International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

# Meteorological Data and Pesticide Usage for Crop Yield Prediction Using Machine Learning

**Mrs. Karthigadevi[1], A. Afrin Mekhanas[2], A. Iswarya[3]**

[1]Associate Professor, Computer Science and Engineering, K. L. N. College Of Engineering, Madurai
[2,3]Student 4th year, Computer Science and Engineering, K. L. N. College Of Engineering, Madurai

**Abstract**
The increasing variability in climate and excessive dependence on chemical inputs have made agricultural productivity highly unpredictable in recent years. Accurate crop yield prediction is therefore essential for ensuring food security, efficient resource utilization, and sustainable farming practices. This study presents a machine learning–based prediction system that integrates meteorological data and pesticide usage records to forecast agricultural yields with higher accuracy and reliability. The proposed model combines both temporal and non-temporal data sources, capturing the intricate relationships between weather conditions, pesticide applications, and crop performance. A variety of machine learning algorithms—XGBoost, Long Short-Term Memory (LSTM), CatBoost, and Graph Neural Networks (GNNs)—were implemented and compared using performance metrics such as $R^2$, MAE, and RMSE. The hybrid LSTM–XGBoost approach achieved the most efficient balance between predictive power and computational performance, providing more stable yield forecasts across different crop and regional datasets. The model also identifies the most influential environmental and chemical factors contributing to productivity changes, offering interpretable insights for decision-making. To enhance accessibility and practical utility, a Streamlit-based interactive dashboard was developed for real-time data visualization, yield prediction, and comparative model analysis. This interface allows farmers and agricultural analysts to explore prediction outcomes, observe climatic trends, and evaluate pesticide management strategies in an intuitive, user-friendly environment. Overall, the proposed system contributes to data-driven precision agriculture by leveraging artificial intelligence to support adaptive crop planning, reduce environmental risks, and promote sustainable production. Future expansions may include IoT-based sensor integration, satellite imagery, and cloud computing for scalable real-time monitoring and region-specific optimization.

**Keywords:** Crop Yield Prediction, Machine Learning, Meteorological Data, Pesticide Usage, LSTM, XGBoost, CatBoost, GNN, Streamlit Dashboard, Sustainable Agriculture.

## 1. INTRODUCTION

Agriculture remains the backbone of the Indian economy, contributing significantly to employment and food security. However, unpredictable climate patterns, excessive pesticide usage, and soil degradation have made it increasingly difficult to achieve consistent crop yields. In this context, **machine learning (ML)** offers powerful data-driven tools to model complex, non-linear relationships between environmental factors and crop performance. This project, titled **"Meteorological Data and Pesticide Usage for Crop Yield Prediction Using Machine Learning,"** aims to forecast crop productivity by analyzing a combination of **meteorological parameters (temperature, rainfall, humidity)** and **pesticide**

**usage records**. By leveraging advanced ML algorithms, the system can identify key parameters influencing yield and provide accurate predictions that enable better planning and resource management. The proposed solution also integrates a real-time dashboard using Streamlit to display predictions, compare model outputs, and provide insights to farmers and agricultural stakeholders. The system's predictive capabilities contribute to climate-resilient agriculture, efficient pesticide management, and sustainable farming practices. Globally, food production systems are facing increasing stress due to climate change, pest outbreaks, and environmental degradation. The Food and Agriculture Organization (FAO) reports that yield losses from improper pesticide use and extreme weather events have risen significantly over the past decade. In such a scenario, developing intelligent predictive systems becomes crucial for enhancing crop management and preventing economic losses. The integration of machine learning models enables proactive planning by providing insights into future yield outcomes based on historical and real-time data. Unlike conventional statistical approaches, ML models adapt to changing environmental patterns, improving accuracy and scalability across multiple crops and regions. Thus, the proposed work serves as a step toward smart agriculture—where data, algorithms, and visualization collectively empower farmers and policymakers to make informed, sustainable, and timely decisions.

## 2. Objective of the System

**The objectives of the proposed system are as follows:**

1. To develop a machine learning–based crop yield prediction model using meteorological and pesticide usage data.
2. To analyze the impact of climate factors and pesticide intensity on crop productivity.
3. To implement and compare multiple ML algorithms such as XGBoost, LSTM, GNN, and CatBoost for model performance evaluation.
4. To enhance prediction accuracy through feature engineering and hyperparameter optimization.
5. To design a Streamlit-based dashboard that visualizes predictions and model comparisons in real time.
6. To support farmers and policymakers in making data-driven agricultural decisions that promote sustainable farming.

## 3. Literature Review

Recent advancements in artificial intelligence and data analytics have significantly influenced agricultural decision systems.

**M.D. Jiabul Hoque et al. (2024)** in their IEEE Access paper "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning" demonstrated that combining weather and pesticide data enhances prediction accuracy across Indian crops using Gradient Boosting and KNN models. **Uppugunduri Nikhil et al. (2024)** proposed a model for South Indian regions using soil and climatic parameters, achieving a 96.15% $R^2$ score with Extra Trees Regressor. Similarly, **Yamuna T. K. et al. (2024)** developed an SVM-based yield prediction model that reduced uncertainty for Maharashtra farmers through accurate soil–climate analysis. Traditional regression models often fail to capture non-linear relationships between multiple factors. Hence, modern ML models like **LSTM (for time-series data)** and **XGBoost (for structured data)** provide higher reliability and scalability. This project extends previous research by incorporating a **hybrid LSTM–XGBoost framework** and adding pesticide-related insights, which are often overlooked in yield prediction systems. **Singh, R. and Mehta, K. (2025)** in their study "Deep Learning Applications in Crop Yield Estimation under Climate Uncertainty" published in

Agricultural Informatics Journal, highlighted the importance of using deep neural networks for multi-variable prediction tasks. Their findings showed that combining meteorological, soil, and chemical data using LSTM-based architectures improved prediction stability under extreme weather conditions. This work reinforces the potential of deep learning to manage data complexity in agriculture and supports the adoption of hybrid ML approaches like the one proposed in this study.

## 4. 4.Methodologies

The proposed methodology involves data acquisition, preprocessing, model training, evaluation, and visualization.

### 4.1 Data Acquisition

Data were collected from agricultural databases and weather sources, including:

1. Meteorological parameters: temperature, rainfall, and humidity
2. Pesticide usage data
3. Soil health indicators
4. Crop yield records

### 4.2 Data Preprocessing

1. Cleaning missing and inconsistent data
2. Feature scaling and normalization
3. Feature selection using correlation analysis
4. Splitting into training and testing datasets

### 4.3 Model Development

Four models were implemented:

1. **XGBoost** for high-accuracy regression
2. **LSTM** for capturing temporal weather trends
3. **CatBoost** for handling categorical variables
4. **Graph Neural Network (GNN)** for spatial relationships

A hybrid LSTM–XGBoost model was built to leverage both temporal learning and boosting-based optimization.

### 4.3 Model Evaluation

Models were evaluated using:

- $R^2$ (Coefficient of Determination)
- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)

### 4.4 Visualization

Results were visualized using a Streamlit dashboard**,** providing:

- Real-time yield predictions
- Model performance comparison charts
- Interactive visual insights for farmers and researchers

## 5. Detailed Analysis

The comparative analysis showed that the hybrid LSTM–XGBoost model achieved the best performance, with high $R^2$ and low RMSE scores.

- **XGBoost** effectively captured non-linear dependencies.
- **LSTM** handled time-series patterns of rainfall and temperature variations.
- **CatBoost** provided strong generalization for pesticide and soil variables.
- **GNN** contributed spatial context for geographically distributed crops.

This combination yielded a robust model capable of forecasting yields with high precision across varying climatic zones. Feature importance analysis indicated that **rainfall, pesticide quantity, and soil moisture** were the top predictors affecting yield. The Streamlit dashboard enabled real-time visualization, simplifying the understanding of yield trends and helping end-users interpret ML outputs for practical use.

## 6. System Functionality

The system architecture seamlessly integrates the frontend, backend, and data processing modules to deliver an efficient crop yield prediction platform. The frontend is built using Streamlit, providing a user-friendly dashboard where users can input relevant data such as weather parameters, soil characteristics, and pesticide usage, while also visualizing the resulting predictions and analytics. The backend comprises Python scripts and Flask APIs that manage model execution and computation, leveraging machine learning libraries such as Pandas, NumPy, Scikit-learn, TensorFlow, and CatBoost for data handling, feature engineering, and model training. The system relies on CSV or Excel files for storing both input data and prediction outputs, ensuring simplicity and flexibility in data management. Once the models are integrated, they generate yield predictions and associated metrics, which are dynamically visualized on the Streamlit dashboard. This includes comparative plots, predicted values, and feature correlation analyses, allowing users to understand not only the expected outcomes but also the driving factors behind them. The overall design ensures a smooth flow from data preprocessing to model computation and real-time visualization, enabling stakeholders to make informed decisions based on accurate, interpretable predictions.

## 7. Conclusion

This project successfully demonstrates the potential of machine learning in enhancing agricultural forecasting by integrating diverse datasets, including meteorological parameters and pesticide usage. By employing advanced algorithms such as XGBoost, LSTM, CatBoost, and GNN, the system is able to deliver highly accurate crop yield predictions, supporting data-driven and sustainable farming practices. The implementation of a real-time Streamlit dashboard empowers farmers, agronomists, and policymakers by providing actionable insights and intuitive visualizations of trends, correlations, and predicted outcomes. Looking forward, the system could be further enhanced through the integration of IoT sensors for live field data, satellite imagery for broader environmental monitoring, and cloud-based deployment to ensure scalability and accessibility. Overall, this work bridges the gap between data science and agriculture, contributing to smart farming initiatives that promote food security, resource optimization, and environmental sustainability.

## Reference

1. M.D. Jiabul Hoque et al., "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning," IEEE Access, 2024.
2. Uppugunduri Nikhil, Athiya Pandiyan, S. Raja, Zoran Stamenkovic, "Machine Learning-Based Crop Yield Prediction in South India," MDPI Computers, Vol.13, 2024.
3. Yamuna T.K., Ashwini T., Rohini N., Vanitha G.N., "Agricultural Crop Yield Prediction for Indian

Farmers Using Machine Learning," IJARESM, 2024.
4. K. Shankar, "Women Safety App to Detect Danger and Prevent Automatically Using Machine Learning," ICCIET 2024 Proceedings, Atlantis Press.
5. Senoy V., "A Holistic Framework for AI-Based Environmental Analysis and Prediction," IEEE Access, 2023.