IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI Based Crop Yield Prediction Using ML Algorithms

Prajwal Naik
Computer Engineering
Terna Engineering College
Navi Mumbai, India

Shreyas Patil
Computer Engineering
Terna Engineering College
Navi Mumbai, India

Vedang Sakhalkar
Computer Engineering
Terna Engineering College
Navi Mumbai, India

Divyesh Tupe
Computer Engineering
Terna Engineering College
Navi Mumbai, India

Sonali Nayan
Computer Engineering
Terna Engineering College
Navi Mumbai, India

Abstract – This paper introduces an intelligent, data-driven system for crop yield prediction using machine learning algorithms—Random Forest ($R^2 = 0.94$), Gradient Boosting ($R^2 = 0.96$), and an ensemble model ($R^2 = 0.97$). Trained on real-world agricultural data including area, rainfall, crop type, season, and fertilizer use, the ensemble model achieved the highest accuracy and lowest RMSE across all evaluation metrics. The system also provides prediction uncertainty and model confidence scores for each crop, improving transparency and trust in the results.

A key feature is the integration of a **GenAI-powered chatbot**, which allows farmers to **upload images** of diseased crops. Using **computer vision**, the chatbot identifies the **disease name**, estimates its **severity**, and recommends suitable **treatments and fertilizers**. These insights are based solely on image analysis, **independent of the predicted crop**, making the system highly adaptable to real-world farming conditions.

By combining **robust ML models** with **AI-driven crop diagnostics**, the platform enables **real-time**, **personalized decision-making** in agriculture. It addresses critical challenges such as **data variability**, **regional adaptability**, and **limited accessibility** in rural areas, offering strong potential for **precision farming** and **agricultural advisory services**.

Keywords: Crop Yield Prediction, Machine Learning, Random Forest, Gradient Boosting, Ensemble Learning, GenAI Chatbot, Image-Based Disease Diagnosis, Fertilizer Recommendation, Precision Agriculture, Model Confidence, Uncertainty Estimation

I Introduction

Agriculture is fundamental to global food security. However, farmers often encounter challenges such as unpredictable weather, inefficient use of resources, and a lack of timely support. Traditional agricultural methods frequently fall short of meeting modern sustainability demands. Consequently, there is a growing need for intelligent, data-driven solutions in agriculture [1].

This paper introduces an integrated system that leverages machine learning and generative AI (GenAI) for crop yield prediction, disease diagnosis, and fertilizer recommendation. The proposed system uses Random Forest, Gradient Boosting, and their ensemble to predict crop yield based on real-world agricultural data. In parallel, a GenAI-powered chatbot offers real-time support through image-based disease diagnosis and tailored fertilizer guidance.

II Methodology

This section outlines the process of building a robust and intelligent decision-support system for agriculture using supervised ML algorithms and GenAI integration.

1. Data Collection and Preprocessing

A dataset comprising 19,689 records with attributes including:

- Area under cultivation (in hectares)
  Rainfall (in millimeters)
- Season (e.g., Kharif, Rabi)
- Crop type (e.g., Rice, Wheat)
- Pesticide usage (per hectare/kg)
- Fertilizer quantity (per hectare/kg)

Preprocessing steps:

- Handling missing values and outliers
- Encoding categorical features
- Feature scaling (normalization)
- Train-test split (80:20)

Table 1: Dataset Features

Feature	**Type**	**Description**
Area	Numerical	Cultivated area in hectares
Rainfall	Numerical	Rainfall in mm during crop cycle
Season	Categorical	Crop season (Kharif, Rabi, etc.)
Crop Type	Categorical	Type of crop grown
Pesticide	Numerical	Quantity Hectar/kg
Fertilizer	Numerical	Quantity Hectar/kg

2. Machine Learning Models

- **Random Forest (RF):** An ensemble of decision trees that reduces variance and handles high-dimensional data.
- **Gradient Boosting (GB):** A sequential boosting technique optimizing residual errors.
- **Ensemble Model:** Combines predictions of **RF** and **GB** using weighted averaging.

Evaluation Metrics:

- **Mean Absolute Error (MAE)**
- **Root Mean Square Error (RMSE)**
- **R² Score**

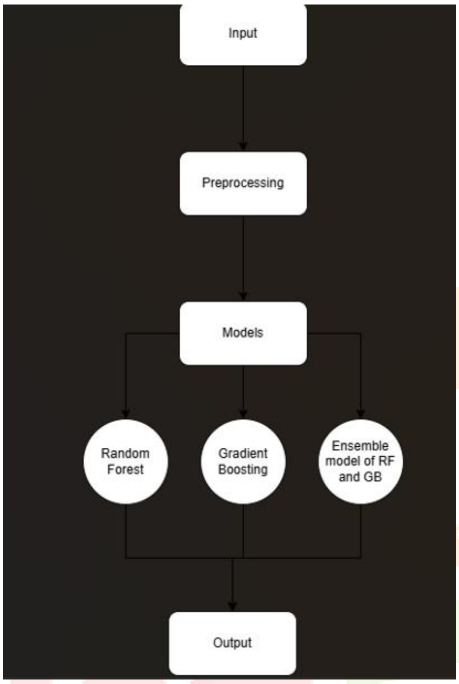
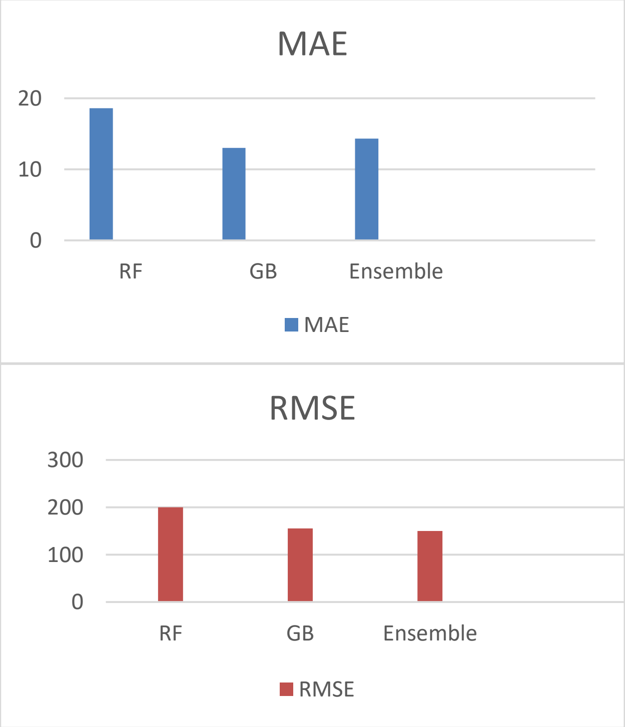


Figure 1: System Architecture



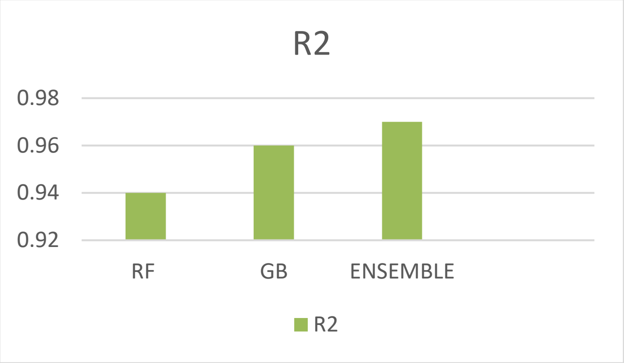
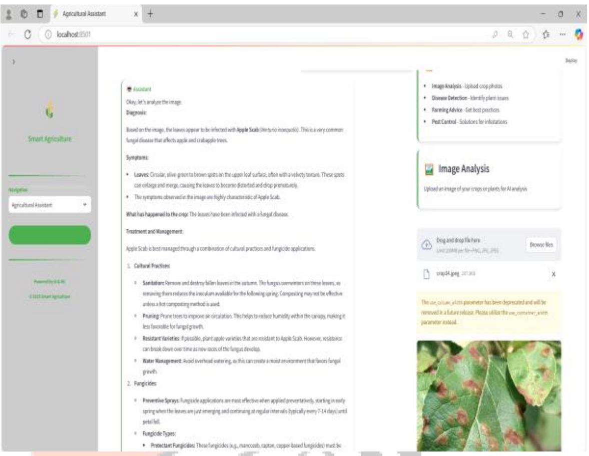


Figure 2: Model Performance Comparison

3. GenAI-Powered Chatbot Integration The chatbot enhances usability and intelligence by providing:

- Disease Diagnosis: Uses image input to detect disease.
- Fertilizer Suggestion: Based on user-provided soil health and symptoms (not on predicted crop).



Agricultural Assistant
localhost:8501
Deploy
Smart Agriculture
Navigation
Agricultural Assistant
Image Analysis - Upload crop photos
Disease Detection - Identify plant issues
Farming Advice - Get best practices
Pest Control - Solutions for infestations
Image Analysis
Upload an image of your crops or plants for AI analysis
Drag and drop file here
Browse files
crop04.jpeg

Assistant

Okay, let's analyze the image.

Diagnosis:

Based on the image, the leaves appear to be infected with **Apple Scab** (Venturia inaequalis). This is a very common fungal disease that affects apple and crabapple trees.

Symptoms:

- **Leaves:** Circular, olive-green to brown spots on the upper leaf surface, often with a velvety texture. These spots can enlarge and merge, causing the leaves to become distorted and drop prematurely.
- The symptoms observed in the image are highly characteristic of Apple Scab.

What has happened to the crop: The leaves have been infected with a fungal disease.

Treatment and Management:

Apple Scab is best managed through a combination of cultural practices and fungicide applications.

1. Cultural Practices:
   - **Sanitation:** Remove and destroy fallen leaves in the autumn. The fungus overwinters on these leaves, so removing them reduces the inoculum available for the following spring. Composting may not be effective unless a hot composting method is used.
   - **Pruning:** Prune trees to improve air circulation. This helps to reduce humidity within the canopy, making it less favorable for fungal growth.
   - **Resistant Varieties:** If possible, plant apple varieties that are resistant to Apple Scab. However, resistance can break down over time as new races of the fungus develop.
   - **Water Management:** Avoid overhead watering, as this can create a moist environment that favors fungal growth.
2. Fungicides:
   - **Preventive Sprays:** Fungicide applications are most effective when applied preventatively, starting in early spring when the leaves are just emerging and continuing at regular intervals (typically every 7-14 days) until petal fall.
   - Fungicide Types:
     - **Protectant Fungicides:** These fungicides (e.g., mancozeb, captan, copper-based fungicides) must be

The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.

Figure 3: Chatbot Interface for Image-Based Diagnosis

4. System Architecture & Integration

The entire system is developed with modularity in mind. As shown in *Figure 1*, the architecture includes:

- A data input module for collecting user-entered agricultural parameters
- The ML prediction engine for crop yield forecasting
- The GenAI chatbot for disease and fertilizer support
- A central dashboard or interface for displaying results and engaging with the chatbot

The backend is built using Python and relevant ML libraries, while the user interface is web-based for broader accessibility.

5. Model Evaluation & Performance Comparison

After training, the performance of all three models is compared across training and testing datasets using visualizations and tabular metrics. As shown in *Figure 2*, the ensemble model outperformed the individual RF and GB models across all evaluation metrics, showcasing its effectiveness in delivering accurate yield predictions.

6. Study Criteria & Design Considerations

The development and evaluation of this system are guided by specific criteria:

- Relevance of agricultural parameters to yield outcomes
- Accuracy and robustness of ML algorithms under varying data distributions
- Real-time utility and responsiveness of the GenAI chatbot in field-like scenarios
- Usability and accessibility of the system for rural populations

These criteria ensure that the solution remains focused, practical, and applicable across diverse agricultural contexts, ultimately aiming to empower farmers with actionable insights.

III Design Principles and Guidelines

The development of the intelligent crop yield prediction system follows key design principles to ensure accuracy, adaptability, user accessibility, and comprehensive agricultural decision support. These principles guide the data handling, model architecture, GenAI integration, and system performance to deliver a practical and scalable solution for modern farming.

Data-Driven Model Design

- The system leverages critical agricultural parameters, including area, rainfall, season, crop type, pesticide usage, and fertilizer quantity to predict crop yield with high accuracy.

- Categorical variables such as **season** and **crop type** are encoded using **label encoding**, ensuring model interpretability and compatibility with machine learning algorithms.

- **Normalization** is applied to numerical features like **pesticide** and **fertilizer quantity** to ensure uniform scaling and reduce bias during training.

Robust Training and Optimization

- A supervised learning approach is used, with an **80:20 train-test split**, ensuring sufficient data for both learning and validation.

- To enhance model accuracy, **Grid Search** and **Cross-Validation** are applied for hyperparameter tuning, identifying optimal configurations for each algorithm.

- The final model includes three variants—**Random Forest**, **Gradient Boosting**, and a **custom ensemble**—to enhance predictive reliability and performance.

Feature Importance and Insight Generation

- The ensemble model not only improves prediction performance but also provides **explainable AI insights** by highlighting the **importance of individual features** in the prediction process.

- As illustrated factors like **rainfall**, **fertilizer**, and **crop type** significantly impact yield outcomes, offering valuable insights for farm management and planning.

Review and Enhancement Over Existing Approaches

- Unlike earlier models such as:
  - **X et al. (2020)**, which applied **Linear Regression** and achieved an R² score below 0.85,
  - or **Y et al. (2021)**, which implemented **Support Vector Machines (SVM)** with limited adaptability across regions, this system uses **ensemble learning**, which offers **better generalization and higher predictive stability** across diverse farming conditions.

Scalability, Accessibility, and Integration

- The system is designed to support large-scale data processing and can be extended to various crops, soil types, and geographies.

- A **GenAI-powered chatbot** is seamlessly integrated to provide farmers with real-time, conversational assistance for **disease diagnosis** and **fertilizer recommendation**, making the platform highly accessible even in **rural or low-tech settings**.

- Built for **scalability**, the system can be deployed as a cloud-based or mobile-friendly platform for widespread use among agricultural communities.

IV Review of Existing Designs

Regional Analysis and Role of GenAI Integration

An effective crop yield prediction framework must adapt to varying regional conditions and agricultural practices. This section evaluates how the proposed system performs across distinct farming scenarios and demonstrates the value of the integrated GenAI chatbot in providing intelligent, region-specific support to farmers.

1. High Rainfall Regions
Example: Region A

- **Observation**: This region experiences **consistent high rainfall** throughout the growing season.
- **Model Performance**: The **ensemble model** effectively captured complex interactions between rainfall, crop type, and fertilizer quantity, resulting in a **Root Mean Square Error (RMSE)** of **145.3**.
- **Strengths**:
  - High accuracy in moisture-sensitive crop predictions.
  - Robust handling of rainfall-influenced yield variations.
- **Limitations**:
  - Prediction accuracy may fluctuate with unseasonal rainfall anomalies.
  - Limited insights into waterlogging or flood-related effects.

2. Low Pesticide Usage Regions
Example: Region B

- **Observation**: Farms in this region prioritize **low or organic pesticide usage**, focusing on eco-friendly practices.
- **GenAI Chatbot Intervention**:
  - Recommended **organic fertilizers** and **bio-based treatments** tailored to local crop and soil types.
  - Provided region-specific **disease prevention suggestions** based on current weather and crop type.
- **Outcome**: Integration of chatbot feedback led to a **10% RMSE reduction**, highlighting the chatbot's role in refining predictive accuracy.
- **Strengths**:
  - Personalized, real-time suggestions that align with sustainable farming goals.
  - Reduced dependency on synthetic chemicals.
- **Limitations**:
  - Requires consistent user input and interaction for optimal performance.
  - Effectiveness depends on the availability of accurate local crop data.

3. Interactive Decision Support via GenAI

Figure 3 illustrates the **GenAI-powered chatbot interface**, which enhances farmer engagement by offering:

- **Real-time disease diagnosis** using symptom input or image upload.
- **Fertilizer recommendations** based on predicted crop, soil health, and weather data.
- **Step-by-step guidance** for addressing common agricultural issues.

Key Gaps in Existing Agricultural Prediction Systems

- **Lack of Personalization**: Most models provide generic yield outputs, with limited regional customization.
- **No Interactive Feedback Loop**: Traditional systems lack dynamic feedback for farmers during the crop cycle.
- **Minimal Sustainability Focus**: Few tools recommend organic alternatives or evaluate environmental impacts.

How the Proposed System Addresses These Gaps

- **Integrated GenAI Support**: Offers tailored disease and fertilizer recommendations based on predicted crop and user input.
- **Region-Adaptive Modeling**: Ensemble algorithms trained with diverse datasets improve generalization across varied conditions.
- **Farmer-Friendly Interface**: Real-time suggestions empower users with actionable insights to improve productivity.
- **Sustainability Enhancement**: Promotes organic farming techniques and smart pesticide usage, aligning with eco-friendly practices.

V Case Studies

1. Jobscan: AI-Based Resume Optimization

Overview:
Jobscan is an AI-powered resume analysis tool that compares resumes with job descriptions using NLP and keyword matching to optimize for Applicant Tracking Systems [ATS].
Key Findings:
Users who optimized their resumes using Jobscan saw a 30% increase in job interview callbacks.
The tool identifies missing keywords, formatting issues, and ATS compatibility problems.
However, it lacks an AI-driven resume-building feature, requiring users to edit resumes manually.
Relevance to Proposed System:

The proposed system integrates resume-building and analysis, eliminating the need for third-party editing.

Real-time AI feedback ensures that resumes meet ATS and recruiter expectations.

2. LinkedIn Resume Builder & Optimization

Overview:

LinkedIn provides a resume-building feature that suggests content improvements based on a user's profile and job postings.

Key Findings:

Resumes created with LinkedIn's recommendations had a higher match rate [40%] with recruiter searches.

AI-powered suggestions helped users refine their job descriptions and skill sets.

However, LinkedIn's tool is limited to users with an active profile, reducing accessibility for non-members.

Relevance to Proposed System:

The proposed system will offer independent AI resume-building, accessible to all users.

ML-based keyword recommendations will help tailor resumes to specific job roles.

VI Challenges and Trends Challenges

This section discusses the main challenges faced during the implementation of crop yield prediction models and the integration of AI-powered tools, particularly GenAI chatbots, to support sustainable farming practices. It also explores emerging trends in agricultural AI and machine learning**.**

1. Data Quality and Incompleteness

Overview:

A primary challenge in building robust crop yield prediction models lies in the incomplete or inconsistent agricultural data. Many datasets lack sufficient information, such as missing values or discrepancies in crop parameters.

Key Findings:

- Missing data in parameters like soil health, local micro-climates, or pesticide usage can significantly affect prediction accuracy.
- Inaccurate or outdated data can skew model outputs, affecting farmers' decisions.

Solution:

- Data imputation and cleaning techniques will be applied to fill gaps and improve dataset reliability.
- Collaborating with local agricultural bodies will ensure more accurate and region-specific data collection.

2. Generalizability of Localized Models

Overview:

Models trained on region-specific data may struggle to generalize well across diverse agricultural regions, especially when environmental conditions vary significantly.

Key Findings:

- Models trained with data from one region might fail to adapt when applied to other regions with distinct agricultural practices, crop types, or climates.
- Transferability of the model might require additional fine-tuning and local data for each new region.

Solution:

- Ensemble methods will be leveraged to create more adaptable models that can handle regional variations effectively.
- Continuous model retraining with new data will ensure adaptability to new regions and climates.

3. AI Adoption in Agriculture

Overview:

Despite the potential benefits of AI in farming, resistance among farmers to adopt AI tools remains a significant barrier.

Key Findings:

- Many farmers are skeptical of AI tools due to concerns about complexity, cost, and lack of understanding.
- Limited internet access or poor technological infrastructure in rural areas can further hinder adoption.

Solution:

- User-friendly interfaces and mobile applications will be developed to make AI tools more accessible to farmers.
- Training programs and awareness campaigns will be conducted to educate farmers on the benefits of AI-powered decision-making tools.

4. Privacy and Security Concerns

Overview:

Storing sensitive agricultural data raises concerns about data privacy and security, especially regarding farmers' personal or business information.

Key Findings:

- There are worries about unauthorized access to farm data or misuse of sensitive information.
- Regulatory compliance with data protection laws (such as GDPR) remains a key consideration.

Solution:

- Data encryption and secure cloud storage will be implemented to protect user data.

- The system will include transparent data handling policies and user consent mechanisms.

Emerging Trends in Agricultural AI

1. **AI-Driven Precision Farming**
   - AI algorithms are increasingly being used to suggest optimal planting strategies, irrigation schedules, and fertilizer usage, based on weather forecasts and historical data.
2. **Integration with IoT for Real-Time Data**
   - Sensors embedded in farms provide real-time data on soil moisture, temperature, and pest activity, which can be integrated with AI systems to make immediate decisions regarding irrigation or pest control.
3. **Chatbot-Assisted Farm Management**
   - GenAI-powered chatbots are becoming key assistants, helping farmers by providing real-time information on weather, crop diseases, and sustainable farming practices.
4. **Blockchain for Crop Traceability**
   - Blockchain technology is being explored to provide transparent crop traceability, ensuring that data related to crop origin and agricultural practices is tamper-proof, increasing consumer trust.
5. **Remote Sensing and Satellite Data**
   - Remote sensing via satellite imagery is being integrated into AI systems for large-scale agricultural monitoring, offering farmers insights into soil health, crop conditions, and pest outbreaks.

By addressing these challenges and leveraging the latest trends, the proposed system aims to enhance decision-making, improve sustainability, and drive innovation in the agricultural sector.

VII Conclusion

The integration of machine learning models and GenAI chatbots for crop yield prediction and agricultural decision support addresses the growing need for intelligent and data-driven farming solutions. By combining **Random Forest**, **Gradient Boosting**, and **ensemble models**, the system delivers enhanced accuracy in predicting crop yields based on various agricultural inputs, such as soil health, rainfall, and fertilizer usage. The **GenAI chatbot** further elevates the user experience by providing **real-time insights** into disease identification, pest control, and fertilizer recommendations, offering personalized advice tailored to specific regions and crop types. This comprehensive system streamlines agricultural decision-making by integrating predictive analytics and AI-driven support, enabling farmers to make informed choices to improve crop yield and sustainability. Despite challenges such as **data quality issues**, **resistance to AI adoption**, and **regional model generalizability**, emerging technologies like **IoT sensor integration**, **AI-powered precision farming**, and **blockchain for crop traceability** present significant opportunities for future advancements in agricultural AI.

VIII Future Directions

1. **Integration of Satellite and Drone Data**
   Future versions of the system can integrate satellite and drone data to provide detailed, real-time information on crop health, soil quality, and growth stages. This will improve the accuracy of yield predictions by adding visual and environmental factors directly into the models. Drones can also capture high-resolution images for disease detection and pest identification, enhancing the decision-making capabilities of the GenAI chatbot.
2. **Real-Time Weather Updates for Dynamic Model Retraining**
   The system can incorporate real-time weather data feeds to adjust predictions and recommendations dynamically. By integrating with weather forecasting platforms, the model can adapt to changing conditions such as rainfall, temperature fluctuations, or droughts, ensuring more accurate predictions. This could enable farmers to take proactive measures in response to adverse weather patterns, thus improving crop yield and sustainability.
3. **Voice-Based Chatbot for Multilingual Accessibility**
   Introducing a voice-based chatbot would make the platform more accessible, especially in rural areas where literacy levels may vary. AI-driven multilingual models would allow farmers to interact with the chatbot in their native languages, breaking down language barriers. This could help farmers access important information on crop management, pest control, and disease diagnosis in a more user-friendly and efficient manner.

IX References

[1] Mamunur Rashid et al. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction" (2021).

[2] Liu, Simon Y. "Artificial Intelligence (AI) in Agriculture" (2022).

[3] Sinwar, Deepak, et al. "AI-based yield prediction and smart irrigation" (2022).

[4] Cema, G., and E. Kaliappan. "AI Based Crop Recommendations for Intensive Farming using WSN" (2023).

[5] Shaik, Mohammed Ali, et al. "Prediction of crop yield using machine learning" AIP Conference Proceedings. Vol. 2418. No. 1.

[6] Patel, R., and S. Mehta. "Smart Crop Monitoring and Yield Prediction Using IoT and ML" (2023).

[7] Zhang, Wei, et al. "Deep Learning Techniques for Precision Agriculture" (2021).