disease Prediction Results: Correct classification of plant diseases from leaf images

5.2 Analytical Framework
5.2.1 Crop Recommendation Performance Analysis
Seven machine learning algorithms, including Decision Tree, Gaussian Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest, XGBoost, and KNN, were assessed using the crop dataset. The Random Forest model achieved the highest accuracy at 99.55% and outperformed all other classifiers. Its ensemble-based learning method effectively managed feature interactions among soil nutrients and climate conditions, leading to stable and consistent predictions across various crop classes.

5.2.2 Comparative Performance Evaluation
The performance comparison revealed that ensemble and probabilistic models generally outperformed linear models. While Naive Bayes and XGBoost also demonstrated high accuracy, Random Forest exhibited superior robustness and consistency. SVM showed significantly lower performance due to sensitivity to parameter tuning and dataset characteristics. The comparative analysis confirms Random Forest as the most reliable model for crop recommendation in this framework.

5.2.3 Plant Disease Identification Analysis
The CNN-based plant disease identification model demonstrated strong learning capability, as indicated by steadily increasing training and validation accuracy and decreasing loss values. The close alignment between training and validation curves suggests good generalization and minimal overfitting. The model successfully identified plant diseases from leaf images, validating the effectiveness of CNNs for automated disease detection in real-world agricultural scenarios.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 2, February 2026

DOI: 10.17148/IJARCCE.2026.15204

5.3 Discussion

The experimental results highlight the effectiveness of integrating machine learning and deep learning techniques in agricultural decision-support systems. Random Forest proved to be the most suitable algorithm for crop recommendation due to its ability to manage complex, high-dimensional data. Similarly, the CNN model demonstrated high accuracy in identifying plant diseases through image analysis, enabling early detection and timely intervention. The combined system enhances precision agriculture by providing reliable crop recommendations and accurate disease identification, contributing to improved yield, sustainability, and reduced crop losses.

VI. PRACTICAL APPLICATIONS

6.1 Use Cases

The proposed system has several practical applications in the agricultural domain, including:

- **Crop Selection:** Assisting farmers in selecting the most suitable crops based on soil nutrients and environmental conditions
- **Fertilizer Recommendation:** Suggesting appropriate fertilizers to address soil nutrient deficiencies and improve yield
- **Plant Disease Detection:** Early identification of plant diseases using leaf image analysis to reduce crop losses
- **Decision Support for Farmers:** Providing data-driven insights to support sustainable farming practices
- **Precision Agriculture:** Enhancing productivity through optimized resource utilization and timely interventions

6.2 Deployment Considerations

The system is designed for real-world deployment through a web-based platform. Model serialization enables:

- Deployment of trained machine learning and CNN models in production environments
- Consistent preprocessing during inference to maintain prediction accuracy
- Scalability for handling multiple users and datasets
- Easy model updates and integration of improved datasets

VII. LIMITATIONS AND FUTURE WORK

7.1 Current Limitations

Despite promising results, the proposed system has certain limitations:

- Dependence on the quality and size of available datasets
- Limited consideration of region-specific market factors such as crop prices
- Sensitivity of predictions to extreme or unseen climatic conditions
- Requirement of image quality consistency for accurate disease detection

7.2 Future Enhancements

Future work may focus on the following improvements:

1. **Dataset Expansion:** Continuous updating of datasets with region-specific and seasonal data
2. **Advanced Imaging Techniques:** Integration of multispectral and hyperspectral imaging for enhanced disease detection
3. **Reinforcement Learning:** Adaptive learning for improved crop and fertilizer recommendations
4. **Real-Time Monitoring:** Integration with IoT sensors for live soil and climate data acquisition
5. **Profit-Oriented Prediction:** Incorporating economic factors such as market price and input cost
6. **Edge Computing:** Deployment on mobile or edge devices for real-time field analysis
7. **Collaborative Platforms:** Crowd-sourced data sharing among farmers, researchers, and agronomists

VIII. CONCLUSION

This research presents an integrated smart agricultural decision-support system that combines crop and fertilizer recommendation with plant disease identification using machine learning and deep learning techniques. By analyzing soil nutrients and environmental parameters, the system effectively recommends suitable crops and appropriate fertilizers, while the Convolutional Neural Network (CNN) enables accurate detection of plant diseases from leaf images. Experimental results demonstrate that the Random Forest algorithm achieves the highest accuracy for crop recommendation, and the CNN model provides reliable disease classification with strong generalization performance.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 2, February 2026
DOI: 10.17148/IJARCCE.2026.15204

The proposed framework offers a practical and scalable solution for precision agriculture by supporting informed decision-making, early disease detection, and optimized resource utilization. Its modular architecture and web-based implementation enhance accessibility and usability for farmers and agricultural stakeholders. Overall, the system contributes to sustainable farming practices, improved crop productivity, and enhanced food security, while providing a strong foundation for future advancements in intelligent agricultural systems.

REFERENCES

[1]. A. Kar, N. Nath, U. Kemprai, and Aman, “Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection,” International Journal of Communications, Network and System Sciences, vol. 17, no. 2, pp. 11–29, Feb. 2024.

[2]. Md. Manowarul Islam et al., “DeepCrop: Deep learning-based crop disease prediction with web application,” Journal of Agriculture and Food Research, vol. 14, pp. 100764–100764, Dec. 2023.

[3]. A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, “A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images,” Cluster Computing, Aug. 2022.

[4]. S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, “Crop Recommender System Using Machine Learning Approach,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021.

[5]. P. Chauhan, Hardwari Lal Mandoria, A. Negi, and R. S. Rajput, “Plant Diseases Concept in Smart Agriculture Using Deep Learning,” Advances in environmental engineering and green technologies book series, pp. 139–153, Oct. 2020.