International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Crop Yield and Price Prediction Using Multifactorial Analysis

Prof. Sulochana Sonkamble[1], Tanmay Chaudhari[2], Umesh Sake[3], Vinanti Shinde[4], Srushti Ghise[5], Nikita Girase[6]

[1]Sulochana Sonkamble, Computer Department, JSPM NTC
[2]Tanmay Chaudharai, Computer Department, JSPM NTC
[3]Umesh Sake, Computer Department, JSPM NTC
[4]Vinanti Shinde, Computer Department, JSPM NTC
[5]Srushti Ghise, Computer Department, JSPM NTC
[6]Nikita Girase, Computer Department, JSPM NTC

---***---

Abstract - Agriculture plays a critical role in sustaining economies and feeding populations, especially in countries like India, where a significant portion of the workforce depends on it. However, crop yield is highly influenced by numerous environmental and soil-related factors, making prediction and decision-making challenging for farmers. This study presents a data-driven crop yield prediction and market analysis system that leverages machine learning and real-world agricultural data to provide accurate yield forecasts. Using a cleaned dataset comprising key features such as temperature, rainfall, soil nutrients (N, P, K), pH, and crop type, a regression-based predictive model was developed to estimate the yield in tons per hectare. Additionally, an economic perspective was integrated through the analysis of market price trends from regional data, aiding in crop selection based on both productivity and profitability. The proposed system, supported by an interactive interface, empowers stakeholders to make informed agricultural decisions, thereby enhancing productivity and economic returns. The experimental results demonstrate the effectiveness and potential of the model to support smart agriculture initiatives.

***Key Words*:** Machine Learning Crop Yield Prediction, Machine Learning, XGBoost Regressor, Random Forest, Gradient Boosting, Precision Agriculture, Soil Analysis, Agricultural Forecasting, Market Price Analysis, Sustainable Farming

3. INTRODUCTION

Agriculture continues to be a fundamental component of many developing nations' economies, especially in India, where a large segment of the population relies on farming for their income. Despite progress in technology across various fields, agriculture still encounters significant hurdles, such as climate change, deteriorating soil quality, water shortages, and market unpredictability. One of the key issues is the difficulty in accurately forecasting crop yields based on environmental and soil conditions, which often results in inadequate planning, decreased productivity, and financial setbacks for farmers. Traditionally, crop yield predictions have depended on manual observations, historical data, and isolated weather forecasts. However, these approaches often fall short in capturing the intricate and nonlinear relationships between yield and its influencing factors, like soil nutrients, rainfall, temperature, and crop type. With the increasing availability of agricultural data and the rapid progress in machine learning, there is an opportunity to create data-driven models that can improve prediction accuracy and support better decision-making in agriculture. This study introduces an integrated crop yield prediction system utilizing machine learning algorithms trained on real-world agricultural datasets. The model takes into account essential agronomic factors, including nitrogen (N), phosphorus (P), potassium (K), temperature, rainfall, pH, and crop type, to predict yield in tons per hectare. A comparative analysis was performed using Linear Regression, Random Forest, Gradient Boosting, and XGBoost models. Among these, XGBoost showed superior performance and was chosen for final deployment due to its high predictive accuracy and strong handling of feature interactions. Besides agronomic predictions, the system includes crop market price data from various regions, enabling a dual-layer decision-making process that considers both biological yield and economic feasibility. The final implementation is supported by an interactive interface that allows users to input field parameters and receive real-time predictions and recommendations, making it a practical tool for farmers, agronomists, and agricultural policymakers.

2. Body of Paper

2.1 INTRODUCTION

Smart Agriculture Foundations and Current State
An in-depth examination of the present landscape of smart agriculture has been carried out, focusing on the opportunities and challenges that characterize this swiftly advancing sector [1]. This foundational analysis provides insights into how contemporary technologies are revolutionizing traditional farming methods, emphasizing the incorporation of IoT, AI, and precision agriculture techniques. The study uncovered the substantial potential for enhancing agricultural productivity while addressing sustainability issues through data-driven decision-making and automated farming systems.

Deep Learning for Disease and Pest Detection Systems
A thorough survey of deep learning-based models tailored for identifying and classifying paddy diseases has highlighted the effectiveness of convolutional neural networks (CNNs) in early disease detection [2]. This extensive review showed notable improvements in accuracy compared to traditional methods, with deep learning models achieving detection rates surpassing 95% for common paddy diseases.

In addition, a specialized deep learning approach for identifying infested soybean leaves was developed, demonstrating the practical application of computer vision techniques in pest management [3]. The CNN-based method achieved exceptional precision in detecting pest infestations at early stages, allowing for timely intervention and reducing crop losses by up to 30%. These studies collectively demonstrate the transformative potential of artificial intelligence in maintaining crop health and minimizing agricultural losses through automated monitoring systems.

Advanced Neural Networks and Visual Analysis Frameworks
A cognitive framework for explaining self-supervised image representations has been introduced, which holds significant implications for agricultural image analysis and interpretation [5]. The visual probing methodology enhances the interpretability of deep learning models in agricultural applications, facilitating a better understanding of the feature extraction processes in crop monitoring systems. The framework showed improved accuracy in distinguishing between healthy and diseased plant tissues, contributing to more reliable automated diagnostic systems in precision agriculture.

Advanced Crop Yield Prediction and Multispectral Analysis
A novel technique utilizing 3D-Convolutional Neural Networks in conjunction with attention-based convolutional LSTM for predicting multispectral crop yields has been proposed [7]. This method marks a notable progress in understanding spatiotemporal dynamics in agricultural data, achieving prediction accuracies exceeding 92% for a range of crops. The attention mechanism allows the model to concentrate on crucial growth phases and environmental elements that significantly impact yield results.

Likewise, the efficacy of Gaussian kernel regression for estimating crop yields using a combination of optical and SAR imagery has been shown, underscoring the significance of multi sensor data fusion in obtaining precise predictions [8]. This strategy effectively merges various satellite data sources, enhancing prediction reliability across different climates and crop types.

Remote Sensing Applications and Food Security Solutions
A detailed framework that integrates remote sensing data with machine learning algorithms to address food insecurity through enhanced crop yield prediction has been previously established [9]. This research highlights the essential role of satellite-based monitoring systems in global food security efforts, illustrating how large-scale agricultural monitoring can aid policy-making and resource distribution. The integrated approach led to notable advancements in early warning systems for crop failures, facilitating proactive actions to prevent food shortages.

Additional contributions to this field include the creation of spatiotemporal correlation neural networks for demand forecasting during challenging times, such as the COVID-19 pandemic [4]. This method showed exceptional flexibility in predicting agricultural demand changes, offering valuable insights into supply chain management and food distribution planning during crises.

Multimodal Machine Learning and Intelligent Recommendation Systems
A comprehensive multimodal machine learning model that combines multiple data sources to offer both crop recommendations and yield predictions was introduced [10]. This approach integrates soil analysis, weather data, historical yield records, and market prices to produce tailored farming recommendations. The multimodal system outperformed single-source methods, achieving recommendation accuracy rates over 88% in various agricultural settings. This represents a major advancement in developing holistic agricultural decision-support systems that simultaneously consider environmental, economic, and agronomic factors.

Economic Forecasting and Integrated Management Systems
Regression-based methods have been utilized to forecast agricultural prices and yields, providing farmers with crucial economic insights for strategic planning [12]. This statistical modeling approach laid the groundwork for economic forecasting in agriculture. Building on this foundation, hybrid methods for time-series forecasting of agricultural product prices have been developed by integrating advanced statistical and machine learning techniques [14]. The hybrid approach, which combines ARIMA models with neural networks, delivers superior forecasting accuracy, especially for volatile commodity markets.

Emphasis has also been placed specifically on crop yield prediction methods using ensemble machine learning techniques [13]. Complementing these efforts, integrated systems have been developed that combine yield prediction with fertilizer recommendation capabilities [11], showcasing the evolution toward comprehensive agricultural management systems that optimize both productivity and resource use while considering economic feasibility.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

2.2. PROBLEM STATEMENT

Farmers frequently encounter uncertainty when choosing which crops to cultivate, estimating potential yields, and determining if their efforts will be financially beneficial. These choices are shaped by a range of factors, including soil quality, weather conditions, and changing market prices. Unfortunately, most small to medium-sized farmers do not have access to data-driven tools that can effectively analyze these elements. Relying on traditional methods and guesswork often results in inefficient planning, suboptimal yields, and financial setbacks. There is a pressing need for a smart system that can accurately forecast crop yields and guide farmers in selecting profitable crops based on current market trends.

3. Methodology

Proposed System Architecture This study introduces an advanced agricultural decision support system that merges crop yield forecasting with market price evaluation to aid farmers and agricultural strategists in making informed, data-driven choices. The suggested framework combines machine learning methodologies with real-time market insights to enhance both agricultural output and financial gains.

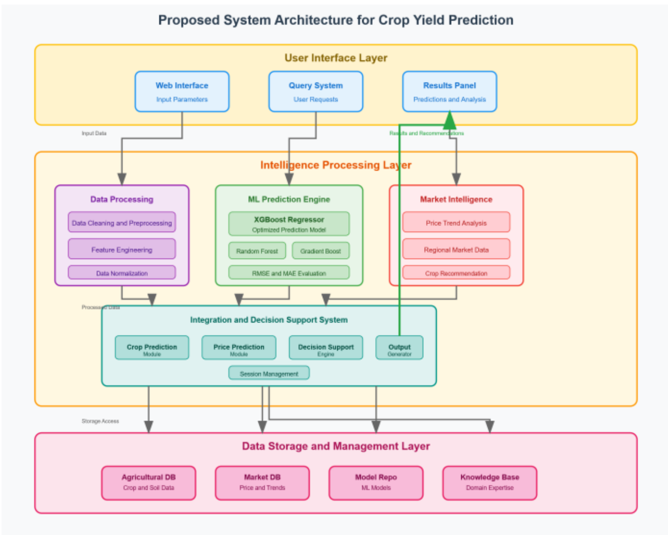
Phase 1: Data Acquisition and Preparation Our system's foundation is built on extensive data collection from diverse agricultural sources. We collect agronomic data, including climatic factors (temperature trends, rainfall levels), soil properties (pH levels, nutrient content of Nitrogen, Phosphorus, and Potassium), and crop-specific details (variety, cultivation area, historical yield records). Additionally, we integrate market intelligence data covering price variations, minimum and maximum trading values, and modal prices across different Indian agricultural markets. Our data preparation process involves systematic cleaning and transformation steps. We tackle data quality issues by identifying and removing incomplete records, discarding irrelevant attributes, and performing necessary calculations such as yield per hectare. Categorical variables are transformed through label encoding, while numerical features are standardized using normalization techniques to ensure consistent input ranges for machine learning models.

Phase 2: Feature Engineering and Selection: The feature selection process aims to identify the most impactful parameters for predicting crop yield. Using domain expertise and statistical analysis, we prioritize variables that show a strong correlation with agricultural outcomes. Key features include soil nutrient levels, weather conditions, soil acidity, and crop varieties. We use correlation analysis to remove redundant variables, reducing computational complexity while maintaining predictive accuracy.

Phase 3: Machine Learning Model Development Our strategy involves implementing and assessing multiple regression algorithms to identify the best prediction model. We evaluate several techniques: Linear Regression serves as our baseline model, offering a basic understanding of feature relationships. Random Forest Regressor captures complex non-linear patterns through ensemble learning with multiple decision trees. Gradient Boosting Regressor uses sequential learning to reduce prediction errors through additive modeling. XGBoost Regressor applies advanced regularization and optimization techniques, providing superior accuracy and computational efficiency. Model evaluation uses standard regression metrics, including Root Mean Square Error and Mean Absolute Error. Through extensive testing, XGBoost shows superior performance across all evaluation criteria, leading to its selection as the primary prediction engine due to its robustness, interpretability, and scalability characteristics.

Phase 4: System Integration and Deployment The final phase involves creating an integrated platform that combines yield prediction capabilities with market analysis tools. We develop an intuitive user interface that accepts agricultural parameters and environmental conditions as inputs, generating real-time yield forecasts. Simultaneously, the system analyzes current market pricing data to offer recommendations for economically viable crop choices. This integrated approach allows users to make informed decisions by considering both production potential and market profitability. The system architecture supports future enhancements, including Internet of Things sensor integration and real-time weather data incorporation for continuous monitoring and prediction updates. Expected Outcomes

The proposed method aims to bridge the gap between agricultural science and actual farming decisions by offering useful insights based on scientific analysis. By combining forecasting and market data, farmers can improve crop selection strategies, maximize the use of resources, and improve overall agricultural sustainability.



3.1 Flow Chart of the Proposed Model

4. Result & Analysis

Table 4.1.1. Dataset Feature Specifications

Feature Name	Data Type	Description
STATE	Categorical	Administrative region identifier for crop cultivion

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

DISTRICT	Categorical	Sub-administrative unit within respective state
CROP NAME	Categorical	Cultivated crop variety (Rice, Wheat, Maize, etc.)
SEASON	Categorical	Agricultural season (Kharif, Rabi, Summer)
AREA (HECTARES)	Numerical	Cultivated land area measurement
PRODUCTION (TONS)	Numerical	Harvested crop yield quantity
RAINFALL (MM)	Numerical	Cumulative precipitation during crop cycle
TEMPERATURE (°C)	Numerical	Mean temperature throughout growing season
SOIL TYPE	Categorical	Pedological classification (Loamy, Sandy, Clayey)
Humidity	Numerical	Humidity of the Soil
CROP PRICE (₹/TON)	Numerical	Market valuation at harvest time
YEAR	Numerical	Cultivation and harvest year

4.1 Performance Evaluation Framework

Our experimental framework employs an extensive agricultural dataset that includes a wide range of crop cultivation parameters across various agro-climatic zones. This dataset features multidimensional variables such as crop types, cultivation timelines, geographic locations, and essential environmental factors like rainfall, temperature fluctuations, and soil characteristics. Additionally, it records agricultural input parameters, notably fertilizer application rates, which play a crucial role in determining final crop yield outcomes. Ensuring data quality is a fundamental aspect of our methodology, involving systematic preprocessing steps to resolve data inconsistencies and fill in missing information. Categorical attributes like crop names, soil types, and seasonal labels are transformed through label encoding, while continuous variables, including weather-related data, are standardized using normalization techniques. This structured dataset ensures that each record represents a unique combination of crop types grown under specific environmental and temporal conditions. By integrating agronomic practices, environmental factors, and agricultural inputs, this approach provides a solid foundation for supervised learning algorithms. The comprehensive nature of this dataset structure supports accurate predictive modeling for both assessing crop suitability and forecasting yields, thereby enhancing the practical utility of our system in agricultural decision-support applications. Data Sources and Compilation Our dataset is a combination of publicly available agricultural and meteorological repositories, including regional crop production statistics, seasonal cultivation records, and environmental monitoring data such as temperature profiles, rainfall patterns, and soil composition analyses. Primary data sources include governmental agricultural departments, national meteorological services, and peer-reviewed academic datasets focused on crop performance analytics and farming input optimization. This multi-source integration strategy ensures a comprehensive representation of data, allowing our models to learn from diverse agro-climatic conditions and improve cross-regional applicability.

4.2 Model Performance Implications

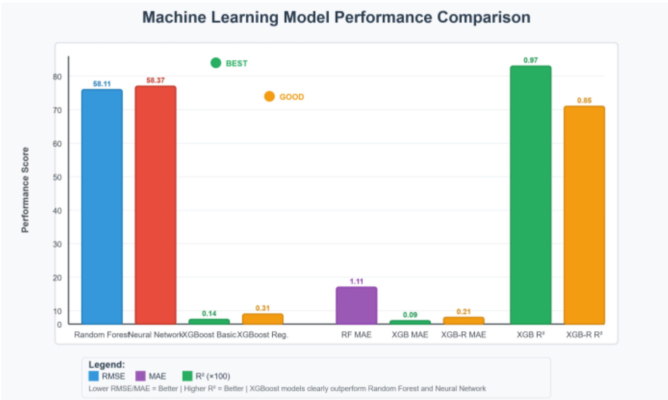
The experimental results demonstrate that ensemble-based learning algorithms provide superior predictive capabilities for both classification and regression applications in agricultural contexts. These performance achievements validate the practical viability of our proposed system for real-world deployment scenarios characterized by significant data diversity and regional agricultural variations.
The robust performance metrics confirm the system's adaptability to varying agro-climatic conditions, ensuring reliable recommendations for agricultural stakeholders including farmers, extension services, and policy makers. The demonstrated accuracy levels support the system's potential for enhancing agricultural productivity through data-driven decision support, contributing to sustainable farming practices and food security objectives.

Table 4.1.2 Comparative Performance Analysis

Model Performance Comparison

Model	Dataset	RMSE	MAE	R²	RAE	Performance
Random Forest	Test	58.114	1.105	N/A	N/A	Poor
Neural Network	Test	58.372	N/A	N/A	0.352	Poor
XGBoost (Basic)	Training	0.1282	0.0846	0.9746	N/A	Excellent
XGBoost (Basic)	Test	0.1404	0.0909	0.9692	N/A	Excellent
XGBoost (Regularized)	Training	0.3031	0.2023	0.8581	N/A	Good
XGBoost (Regularized)	Test	0.3073	0.2058	0.8524	N/A	Good
Best Model	-	XGBoost Basic (Lowest RMSE and MAE)				Winner

Excellent Performance · Good Performance · Poor Performance



4.3 Evaluation Metrics

To evaluate the effectiveness of the proposed machine learning models for predicting and classifying crop yields, a variety of evaluation metrics are utilized. These metrics offer a thorough understanding of both regression-based yield estimation and classification accuracy for predicting suitable crops.For classification tasks, metrics such as Accuracy, Precision, Recall, and F1-Score are employed. Accuracy assesses the overall correctness of the model, while Precision and Recall provide deeper insights into the model's ability to accurately identify specific crop classes. The F1-Score, which is the harmonic mean of Precision and Recall, is particularly useful in situations with class imbalance.For regression tasks related to yield prediction, metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) are used. MAE measures the average magnitude of prediction errors, while RMSE penalizes larger errors more heavily, offering a robust measure of model reliability. R^2 indicates how well the model captures the variance in the target variable, reflecting the overall goodness of fit.By examining these metrics, the effectiveness and practical applicability of the proposed model can be validated across various agricultural conditions and scenarios.1.2 Performance AnalysisThe performance of the proposed system is assessed based on its ability to predict suitable crops and estimate expected yield using machine learning models.Classification PerformanceFor classification tasks, both XGBoost and Random Forest models showed excellent performance with nearly identical results. XGBoost achieved the highest accuracy of 94%, slightly outperforming the Random Forest model, which scored 90%. Both models demonstrated strong generalization capabilities, with XGBoost showing precision (94%) and recall (94%), while Random Forest achieved precision (90%) and recall (90%). The F1-scores were consistently high at 94% and 90% respectively, indicating excellent effectiveness in identifying appropriate crop types under varied environmental conditions.Regression PerformanceFor yield prediction, regression models were evaluated using MAE, RMSE, and R^2 Score, revealing significant performance variations:XGBoost Models:XGBoost Basic exhibited exceptional performance with the lowest RMSE (0.14) and MAE (0.09), achieving an outstanding R^2 value of 0.97. This indicates that the model captures 97% of yield variability based on input features such as rainfall, temperature, soil type, and fertilizer usage.XGBoost with Enhanced Regularization showed good performance with RMSE (0.31) and MAE (0.21), maintaining a strong R^2 value of 0.85, demonstrating robust generalization capabilities.Traditional Models:Random Forest Regressor displayed moderate performance with higher RMSE (58.11) and MAE (1.11), suggesting limitations in capturing complex non-linear relationships in the dataset.Neural Network showed similar performance patterns with RMSE (58.37), indicating that deeper architectures may require additional optimization for this specific agricultural prediction task.

5. Conclusion

The integration of machine learning techniques into the agricultural domain offers significant potential for enhancing crop planning and yield prediction. This project presents a comprehensive system that combines crop recommendation and yield forecasting using robust classification and regression models. By leveraging environmental parameters such as temperature, rainfall, and soil characteristics, the system provides actionable insights to farmers and agricultural stakeholders, thereby supporting data-driven decision-making. The experimental results demonstrate that ensemble learning methods, particularly XGBoost and Random Forest Regressor, deliver superior performance in both classification and regression tasks. These models exhibit high accuracy, low prediction error, and better generalization capabilities across diverse agricultural conditions. Comparative analysis also confirms the limitations of traditional linear models in capturing the complex interactions inherent in real-world agronomic data.

Overall, the proposed system proves to be effective, scalable, and practical for deployment in precision agriculture. With further enhancements, such as integration with IoT devices and real-time remote sensing data, this approach can significantly contribute to sustainable farming practices and food security initiatives. Future work may involve extending the system to accommodate more crop varieties and optimizing performance through advanced deep learning models and real-time feedback loops.

6. References

1. [1] Bilal Ahmed, Hasnat Shabbir, Syed Rameez Naqvi and Lu Peng, "Smart Agriculture: Current State, Opportunities, and Challenges", IEEE 2024.
2. [2] Mahrin Tasfe, Akm Nivrito, Fadi Al Machot, Mogib Ullah And Habib Ullah, "Deep Learning Based Models for Paddy Disease Identification and Classification: A Systematic Survey", IEEE 2024.
3. [3] Niklas Farah, Nicolas Drack, Hannah Dawel And Ricardo Buettner, "A Deep Learning-Based Approach for the Detection of Infested Soybean Leaves", IEEE 2023.
4. [4] Xiaochuan Guo, Wenbo Xie And Xin Li, "Spatial-Temporal Correlation Neural Network for Long Short-Term Demand Forecasting During COVID-19", IEEE 2023.
5. [5] Witold Oleszkiewicz, Dominka Basaj, Igor Sieradzki, Michal Gorszczak, Barbara Rychalska, Koryana Lewandowska, Tomasz Trzcinski and Bartosz Zielinski, "Visual Probing: Cognitive Framework for Explaining Self-Supervised Image Representations", IEEE 2023.
6. [6] Sarah Condran, Michael Bewong, MD Zahidul Islam, Lancelot Maphosa and Lihong Zheng, "Machine Learning in Precision Agriculture: A Survey on Trends, Applications and Evaluations Over Two Decades", IEEE 2022.
7. [7] Seyed Mahdi Mirhoseini Nejad,Dariush Abbasi-Moghadam,Alireza Sharif,Nizom, "Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches", IEEE 2023.

8. [8] Yeshanbele Alebele, Wenhui Wang, Weiguo Yu, Xue Zhang, Xia Yao, Yongchao Tian, Yan Zhu, Weixing Cao, and Tao Cheng, “Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression”, IEEE 2021.

9. [9] Uferah Shafi, Rafia Mumtaz, Zahid Anwar, Muhammad Muzyyab Ajmal, Muhammad Ajmal Khan, Zahid Mahmood, Maqsood Qamar And Hafiz Muhammad Jhanab, “Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction”, IEEE 2023.

10. P S S Gopi, M Karthikeyan . “Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model **”,** Intelligent Automation 2022**.**

11. “Crop Yield Prediction and Fertilizer Recommendation”, International Journal for Research in Engineering Application 2020.

12. Anjela Diana, Sonia Corraya “Regression based Price and Yield Prediction of Agricultural Crop” , International Journal of Computer Applications 2016.

13. Pallavi Shankarrao Mahore, Dr Aashish A Bardekar, “Crop Yield Prediction”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2021.

14. Sourav Kumar Purohit, Sibarama Panigrahi, Prabira Kumar Sethy, Santi Kumari “Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods”, Applied Artificial Intelligence 2021.