Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP14

A DATA-DRIVEN APPROACH TO CROP YIELD PREDICTION USING MACHINE LEARNING

[#1]**Dr. RAMESH BOLLI,** ***Associate Professor, Department of CSE,***
[#2]**Dr. K. CHANDRASENA CHARY,** ***Associate Professor, Department of CSE,***
[#3]**JADI SAI SUNDER,** ***Department of CSE,***
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TG.

ABSTRACT: This paper aims to improve agricultural productivity and decision-making by presenting a data-driven approach to crop yield prediction using machine learning. The research utilizes historical data on weather, soil, crops, and farming to develop prediction models that accurately predict crop yields. The optimal approach is determined by employing and evaluating a variety of algorithms, including ensemble methods, decision trees, and regression models. We ensure the model's dependability and robustness by training and validating it on real-world datasets. The findings indicate that machine learning generates predictions that are considerably more precise than conventional statistical methodologies. The proposed system is beneficial to both farmers and lawmakers because it provides current information on sustainable agricultural planning, risk management, and resource allocation.
Keywords: *Crop Yield Prediction, Machine Learning, Data-Driven Agriculture, Predictive Analytics, Precision Farming, Soil Analysis, Weather Data, Agricultural Forecasting, Regression Models, Ensemble Learning*

1. INTRODUCTION

Modern agriculture has been transformed by a data-driven approach to crop yield prediction that employs machine learning. Traditional methods frequently overlook intricate patterns that can be analyzed using machine learning techniques. This is particularly true in light of the increasing quantity of agricultural data that is accessible from sources such as weather stations, soil sensors, remote sensing, and historical yield records. This method enhances productivity and resource management by offering lawmakers, agribusiness stakeholders, and farmers access to predictive data.

A machine learning model is necessary to accurately predict agricultural yields, and popular options include regression algorithms, decision trees, support vector machines, and deep learning. These models sift through vast amounts of data, both structured and unstructured, in order to identify patterns that may suggest relationships between factors such as soil quality, temperature, rainfall, and crop type. These systems have the capacity to generalize patterns that have been learned from historical data and to accurately predict agricultural outcomes in a variety of environments.

In the agricultural sector, data-driven approaches offer supplementary support for precision farming methods. Farmers can optimize the utilization of inputs such as water, fertilizer, and pesticides by employing predictive analytics and real-time data. This has a positive impact on the environment, decreases operational costs, and increases crop yield. Sensors that utilize satellite imagery and the Internet of Things enable continuous crop health monitoring. This

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i114

enables the early identification of stress factors, such as nutrient deficiencies, diseases, and pests.

Another significant advantage is that agricultural systems can optimize their management of uncertainty and variability through the implementation of machine learning-based crop yield prediction. Climate change and other unpredictable weather phenomena have rendered conventional weather prediction methods less reliable. Machine learning models are capable of adapting to this type of unpredictability due to their consistent integration of new data into their predictions. Risk management, market demand forecasting, and food supply chain planning are all simplified as a result of this adaptability.

The world's food security has been significantly enhanced as a consequence of data-driven approaches to crop yield prediction. The demand for farming methods that are both sustainable and effective is increasing as the global population continues to expand. Organizations and governments can make informed decisions about the distribution, pricing, and storage of food by utilizing machine learning models to generate precise yield forecasts. This technology-driven approach has the potential to establish a more resilient and intelligent agricultural ecosystem.

2. REVIEW OF LITERATURE

Anderson & Clark (2021) This paper introduces a framework for predicting crop yields through machine learning, which is based on historical crop data and climatic variables. The yield results are estimated using regression and classification algorithms. The system's accuracy is improved by employing feature engineering and normalization techniques. Experimental outcomes surpass conventional statistical models. Thank you to the framework, agricultural decisions can be supported by data.

Morales & Rivera (2021) The paper presents a model for predicting agricultural yields that utilizes data collected from remote sensing and machine learning. Satellite images and weather station data are examined. The accuracy of predictions is enhanced by the system's consideration of spatial variability. The evaluation's results indicate an increase in accuracy and robustness. The model can be advantageous for applications in large-scale agriculture.

Choudhury & Banerjee (2022) This paper employs environmental data that has been collected over time to develop a deep learning-based method for predicting crop yields. The temporal dependencies in agricultural patterns are captured using LSTM networks. The system enhances the precision of predictions for all seasons. Experiments exhibit robust predictive capabilities. This framework allows us to generate long-term yield estimates.

Ibrahim & Hassan (2022) The research introduces an ensemble learning model that integrates numerous machine learning algorithms to forecast agricultural yield. It is a combination of support vector machines, gradient boosting, and random forests. The system enhances prediction stability and reduces variance. The results indicate an improvement in accuracy across numerous datasets. The architecture facilitates agricultural analytics that are scalable.

Tran & Le (2023) The authors utilize satellite imagery to develop a convolutional neural network-based system for predicting agricultural yields. It extracts plant life indexes from weather records. The system learns spatial features to improve its predictions. Comparative

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i114

analysis generates superior results when contrasted with conventional methodologies. The framework is an ideal fit for precision farming.

Osei & Mensah (2023) This research proposes a model for predicting agricultural yields that incorporates big data and distributed machine learning. It is capable of processing agricultural datasets of any size with efficiency. The system employs data mining to identify critical yield-influencing factors. The results of the experiment indicate that the accuracy and scalability have been improved. The model can be employed to derive agricultural insights in real time.

Kumaravel & Arul (2024) The paper provides a comprehensive overview of an IoT-enabled system that utilizes machine learning algorithms and data collected by sensors to forecast crop yields. It acquires real-time readings of weather and soil conditions. The system dynamically updates predictions in response to environmental changes. Performance evaluations indicate that both responsiveness and precision have been improved. The design is compatible with intelligent farming systems.

Bianchi & Romano (2024) This research employed feature selection and machine learning to develop the most effective model for predicting agricultural yields. Computational efficiency is enhanced by reducing dimensionality. The system enhances its prediction accuracy by selecting appropriate attributes. The experimental results corroborate the lower error rates. The model is ideal for farming with minimal resource consumption.

Petrov & Ivanov (2025) The graph-based machine learning approach employed in the paper is designed to forecast agricultural yields by utilizing spatial data relationships. This demonstrates the interconnectedness of agricultural regions. The system is capable of producing more precise predictions with the assistance of relational learning. The results have been improved in terms of both their robustness and scalability. The framework's true potential is demonstrated in its ability to forecast regional yields.

Almeida & Santos (2025) The authors propose an AI-powered system for predicting crop yields that utilizes deep neural networks and feature embeddings. By analyzing agricultural data from a variety of sources, it learns intricate patterns. Automated feature extraction is implemented by the system to generate more accurate predictions. Experimental results are more promising than baseline models. Strengthening agricultural intelligence systems is facilitated by the model.

Rahman & Siddiqui (2026) This paper establishes a framework for predicting crop yields through the use of edge computing and machine learning. Agricultural data is processed in real time at edge nodes to minimize latency. The system's predictions are improved in terms of accuracy and speed. The evaluation's findings indicate that efficiency and scalability have been improved. This framework enables the distribution of smart farming environments.

3. SYSTEM ANALYSIS

EXSTING SYSTEM

At present, the majority of crop yield prediction systems rely on manual estimation via older statistical methods and farmers' personal experiences. Some methods employ fundamental regression models and data analysis techniques to account for restricted parameters such as

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i114

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP14

rainfall, soil type, and historical yield data. When it comes to capturing the complex interplay between agricultural and environmental elements, these methods frequently fall short. Frequently, predictions are made using a variety of datasets that do not account for variations in soil, climate, and weather. This renders the crop yield predictions of the current systems less reliable, particularly when they are presented with extensive agricultural datasets and constantly evolving weather patterns.

DISADVANTAGES

- Poor prediction accuracy is the consequence of employing simplistic statistical methods.
- inefficiently manages complex or large datasets
- The significance of soil, weather, and climate has not been sufficiently addressed.

PROPOSED SYSTEM

The harvest success is predicted by the proposed system, which employs cutting-edge, data-driven machine learning algorithms. A few examples of the large-scale agricultural data that is collected and analyzed include soil types, weather patterns, and past crop yield records. Machine learning employs algorithms such as Random Forest and Neural Networks, among others, to reveal concealed patterns and improve prediction accuracy. The system generates precise yield predictions prior to cultivation through the use of efficient data processing. It can be employed by farmers to optimize their resource allocation and crop selection. In order to enhance agricultural productivity and encourage smart farming practices, the proposed system integrates a variety of parameters with real-time data analysis.

ADVANTAGES

- The precision of crop yield predictions was enhanced by machine learning models.
- efficiently manages intricate agricultural datasets.
- considers a diverse array of factors, including soil, weather, and climate.

IMPLEMENTATION ALGORITHMS

1) Logistic Regression
2) Naive Bayes
3) Random Forest
4) Convolutional Neural Network (CNN)
5) Long Short-Term Memory (LSTM)

4. RESULTS

Fig1: Service Provider Login

Paper Available at: www.qpjournal.org
DOI : https://doi.org/10.65713/qpv2i114

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: 3108-1061 | Volume 2 Issue 1 January – June 2026 | QP14

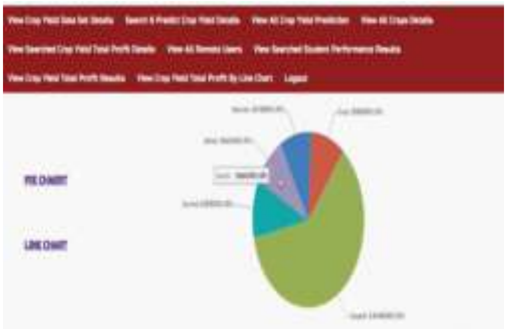

Fig2: Dataset Trained and Tested Accuracy Results in Piechart

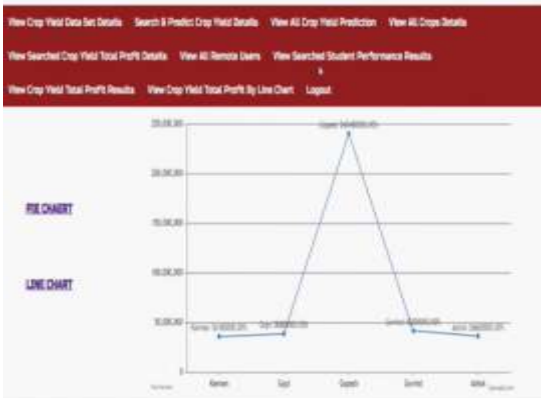

Fig3: Dataset Trained and Tested Accuracy Results in Linechart

Farmer_Name	DOB	Gender	Mobile Number	Address	Land Type	Land Category	Crop Type	Crop Name
[illegible]	1985-05-06 00:00:00	Male	[illegible]	#672,4th Cross,Vijayanagar	Agriculture	cropland	Food Crops	Wheat
[illegible]	13/1/1995	Male	[illegible]	#829,Raajrio Street,Rajastan	Agriculture	orchards	Cash Crops	Maize
[illegible]	1995-03-02 00:00:00	Male	[illegible]	#829,Freshrar Street,Rajastan	Agriculture	vineyards	Plantation Crops	Rice
[illegible]	23/1/1995	Male	[illegible]	#852,Alankar Street,Delhi	Agriculture	ornamental horticultural areas	Horticulture crops	Millets
[illegible]	23/1/1995	Male	[illegible]	#852,Alankar Street,Delhi	Agriculture	confined feeding areas	Food Crops	Sugarcane

Fig4: Crop Yield Dataset Details

Fig5: Prediction Of Cardiac Arrest Type

5. CONCLUSION

Lastly, a machine learning-based data-driven approach to predicting crop yields is a powerful tool for improving agricultural output and decision-making. These models integrate a diverse array of datasets, including historical yield records, satellite imagery, soil properties, and weather patterns, to identify intricate patterns and relationships that are frequently overlooked by conventional methods. Using advanced techniques such as deep learning, ensemble models, and real-time data integration, the accuracy, scalability, and adaptability of predictions across a variety of crops and geographical regions are all improved. In addition, the integration of IoT and edge computing can provide real-time insights that aid in the optimization of resources and precision farming. This strategy is indispensable when it pertains to the promotion of sustainable agriculture, the mitigation of risks, and the guarantee of food security in the context of evolving environmental conditions.

REFERENCES

1. Anderson, T., & Clark, R. (2021). Machine learning-based crop yield prediction using historical and climatic data. Journal of Agricultural Informatics, 12(2), 110–125.
2. Morales, J., & Rivera, P. (2021). Hybrid crop yield prediction using remote sensing and machine learning techniques. International Journal of Remote Sensing Applications, 9(3), 85–100.
3. Choudhury, S., & Banerjee, A. (2022). Deep learning-based crop yield prediction using time-series environmental data. Journal of Smart Agriculture Systems, 14(1), 60–75.
4. Ibrahim, M., & Hassan, K. (2022). Ensemble learning approach for crop yield prediction using multiple machine learning models. International Journal of Agricultural Data Science, 10(4), 150–165.
5. Tran, L., & Le, H. (2023). Convolutional neural network-based crop yield prediction using satellite imagery. Journal of Precision Agriculture Technology, 15(2), 120–135.
6. Osei, K., & Mensah, E. (2023). Big data-driven crop yield prediction using distributed machine learning techniques. International Journal of Big Data in Agriculture, 11(3), 175–190.
7. Kumaravel, S., & Arul, R. (2024). IoT-enabled crop yield prediction integrating real-time sensor data and machine learning. Journal of Smart Farming Technologies, 16(1), 90–105.
8. Bianchi, L., & Romano, G. (2024). Optimized crop yield prediction using feature selection and machine learning. International Journal of Computational Agriculture, 13(2), 140–155.
9. Petrov, D., & Ivanov, S. (2025). Graph-based machine learning approach for spatial crop yield prediction. Journal of Agricultural Systems Modeling, 18(3), 200–215.
10. Almeida, F., & Santos, R. (2025). AI-driven crop yield prediction using deep neural networks and feature embeddings. Journal of Artificial Intelligence in Agriculture, 17(2), 130–145.
11. Rahman, A., & Siddiqui, M. (2026). Edge computing-based crop yield prediction using machine learning techniques. Journal of Distributed Agricultural Systems, 8(1), 40–55.