



Crop Disease Prediction & Solution Recommendation System Using Chatbots

Sakshi Sugriv Kale¹, Sneha Nilesh Morey², Prajakta Rajesh Murade³, Pooja Kailashsingh Rajput⁴,
Vaishnavi Nandakishor Darane⁵, Prof. Amit Sahu⁶

^{1,2,3,4,5}Student, CSE, School of Engineering & Technology, G. H. Rasoni University, Amravati

⁶Assistant Professor, CSE, CSE, School of Engineering & Technology, G. H. Rasoni University, Amravati

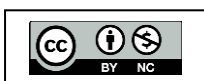
Abstract: Agriculture plays a crucial role in economic development, yet farmers often face challenges in selecting suitable crops, managing soil fertility, and preventing diseases, which ultimately affect productivity and profitability. This paper presents an intelligent Crop Prediction and Solution Recommendation System that integrates Machine Learning (ML) techniques to assist farmers in making data-driven decisions. The system focuses on analyzing soil properties, predicting appropriate crops, and recommending fertilizers and preventive measures for diseases and pests. The proposed system collects real-time soil data such as moisture, temperature, and nutrient content using embedded sensors and microcontrollers. This data is combined with historical datasets obtained from agricultural sources and undergoes preprocessing steps including cleaning, normalization, and feature scaling to ensure accuracy. Machine learning algorithms such as Random Forest and Linear Regression are applied to predict crop suitability and expected yield. Additionally, pattern recognition techniques are used to identify potential disease occurrences and provide preventive recommendations.

Keyword: Crop Prediction, Precision Agriculture, Machine Learning, Soil Analysis, Random Forest, Linear Regression, Data Preprocessing, Soil Fertility, Smart Farming, Yield Prediction, Fertilizer Recommendation, Agricultural Decision Support System.

I. INTRODUCTION

Agriculture remains the backbone of many economies, especially in countries like India, where a significant portion of the population depends on farming for their livelihood. However, traditional agricultural practices often rely on manual observation, experience-based decision-making, and inconsistent environmental conditions, which can lead to reduced crop productivity and financial losses. Factors such as soil fertility, climatic variations, improper crop selection, and increasing incidence of pests and diseases pose major challenges to farmers [1].

With the rapid advancement in technologies such as Machine Learning, agriculture is undergoing a transformation toward smarter and more data-driven practices. These technologies enable the collection and analysis of real-time data from agricultural fields, allowing farmers to make informed decisions regarding crop selection, irrigation, and fertilizer usage. Precision agriculture, which focuses on optimizing inputs and maximizing output, has emerged as a promising solution to address the inefficiencies of traditional farming systems [2].



In this context, the proposed Crop Prediction and Solution Recommendation System aims to assist farmers by providing accurate and timely recommendations based on soil and environmental conditions. The system utilizes IoT-based sensors to collect soil parameters such as moisture, temperature, and nutrient levels. This data is then processed using machine learning algorithms like Random Forest and Linear Regression to predict suitable crops and expected yield. Additionally, the system offers recommendations for fertilizers and identifies potential risks such as diseases and pest attacks [3].

The integration of intelligent data analysis with real-time monitoring not only enhances productivity but also promotes sustainable farming practices. By reducing dependency on guesswork and improving resource utilization, the system helps farmers achieve better outcomes and economic stability. This research contributes to the development of an efficient agricultural decision support system that bridges the gap between traditional farming and modern technological advancements [4].

II. LITERATURE ANALYSIS

The literature survey highlights the significant advancements in smart agriculture through the integration of modern technologies such as Machine Learning and remote sensing techniques. Various studies have focused on improving soil analysis, crop prediction, and agricultural monitoring systems. For instance, geo-object based machine learning methods have been effectively used for soil organic matter mapping, while IoT-based systems enable real-time monitoring of soil parameters such as moisture, temperature, and pH. Several researchers have applied algorithms like Random Forest, Decision Trees, K-NN, and Neural Networks to predict crop yield, pest population, and agricultural output.

Additionally, satellite imagery and NDVI-based approaches have been utilized for crop identification and yield estimation in complex farming regions. Statistical models and regression techniques have also been employed to forecast crop trends and optimal cultivation periods. Furthermore, innovative approaches such as wearable sensors for activity recognition and GSM-based greenhouse monitoring systems demonstrate the diversity of technological applications in agriculture. Overall, the reviewed studies emphasize the importance of data-driven decision-making and highlight the potential of machine learning techniques to enhance agricultural productivity, accuracy, and sustainability.

TABLE I: LITERATURE WORK

| Ref. | Authors & Year | Objective / Focus Area | Method / Algorithm | Key Findings / Contribution |
|------|---|-----------------------------------|--|---|
| [1] | Tianjun Wu , Jiancheng Luo, Wen Dong, Yingwei Sun , Liegang Xia , and Xuejian Zhang(2019) | Soil Organic Matter (SOM) Mapping | Geo-Object Based, Machine Learning (ML) | Effective fine-scale mapping of SOM using multi-source geospatial data. |
| [2] | .N.ANANTHI Divya J. Divya M. Janani V (2017) | Smart Soil Monitoring System | Internet of Things (IoT), Sensors, Microcontroller | Real-time collection and remote monitoring of soil parameters such as |

| | | | | |
|------|--|---|---|---|
| | | | | moisture, temperature, and pH. |
| [3] | Jyothi patil Dr A.Govardhan Dr V.D.Mytri | Predicting Pest Population Dynamics (Thrips Tabaci) | Feed Forward Multi-Layer Perceptron (MLP) Neural Network | Developed an intelligent system for predicting pest population dynamics in cotton crops. |
| [4] | Nidhi H Kulkarni Dr. G N Srinivasan Dr. B M Sagar Dr.N K Cauvery(2018) | Crop Recommendation System | Ensembling Technique (Combining ML Models) | Improved crop productivity by providing accurate crop recommendations based on soil parameters. |
| [5] | Asolkar & Bhadade (2015) | Greenhouse & Crop Monitoring | GSM (Global System for Mobile Communication) | Proposed a low-cost and efficient system for monitoring greenhouse parameters via mobile communication. |
| [6] | Md. Tahmid Shakoar, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty (2017) | Agricultural Production Output Prediction | Supervised Machine Learning (Decision Tree, K-NN) | Predicted crop production output to support cost-effective farming decisions. |
| [7] | Zhang Miao ^{1, 2} , Li Qiangzi ¹ , Wu Bingfang ¹ | Crop Identification in Fragmented Farmland | Multi-temporal Landsat Images, Object-Oriented Classification | Demonstrated accurate crop identification using satellite image time-series data. |
| [8] | D. Turgay AltılarAnıl Suat Terliksiz | Predicting Wheat Yield Trends | Statistical Methods (Holt-Winters, Dynamic Linear Models, Regression) | Compared statistical models for analyzing and predicting wheat yield trends. |
| [9] | Sanat Sarangi, Somya Sharma, Bhushan Jagyasi (2015) | Agricultural Activity Recognition | Smart-shirt, Wearable Sensors | Introduced wearable technology to recognize activities of agricultural workers. |
| [10] | Luminto a. Harlili, M, Weather | Predict Rice Cultivation Time | Multiple Linear Regression (MLR) | Predicted optimal rice planting time using weather parameters like temperature, humidity, and rainfall. |

III. CNN ALGORITHM

Algorithm Steps:-

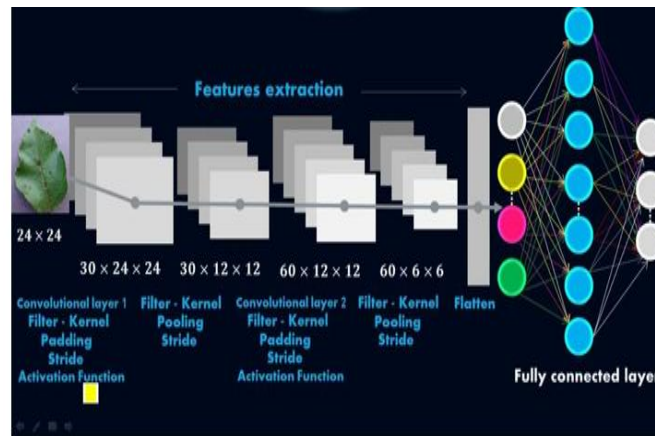


Figure 3.1: CNN Architecture

1. Data Collection:

We need a dataset of Crop images to train CNN. Popular datasets for Crop diseases detection include:

The PlantVillage dataset is a large, open-access collection of images of healthy and diseased plant leaves used extensively in machine learning research for disease detection. It contains over 54,000 high-quality RGB images, divided into 38 different classes representing both various plant species and specific plant diseases. Each image is labeled with its species and health status (healthy or Crop Disease Prediction and Solution Recommendation System Using Chatbots infected by a certain disease), enabling the training of deep learning models for automated plant disease classification.

2. Data Preprocessing:

Before feeding images into the CNN, preprocessing steps are needed:

- Image Resizing: Resize all images to a fixed dimension (e.g., 224x224 pixels) so they can be fed into the CNN uniformly.
- Normalization: Normalize pixel values to a range (0,1) by dividing each pixel value by 255.
- Data Augmentation: Apply transformations like rotation, flipping, and zooming to create more diverse data and avoid overfitting. This helps the model generalize better.
- Splitting the Dataset: Split the data into training, validation, and test sets (typically 70%20%-10%).

3. Building the CNN Model:

A CNN typically consists of several layers. Here's a breakdown of each layer and its function:

- Input Layer: The input will be an image (e.g., 224x224x3 for RGB images).
- Convolutional Layers: These layers extract features like edges, textures, and patterns from the image using filters.

- Convolution Operation: The image is passed through filters (e.g., 3x3) to extract feature maps.
- Pooling Layer: This reduces the spatial dimensions of the feature maps.
- Max Pooling: Retains the most important information by taking the maximum value in a filter region (e.g., 2x2).
- Dropout (Optional): To prevent overfitting, apply dropout, which randomly turns off some neurons during training.
- Fully Connected Layers: After several convolution and pooling layers, flatten the feature maps and connect them to fully connected (dense) layers. This is used for classification.
- Output Layer: The output layer will use a softmax or sigmoid activation function for disease detection.

Crop Disease Prediction and Solution Recommendation System Using Chatbots

Here is a basic CNN architecture for binary classification:

```
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout

model = Sequential()

# Convolution Layer 1
model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(224, 224, 3)))
model.add(MaxPooling2D(pool_size=(2, 2)))

# Convolution Layer 2
model.add(Conv2D(64, kernel_size=(3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

# Flatten the layers
model.add(Flatten())

# Fully Connected Layer
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))

# Output Layer
model.add(Dense(1, activation='sigmoid')) # For binary classification
```

Figure 3.2: CNN Architecture for Binary Classification

4. Training the CNN:

After building the model, we can train it on our dataset:

- Loss Function: Use binary cross-entropy since it is a binary classification task.
- Optimizer: Use Adam optimizer for faster convergence.
- Metrics: Track accuracy as the evaluation metric.
- Batch Size & Epochs: Choose batch size (e.g., 32) and the number of epochs (e.g., 20) depending on our hardware and dataset size.

Example Code:

```
history = model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=20,
```

Figure 3.3: CODE 1

5. Evaluation and Testing:

Once the model is trained, evaluate it on the test set:

- Accuracy: Check how accurate the model is in classifying the test images.
- Confusion Matrix: To check True Positives, True Negatives, False Positives, and False Negatives.
- Precision, Recall, F1-Score: Calculate precision and recall to ensure a good balance between Crop Disease Prediction and Solution Recommendation System Using Chatbots false positives and false negatives (important for medical diagnosis).

Example Code:

```
from sklearn.metrics import classification_report, confusion_matrix

y_pred = model.predict(X_test)
y_pred = (y_pred > 0.5)

print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

Figure 3.4: CODE 2

6. Deployment:

Once the model is trained and evaluated, we can deploy it:

- Model Serialization: Save the trained model using `model.save()`.
- Deployment Web: We can deploy the model using a web framework like Flask.

IV. WORKING METHODOLOGY

The working methodology of the proposed Crop Prediction and Solution Recommendation System is designed to provide accurate and real-time agricultural insights by integrating Machine Learning techniques. The system follows a systematic pipeline that includes data acquisition, preprocessing, model training, prediction, and recommendation.

Initially, soil data is collected from agricultural fields using IoT-based sensors. These sensors measure important parameters such as soil moisture, temperature, humidity, and nutrient levels (e.g., nitrogen, phosphorus, potassium). The collected data is transmitted through microcontrollers (such as Arduino) and stored in a centralized dataset. This real-time data acquisition ensures that the system reflects current field conditions.

Once the data is collected, it undergoes preprocessing to improve quality and consistency. This step includes handling missing values, removing noise, normalizing data, and performing feature scaling.



Data preprocessing is essential to ensure that the machine learning models perform efficiently and produce accurate results.

After preprocessing, the dataset is divided into training and testing sets. The training set is used to build predictive models using supervised learning algorithms such as Random Forest and Linear Regression. These models learn the relationship between soil parameters and crop yield. During this phase, the system also applies techniques such as pattern recognition to identify trends and correlations within the data.

In the prediction phase, the trained model analyzes new input data from the field and predicts the most suitable crop for cultivation along with the expected yield. The system further evaluates soil conditions to provide fertilizer recommendations and identify potential issues such as nutrient deficiencies, pests, or disease risks.

Finally, the output is presented to the user in an understandable format, often through a user interface or chatbot system. Farmers receive actionable insights, including crop suggestions, soil health status, and preventive measures. This end-to-end methodology ensures efficient decision-making, improved productivity, and supports the adoption of precision agriculture practices.

Dataset Description:

The dataset used in this system plays a crucial role in training and evaluating the machine learning models for accurate crop prediction and recommendation. It consists of both real-time sensor data and historical agricultural data collected from reliable sources.

The primary dataset includes soil parameters such as moisture, temperature, humidity, and essential nutrients like nitrogen (N), phosphorus (P), and potassium (K). These values are obtained using IoT-based sensors embedded in agricultural fields. In addition to real-time data, historical datasets are collected from agricultural universities, government repositories, and research institutions, which contain information about crop yield, soil fertility, rainfall, and environmental conditions.

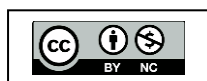
The dataset is structured with multiple attributes, where each record represents a specific soil sample or field condition. Common features include soil nutrient levels (N, P, K), pH value, moisture content, temperature, humidity, rainfall, and crop type. The target variable is typically the suitable crop or predicted yield corresponding to the given input conditions.

Before using the dataset, it undergoes preprocessing steps such as handling missing values, removing inconsistencies, encoding categorical variables, and feature scaling. The cleaned dataset is then divided into training and testing sets, usually in an 80:20 ratio, to build and evaluate machine learning models.

The combination of real-time and historical data ensures that the system captures both current field conditions and long-term agricultural patterns, thereby improving prediction accuracy and reliability. This comprehensive dataset enables the system to provide effective crop recommendations, fertilizer suggestions, and insights into soil health for better agricultural decision-making.

We are going to use "PlantVillage Dataset" from kaggle.com

DatasetLink: <https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated>





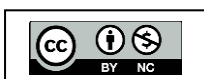
The PlantVillage dataset is a large, open-access collection of images of healthy and diseased plant leaves used extensively in machine learning research for disease detection. It contains over 54,000 high-quality RGB images, divided into 38 different classes representing both various plant species and specific plant diseases. Each image is labeled with its species and health status (healthy or Crop Disease Prediction and Solution Recommendation System Using Chatbots infected by a certain disease), enabling the training of deep learning models for automated plant disease classification.

The dataset, crop types are as follows,

1. Apple__Apple_scab
2. Apple__Black_rot
3. Apple__Cedar_apple_rust
4. Apple__healthy
5. Blueberry__healthy
6. Cherry_(including_sour)__Powdery_mildew
7. Cherry_(including_sour)__healthy
8. Corn_(maize)__Cercospora_leaf_spot Gray_leaf_spot
9. Corn_(maize)__Common_rust_
10. Corn_(maize)__Northern_Leaf_Blight
11. Corn_(maize)__healthy
12. Grape__Black_rot
13. Grape__Esca_(Black_Measles)
14. Grape__Leaf_blight_(Isariopsis_Leaf_Spot)
15. Grape__healthy
16. Orange__Haunglongbing_(Citrus_greening)
17. Peach__Bacterial_spot
18. Peach__healthy
19. Pepper_bell__Bacterial_spot
20. Pepper_bell__healthy
21. Potato__Early_blight
22. Potato__Late_blight
23. Potato__healthy
24. Raspberry__healthy
25. Soybean__healthy
26. Squash__Powdery_mildew

Crop Disease Prediction and Solution Recommendation System Using Chatbots

27. Strawberry__Leaf_scorch
28. Strawberry__healthy
29. Tomato__Bacterial_spot
30. Tomato__Early_blight



31. Tomato__Late_blight
32. Tomato__Leaf_Mold
33. Tomato__Septoria_leaf_spot
34. Tomato__Spider_mites Two-spotted_spider_mite
35. Tomato__Target_Spot
36. Tomato__Tomato_Yellow_Leaf_Curl_Virus
37. Tomato__Tomato_mosaic_virus
38. Tomato__healthy

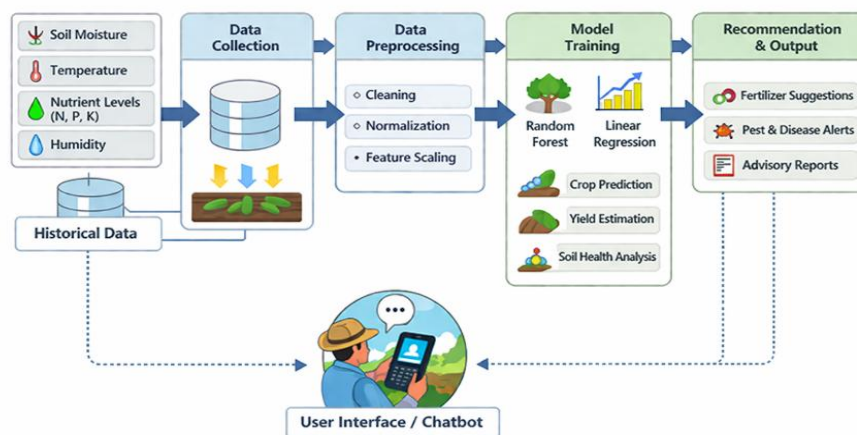


Figure 4.1: System Diagram

V. RESULTS AND DISCUSSION

The proposed Crop Prediction and Solution Recommendation System were evaluated using a combination of real-time soil sensor data and historical agricultural datasets. The system was implemented by integrating Machine Learning algorithms based data collection to ensure accurate and reliable predictions.

The performance of the system was analyzed using supervised learning models such as Random Forest and Linear Regression. Among these, the Random Forest algorithm demonstrated higher accuracy in predicting suitable crops due to its ability to handle complex and non-linear relationships between soil parameters and crop yield. Linear Regression, while effective for continuous value prediction such as yield estimation, showed comparatively lower accuracy in classification tasks.

The dataset was divided into training and testing sets, and the models were evaluated using performance metrics such as accuracy, precision, and error rate. The results indicated that the system achieved a significant improvement in prediction accuracy compared to traditional methods that rely on manual analysis or basic statistical techniques. The data pre-processing techniques such as normalization and handling missing values further enhanced model performance.

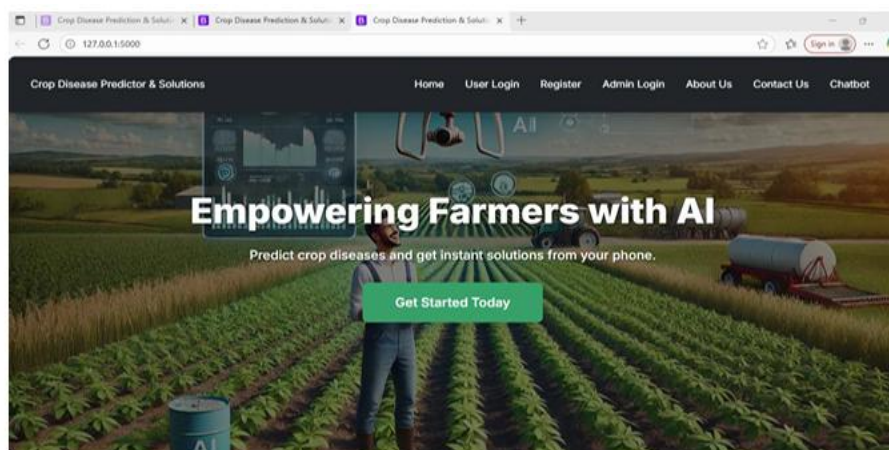
In addition to crop prediction, the system successfully provided fertilizer recommendations based on soil nutrient analysis. It was also capable of identifying potential risks such as nutrient deficiencies, pest attacks, and disease occurrence using pattern recognition techniques.

The integration of real-time sensor data improved the responsiveness of the system, enabling dynamic decision-making based on current field conditions.

The discussion highlights that the proposed system not only improves prediction accuracy but also enhances usability for farmers by providing actionable insights in a simple and understandable format. The machine learning supports precision agriculture by optimizing resource utilization, reducing wastage, and increasing crop productivity.

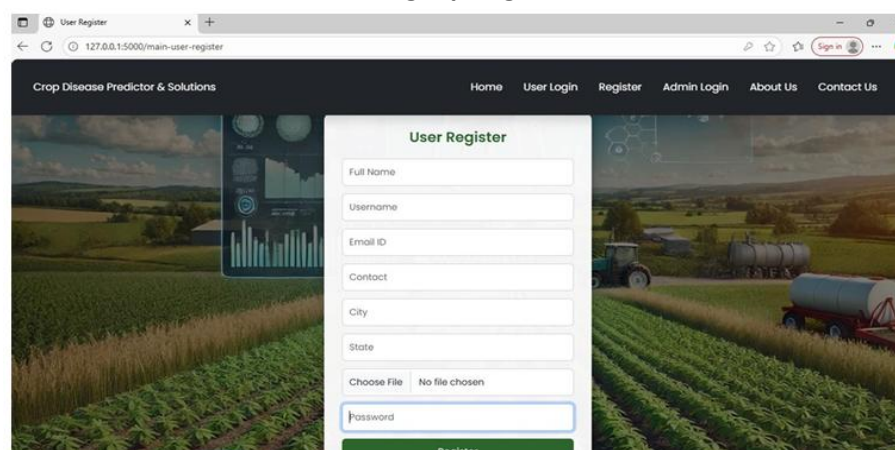
Overall, the results demonstrate that the system is efficient, reliable, and scalable for real-world agricultural applications. It provides a practical solution to modern farming challenges and contributes toward sustainable agriculture and improved farmer livelihoods.

Home Page



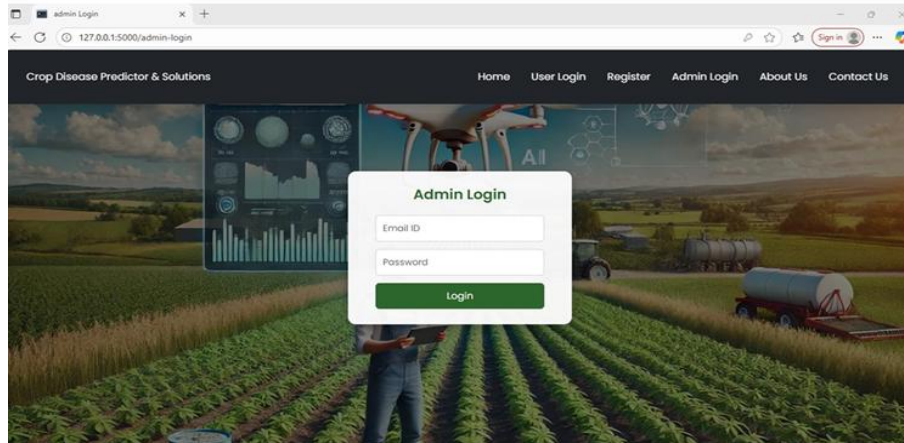
Screenshot 5.1: Home Page

Signup Page



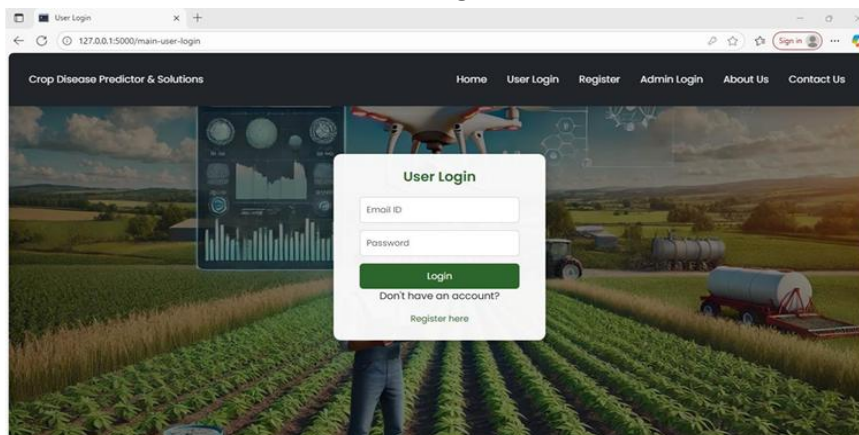
Screenshot 5.2: Signup Page

Admin Login Page



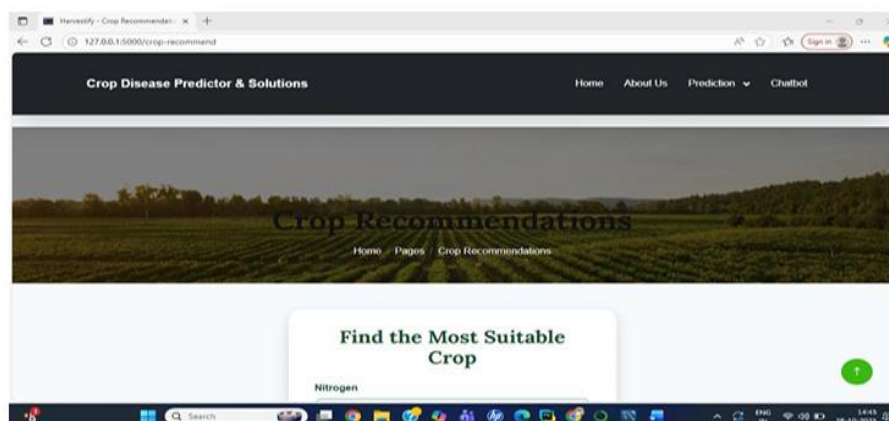
Screenshot 5.3: Admin Login Page

User Login



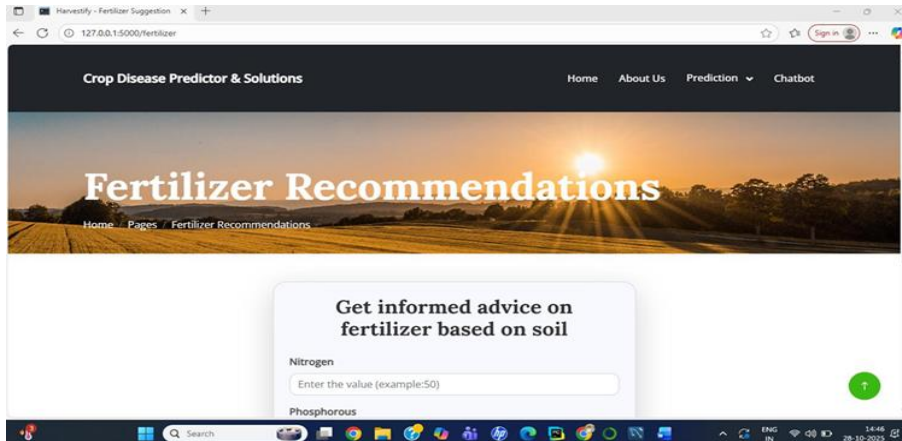
Screenshot 5.4: User Login Page

Crop Prediction



Screenshot 5.5: Crop Prediction Page

Fertilizer-Prediction



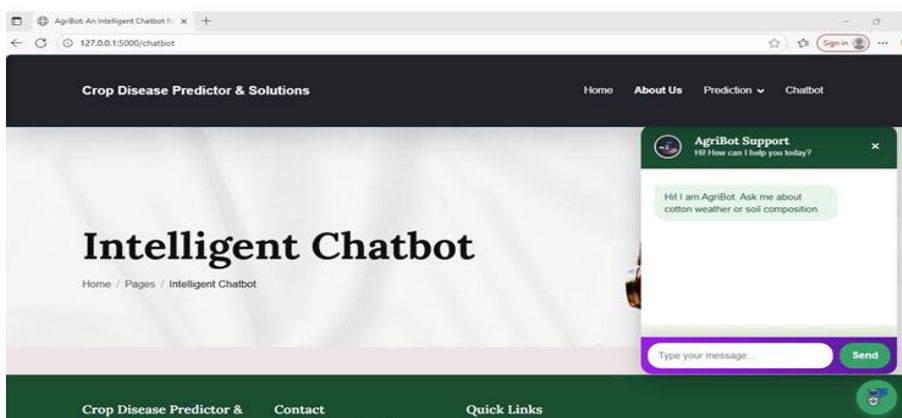
Screenshot 5.6: Fertilizer-Prediction

Disease Prediction

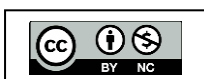


Screenshot 5.7: Disease Prediction

ChatBot



Screenshot 5.8: ChatBot





VI. CONCLUSION

This paper presents an intelligent Crop Prediction and Solution Recommendation System that leverages advanced technologies such as Machine Learning to enhance agricultural decision-making. The system effectively integrates real-time soil data acquisition through sensors with predictive analytics to recommend suitable crops, estimate yield, and provide fertilizer suggestions.

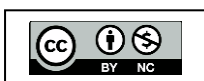
The implementation demonstrates that the use of algorithms such as Random Forest and Linear Regression significantly improves prediction accuracy compared to traditional farming practices. Data preprocessing and proper model training further enhance the reliability and efficiency of the system. Additionally, the system is capable of identifying potential risks such as soil nutrient deficiencies, pests, and diseases, thereby enabling preventive measures.

By providing farmers with accurate, timely, and easy-to-understand recommendations, the proposed system supports precision agriculture and promotes optimal utilization of resources. It reduces dependency on manual judgment and helps in increasing productivity and profitability. Moreover, the integration ensures continuous monitoring of field conditions, making the system dynamic and adaptable.

In conclusion, the proposed approach offers a scalable and efficient solution for modern agriculture. It not only improves crop yield and soil health management but also contributes to sustainable farming practices. Future enhancements may include the integration of advanced deep learning models, weather forecasting systems, and mobile-based applications to further improve system performance and accessibility for farmers.

REFERENCES

- [1] Tianjun Wu , Jiancheng Luo, Wen Dong, Yingwei Sun , Liegang Xia , and Xuejian Zhang, Geo Object-Based Soil Organic Matter Mapping Using Machine Learning Algorithms With Multi-Source Geo-Spatial Data, IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, 2019.
- [2] N.ANANTHI Divya J. Divya M. Janani V, IoT based Smart Soil Monitoring System for Agricultural Production, IEEE International Conference On Technological innovations in ICT for Agriculture and Rural development (TIAR_2017)
- [3] Jyothi patil Dr A.Govardhan Dr V.D.Mytri, An Intelligent System for Predicting Thrips Tabaci Linde Pest Population Dynamics Allied To Cotton Crop,
- [4] Nidhi H Kulkarni Dr. G N Srinivasan Dr. B M Sagar Dr.N K Cauvery, Improving Crop, Productivity Through A Crop Recommendation System Using Ensembling Technique, IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions(2018)
- [5] P. S. Asolkar Prof.Dr. U. S. Bhadade, An Effective Method of Controlling the Greenhouse and Crop Monitoring Using GSM, (2015).
- [6] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty, Agricultural Production Output Prediction Using Supervised Machine Learning Techniques,(2017).
- [7] Zhang Miao^{1, 2}, Li Qiangzi¹ , Wu Bingfang^{1,*}, Investigating the capability of multi- temporal Landsat images for crop identification in high farmland fragmentation regions.
- [8] D. Turgay Altılar Anıl Suat Terliksiz, Comparison of Statistical Methods for Predicting Wheat Yield Trends in Turkey.
- [9] Sanat Sarangi, Somya Sharma, Bhushan Jagyasi, Agricultural Activity Recognition with Smart-shirt and Crop Protocol, Global Humanitarian Technology Conference(2015).
- [10] Luminto a. Harlili, M, Weather Analysis to Predict Rice Cultivation Time Using Multiple Linear Regression to Escalate Farmer's Exchange Rate.





- [11] Bert LittleMichael SchuckingKenton Ross, High Granularity Remote Sensing and Crop Production over Space and Time: NDVI over the Growing Season and Prediction of Cotton Yields at the Farm Field Level in Texas, IEEE International Conference on Data Mining Workshops, (2008)
- [12] R. Sujatha, P. Isakki, A Study on Crop Yield Forecasting Using Machine Learning Techniques, International Journal of Advanced Research in Computer Science, 2016.
- [13] K. R. Patil, S. K. Patil, Crop Prediction System Using Machine Learning, International Journal of Innovative Research in Computer Science, 2018.
- [14] A. Khanna, S. Kaur, Evolution of Internet of Things (IoT) and its Significant Impact in the Field of Precision Agriculture, Computers and Electronics in Agriculture, 2019.
- [15] J. Jeong, N. Resop, N. Mueller, et al., Random Forests for Global and Regional Crop Yield Predictions, PLOS ONE, 2016.
- [16] S. Ramesh, B. Vishnu Vardhan, Analysis of Crop Yield Prediction Using Data Mining Techniques, International Journal of Computer Science and Information Technologies, 2015.
- [17] Y. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine Learning in Agriculture: A Review, Sensors Journal, 2018.
- [18] K. Mekala, P. Viswanathan, A Survey: Smart Agriculture IoT with Cloud Computing, International Conference on Microelectronics, Computing and Communications, 2017.
- [19] S. Wolfert, L. Ge, C. Verdouw, M. J. Bogaardt, Big Data in Smart Farming – A Review, Agricultural Systems, 2017.
- [20] A. Kamilaris, F. X. Prenafeta-Boldú, Deep Learning in Agriculture: A Survey, Computers and Electronics in Agriculture, 2018.

