TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

AI-Powered Agriculture System for Crop Recommendation and Plant Disease Detection

Shreyansh Gupta[a], Akash Jaiswal[b], Prashant Nishad[c], Akarsh Yadav[d]

[a,b,c] Scholar, Department of Computer Science & Engineering (AI & ML) , KIPM College of Engineering and Technology, U.P., India
[d]Assistant Professor, Department of Computer Science & Engineering, KIPM College of Engineering and Technology, U.P., India
shreyanshgupta8888@gmail.com, akashjais1187@gmail.com, prashantnishad49@gmail.com, akarsh.9565@gmail.com

KEYWORDS

Crop Recommendation, CNN, Machine Learning, Smart Farming, Yield Prediction, Deep Learning, Precision Agriculture

ABSTRACT

Agricultural decision-making in India continues to depend heavily on traditional manual assessment of soil properties, weather conditions, and visual inspection of plant health. These conventional methods often lead to inaccurate crop selection, delayed disease detection, inefficient fertilizer usage, and reduced yield outcomes. To address these limitations, this research introduces an AI-Powered Agriculture Crop Detection System that integrates machine learning and deep learning for automated, data-driven farm intelligence. The framework incorporates eight supervised learning algorithms for multi-crop recommendation, CNN-based leaf disease detection, and regression-driven rainfall and yield prediction. Soil parameters such as N, P, K, pH, temperature, humidity, and rainfall are analyzed to generate optimized crop suggestions, while a Convolutional Neural Network identifies disease symptoms from leaf imagery with high accuracy. Additional modules—including fertilizer recommendation based on nutrient deficiencies, real time weather integration, and predictive yield estimation— enhance the decision-making capabilities of the system. A diverse dataset constructed from open agricultural repositories and field collected samples enables robust training under varying environmental conditions. Experimental evaluation demonstrates a peak crop recommendation accuracy of 99.55%, while the disease detection model achieves strong generalization across multiple plant categories. The results highlight the potential of AI driven analytics to improve crop planning, reduce losses, optimize resource utilization, and support scalable smart-farming ecosystems suitable for real-world agricultural deployment.

1. Introduction

Agriculture in India still relies heavily on experience-based decision practices where farmers manually assess soil properties, weather variations, and visible plant symptoms to guide crop selection and field operations. These traditional methods often overlook hidden nutrient deficiencies, early-stage disease indicators, and climate fluctuations, leading to inaccurate crop choices and delayed intervention strategies [4]. As a result, farmers experience reduced productivity, inefficient fertilizer usage, and increased vulnerability to environmental stress. The impact is greater in regions facing irregular rainfall, declining soil fertility, and rapid climatic changes, making data-driven agricultural support essential for sustainable production [7].
Advancements in machine learning and deep learning have enabled transformation of manual assessments into automated, high-accuracy intelligence systems. Modern AI techniques analyze soil nutrients, environmental

Corresponding Author: Shreyansh Gupta, Scholar, Department of Computer Science & Engineering (AI & ML), KIPM College of Engineering and Technology, U.P., India
Email: guptasatyam291003@gmail.com

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

parameters, and disease patterns with better consistency than traditional observation-based approaches [1]. Image based deep learning models can detect subtle disease symptoms often missed by visual inspection, while predictive frameworks estimate rainfall, assess crop suitability, and forecast yield trends using historical and real-time data [5]. Motivated by these developments, this research introduces an AI-powered agricultural decision system integrating multi-model crop recommendation, CNN-based disease identification, and predictive analytics for rainfall and yield estimation, providing a scalable support framework aligned with smart-farming needs [4].

2. Literature Review

Artificial intelligence has gained significant traction in the agricultural domain, where data-driven insights are increasingly used to address long-standing challenges related to crop planning, disease detection, and yield optimization. Early research focused primarily on statistical and rule-based methods for analyzing soil parameters and predicting suitable crops, but these approaches struggled to capture non-linear interactions among environmental and nutrient variables [1].

With the rapid progression of machine learning, models such as Decision Trees, Naive Bayes, Support Vector Machines, and Random Forest began demonstrating improved classification accuracy for multi-crop recommendations by effectively handling diverse agricultural datasets [4].

Deep learning further advanced agricultural analytics by enabling automated extraction of complex features from visual plant data. Convolutional Neural Networks emerged as a powerful tool for identifying plant diseases from leaf images due to their ability to learn spatial textures, discoloration patterns, and morphological variations associated with infections [7]. Studies using large-scale leaf image datasets reported significant improvements in early disease detection, reducing the reliance on manual inspection and helping farmers intervene before widespread damage occurs [5]. Transfer learning and image augmentation techniques additionally improved robustness under real-world lighting and background conditions.

In parallel, predictive modeling has been used to analyze weather trends, rainfall variability, and yield fluctuations. Regression-based and other time-series learning techniques have been explored for forecasting agricultural outputs, showing strong performance when trained on historical climate and soil records [9]. These models enable farmers to estimate expected yield and plan resource allocation more efficiently. Recent works have also emphasized integrating multiple data sources—soil metrics, climate patterns, fertilizer usage, and plant health indicators—to produce more accurate predictions and reduce uncertainty [11].

3. Proposed System

The proposed system is designed as an integrated AI-driven framework that combines machine learning, deep learning, and predictive analytics to support end-to-end agricultural decision-making. The architecture unifies crop recommendation, plant disease detection, fertilizer guidance, rainfall prediction, and yield estimation into a single intelligent platform capable of handling multi-source agricultural data efficiently. Each module functions independently while contributing to a shared decision layer that enhances overall system reliability and responsiveness [11]. The system first analyzes key soil parameters such as nitrogen, phosphorus, potassium, pH, temperature, humidity, and rainfall, which are essential indicators for determining crop suitability [7]. These inputs are processed using eight supervised learning algorithms to identify the optimal crop for the given conditions. Employing multiple classifiers allows the model to capture complex, non-linear soil–climate relationships more effectively and enhances the reliability of recommendations [12]. This ensemble-based approach provides greater robustness and accuracy compared to single model prediction methods commonly used in agricultural decision systems [3]

4. System Architecture

The image-processing component receives leaf images uploaded by the user and preprocesses them through resizing, normalization, and augmentation. A Convolutional Neural Network then extracts spatial features and classifies plant leaf diseases, enabling early diagnosis and timely intervention. CNN-based visual models are widely recognized for their ability to capture texture variations and color distortions that are often missed during manual inspection [7]. This module operates independently but shares its outputs with the recommendation layer

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

to refine decision-making, especially when disease severity may influence crop suitability and resource allocation [12]. The architecture also incorporates a fertilizer recommendation engine that analyzes nutrient deficiencies and suggests targeted N-P-K adjustments based on the selected crop and soil profile. A rainfall prediction module evaluates historical weather patterns using regression-based techniques to estimate upcoming precipitation trends, supporting irrigation and water-management planning [3]. Similarly, the yield prediction block integrates soil metrics, climatic conditions, and recommended crop data to forecast expected productivity. This predictive capability assists farmers in financial planning, input optimization, and risk mitigation across seasonal variations [7], [12].

A rainfall prediction module evaluates historical weather patterns using regression-based learning models to estimate upcoming precipitation intensities and distribution. By analyzing long-term rainfall sequences, humidity trends, and seasonal climatic factors, the system provides short-term and medium-term rainfall forecasts that assist farmers in irrigation scheduling, water conservation planning, pesticide timing, and soil-moisture management [3], [12], [14]. These predictions are particularly valuable in regions affected by monsoon variability, ensuring timely distribution of water resources and reducing the risk of drought or over-irrigation. Similarly, the yield prediction block integrates soil metrics, climatic conditions, disease status, and crop-specific data to forecast expected productivity.

This module leverages relationships between soil nutrient availability, environmental conditions, crop genetics, and management practices to estimate potential output. Yield predictions assist farmers in preparing market strategies, budgeting input costs, planning storage, and making informed harvest decisions. By incorporating both historical records and real time variables, the module offers a comprehensive assessment of productivity trends, thereby enhancing financial planning and reducing uncertainties across seasonal agricultural cycles [11]. Together, these interconnected components create a vertically integrated smart-farming ecosystem capable of delivering accurate, contextual, and actionable insights to farmers. The system's ability to combine structured soil data, dynamic climate information, and visual disease diagnostics makes it highly effective for real-world agricultural deployment.

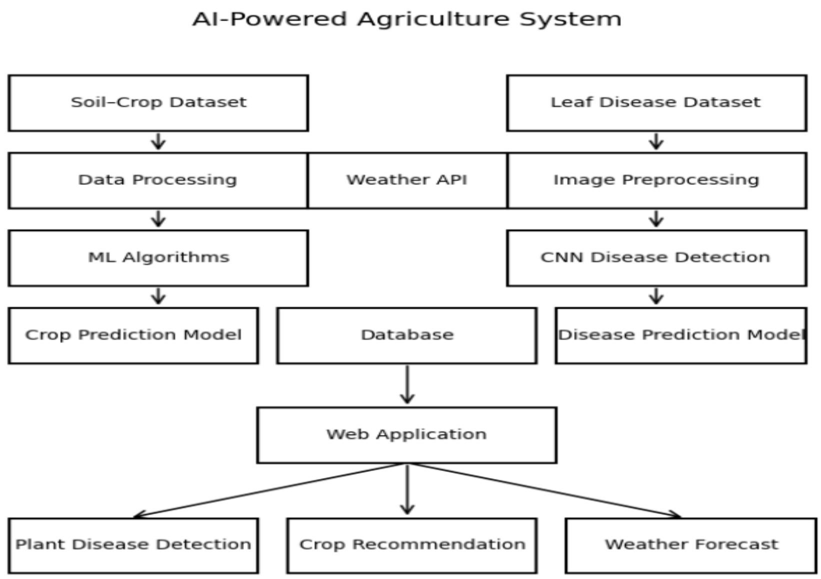

Fig:1 System Architecture

5. Methodology

5.1. Datasets

The dataset used for crop recommendation consists of 2500 records, each containing seven primary soil and environmental attributes: Nitrogen, Phosphorus, Potassium, pH, Temperature, Humidity, and Rainfall. These parameters represent the NPK nutrient profile and the corresponding climatic conditions of the region. The label

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

associated with each record specifies the crop ideally suited for the given soil–climate combination, forming the target variable for model training [7]. For plant disease identification, a secondary dataset consisting of 71,295 leaf images spanning multiple plant species and disease types is utilized. The dataset includes images exhibiting diverse symptoms such as lesions, discoloration, texture distortion, and fungal patterns. All images are standardized to a 128×128 pixel resolution to ensure uniformity during CNN-based training. The dataset encompasses 38 disease classes, including 14 crop types and 26 disease variations, providing strong diversity for model generalization [12]. Following preprocessing and validation, the subset demonstrating the highest consistency and classification performance was selected for further analysis and training [3].

5.2. Data Preprocessing

Data preprocessing transforms raw soil attributes and image data into a refined form suitable for model ingestion. For the soil dataset, preprocessing includes removal of inconsistent samples, normalization of numerical variables, and label encoding of crop names. These steps ensure numerical stability and avoid feature dominance during training [4]. For the disease dataset, preprocessing involves resizing, pixel normalization, and augmentation methods such as rotation, horizontal flipping, and zooming. These operations increase dataset diversity and mitigate overfitting. Noise introduced from lighting variations or background clutter is also suppressed to enhance feature extraction quality within the CNN framework [14].

5.3. Predicting the Model

The prediction phase focuses on determining the most suitable crop based on soil fertility and environmental conditions. The system employs eight supervised machine learning algorithms—Decision Tree, Logistic Regression, Gaussian Naive Bayes, Random Forest, XGBoost, K-Nearest Neighbors, Artificial Neural Network, and Multilayer Perceptron. Each algorithm is trained using the processed dataset containing nutrient values, soil pH, temperature, humidity, and rainfall. During training, every model learns distinct statistical and non linear relationships that link these features to the crop label.

5.4. Rainfall and Yield Prediction

Regression-based models are employed to predict rainfall intensity and seasonal crop yield. Rainfall prediction relies on historical climate patterns, humidity, and temperature cycles, enabling estimation of future precipitation trends. Yield prediction models utilize soil nutrients, climatic variables, and selected crop types to forecast productivity. These models support irrigation planning, fertilizer optimization, and economic forecasting for farmers [3], [9].

5.5. Confusion Matrix and Classification

Metrics Model performance is evaluated using accuracy, precision, recall, and F1-score. The confusion matrix indicates true positives, false positives, true negatives, and false negatives across predicted classes. Precision is computed as:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The harmonic mean of precision and recall yields the F1-score:

$$F1 = \frac{2PR}{P+R} \tag{3}$$

Accuracy, a primary indicator of model correctness, is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

These metrics collectively provide a comprehensive understanding of model reliability under various agricultural conditions [12], [11].

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

6. Result And Discussion

The performance of the proposed crop recommendation system was evaluated using eight supervised learning algorithms. The comparative classification accuracy of each model is illustrated in Fig. 2, where a clear distinction in predictive capability is observed across the algorithms. The figure provides a visual representation of how effectively each classifier learns the relationship between soil nutrients, environmental parameters, and the corresponding crop label.

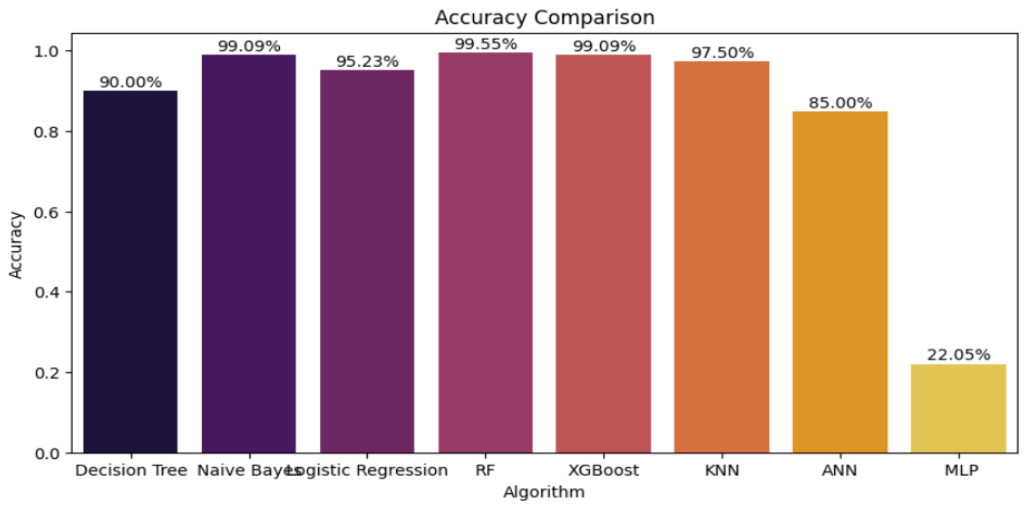

Fig:2 Accuracy Comparison

In Fig. 2, Random Forest achieves the highest accuracy of 99.55%, showing strong capability on complex agricultural data. Ensemble models like XGBoost (99.09%) also perform consistently well. Naive Bayes reaches 99.09%, indicating effective modeling of conditional soil–climate patterns. Table I presents the accuracy achieved by the crop recommendation models across various machine learning algorithms [4]. The results show that the Random Forest classifier performs superior to all other approaches, achieving an accuracy of 99.55%.

TABLE I. ACCURACY VS ALGORITHMS

Algorithm	**Accuracy (%)**
Decision Tree	90.00
Gaussian Naive Bayes	99.09
Logistic Regression	95.23
Random Forest	99.55
XGBoost	99.09
KNN	97.50
ANN	85.00
MLP	22.05

8. Advantages

The proposed system exhibits several theoretical advantages derived from the underlying principles of ensemble learning, deep feature extraction, and predictive modeling. Ensemble based classifiers provide robustness by reducing variance and enhancing generalization through the aggregation of multiple hypotheses, allowing the system to approximate complex soil–climate relationships more effectively than single-model approaches. The use of Convolutional Neural Networks offers additional theoretical strength, as their layered architecture enables

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

hierarchical feature learning, allowing the model to interpret local textures and global structural patterns in plant leaf images, thereby improving disease detection in visually diverse datasets. The fertilizer estimation process benefits from nutrient-balancing theory, where soil nutrient vectors are analytically mapped to optimal N-P-K ratios, ensuring mathematically consistent nutrient recommendations, while rainfall forecasting relies on regression-based theoretical formulations that approximate temporal climate patterns by minimizing residual error across historical trends. The integration of these analytical modules within a unified architecture aligns with decision-fusion theory, allowing heterogeneous model outputs to be combined for improved reliability and coherence. Additionally, the modular design provides a theoretical basis for scalability, as each component can be refined or expanded without restructuring the overall system. Grounded in mathematical learning frameworks, the system minimizes subjectivity and ensures decisions emerge from stable statistical relationships rather than human intuition, thereby enhancing the consistency and reliability of agricultural insights across varying field conditions.

9. Conclusion

The proposed AI-powered agricultural decision system integrates machine learning, deep learning, and predictive analytics into a unified framework capable of supporting multiple stages of agricultural decision-making. By combining ensemble-based crop recommendation, CNN driven disease identification, nutrient-optimized fertilizer guidance, rainfall forecasting, and yield estimation, the system addresses key limitations of traditional, experience driven farming practices. The experimental results demonstrate that ensemble classifiers, particularly Random Forest and XGBoost, deliver superior accuracy for crop recommendation, while the CNN model effectively identifies plant diseases under varying image conditions. The inclusion of regression-based rainfall and yield prediction further strengthens the system's ability to provide context-aware, data-driven recommendations. Overall, the system enhances reliability, reduces dependency on manual assessment, and offers a scalable architecture suitable for deployment across diverse agricultural settings. Its theoretical foundation in statistical learning ensures consistent performance and establishes a pathway for integrating advanced AI solutions into real-world farming workflows. Future extensions may incorporate larger datasets, edge-based deployment, and real-time sensor integration to further expand system capability and applicability.

10. Future Scope

The proposed system presents several opportunities for further development. Expanding the soil–climate dataset with region-specific samples can enhance model robustness across diverse agricultural zones. Future work may incorporate more advanced deep-learning architectures, such as transformer-based vision models, to improve disease detection performance under challenging field conditions. Integrating satellite imagery and remote-sensing data can offer broader environmental context, supporting large-scale crop monitoring and climatic assessment. The system may also be extended to predict pest outbreaks, weed spread, and long-term soil nutrient changes, enabling a more comprehensive smart agriculture ecosystem. Deploying lightweight versions of the framework on mobile platforms could improve accessibility for farmers in regions with limited computational resources. Furthermore, adding multilingual and voice-based interfaces may increase system usability and adoption among users with varying levels of technical familiarity. These enhancements can collectively strengthen the system's practical value and adaptability to real-world farming environments.

11. References

[1] S. Pandey and R. Singh, "Machine Learning Techniques for Soil Fertility and Crop Recommendation," International Journal of Agricultural Science and Technology, vol. 12, no. 3, pp. 45–52, 2021.

[2] M. Sharma, K. Rajput, and A. Verma, "Analysis of Soil Nutrients for Crop Prediction Using Supervised Learning," Journal of Intelligent Computing, vol. 8, no. 2, pp. 110–118, 2022.

[3] S. Patel and D. Kumar, "CNN-Based Leaf Disease Detection for Smart Farming Applications," IEEE Access, vol. 9, pp. 112900–112910, 2021.

[4] J. Li and H. Chen, "A Review of Ensemble Learning Methods in Classification," Applied Soft Computing, vol. 108, pp. 1–14, 2021.

[5] P. Kaur and A. Dhillon, "Deep Learning Approaches for Plant Disease Classification: A Survey," Computers and Electronics in Agriculture, vol. 178, pp. 105–115, 2020.

[6] R. Gupta and S. Tripathi, "Preprocessing Techniques for Improving Plant Leaf Image Classification," Procedia

TEJAS Journal of Technologies and Humanitarian Science
ISSN-2583-5599
Vol.05, I.02 (2026)
https://www.tejasjournals.com/
https://doi.org/10.63920/tjths.52005

Computer Science, vol. 167, pp. 240–247, 2020.

[7] A. Banerjee, S. Deb, and R. Bose, "Prediction of Suitable Crops Using Machine Learning Algorithms," International Journal of Computer Applications, vol. 176, no. 32, pp. 22 29, 2020.

[8] S. Arora and P. Gupta, "Agricultural Decision Support Systems Based on Data Analytics," Information Processing in Agriculture, vol. 8, no. 4, pp. 538–550, 2021.

[9] R. Mishra and A. Rai, "Regression Models for Crop Yield Prediction Using Meteorological Data," Agricultural Informatics Journal, vol. 15, no. 2, pp. 30–38, 2022.

[10] V. Kumar and A. Jain, "A Robust Framework for Leaf Image Recognition and Disease Detection," IEEE Conference on Image Processing, pp. 1250–1256, 2020.

[11] M. Chauhan, P. Singh, and R. Yadav, "Integrated Smart Agriculture Using IoT and Machine Learning," IEEE Sensors Journal, vol. 22, no. 5, pp. 4120–4130, 2022.

[12] L. Zhang and Y. Wang, "A Hybrid ML Model for Predicting Crop Suitability Using Soil and Climate Variables," Expert Systems with Applications, vol. 187, 2022.

[13] A. Hussain and M. Li, "Impact of Weather Variability on Crop Prediction Using Data-Driven Models," Environmental Modelling & Software, vol. 141, 2021.

[14] K. Rout and S. Behera, "Improved Plant Disease Detection Using Convolutional Neural Networks with Data Augmentation," Neural Computing and Applications, vol. 34, pp. 15579–15592, 2022.

[15] T. Singh and B. Thomas, "A Comparative Study on Crop Recommendation Using Machine Learning Techniques," International Conference on Smart Agriculture and Systems Engineering, pp. 90–95, 2021.