Crop Yield Prediction Driven by ML & DL

Prof. Swati P. Gade[1], Nikhil Shelke[2], Rohan Pawar[3], Manas Torane[4], Pranay Satpute[5]
[1,2,3,4,5]*Department of Computer Engineering PDEA'sCollege of Engineering Manjri Bk Pune*

***Abstract*—The approach and implementation of an integrated Machine Learning system are presented for modernizing agriculture. It uses Machine Learning to overcome various challenges faced by farmers due to unfavourable weather, pest diseases, resource constraints, and methods that are inefficient. The proposed solution aims to provide data-driven insights to help farmers with the selection of crops. At the core, Random Forest Regression is used here to estimate potential yield and suggest the best crop for a given soil type at any specific real-time weather condition. It also features a CNN for accurate detection of diseases from user-uploaded images of crops. The application integrates the Open Weather API to fetch real-time climate data and Gemini API for contextual understanding and personalized fertilizer recommendations. This presented system combines these intelligent modules to come up with reliable, context-aware, and data-driven suggestions for farmers, thereby increasing productivity and ensuring sustainability.**

***Index Terms*—Random Forest, CNN, Crop Yield Prediction, Fertilizer Recommendation, Disease Detection, Smart Agriculture.**

I. INTRODUCTION

The agricultural sector is facing increasing challenges due to climate change, resource limitations, and rising demand for food globally. Unpredictable weather, pest outbreaks, soil degradation, and inefficient farming methods further challenge the sector. Thus, intelligent and data-driven solutions are required to help farmers make informed decisions. The project, Crop Yield Prediction using ML, forms a foundational bridge between traditional agriculture and smart agriculture. It develops an integrated, complete platform for farmers using Machine Learning. The main goal of the project is providing a unified platform for users in the following ways: Crop Recommendation: Suggest the most suitable crop by analysing user inputs, soil data, weather condition, and regional climate trend. Fertilizer Suggestion: Calculation and suggestion regarding the optimal fertilizer composition and amount, considering the type of crop, soil health, and real-time data uploaded. Disease Detection: Deep learning models were used, namely CNN, for proper classification of diseases and time warnings based on user-uploaded images. Yield Prediction: Estimation of potential yield for a given crop using machine learning models trained on historical datasets pertaining to agriculture and weather. The system leverages models like Random Forest, CNN, and APIs such as Open Weather and Gemini for context-aware suggestions to help improve productivity and sustainability. Following is the detail of the methodology, literature supporting choices for the models, and the integrated architecture of the ML-based smart farming system in subsequent sections.

II. LITERATURE SURVEY

This proposed ML-based smart farming system has been greatly supported by existing academic studies that validate the application of the selection of machine learning models and also pinpoint areas where critical innovation is needed in agricultural technology.

2.1. Crop Yield Prediction and Recommendation Several research pieces have established the efficiency of the RF models in field-level agricultural decision-making: Yield Prediction Accuracy: Studies have identified Random Forest regression as the best model for yield prediction, exhibiting high performance with a high R-squared value of 0.961. Another project that used the same algorithm reportedly presented high yield prediction accuracy of more than 99%. A specialized application using an

RF model with mixed field and synthetic data successfully predicted cotton yield with $97.75\\%$ accuracy. Performance on Crop Selection: For crop suitability, studies have established that Random Forest is a topper forming model, reaching up to $99.7%$ accuracy. The Naïve Bayes model also showed high performance at $99.39\\%$ for crop selection.

2.2. Disease Detection & Deep Learning Application of deep learning models, particularly Crop Convolution Neural Network (CNN) is validated for plant health Monitoring. Disease Classification: The CNN technique is proven for disease classification from user-uploaded images. The current system leverages this technique, and has proven effective for disease classification from user-uploaded images. The current system leverages this technique and takes it a step further by verifying the results on the Plant.id API. Data Challenges: While effective, these complex deep learning models, such as Conv LSTM-VIT, require a lot of data and computational power.

2.3. Addressing Gaps in Existing Systems The proposed ML-based system directly addresses critical limitations identified from prior studies: Real-time Data and Fertilizers: Previous works have indicated that "real time data through IoT devices" need to be included in future work, and "fertilizers use" should be considered. Our system covers this by the inclusion of the Open Weather API for real-time data, along with a module for optimal fertilizer suggestion. System Accessibility and Interpretability: A systematic review indicated that the models should be accessible and easy to use, particularly for small farmers, while strongly advocating for the development of "user-friendly interfaces and mobile applications". It also highlighted the necessity for more "interpretable models".

III. PROBLEM STATEMENT

The basic problem that this research tries to address is the increasing inability of conventional agricultural practices to cope with the complexities and volatility of the present environmental and economic world. Agriculture is increasingly being challenged due to changes in climate, resource limitations, and increased global demand. It is increasingly challenged by unpredictable weather, pest outbreaks, soil degradation, and inefficient farming practices. Despite the immense historical and cultural importance of monuments and heritage, these places are disconnected from modern society. This demands an intelligent, data-driven decision-making solution for farmers. In more specific terms, the key deficiencies which cause suboptimal agricultural outcomes include:

I. Environmental Volatility and Risk: The farmer needs help to deal with the various environmental risks caused by unstable weather conditions or frequent pest manifestations.

II. Information Deficiency in Decision-Making: There is a critical lack of solutions to suggest the most suitable crop for a given type of soil and weather and to calculate the optimal fertilizer composition and amount.

III. Disease Management: Farmers need timely and valid disease classification and warning for optimizing resources to increase crop yield.

IV. PROPOSED SYSTEM

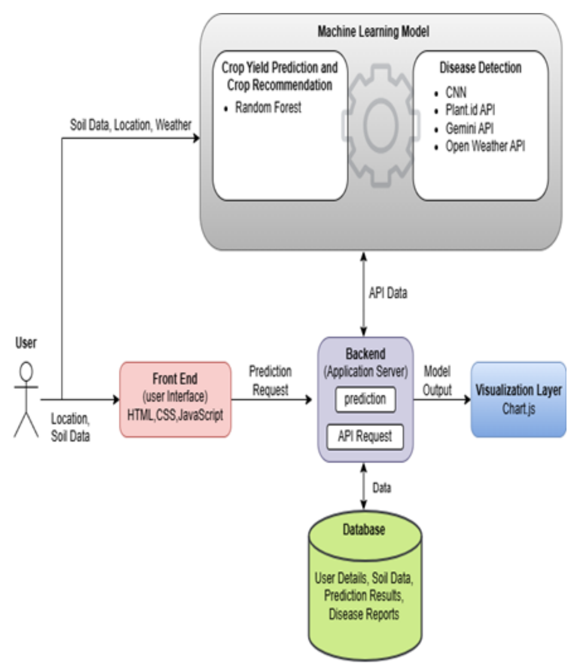

Fig: System Architecture

The proposed system is designed as an integrated, intelligent platform that will bridge the gap between traditional farming and smart agriculture effectively. The methodology is structured into three distinct phases: Data Management, Core Intelligent Processing, and Output and Delivery Platform.

4.1. Data Management (Phase 1)

This phase focuses on the acquisition, preparation, and management of all inputs required by the predictive module:

Data Acquisition & Sources: The system integrated real-time weather data through the OpenWeather API. It also uses historical agricultural datasets and user input, like location and soil data. The images uploaded by users for the disease detection module are collected separately.

Data Preparation: It incorporates Data Preprocessing: Cleaning, Normalization, and Encoding. Most importantly, Feature Selection includes key variables like Rainfall, Temperature, Fertilizer, and Crop Type.

4.2. Core Intelligent Processing - Phase 2

a. Yield Prediction:

- Module The module uses a Random Forest machine learning model.
- The model was trained and validated using historical data, whose performance was tracked by metrics such as RMSE, SR^2$, and MAE.
- Expected Yield is the output.

b. Crop & Fertilizer Recommendation

- Module: The module performs the analysis of soil data and weather and suggests the best crop suitable with optimal fertilizer composition/amount.
- It uses the Gemini API for contextual understanding and personalized suggestions.
- The output is the Best Crop and the optimum fertilizer suggestions.

c. Disease Detection Module

- This module uses a CNN to classify diseases from images provided by the user.
- It uses the Plant.id API for external verification and classification of diseases in plants.
- The output is Disease Classification and timely warning.

4.3. Output and Delivery Platform - Phase 3

This phase is responsible for presenting the results to the farmer in an actionable format:

Visualization and Insights: The system displays actionable insights, such as Yield, Crop, and Fertilizer Advice, through an intuitive dashboard. Data Visualization: The Chart.js library is used to provide two dynamic, interactive charts (one line graph and one pie chart) to visualize trends.

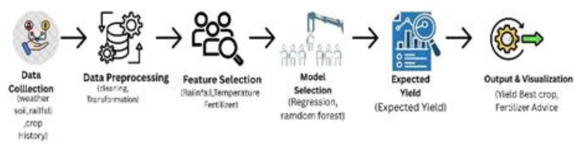

Fig: System Flow

V. FLOW PROCESS

This phase includes three major intelligent modules that are responsible for the core functionality of the system.

VI. CONCLUSION

The developed ML-Based Smart Farming System will provide an integrated intelligent platform that effectively bridges the gap between traditional farming and smart agriculture. It will deliver reliable and context-aware suggestions by performing crop and fertilizer recommendations, disease detection, and yield prediction. Module A random forest machine learning model is used in the module.

The strength of the system is in how it integrates comprehensive, proven machine learning techniques with sources of external data.

- It uses the high accuracy of Random Forest for yield prediction and crop suitability.
- It uses a CNN for correct image-based disease detection.
- It utilizes the Open Weather API for necessary real-time climate information as well as the Gemini API for contextual, personalized recommendations.

REFERENCES

[1] PRITESH PATIL, PRANAV ATHAVALE, MANAS BOTHARA, SIDDHI TAMBOLKAR, and ADITYA MORE. "Crop Selection and Yield Prediction using Machine Learning Approach." ISSN 2347-4688, 31 October 2023.
[2] Dr. R.C. Dyana Priyatharsini, Budha Venkata Raman, Chandu M N V L Saipraneeth, Bogireddy Balakrishna Reddy."Crop Yield Predication using Random Forest Regression Algorithm." IEEE 123456, 4 April 2025.
[3] ABID BADSHAH KAMAL Z. ZAMLI 1, BASEMYOUSEF ALKAZEMI 2, FAKHRUD DIN 3,4, AND MUHAMMAD HARIS 5. "Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability." IEEE 789101, 25 October 2024.
[4] ALAKANANDA MITRA VANGIMALLA R. REDDY, SAHILA BEEGUM, DAVID FLEISHER, WENGUANGSUN, CHITTARANJAN RAY DENNIS TIMLIN, ANDARINDAMMALAKAR. "Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data." 2169-3536, 24 June 2024.
[5] DONG-WOOK KIM, GYUJIN JANG, AND HAK JIN KIM. "Development of CNN-Based Semantic Segmentation Algorithm for Crop Classification of Korean Major Upland Crops Using NIA AI HUB." 2169-3536, 9 January 2025.
[6] Seyed Mahdi Mirhoseini Nejad, Dariush Abbasi Moghadam and Alireza Sharif. "ConvLSTM-ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data." 2151 1535, 18 September 2024.
[7] CLAUDIA R. SCREPNIK, EDUARDOZAMUDIO GIMENEZ. "Artificial Intelligence in Agriculture: A Systematic Review of Crop Yield Prediction and Optimization."15 April 2025. 3536,24 June 2024.