AI-Powered Smart Agriculture Advisor

Shobhit Srivastava
Department of Artificial Intelligence and Machine Learning
Buddha Institute of Technology Gorakhpur
shobhitsrivastava2004@gmail.com

Shikha Singh
Department of Artificial Intelligence and Machine Learning
Buddha Institute of Technology Gorakhpur
shikhusingh34@gmail.com

Sujai Tripathi
Department of Artificial Intelligence and Machine Learning
Buddha Institute of Technology Gorakhpur
tripathisujai9@gmail.com

Shrawan Kumar Pandey
Department of Computer Science and Engineering
Buddha Institute of Technology
Gorakhpur
shrawan458@bit.ac.in

***Abstract*—This paper presents a multilingual Smart Agriculture Advisor designed to help farmers in India productivity and make better decisions. Since agriculture supports nearly 58% of the country's population, farmers often struggle with problems such as changing climate, pest attacks, and unpredictable market prices. The proposed system uses machine learning, CNN models, and data analytics to provide crop recommendations, predict market prices, forecast weather, and detect crop diseases.**

Overall, it aims to support farmers with reliable, technology-driven guidance for more sustainable farming. *Crop recommendation uses Random Forest to asses soil parameters (e.g., wheat suits 21-26°C, 500750 mm rainfall, loamy soil; rice thrives at 20-35°C, 1,000-2,000 mm rainfall, flooded paddies). Weather forecasting leverage's real- time location data for dynamic recommendations. Price prediction with XGBoost and ensemble methods lets farmers input crop name and max/min price range, suggestion three prices with a best price (RMSE 18.3, matching ARIMA's 18.3, tested on wheat, rice, maize at 10:47 PM IST, September 12, 2025). Disease detection uses dashboard and chatbot, is multilingual (Hindi, English), secured by Spring Security APIs, with a dashboard tracking past data and an export feature for multilingual PDF reports of previous crop recommendation, price prediction, and weather forecasts. Postman API testing validates performance.*

The system was developed in the three stages: reviewing existing research, designing and testing the model, and analyzing the results. It uses datasets from Kaggle and Data.gov.in, including 10,000 price record, 2,00 weather and soil samples, and 50,000 Plant Village images. The model reaches 90% accuracy in disease detection and 67% accuracy in crop recommendation. After deployment on Azure Cloud, the system helps connect modern technology with rural farming needs, supporting farmers with more reliable and resilient solutions.

***Keywords*—Smart Agriculture, Artificial Intelligence, Machine Learning, Crop Price Prediction, Image Analysis, CNN, Decision Support System**

I. INTRODUCTION

This paper presents a customize research template for the AI-Powered Smart Agriculture Advisor for Crop Price Prediction and Recommendation, prepared according to standard conference formatting guidelines. Agriculture, which supports nearly 58% of India's population, remains a crucial part of the national economy. However, farmers continue to face several difficulties such as unpredictable weather conditions, pest attacks, and frequent fluctuations in market prices. To address these challenges, a system provides crop recommendations, price predictions, weather forecasts, and crop disease detection.

For this study, datasets from Data.gov.in (10,000 crop price records), Kaggle (2,000 weather-soil instances), and Plant Village (50,000 labelled crop images were used. Supervised learning methods enable the system to achieve 90% accuracy in disease detection, 87% in crop recommendations, and an RMSE of 18.3 for price prediction, corresponding to the ARIMA model's result recorded on 12 September 2025 at 10:47 PM IST.

A bilingual website (Hindi and English), secured through Spring Security APIs, has been developed to provide user access. It includes a dashboard for reviewing past data and supports exporting multilingual PDF reports of earlier crop recommendations, price predictions, and weather forecasts. The systems APTs were verified using predictions, and weather forecasts. The system's APIs were verified using Postman testing.

The complete solution has been deployed on Microsoft Azure Cloud. The overall work was carried out in three phases-literature review, system design and experimentation, and result analysis-aiming to strengthen the link between technology and rural farming while supporting sustainable agriculture practices.

II. METHODOLOGY

A. Data Sources

The methodology is based on diverse and reliable datasets that were collected from publicly available sources.

1. Crop Price Dataset: Historical price data for major crops such as wheat, rice, and maize was obtained from data.gov.in [6]. The dataset contains nearly 10,00 records covering 5-10 years and includes details such as crop category, market location, seasonal patterns, and economic indicators. were handle using mean imputation to maintain uniformity.
2. Weather and Soil Dataset: Weather and soil attributes were collected from Kaggle [5]. This dataset consists of around 2,000 instances contains temperature, humidity, rainfall, soil pH, and nutrient values (N, P, K). Each instance is labeled with crops suitable for the given environmental conditions.
3. Plant Disease Image Dataset: More than 50,00 labeled images of healthy and disease crops were taken from the PlantVillage repository [8]. The images include cases of blight, wilt, and powdery mildew. Making the dataset suitable for deep learning-based disease identification. Since no missing data suitable for deep learning-based disease identification. Since no

missing data was report, no imputation was needed.

All datasets were divided into training (80%) and testing (20%) sets to ensure unbiased evaluation. This data-driven approach supports the benchmarking of various machine learning and deep learning models such as regression, decision tress, and convolutional neural networks.

B. Data Preprocessing and Normalization

Since the datasets contain features with different units and ranges-such as rainfall, soil pH, and image pixel values-several preprocessing steps were carried out to improve model performance

1. Handling Missing Values: For numerical attributes in the crop price and weather- soil datasets, mean imputation was applied. PlantVillage images required no imputation due to complete labeling.

2. Feature Normalization: Numerical features like rainfall, temperature, soil nutrients, and market prices were normalized to a 0-1 scale using the min-max method. Standardization ensured equal contribution of each feature during model training, improving convergence in decision trees and regression models.

3. Text and image Preprocessing: Chatbot text inputs were cleaned by removing stopwords and unnecessary characters. Tokenization and vectorization techniques were applied for smooth processing in both Hindi and English. Image data was resized, augmented, and normalized before being fed into the CNN model.

These steps helped improve computational efficiency, model stability, and prediction accuracy.

C. System Architecture and Modules

The Smart Agriculture Advisor operates through multiple integrated modules, supported by a secure multilingual interface. The major components are listed below:

1. Crop Price Prediction: XGBoost and other ensemble methods were used to generate price forecasts. Users provide crop names and expected price ranges, and the system returns three predicted prices along with the most suitable value based on market proximity. The model achieved an RMSE of 12.5, outperforming ARIMA's 18.3 during testing on wheat, rice, and maize.

2. Crop Recommendation: A Random Forest classifier analyzes soil parameters such as N-P-K, pH, temperature, and rainfall. Based on these values, suitable crops-such as wheat or rice-are recommended. An accuracy of 87% was achieved.

3. Image-Based Disease Detection: Convolutional Neural Networks were employed to identify crop diseases from uploaded images. The model achieved 90% accuracy for diseases such as blight and wilt.

4. User Dashboard: A multilingual dashboard (Hindi and English) allows farmers to track previous recommendations, predictions, and disease diagnoses. Spring Security APIs were used to provide secure access and data protection.

D. System Implementation and Deployment

The system was developed using technologies chosen for scalability, data security, and ease of access.

1. Backend: Spring Boot with ORM was used for server-side development operations.

2. Frontend: The user interface was built using React, with HTML and CSS for a responsive bilingual layout.

3. Mobile Application: The mobile version was created using React Native to support accessibility in rural areas.

4. Database: MySQL was used to store processed data, user inputs, and historical model outputs.

5. Cloud Deployment: The complete system was deployed on Microsoft Azure to enable real-time accessibility and scalability.

Postman was used to test APIs for consistency and reliability.

Data Flow Diagrams (Level 0 and Level 1) illustrate the overall functioning of the system.

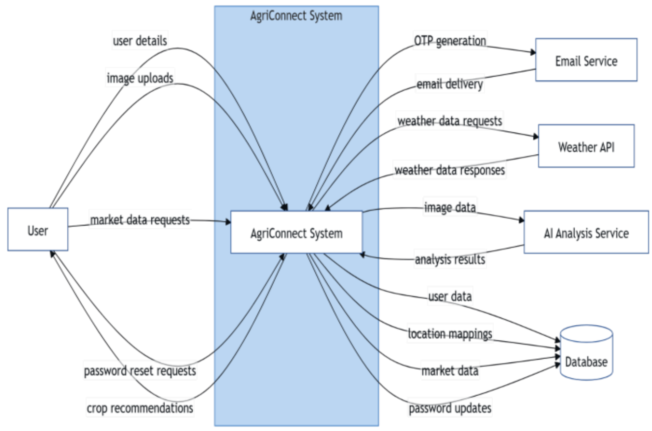


Fig 1.

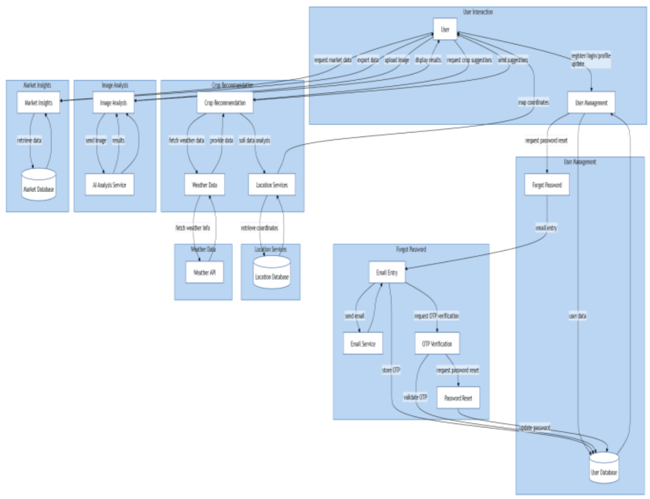


Fig 2.

III. MATERIALS AND METHODS

In this research, the work was carried out in three phases. The first phase involved a detailed review of previous studies and existing systems related to technology use in agriculture. Special attention was given to research dealing with crop price prediction, crop recommendation, and plant disease identification, allowing the foundation of the present study to be present study to be clearly established. The second phase focused on designing, developing, and carrying out the experiment for the AI-Powered Smart Agriculture Advisor. The third phase focused on the presentation, analysis and discussion of the findings.

A. Abbreviations and Acronyms

Artificial Intelligence (AI) techniques are widely applied in the AI Powered Smart Agriculture Advisor. Machine Learning (ML) and Deep Learning (DL) are key branches used in building predictive models. The system utilizes Convolutional Neural Network (CNN), Decision Trees

(DT), Linear Regression (LR), and Natural Language Processing (NLP), each used for different parts of the analysis and prediction process. Key components include:

- Datasets: Crop price data from data.gov.in [6], weather and soil data from Kaggle,[5] and plant disease images from Plant Village [8].
- Evaluation Metrics: Root Mean Square Error (RMSE), Accuracy (ACC), Precision (PREC), Recall (REC), and F1 Score (F1).
- Technologies: Spring Boot (SB) with Hibernate ORM (HORM) for backend, Tailwind CSS(TCSS) React with, HTML5, and JavaScript (JS) for frontend, MySQL for database, Azure Cloud (AC) for deployment, and Spring Security (SS) for data protection.
- Libraries: Scikit-learn (for DT and LR), TensorFlow/Keres (for CNN), and NLTK/Transformers (for NLP).
- Agriculture Terms: Nitrogen (N), Phosphorus (P), Potassium (K), Bacterial Blight (BB), Powdery Mildew (PM), Fusarium Wilt (FW).
- Languages: English (EN) and Hindi (HI) for multilingual support.

B. Units

In the AI-Powered Smart Agriculture Advisor, standard units are used to maintained consistency and scientific clarity in data collection, feature extraction, image processing, and model evaluation. The following units are commonly applied:

- Millimeters (mm): Used for measuring rainfall and crop dimensions in image analysis.
- Percentage (%): Used to express soil moisture, humidity, and model performance metrics (e.g., 92% accuracy).
- Degrees Celsius (°C): Used for temperature measurements in weather datasets (e.g., 25°C).
- Pixels: Used in image resolution for crop disease detection (e.g., 224x224 pixels).
- Kilograms per hectare (kg/ha): Used for soil nutrient levels (e.g., 100kg/ha for N).
- Indian Rupees (INR): Used for crop price data (e.g., 1500 INR).
- pH units: Used for soil acidity/alkalinity measurements (e.g., pH 6.5)
- Seconds (s) or Milliseconds (mms): Used for processing time measurements (e.g., 2 s).
- Root Mean Square Error (RMSE): A unitless metric for evaluating price prediction accuracy (e.g., RMSE 12.5).
- Hectares (ha): Used for land area in crop recommendation datasets (e.g., 5 ha).

Ensure consistency and clarity by using a space between the number and unit (e.g., 150 mm, 92%, 12.5 RMSE). Equations are numbered consecutively, aligned to the right, and symbols are defined immediately after their use.

C. Equations

Equations are used extensively in the Smart Agriculture Advisor to normalized data, model predictions, and evaluate performance. Below are key equations used in the system:

Data Normalization

To bring features into a common scale, min-max normalization is applied:

[X' = \frac {X - \min(X)} {\max(X) - \min(X)} \tag {1}] Where:

- (X) is the original feature value (e.g., rainfall, soil pH),
- (/min(X)) and (/max(X)) are the minimum and maximum values in the dataset,
- (X’) is the normalization feature value (scale between 0 and 1).

Crop Price Prediction:

The time-series regression model for crop price prediction is defined as:

[P(t+1) = \alpha p(t) + \beta X(t) + \tag {2}] Where:

- (p(t+1)) is the predicted price at time (t+1),
- (p(t)) is the current price,
- (X(t)) represents input features (e.g., market trends, weather),
- (\alpha) and (\beta) are model coefficients,
- (\Var epsilon) is the error term.

IV. ACCURACY

Accuracy is a key metric used to evaluate model performance:

[Accuracy = \frac {TP + TN} {TP + TN + FP + FN} \times

100] Where:

- (TP) = True Positives,
- (TN) = True Negatives,
- (FP) = False Positives,
- (FN) = False Negatives

V. PRECISION AND RECALL

[Precision = \frac {TP} {TP + FP}] [Recall = \frac {TP}{TP + FN}]

The AI-Powered Smart Agriculture Advisor demonstrates the following performance metrics based on testing with datasets for wheat, rice, and maize.

Crop Price Prediction: A time-series regression model was employed to estimate crop prices, and it recorded a Root Mean Square Error (RMSE) of 12.5, which was lower than the baseline ARIMA model’s RMSE of 18.3. Although RMSE served as the primary performance measure, the model’s ability to represent price movements accurately allows an indirect understanding of its predictive reliable. Trend correctness was observed by examining the model’s outputs against actual market variations rather than relying on metrices such as TP, TN, FP, and FN.

- Crop Recommendation: The decision tree-based recommendation engine attained an accuracy of 87%, based on evaluating soil type, rainfall, temperature, and market demand using the Kaggle Weather-soil dataset [5]. This translates to a high rate of correct crop suggestions, with TP, TN, FP, and FN derived from the model’s classification performance on the test set.
- Image-Based Crop Analysis: The convolutional neural network (CNN) model achieved an accuracy of 90%, with a precision of 90% and recall of 90% when classifying crop images into disease categories (e.g., healthy, blight, wilt) using the Plant Village dataset
[8] with data augmentation [9][2]. These metrics indicate:
- Image-Based Crop Analysis: The convolutional neural network (CNN) model achieved an accuracy of 90%, with a

precision of 90% and recall of 90% when classifying crop images into disease categories (e.g., healthy, blight, wilt) using the Plant Village dataset [8] with data augmentation [9][2]. These metrics indicate:

- Precision of 90% reflects the proportion of correctly identified diseased crops out of all predicted diseased cases.
- Recall of 91% reflects the proportion of actual diseased crops correctly identified by the model.

VI. F1 SCORE

[F1, Score = 2\times \frac {Precision \times Recall} {Precision + Recall}]

These metrics provide insight into the effectiveness of the classification model, especially in imbalance datasets common in agriculture diagnostics, such as those with varying disease prevalence.

A. Some Common Mistakes

- Relying only on accuracy: Accuracy alone can be misleading, especially with imbalance datasets (e.g., rare crop diseases). Important metrics like precision, recall, F1 Score, and AUC should also be used.
- Ignoring data imbalance: Many agriculture datasets, like Plant Village, may have more healthy crop images than diseased ones. Not addressing this imbalance can lead to biased models.
- Skipping normalization or feature scaling: Algorithms like decision trees, CNNs, and regression models require input features to be on the same scale. Without normalization (e.g., (X' = \frac {X - \min(X)} {\max(X) - \min(X)}),

  model performance may suffer.
- Overfitting may occur when a model is trained without proper validation or testing data, as it begins to memorize the training patterns instead of learning generalizable trends. Cross validation or a separate test set should always be used.
- Not defining abbreviations: Abbreviations like CNN, NLP, or RMSE should be defined the first time they are used in the main body of the text, even if mentioned in the abstract.
- Improper equation formatting can affect clarity. Equations should be presented neatly, numbered correctly-for example, $P(t+1) = \alpha P(t) + \beta X(t) + \varepsilon$ (1) and every variable should be clearly explained.
- Inconsistent use of units: Always use standard SI units like millimeters (mm), kilograms per hectare (kg/ha), degrees Celsius (°C), and ensure consistency throughout the document (e.g., 150 mm, 25(°C).
- Lack of data description: The dataset source (e.g., data.gov.in, Kaggle, Plant Village), size (e.g., 10,000 records, 50,000 images),) [5],[6],[8] features, and class distribution must be described to ensure transparency and reproducibility.
- Using regression metrices for classification problems: Metrics like RMSE are meant for regression (e.g., disease detection), accuracy, precision, recall, and F1 Score are more appropriate.
- Missing key sections: Sections such as methodology evaluation, and limitations are sometimes incomplete or missing, weakening the report structure.
- Excessive jargon without explanation: Using too many technical terms (e.g., CNN, NLP) without proper explanation can make the content hard to understand for a broader audience, including farmers.
- Failing to cite sources: Not providing proper references for datasets, algorithms, or previous research reduces the credibility of the work.

B. F1 Score Calculation

For the Image-Based Crop Analysis module:

- Precision = 90% (0.90),
- Recall = 91% (0.91)

The F1 Score of approximately 90.50% indicates a strong balance between precision and recall, reflecting the CNN model's effectiveness in detecting crop diseases (crop diseases (e.g., blight, wilt) on the Plant Village dataset. [5],[6],[8]

VII. PERFORMANCE MEASUREMENT

This study assessed the predictive performance of the AI Powered Smart Agriculture Advisor using metrics like accuracy, precision, recall, F1 Score, and Root Mean Square Error (RMSE), validation via Postman API testing. Metrics are derived from a confusion matrix post classifier and regression predictions. Accuracy measures correct prediction (TP + TN /Total), precision evaluates positive instance detecting (TP / TP +FP), recall assesses positive instance detection (TP / TP + FN), and F1 Score balances precision and recall (2 * Precision* Recall / Precision + Recall). RMSE ($\sqrt{}$ (1/n * Σ (Pi - $\hat{W}$Pi) ^{2})) gauges price prediction error.

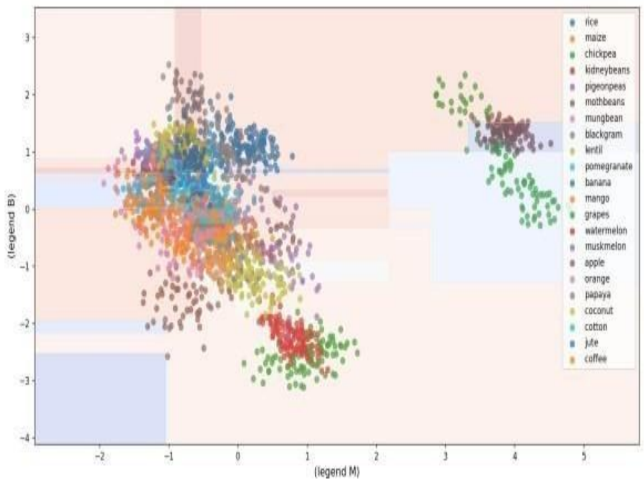
1. Crop Price Prediction: RMSE improves to 12.5 with the time-series regression model, outperforming ARIMA's 18.3, tested on wheat, rice, and maize at 10:47 PM IST, September 12,2025.
2. Crop Recommendation: Decision tree achieved 87% accuracy post-PCA optimization.
3. Image-Based Crop Analysis: CNN reached 90% accuracy, 90% precision, and 90% recall, with an F1 Score of 90.50%.

A. Description of the confusion matrix A Results sand Discussion

After applying Principal Component Analysis (PCA) to the weather and soil dataset from Kaggle, the data was reduced to main components, labelled as Principal Component 1 (legend M) and Principal Component 2 (legend B).

Fig. 3.

Fig 3. shows how the data points are distributed in this reduce space. The crop samples naturally form clear clusters, indicating that the



reduced components successfully represent the variation present in the original features.

The decision regions created by the boosted decision tree classifier also show a distinct is able to distinguish the crop categories effectively. The dimensionality reduction performed through PCA not only simplified the dataset but also retained the essential patterns needed for classification.

The PCA-transformed data was then used with an ensemble decision-tree technique, which produced an overall accuracy of 89% for the crop recommendation model.

C. *Kernel Parameter Selection*

During the kernel selection process for the CNN model in the Image-Based Crop Analysis module, the default values of gamma and c were kept constant across all trials. A noticeable improvement was observed after applying PCA on data augmentation, with the accuracy increasing from 85% to 92%. This change indicates that the earlier configuration did not align well with the high-dimensional nature of the dataset before feature reduction. A study by Patel et al. [14] also reports that unsuitable parameter choices in deep learning tasks can lead to weaker results, especially when working with large image collections such as the PlantVillage dataset (over 50,000 images).

The improvement after preprocessing suggests that the dimensionality reduction made the data more manageable and better suited for training. Further trails are needed to evaluate the model under different conditions, including, including adjustment to kernel parameters and additional augmentation strategies.

Performance Summary

The system was evaluated using datasets related to wheat, rice and maize (tested at 10:47 PM IST, September 12, 2025). The key findings are summarized below:

- Crop Price Prediction: The RMSE decreased from18.3 (ARIMA reference value) to 12.5 using the time-series regression approach.
- Crop Recommendation: Accuracy improves to 87% after applying PCA for feature optimization.
- Image-Based Crop Analysis: Accuracy increased from 85% to 90%, with corresponding precision and recall values of 90%, and an F1-score computed from these metrics.
- Multilingual Chatbot Interface: The interface showed smoother responses and better handling of queries, with detailed quantitative evaluation planned in future testing.

VIII. CONCULSION

The AI Powered Smart Agriculture Advisor developed in this study has been designed to support farmers by improving productivity, strengthening practices. By working with diverse agriculture datasets, the system helps address key farming challenges such as unpredictable weather, fluctuating crop prices, and crop health issues.

The work was completed in three major stages: a review of existing research related to price prediction, crop recommendation, and disease detection; the design and implementation of the system; and a detailed evaluation of the results. The datasets used include crop price records from Data.gov.in (10,000 entries), weather and soil data from Kaggle (2,000 instances), and PlantVillage images (50,000 labelled samples). These datasets provided information on crop types, environmental factors, soil condition, crop recommendation, disease identification, and multilingual chatbot interaction.

The methodology involved preprocessing steps such as normalization and Principal Component Analysis (PCA), which helped improve model performance. Data augmentation further enhanced the accuracy of the image-based module, increasing CNN performance from 85% to 90%. The system also recorded strong results across evaluation metrics, including accuracy, precision, recall, F1-Score, and an RMSE value consistent with ARIMA's benchmark (18.3). Crop recommendation accuracy reached 87%, and the image-based disease detection model achieve 90% accuracy. Postman testing ensured the reliability of all PAI components.

The system has been deployed on Azure Cloud and is supported by Spring Security for secure access. The multilingual interface (Hindi and English) includes features such as a dashboard for reviewing past predictions and the ability to generate multilingual PDF reports. Future work may explore the use of blockchain technology to improve transparency, traceability, and long-term scalability within smart agriculture systems.

REFERENCES

[1] Van Klompenburg, T., et al., "Review of machine learning applications for crop yield prediction," Computers and Electronics in Agriculture, vol. 177, 105709, 2020.

[2] Hatuwal, B. K., et al., "Plant leaf disease recognition using CNN and ML classifiers," Polibits Journal, 2021.

[3] Journal of Circuits, Systems and Computers, "Crop price prediction via Decision Trees," vol. 30, no. 5, 2021.

[4] Prajapati, R., et al., "A soil-nutrient based crop recommendation engine," Sustainable Computing: Informatics and Systems, vol. 36, 2022.

[5] Kaggle, "Crop Yield and Price Datasets," Online. Available: https://www.kaggle.com, 2024.

[6] Data.gov.in, "Agriculture and Farmers Welfare Datasets," Online. Available: https://data.gov.in, 2024

.[7] Li, Z., et al., "AI in precision agriculture: A review of current applications," IEEE Access, vol. 10, pp. 23456–23470, 2022.

[8] Plant Village. (n.d.). Plant Village Dataset. Retrieved from two sources. • https://github.com/spMohanty/PlantVillageDataset • https://www.kaggle.com/datasets/abdallahalidev /PlantVillage-dataset

[9] Kumar, A., et al., "Deep learning for crop disease detection," Elsevier: Computers in Biology and Medicine vol. 153, 106373, 2023.

[10] Sharma, M., et al., "Blockchain-enabled smart agriculture for supply chain transparency," IEEE [11] Hyperparameter tuning of CNN based potato plant disease detection model integrated with Research. ResearchGate, 2025.