AI application for crop management and yield prediction

Ms. Ritu Jaiswal[1], Dr. Ranjana Rajnish[2]
[1,2] *Amity University Lucknow Campus*

***Abstract*—Agriculture, practiced on 60.45% of India's land, has evolved with AI technologies like ML, computer vision, and big data to improve crop management and yield prediction. Tools such as satellite imagery, drones, and IoT devices monitor soil, weather, and pests. ML algorithms process this data to assess risks and suggest actions, with CNNs and Random Forests aiding in disease detection and yield prediction [1].These advancements boost productivity and sustainability by reducing chemical use and conserving water. However, challenges like limited data and unequal tech access remain. This study aims to identify suitable crops based on soil type, humidity, and temperature, and forecast yield.**

***Index Terms*—AI, ML, Random Forest, SVM, Decision Tree, KNN**

I. INTRODUCTION

Crop yield prediction is vital amid growing food security concerns. In 2023, around 733 million people faced hunger—one in eleven globally and one in five in Africa—according to the SOFI report by UN agencies [2]. The WHO also estimates that 820 million people still lack adequate food [3] and the FAO predicts a 60% rise in food demand by 2050 for a population of 9.3 billion.

AI-driven crop management helps farmers make informed decisions, optimize resources, and boost productivity. Real-time data from sensors, drones, and satellites monitor soil, climate, and pests. ML models like KNN, SVM, and Random Forest forecast yields using historical and live data. Deep learning, with its layered neural networks, further improves

accuracy [4]. Integrating AI with traditional methods supports more efficient and sustainable agriculture.

ML Task Classification-Machine Learning (ML) tasks are categorized based on learning approaches (supervised or unsupervised), objectives (classification, regression, clustering, dimensionality reduction), and models used for specific tasks.

Types of Learning

Machine Learning (ML) includes supervised learning, where models learn from labelled data to make accurate predictions, and unsupervised learning, which finds patterns in unlabelled data. Reinforcement learning works through feedback, helping models learn by trial and error.

Learning Analysis-:

Dimensionality Reduction (DR)- DR is a technique used to simplify complex data by reducing the number of features while keeping important information. This makes machine learning models faster and more accurate.

Learning Models

1. Random Forest Classifier (RFC)- RFC is a powerful model that combines many decision trees to make reliable predictions and avoid overfitting. In agriculture, it analyses factors like nitrogen, temperature, and rainfall to suggest suitable crops and estimate yields, helping farmers plan better.
2. Decision Tree Classifier (DTC)-A hierarchical model that splits data into branches based on decision rules [5]. It predicts crops by evaluating soil, climate, and irrigation conditions. It also estimates yield by analysing historical patterns under similar conditions.
3. K-Nearest Neighbours (KNN)-A distance-based model that classifies new data points by comparing them with 'k' nearest neighbours. It recommends crops based on similarity with past soil and climate conditions. It also predicts yields using averaged historical data, making it ideal for real-time recommendations.
4. Support Vector Machine (SVM)-A classification model that finds an optimal decision boundary (hyperplane) for separating different crop categories [6]. It handles complex datasets using kernel functions, analysing soil composition, climate, and past agricultural trends to predict crop suitability and yield outcomes effectively.

In short, machine learning has improved crop prediction and yield estimation by giving farmers useful, data-driven insights. Models like Decision Trees, SVM, Random Forests, and deep learning have made crop selection and forecasting more accurate. IoT helps with real-time monitoring for better decisions. While challenges like limited data and complex models remain, ongoing research is working to solve them. Overall, ML is set to play a key role in making farming more efficient and sustainable.

Objective

This web app helps farmers make smart choices by analysing data like humidity, temperature, soil type, and nutrients. It uses AI to recommend the best crop to grow and predict its yield, improving accuracy and planning. The tool boosts efficiency, sustainability, and profits by helping farmers reduce waste and increase harvests. By offering reliable advice, it supports food security, better land use, and stable incomes, enabling farmers to plan for long-term success.

II. LITERATURE REVIEW

Crop prediction and yield estimation are vital for modern farming, helping farmers use resources better and increase productivity. ML improves these predictions by analysing factors like soil, climate, and geography. Data such as temperature, humidity, soil nutrients, and rainfall help determine the best crops for different regions. Using historical and real-time data, ML reduces uncertainties caused by changing weather and soil conditions [7]. Various ML methods are used for crop prediction. Decision Trees split data based on soil and weather to classify crops. SVM find boundaries between crop types, capturing complex patterns [8].

Deep learning improves accuracy further. CNNs analyse satellite images to study vegetation, while RNNs and LSTMs process time-based data like weather to predict yields. Combining GIS with CNNs helps map crop distribution for better recommendations. Ensemble learning mixes models like Decision Trees, SVM, and Neural Networks to boost accuracy. Fuzzy logic and genetic algorithms help pick important features while keeping models efficient [9]. Yield estimation also relies on ML using historical and live data. Regression models (linear, polynomial, SVR) find links between environment and yield. Random Forest Regression handles big data well and boosting methods like XG-Boost refine predictions.

IoT sensors provide real-time soil and weather info, improving ML yield estimates. Challenges remain in data access and model transparency, needing better datasets and explainable AI. Future work focuses on combining multiple data sources and creating easy-to-use AI tools to help farmers. Overall, ML is transforming agriculture towards smarter, sustainable farming.

III. PROBLEM STATEMENT

1.Uncertainty in Crop Selection-Farmers often struggle to determine the most suitable crop for their land due to varying environmental factors such as soil composition, weather conditions, and nutrient levels. This leads to inefficient resource utilization and reduced agricultural productivity.

2.Inaccurate Yield Estimation-Traditional yield estimation methods rely on manual observations and historical data, which may not account for real-time climate changes, soil conditions, and other dynamic factors affecting crop productivity.

3.Lack of Data-Driven Decision Making-Many farmers lack access to advanced tools that utilize machine learning (ML) for predictive analysis, making it difficult to make informed decisions regarding crop selection and expected yield.

IV. METHODOLOGY

The methodology of this project involves several steps, including data preprocessing, feature engineering, model training, and deployment of a Flask-based web application for real-time crop recommendation and yield prediction.

1. Data Preprocessing

1. Dataset Preparation

The dataset (Crop_recommendation.csv) is loaded, and duplicate entries are removed to maintain data integrity. Key environmental features used for prediction include humidity, temperature, potassium, nitrogen, phosphorus, pH, and rainfall.The categorical feature "soil type" is one-hot encoded, converting each type (sandy, clayey, loamy, silty, peaty, saline, laterite) into separate binary columns.

2. Splitting the Dataset

1.The dataset is split into features (X) and target labels (y):

- y_crop: Crop type
- y_yield: Yield output

2.Crop labels are factorized into numerical values.

3.The dataset is split into 80% training and 20% testing sets using train_test_split() for effective model training.

3. Model Selection and Training

3.1. Crop Classification Models

Four machine learning classifiers are implemented and compared:

1.Random Forest Classifier (RFC) – Uses multiple decision trees and averages their outputs.

2.Decision Tree Classifier (DTC) – Classifies crops through hierarchical splitting of features.

3.K-Nearest Neighbours (KNN) – Assigns labels based on nearby data points.

4.Support Vector Machine (SVM) – Finds the best decision boundary using margin maximization.

5.Each model is trained using X_train and y_crop_train.

6. Model performance is evaluated using accuracy_score(), and the most accurate model is selected as the final classifier.

• Model	Accuracy (%)
SVM	96.82%
KNN	95.68%
Decision Tree	98.41%
Random Forest	99.32%

[Table-1]

Random Forest performed the best with 99.32% accuracy.

3.2. Yield Prediction Model

1.A Random Forest Regressor (RFR) is trained to predict crop yield using the same input features.

2.The model learns from historical data and estimates the expected yield based on soil and weather conditions.

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
SVM	67,151.89	11.64 billion
KNN	75,341.03	10.27 billion
Decision Tree	94,709.09	18.84 billion
Random Forest	77,397.06	10.01 billion

[Table-2]

Random Forest had the lowest MSE (10.01B), making it the best choice for yield prediction.

4. Model Deployment Using Flask

- The trained models are saved using Pickle (pickle. dump()) for later use.
- A Flask web application is developed to allow users to input environmental data and receive predictions.
- The HTML frontend (app3.html) takes user inputs and sends them to the backend for processing.
- The backend processes the input, makes predictions using the trained models, and returns the recommended crop and expected yield to the user.

5. Real-Time Prediction Process

When a user submits input data via the web interface:

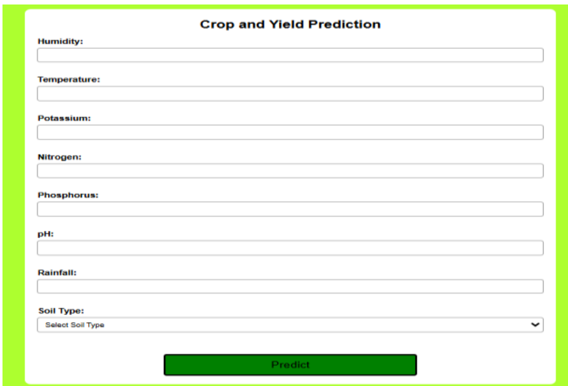
- The form data is retrieved and processed.
- The model predicts the most suitable crop and expected yield.
- The predictions are displayed on the webpage**.**

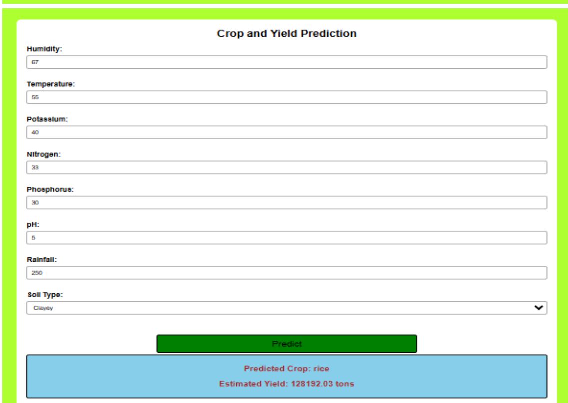
V. RESULT AND DISCUSSION

The results show that machine learning effectively predicts the best crop based on environmental factors and estimates its yield. Key inputs like humidity, temperature, nutrients, soil pH, rainfall, and soil type improve accuracy, with soil type playing a crucial role since it affects water retention and nutrients. Allowing users to select soil type makes recommendations more precise. During data processing, duplicates were removed, and missing values handled to keep data clean. The dataset was split into training and testing sets for proper evaluation. Several ML models Random Forest, Decision Tree, K-Nearest Neighbours, and SVM were tested for crop classification. Random Forest performed best due to its ensemble learning, which reduces overfitting and improves results.

For yield prediction, Random Forest Regression captured complex relationships between factors. It accurately estimated crop yield, helping farmers plan resources and forecast production. Metrics like mean absolute error and root mean square error confirmed reliability. Including soil type as a one-hot encoded feature improved classification for example, rice grows well in clayey or loamy soil, maize in sandy or loamy soil matching real farming practices.

Challenges remain with data availability, soil variability, and real-time updates. Future improvements could use IoT soil sensors and satellite data for continuous learning. A mobile-friendly version with voice and multilingual support could increase access. Overall, this system offers data-driven advice, reducing farming uncertainty and supporting sustainable agriculture.





VI. FUTURE ENHANCEMENTS

In the future, we want to make the system more accurate and easier to use by adding real-time data from soil sensors and satellites. This will help the model update with current soil and weather conditions for better crop predictions, including factors like altitude, wind speed, and pest issues.

Since soil quality can vary even within a farm, soil testing kits or mobile tools will provide more precise data. A mobile app with voice commands and multiple languages will make it easier for farmers, especially in rural areas. We'll also explore deep learning and advanced AI. Collaborating with experts and governments will help customize the system for different regions, helping farmers get better harvests and reduce losses.

REFERENCES

[1] Priyanga Muruganantham, Santosh Wibowo, Srimannarayana Grandhi, Nahidul Hoque Samrat and Nahina Islam. A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. (Google scholar)https://www.mdpi.com/2072-4292/14/9/1990

[2] World Health Organization. (2024, July 24). Hunger numbers stubbornly high for three consecutive years as global crises deepen: UN report. https://www.who.int/news/item/24-07-2024-hunger-numbers-stubbornly-high-for-three-consecutive-years-as-global-crises-deepen--un-report

[3] WHO. World Hunger Is Still Not Going Down after Three Years and Obesity Is Still Growing—UN Report. https://www.who.int/news/item/15-07-2019-world-hunger-is-still-not-going-down-after-three-years-and-obesity-is-still-growing-un-report

[4] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90.https://doi.org/10.1016/j.compag.2018.02.016

[5] Belson, W.A. Matching and Prediction on the Principle of Biological Classification. *Appl. Stat.* 1959, *8*, 65–75. [Google Scholar]

[6] Vapnik, V. Support vector machine. *Mach. Learn.* 1995, *20*, 273–297. [Google Scholar]

[7] Liakos, K.G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. Sensors, 18(8), 2674. mdpi.com

[8] Deo, R.C., & Şahin, M. (2019). Forecasting Agricultural Output Using Machine Learning Techniques and Satellite Imagery: A Review. Computers and Electronics in Agriculture, 165, 104968. researchgate.net

[9] Singh, P., Jain, M., & Gupta, R.K. (2020). Application of Machine Learning for Crop Yield Prediction and Agricultural Planning: A Review. Artificial Intelligence in Agriculture, 4, 58-72. researchgate.net

[10] M. Goel and M. Pandey, "Crop Yield Prediction Using AI: A Review," *2024 2nd International Conference on Disruptive Technologies (ICDT)*,

Greater Noida, India, 2024, pp. 1547-1553, doi: 10.1109/ICDT61202.2024.10489432. (Google scholar)https://ieeexplore.ieee.org/abstract/document/10489432/references

[11] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms", In Journal of Physics: Conference Series, vol. 1714,no.1,pp.012012,2021.(Google-scholar)https://iopscience.iop.org/article/10.1088/1742-6596/1714/1/012012/meta

[12] "AI in Agriculture: Smart Farming and Precision Agriculture," https://relevant.software/blog/ai-in-agriculture.