Intelligent Decision-Making for Sustainable Farming Using Machine Learning

[1]Shiwani Wasnik, [2]Mayuri Sawant, [3]Sakshi Shirishkar, [4]Prof. Sujata Kullur

[1]Student, [2]Student, [3]Student, [4]Professor
[1]Information Technology,
[1]Usha Mittal Institute of Technology SNDT Women's University Mumbai, India
[1]shiwaniwasnik@gmail.com, [2]mayurisawant0714@gmail.com, [3]sakshirishkar@gmail.com, [4]Sujata.Kullur@umit.sndt.ac.in

***Abstract*——Sustainable agriculture demands precise, data-driven strategies to maximize crop productivity while minimizing environmental harm. This study presents an innovative, scalable, and intelligent decision- making framework that leverages Machine Learning (ML) algorithm such as random forest regression (RFR) to estimate the nutrients requirement. Decision Tree classifier for smart crop recommendation. Predictive analytics for fertilizer recommendation on the basis of N, P, K values in the soil. Also, Convolutional Neural Network (CNN) with RaceNet for diseases detection, to help farmer in assisting proper plant health. Beyond enhancing farming efficiency, this research aspires to drive technological advancements in agriculture by demonstrating the transformative power of AI, ML, and data analytics. By integrating cutting-edge technology with traditional farming methods, this study envisions a future where precision agriculture empowers farmers, strengthens global food security, and supports eco-friendly agricultural evolution.**

I. Introduction

Agriculture serves as a cornerstone for food security, economic stability, and environmental sustainability, underpinning economies worldwide by supplying essential resources for human survival and industrial progress. However, conventional farming practices face pressing challenges such as excessive fertilizer use, inefficient crop selection, and delayed disease diagnosis, leading to diminished yields, soil degradation, and environmental damage. Additionally, the escalating impact of climate change and unpredictable weather patterns complicate decision-making for farmers, necessitating data-driven strategies to enhance agricultural efficiency and sustainability.

Sustainable agricultural practices aim to utilize natural resources without compromising future generations' needs, fostering long-term environmental conservation [1](Kumar et al., 2022). By adopting innovative approaches, sustainable farming mitigates adverse effects on water resources, soil health, biodiversity, and energy consumption.

Advancements in machine learning (ML) and deep learning (DL) offer promising solutions to contemporary agricultural challenges. Integrating ML and DL algorithms can significantly improve agricultural productivity through precise nutrient recommendations, fertilizer optimization, crop selection, and early disease detection [2](Senthil Kumar et al., 2022). These technologies enable data-driven precision farming, enhancing overall farm management.

Among ML algorithms, Random Forest (RF) demonstrates superior predictive accuracy compared to traditional statistical models like logistic regression [3] (Couronne et al., 2018). RF outperforms logistic regression in classification and regression tasks, establishing itself as a preferred model for agricultural decision-making systems that demand high accuracy.

Furthermore, the incorporation of real-time climate data elevates the precision of agricultural recommendations. Integrating parameters such as temperature, humidity, and rainfall into decision-making systems allows for dynamic adaptation, promoting sustainable resource management [4] (Ramakrishna Parama, 2017).

The primary objective is to address the limitations of conventional farming practices, including excessive fertilizer application, inefficient crop selection, and delayed disease diagnosis, all of which contribute to reduced yields, soil depletion, and environmental degradation. The remainder of this paper is structured as follows: Section II discusses the study area and reviewed literature, Section III elaborates on the methodology and ML model implementation, Section IV presents experimental findings and validation results, and Section V concludes with insights on system impact and future advancements.

II. Study Area

This study focuses on sustainable crop management through advanced machine learning techniques, with a particular emphasis on optimizing nutrient recommendations (NPK), fertilizer selection, crop recommendation, and disease detection. The research draws from a combination of online agricultural datasets and field insights obtained through direct engagement with agricultural institutions.

The primary data sources include publicly available datasets from platforms such as Kaggle, providing extensive information on soil composition, crop yield patterns, and environmental factors across various regions of India. Weather API is used for real- time weather data. These datasets offer a comprehensive foundation for training and validating the Random Forest Regression model employed in the project. To deepen our understanding of agricultural practices, field visits were conducted at agricultural colleges in Wardha and Satara, Maharashtra. These visits facilitated in-depth discussions with agriculture experts and local farmers, providing valuable qualitative insights into regional soil characteristics, common crop diseases, and fertilizer application practices. This dual approach—combining empirical data and field observations—ensures a holistic analysis and enhances the practical relevance of our machine learning models.

The study area encompasses diverse agro-climatic zones of India, as reflected in the datasets and fieldwork, allowing the system to provide tailored recommendations across a wide range of agricultural environments. By integrating multisource data and domain expertise, this research aims to deliver an intelligent decision-making system that promotes precision agriculture and sustainable resource management.

III. METHODOLOGY

The research methodology consists of several key steps:

- Data Acquisition and Preprocessing
- Machine Learning Model Implementation
- Weather API Integration for Forecasting
- Fertilizer and Crop Recommendation Framework
- Crop Disease Detection System

1. Data Acquisition and Preprocessing

The dataset used in this research incorporates various agricultural parameters sourced from reliable databases such as the Indian Council of Agricultural Research (ICAR), Kaggle, and governmental agricultural repositories. The primary data attributes include:

- **Soil Characteristics:** Nitrogen (N), Phosphorus (P), Potassium (K), pH value, moisture levels, and organic matter.
- **Climatic Factors:** Temperature, humidity, precipitation, and wind velocity.
- **Plant Health Indicators:** Leaf color variations, disease symptoms, and historical crop yields.

1.1 Data Cleaning and Normalization

To maintain data consistency, min-max normalization is applied using the following transformation:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

(Equation 1)

Where:

- X′ = Normalized value
- X = Original value
- Xmin, Xmax = Minimum and maximum dataset values

1.2 Feature Engineering

To optimize model performance, Principal Component Analysis (PCA) is employed for dimensionality reduction:

$$Z = XW$$

(Equation 2)

Where:

- X = Standardized dataset matrix
- W = Eigenvector matrix of covariance

This technique eliminates redundant data points, ensuring efficient processing.

2. Machine Learning Model Implementation

2.1 Nutrient Prediction using Random Forest Regression (RFR)

A Random Forest Regression (RFR) model is trained to estimate optimal NPK levels based on soil conditions and climatic variables. The model's final prediction is computed as:

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} f_i(X)$$

(Equation 3)

Where:

- Yˆ = Predicted nutrient level
- N = Number of decision trees
- fi(X) = Prediction from the i th decision tree

To fine-tune the model, GridSearchCV is utilized for hyperparameter optimization, focusing on:

- Number of estimators (N)
- Maximum tree depth
- Minimum data samples per split

3. Weather API Integration for Forecasting

A real-time weather forecasting API is integrated to dynamically adjust agricultural recommendations. The system considers:

- **Temperature (T)** - Affects crop growth and irrigation needs.
- **Precipitation (P)** - Influences fertilizer application timing.
- **Humidity (H)** - Impacts plant disease susceptibility.

A weighted prediction formula is applied:

$$W_{pred} = \alpha T + \beta P + \gamma H$$

(Equation 4)

Where α, β, γ are optimized coefficients derived from historical climate data.

4. Fertilizer and Crop Recommendation Framework

4.1 Fertilizer Recommendation System

The required amount of each nutrient (N, P, K) is estimated based on expected crop yield:

$$N_{opt} = \frac{Y_{target} - Y_{current}}{E_f}$$

(Equation 5)

Where:

- Nopt = Required nitrogen level for the crop
- Ytarget = Desired yield output
- Ycurrent = Current yield from soil analysis
- Ef = Fertilizer efficiency factor

Similar calculations are applied for phosphorus (Popt) and potassium (Kopt).

4.2 Crop Selection Model

A Decision Tree Classifier recommends suitable crops based on soil and climate factors. The classification decision is determined using the Gini Index, defined as:

$$\text{Gini} = 1 - \sum p_i^2 \qquad \text{(Equation 6)}$$

Where pi represents the probability of a given crop being the best fit for the conditions.

5. Crop Disease Detection System

A Convolutional Neural Network (CNN) is developed to diagnose plant diseases based on leaf images. The architecture includes:
• Convolutional Layers – Extract spatial patterns from images.
• Max Pooling Layers – Reduce feature dimensions while preserving critical information.
• Fully Connected Layers – Perform final classification.

The activation function for hidden layers is ReLU:

$$f(x) = \max(0, x) \qquad \text{(Equation 7)}$$

Final classification probabilities are computed using the Softmax function:

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \qquad \text{(Equation 8)}$$

Where:
• P(yi) = Probability of the leaf image belonging to disease category i.
• zi = Model output before activation.

The CNN model is trained on publicly available agricultural image datasets, enabling robust disease identification.

IV. RESULT AND DISCUSSION

The successful implementation of our project involved the integration of machine learning models with real-time agricultural data. This section outlines the outcomes of each module, emphasizing the technical process and practical applications for sustainable crop management.

- **N, P, K Recommendation Module:** This module is designed to provide nutrient recommendations based on crop type, state, and city. We collected and preprocessed data on soil properties, climate conditions, and historical agricultural records. Utilizing the Random Forest Regression model, we developed a system capable of analyzing these inputs to suggest optimal levels of nitrogen (N), phosphorus (P), and potassium (K). This data-driven approach enhances the accuracy of nutrient recommendations, promoting efficient fertilizer usage and improved crop yields.

Fig. 1. N, P, K estimation

```
class NPKEstimator:
    def __init__(self):
        # Get the current directory of this script
        base_dir = os.path.dirname(os.path.abspath(__file__))

        # Define the correct path to the CSV file
        data_path = os.path.join(base_dir, "Nutrient_recommendation.csv")

        # Check if the file exists before loading
        if not os.path.exists(data_path):
            raise FileNotFoundError(f"CSV file not found: {data_path}")

        # Load the dataset
        self.df = pd.read_csv(data_path, header=None)
        self.X_train = None
        self.X_test = None
        self.y_train = None
        self.y_test = None
```

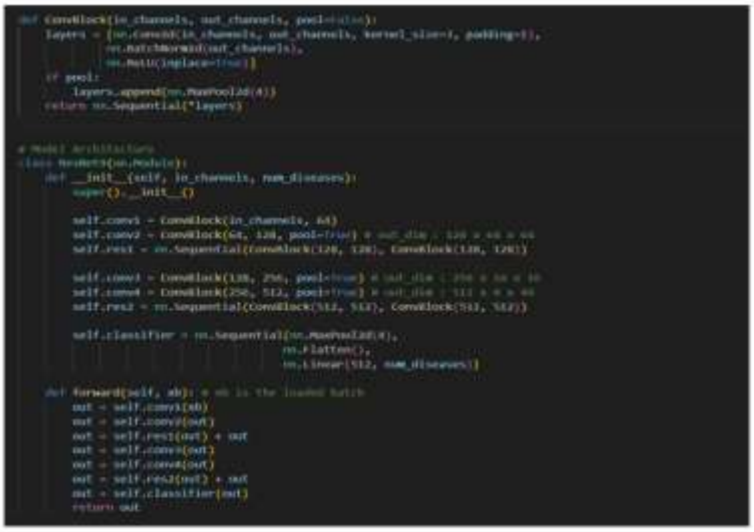
Fig. 2. Random Forest algorithm

```
    def accuracyCalculator(self):
        model = RandomForestRegressor(n_jobs=-1)
        estimators = np.arange(10, 200, 10)
        scores = []
        for n in estimators:
            model.set_params(n_estimators=n)
            model.fit(self.X_train, self.y_train)
            scores.append(model.score(self.X_test, self.y_test))

        scores_arr = [round(sc, 3) for sc in scores]
        unique, counts = np.unique(scores_arr, return_counts=True)

        max_count = max(counts)
        accuracy = -1
        for uni, count in zip(unique, counts):
            if count == max_count:
                accuracy = uni

        return accuracy
```

- **Fertilizer Recommendation Module:** The fertilizer recommendation module maps the predicted nutrient deficiencies to appropriate fertilizer types. The system considers nitrogen, phosphorus, potassium, and crop type to suggest the required fertilizers. This targeted recommendation approach optimizes resource utilization while minimizing environmental impact. By aligning fertilizer suggestions with specific nutrient needs, the system reduces the likelihood of over-application and subsequent ecological harm.

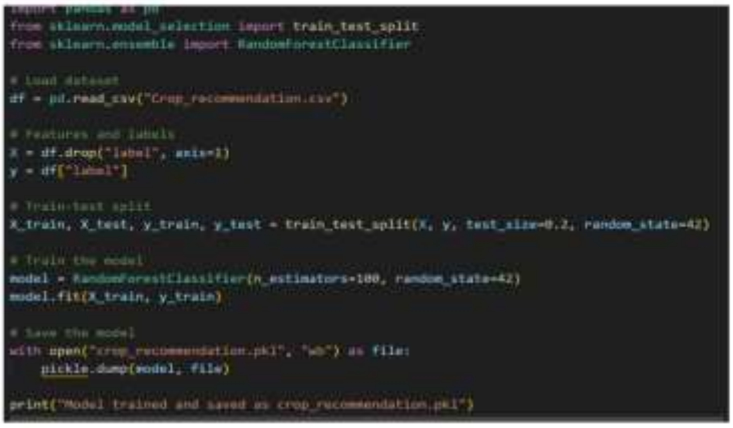
- **Disease Detection Module:** The disease detection module is implemented using image-based analysis. We utilized a Convolutional Neural Network (CNN) to classify and identify common crop diseases. By preprocessing agricultural images and feeding them through the trained model, the system detects and categorizes signs of disease for early intervention. This module supports image inputs, allowing users to upload crop images for real-time analysis and diagnosis. This feature facilitates early detection, enabling prompt action to mitigate potential crop loss.

Fig. 3. resNet9 algorithm



- **Crop Recommendation Module:** This module suggests suitable crops based on nitrogen, phosphorus, potassium levels, pH level, rainfall (in mm), state, and city. We integrated climate factors (such as rainfall) with soil characteristics to identify the most suitable crop options for a given location. The system analyzes these parameters and provides data-driven crop recommendations that align with both current environmental conditions and long-term sustainability goals. This assists in selecting crops that are best suited for the region, thereby increasing productivity and reducing environmental degradation.

Fig. 4. Crop Recommendation



The implementation of these modules demonstrates a practical and scalable approach to precision agriculture. The system integrates diverse datasets and applies machine learning to provide tailored recommendations. Throughout the development, a key challenge was ensuring consistency across data sources, which required thorough preprocessing and validation steps. vide tailored recommendations. Throughout the development, a key challenge was ensuring consistency across data sources, which required thorough preprocessing and validation steps.

The modular architecture allows each component to function independently while maintaining seamless integration. This design enables users to access individual modules or utilize the entire system based on their specific needs. Furthermore, the incorporation of real-time environmental data enhances the system's adaptability to changing conditions, improving the reliability and accuracy of the recommendations.

By leveraging advanced machine learning methodologies and real-time inputs, the project effectively addresses critical agricultural challenges, including optimizing fertilizer use, improving crop selection, and facilitating early disease detection. Future enhancements may include expanding the dataset, integrating advanced deep learning techniques, and providing multilingual user interfaces to enhance accessibility and overall system performance.

V. Conclusion

This project successfully integrates machine learning models with real-time agricultural data to address critical challenges in sustainable crop management. Through the development of four key modules—N, P, K recommendation, fertilizer recommendation, disease detection, and crop recommendation—a comprehensive system was implemented to enhance agricultural decision-making. The modular architecture supports independent operation of each component while maintaining seamless integration.

This design allows users to either utilize specific modules based on their needs or employ the entire system for holistic agricultural guidance. One of the primary challenges during implementation was ensuring data consistency across diverse sources, which required rigorous preprocessing and validation. The incorporation of real-time weather data significantly improved the system's responsiveness and accuracy, enabling dynamic adjustments based on changing environmental conditions.

Future Enhancements: To further improve the system's efficiency and accessibility, the following advancements are proposed:

- IoT Integration: Implementing real-time monitoring of soil health through smart sensors to provide more precise data.
- Multilingual Support: Expanding language options to facilitate broader adoption across diverse agricultural communities.
- Advanced Deep Learning Techniques: Enhancing the accuracy of disease detection through the integration of more sophisticated deep-learning models
- Expanded Crop Coverage: Increasing the system's applicability by incorporating a wider range of crop varieties and soil conditions.

By combining advanced machine learning methodologies with real-time environmental data, this project delivers a scalable and intelligent solution for sustainable agriculture. The system's ability to optimize fertilizer usage, enable early disease detection, and recommend suitable crops enhances agricultural productivity while minimizing environmental impact. As technology continues to evolve, the integration of more advanced methodologies will further refine precision agriculture, empowering farmers to make informed decisions and promoting sustainable agricultural practices.

Looking forward, this research aims to contribute to the advancement of smart farming by bridging the gap between technology and agriculture. Promoting the adoption of intelligent systems will empower farmers, researchers, and policymakers, fostering a more efficient, resilient, and sustainable agricultural future.

VI. Acknowledgment

We would like to acknowledge the various online resources, research papers, and open-source libraries that played a crucial role in the development of this project. Our research was self-driven, relying on extensive literature review and hands-on experimentation. We also extend our appreciation to our guide for their suggestions during the project. Lastly, we thank us team members for their dedication and collaborative efforts in bringing this project to completion.

References

[1] D. K. M. A. S. Y. Mandeep Kumar, Ambrish Kumar, "Sustainable agriculture," Vidya Varta SPI: 28- 31, 2022. [Online]. Available: https: //www.researchgate.net/publication/363266859 Sustainable Agriculture

[2] M. D. S. b. Senthil Kumar, Swami Durai, "Smart farming using machine learning deep learning techniques," Decision Analytics Journal Volume 3, 2022. [Online]. Available: https://www.sciencedirect. com/science/article/pii/S277266222200011X

[3] A.-L. B. Raphael Couronne, Philiph Probst, "Random forest versus logistic regression a large- scale benchmark experiment," BMC bioinformatics 19(1), 2018. [Online]. Available: https://www.researchgate.net/publication/326453804 Random forest versus logistic regression A large-scale benchmark experiment

[4] V. R. Parama, "Effects of climate on soil properties crop nutritious," MaysoreJ. Agri. Sci 51(1):1- 11, 2017. [Online]. Available: https://www.uasbangalore.edu.in/images/2017-1st-Issue/1.pdf

[5] ResearchGate, "Analysis of crop yield prediction using random forest regression model," 2023. [Online]. Available: https://www.researchgate.net/publication/368937696 Analysis of Crop Yield Prediction Using Random Forest Regression Model

[6] S. Kumar, "Comparison of random forest and multiple linear regression models," 2023. [Online]. Available: https://www.sciencedirect.com/ science/article/pii/S1470160X20308645

[7] ResearchGate, "Crop yield predictions and recommendations using random forest regression in agroclimatic zone rajasthan," 2023. [Online]. Available: https://www.researchgate.net/publication/370029230 CROP YIELD PREDICTIONS AND RECOMMENDATIONS USING RANDOM FOREST REGRESSION IN 3A AGROCLIMATIC ZONE RAJASTHAN

[8] F. India. (2023) India at a glance: Agriculture and food security. [Online]. Available: https://www.fao.org

[9] I. Conferences, "Agriculture data analysis using ai," ITM Conferences, 2023. [Online]. Available: https://www.itm-conferences.org/articles/ itmconf/pdf/2023/06/itmconf icdsac2023 02007.pdf

[10] B. Journals, "Plant pathology analysis," Plant Pathology, 2023. [Online]. Available: https://bsppjournals.onlinelibrary.wiley.com/doi/10. 1111/ppa.14006

[11] Reuters, "Empowering smallholder farmers with ai tools to bolster global food supply," Reuters, 2025. [Online]. Available: https: //www.reuters.com

[12] F. in Remote Sensing, "Remote sensing applications in agriculture," Frontiers in Remote Sensing, 2022. [Online]. Available: https: //www.mdpi.com/2076-3417/15/8/4537

[13] F. in Plant Science, "Plant science and crop yield prediction," Frontiers in Plant Science, 2023. [Online]. Available: https://www.frontiersin.org/ journals/plant-science/articles/10.3389/fpls.2023.1158933/full

[14] Nature, "Impacts of climate on crop production," Scientific Reports, 2023. [Online]. Available: https://www.nature.com/articles/ s41598-023-46218-5