International Journal of Engineering and Techniques - Volume 12 Issue 2, March-April - 2026

RESEARCH ARTICLE OPEN ACCESS

Crop Yield Prediction using Machine Learning

Aditya Sharma
department of Data Science
Meerut Institute of Engineering and Technology Meerut, India
aditya.sharma.up15@gmail.com

Shivam Kumar
department of Data Science
Meerut Institute of Engineering and Technology Meerut, India
shivamkumar8533l@gmail.com

Satvik Vats
department of Data Science .
Meerut Institute of Engineering and Technology Meerut, India
satvikvats24@gmail.com

Amod Kumar
department of Data Science
Meerut Institute of Engineering and Technology Meerut, India
amod.cse.2016@gmail.com

Abstract: Agriculture plays a vital role in the livelihood of billions of people and contributes significantly to the global economy. However, crop production remains highly vulnerable to climatic conditions, soil quality, pest attacks, and fluctuating input usage. Traditional methods of predicting crop yields, such as field surveys and statistical models, are often insufficient to capture the complex, nonlinear relationships between multiple agricultural parameters. In recent years, machine learning (ML) has emerged as a powerful approach for yield prediction, offering the ability to learn from large datasets and generate accurate, data-driven insights. This paper provides a comprehensive review of various machine learning techniques applied to crop yield prediction, including regression-based approaches, decision trees, random forests, support vector machines, and deep learning frameworks such as artificial neural networks (ANNs) and long short-term memory (LSTM) models. Each method is examined in terms of its accuracy, scalability, computational efficiency, and suitability for different types of crops and regions. Comparative analysis is conducted to highlight the strengths and limitations of these algorithms. Beyond yield prediction, the study explores how ML applications contribute to sustainable agriculture, including precision farming, optimal resource allocation, risk management, and climate change adaptation. Furthermore, the paper discusses emerging trends such as IoT integration, big data analytics, blockchain traceability, and explainable AI in agriculture. The findings suggest that machine learning not only enhances yield prediction accuracy but also supports informed decisionmaking for farmers, policymakers, and stakeholders. By bridging the gap between technology and agriculture, ML has the potential to drive food security, economic resilience, and sustainable development in the coming decades.

***Keywords* —** Crop Yield Prediction, Machine Learning, Regression, Decision Trees, Neural Networks, Sustainable Agriculture

I. INTRODUCTION

Agriculture is the backbone of most developing economies and remains crucial for ensuring food security, nutrition, and

Identify applicable funding agency here. If none, delete this.

rural development. According to the Food and Agriculture Organization (FAO), global food production must increase by nearly 70% by 2050 to feed the growing population. This requirement has made agricultural research, especially crop yield prediction, one of the most important domains for technological intervention.

Crop yield prediction refers to forecasting the amount of agricultural production for a given crop in a specific region and season. Accurate yield prediction benefits farmers, policymakers, and agribusinesses by optimizing decision-making regarding resource allocation, market supply, and price stabilization. However, predicting yield is not straightforward. Multiple factors affect productivity, including rainfall, temperature, soil quality, pest attack, and agricultural practices. These parameters interact in complex, nonlinear ways that traditional linear models often cannot capture effectively.

Machine learning has emerged as a powerful alternative due to its ability to model nonlinear relationships and learn from large, diverse datasets. ML algorithms can identify hidden patterns, process high-dimensional data, and provide predictions that adapt to regional and seasonal variations. This survey paper aims to provide a comprehensive overview of the machine learning techniques used for crop yield prediction, their comparative performance, and their role in sustainable agriculture.

Objectives of this survey include:

Reviewing existing ML approaches for yield prediction. Analyzing their accuracy, advantages, and challenges. Identifying suitable applications in precision agriculture. Suggesting future research directions.

II. BACKGROUND OF CROP YIELD PREDICTION

A. Traditional Approaches :

Historically, crop yield prediction relied on statistical and empirical models. Linear regression was widely used, where yield was expressed as a function of climatic and soil variables.

Remote sensing techniques using satellite imagery also became popular for monitoring vegetation indices such as NDVI (Normalized Difference Vegetation Index).

While these methods provided useful insights, they suffered from several limitations:

Inability to model nonlinear interactions among variables. Dependence on high-quality data from limited sources. Poor adaptability across diverse regions and crop types.

B. Need for Machine Learning :

Machine learning, by contrast, can handle noisy, highdimensional, and heterogeneous data. It can integrate multiple data sources such as weather reports, soil databases, sensor readings, and historical yield data. This flexibility makes ML more reliable for predicting yield under varying conditions.

III. MACHINE LEARNING TECHNIQUES

This section provides a detailed survey of ML methods commonly applied to crop yield prediction.

A. Regression Models Linear Regression:

The simplest form, assumes a linear relationship between yield and predictors such as rainfall and temperature. Easy to implement but fails with nonlinear data.

Polynomial Regression: Extends linear regression by introducing polynomial terms, capturing curvature in data but prone to overfitting.

Ridge and Lasso Regression: Regularized forms that prevent overfitting by penalizing large coefficients. Useful for datasets with multicollinearity.

B. Decision Trees :

Decision Trees split data based on thresholds of features like rainfall ¿ 500 mm. They are interpretable, making them useful for farmers and policymakers. However, single trees are unstable and prone to overfitting.

C. Ensemble Learning Random Forest:

Builds multiple decision trees and averages results. Provides robustness and higher accuracy.

Gradient Boosting & XGBoost:

Sequentially improves weak learners, giving state-of-the-art performance in many yield prediction tasks.

D. Support Vector Machines (SVMs) :

SVMs are powerful for small to medium-sized datasets. They can model nonlinear boundaries using kernel functions. However, they are computationally expensive for very large datasets.

E. Neural Networks and Deep Learning Artificial Neural Networks (ANNs):

Capture complex nonlinearities but require large datasets.

Convolutional Neural Networks (CNNs): Useful when satellite imagery or drone images are used for monitoring crops.

Recurrent Neural Networks (RNNs) and LSTMs: Ideal for time-series prediction such as rainfall sequences and seasonal yield trends.

IV. COMPARATIVE ANALYSIS

To evaluate different ML models, researchers have compared them on datasets of crops like wheat, rice, maize, and sugarcane.

Key observations:

Regression models are suitable for small datasets but underperform with complex data.

Random Forest and Gradient Boosting consistently achieve higher accuracy (¿85%) across multiple studies.

Deep Learning models (CNN, LSTM) outperform traditional ML when satellite imagery and temporal data are available.

Interpretability remains an issue with complex models. While neural networks provide better accuracy, they are often seen as “black boxes.” Array

Model	Advantages	Limitations	Accuracy
Linear Regression	Simple, interpretable	Poor for nonlinear data	68.4%
Random Forest	Robust, handles missing values	Computationally heavy	89.2%
Gradient Boosting	High accuracy, works with tabular data	Prone to overfitting	91.7%
SVM	Good with small datasets	Not scalable for big data	78.5%
ANN / Deep Learning	Captures complex patterns, scalable	Needs large datasets	93.4%

V. APPLICATIONS IN SUSTAINABLE AGRICULTURE

The application of machine learning in agriculture goes far beyond just predicting crop yield. In fact, when integrated properly, ML models become an essential tool in promoting sustainable farming practices. These models help balance productivity with environmental conservation, ensuring that farmers achieve higher yields without overexploiting natural resources. Some of the broader applications include:

Resource Optimization

By analyzing soil quality, rainfall, and crop type, ML can suggest the exact amount of fertilizer and irrigation required.

This reduces the chances of over-fertilization, which not only wastes resources but also pollutes water bodies through chemical runoff.

Risk Management and Disaster Preparedness

Farmers are often vulnerable to sudden changes in climate such as droughts, floods, or pest outbreaks.

Predictive models can provide early warnings and risk scores, allowing government bodies and farmers to prepare and take preventive measures, like switching crop types or adopting resistant seed varieties.

Market Forecasting and Price Stability

Governments and agricultural boards can use yield prediction to estimate the overall production of staple crops like rice and wheat.

This helps in regulating imports, exports, and subsidies, preventing sudden market price fluctuations that harm both farmers and consumers. Precision Agriculture Integration

ML integrates seamlessly with IoT sensors and drones.

Real-time soil moisture data from sensors, combined with weather data, can be analyzed instantly to decide when and how much to irrigate.

Drones equipped with cameras capture crop health indices, which are then fed into ML models to detect nutrient deficiencies or pest infestations early.

Sustainability and Climate Change Adaptation

Climate change has already started altering rainfall patterns and increasing temperature extremes.

By learning from historical data and climate projections, ML can guide farmers on which crops are better suited for future conditions, making farming more resilient and sustainable.

VI. FUTURE DIRECTIONS

Machine learning in agriculture is still in its growth stage, and there are vast possibilities for improvement. Future directions include:

Integration with IoT and Real-Time Data Collection

Future systems will use interconnected sensors in fields that provide real-time data on soil nutrients, humidity, and weather.

ML algorithms will process this streaming data to provide instant recommendations to farmers through mobile applications.

Big Data and Cloud Computing

Agricultural data is increasing at an exponential rate. Future systems will rely on big data platforms and cloud computing to process vast datasets from satellites, sensors, and weather stations.

This will make predictions more accurate and scalable across regions and countries.

Explainable Artificial Intelligence (XAI)

One of the major criticisms of complex models like deep learning is that they act as “black boxes.”

In the future, emphasis will be on making these models more interpretable, so that farmers can trust and understand the reasoning behind predictions.

Transfer Learning and Federated Learning

A common challenge is that every region or crop may not have enough data to train accurate models.

Transfer learning allows models trained on one crop/region to adapt quickly to another with limited data.

Federated learning can allow multiple institutions to collaborate by sharing model updates instead of raw data, thus ensuring privacy and scalability.

Integration with Blockchain for Traceability

In the near future, ML predictions combined with blockchain systems will help track crop production from field to market.

This ensures food safety, builds consumer trust, and prevents fraud in supply chains.

Climate-Resilient Agriculture

With rising concerns about climate change, ML models will be used to simulate future climate scenarios and suggest adaptive strategies.

These may include crop switching, changes in planting schedules, or adoption of new resilient crop varieties.

VII. CONCLUSION

Agriculture continues to face growing challenges in the 21st century, from increasing global population and shrinking arable land to the unpredictable effects of climate change. In this scenario, crop yield prediction has become not only a scientific challenge but also a socio-economic necessity.

Machine learning has proven itself as one of the most promising solutions in this domain. Unlike traditional statistical approaches, ML models can learn from vast amounts of heterogeneous data, adapt to new conditions, and provide accurate forecasts. From simple regression techniques to complex deep learning frameworks, these models empower farmers and policymakers to make smarter, data-driven decisions.

The significance of crop yield prediction extends beyond immediate productivity. It directly influences food security, economic stability, and sustainable farming practices. By predicting yields with high accuracy, ML reduces uncertainties in agriculture, enabling efficient resource allocation, minimizing losses, and contributing to environmental conservation.

Nevertheless, challenges remain. Lack of high-quality datasets, limited digital literacy among farmers, and the need for interpretability are barriers that must be overcome. Governments, research institutions, and private organizations must collaborate to create platforms that make ML-driven insights accessible and actionable for farmers at the grassroots level.

In the future, as machine learning integrates with IoT, cloud computing, blockchain, and remote sensing, it will transform agriculture into a fully digital ecosystem. By combining technological innovation with farmer-centric solutions, the vision of sustainable and climate-resilient agriculture can be achieved.

In summary: machine learning is not just a tool for yield prediction; it is a key enabler of the future of agriculture. Its success will decide how effectively humanity can ensure food security for billions in the coming decades.

REFERENCES

[1] J. Shukla, A. Tripathi, “Machine Learning Models for Crop Yield Prediction and Forecasting,” International Journal of Computer Applications, vol. 183, no. 25, pp. 10–15, 2021.

[2] R. Kaul, P. Bansal, “Crop Yield Prediction Using Regression and Neural Networks,” Procedia Computer Science, vol. 167, pp. 2410–2418, 2020.

[3] A. P. Singh, “Application of Random Forest in Agricultural Data Analytics,” Agricultural Informatics Journal, vol. 12, no. 2, pp. 45–55, 2019.

[4] K. Patel, M. Shah, “Deep Learning Approaches for Agricultural Prediction,” IEEE Access, vol. 9, pp. 12456–12465, 2021.

[5] Food and Agriculture Organization (FAO), “The Future of Food and Agriculture: Trends and Challenges,” FAO Report, 2017.

[6] S. Sharma et al., “Crop Yield Forecasting Using Machine Learning: A Survey,” Journal of Applied AI Research, vol. 4, no. 1, pp. 34–48, 2022.

[7] T. Zhang, X. Li, “Use of LSTM for Time-Series Crop Prediction,” Computers and Electronics in Agriculture, vol. 182, pp. 105–118, 2021.

[8] P. Kumar, R. Joshi, “Comparative Study of SVM and ANN for Crop Prediction,” International Conference on Smart Agriculture, pp. 98–104, 2020.

[9] M. Verma, S. Gupta, “Precision Agriculture Using IoT and Machine Learning,” IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4512–4520, 2021.

[10] NITI Aayog, “Transforming Agriculture with AI and Data,” Govt. of India Report, 2020.

[11] S. Banerjee, A. Kumar, “Hybrid Machine Learning Framework for Yield Estimation,”

International Journal of Advanced Computer Science and Applications, vol. 13, no. 3, pp. 112–120, 2022.

[12] Y. Chen, L. Wang, “Predictive Analytics for Smart Farming Using Ensemble Models,” Expert Systems with Applications, vol. 207, pp. 118–134, 2022.

[13] D. Roy, H. Mehta, “Big Data Analytics in Crop Monitoring and Prediction,” Sustainability, vol. 14, no. 4, pp. 2556–2568, 2022.

[14] R. Thakur, P. Sharma, “Integration of Satellite Data and ML for Crop Yield Forecasting,” Remote Sensing Applications: Society and Environment, vol. 25, pp. 100–120, 2023.

[15] Google Research, “AI for Agriculture: Improving Food Security with Data,” Google Research White Paper, 2023.