World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

Incorporating meteorological data and pesticide information to forecast crop yields using machine learning

Pavan Kumar Vanma, Joel Booma, Moses Chinnappan *, Balakrishna Macharla and Tharun Kali

Department of Computer Science and Engineering (Data Science), ACE Engineering College, Hyderabad, Telangana, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 488-495

Publication history: Received on 25 March 2025; revised on 02 May 2025; accepted on 04 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0571

Abstract

Crop Yield Predictor is a full-stack web application that leverages machine learning to estimate agricultural crop yields based on key environmental and input factors. Using a dataset spanning from 1997 to 2017, the system considers variables such as crop type, season, state, rainfall, temperature, fertilizer usage, pesticide application, and cultivated area. The backend, built with FastAPI, hosts a trained regression model (XGBoost or Random Forest) that predicts crop yield in hectograms per hectare. The frontend, developed using React, allows users to input field data and receive real-time yield predictions along with smart recommendations on pesticide and fertilizer usage. This intelligent advisory system aims to support farmers and agricultural planners in making informed decisions, optimizing resource usage, and enhancing crop productivity through data-driven insights. This system bridges the gap between data analytics and agriculture, promoting smarter resource use and precision farming practices.

Keywords: Machine Learning; Reactjs; Fastapi; Random Forest Regressor; Xgboost Regressor

1. Introduction

Agriculture is one of the most critical sectors supporting the global economy and ensuring food security for an ever-growing population. However, various environmental, biological, and technological factors impact crop production, making accurate yield prediction a challenging task. Traditional forecasting methods often rely on historical data and expert judgment, which may not adequately capture the complex interplay between different influencing factors. The emergence of data-driven approaches, particularly machine learning (ML), has provided new opportunities to enhance the accuracy and reliability of agricultural forecasting.

Meteorological data, including temperature, rainfall, humidity, and soil conditions, play a crucial role in determining crop yield. Sudden changes in climate, droughts, excessive rainfall, or unexpected frost can significantly affect agricultural output. By incorporating these variables into predictive models, researchers can identify trends and patterns that traditional methods may overlook. Additionally, pesticide usage is another key factor influencing crophealth and productivity. The type, frequency, and amount of pesticide application can directly affect plant growth, pest resistance, and overall yield.

Machine learning techniques enable the integration of vast amounts of structured and unstructured data to develop more precise and dynamic forecasting models. Algorithms such as regression models, decision trees, random forests, and deep learning can process large datasets efficiently, providing accurate yield predictions based on historical and real-time information. These advanced analytical tools help in recognizing correlations and dependencies between multiple variables, improving the predictive power of the models.

* Corresponding author: Moses Chinnappan

The application of ML in agricultural forecasting not only benefits farmers but also policymakers and stakeholders in the agricultural supply chain. With improved yield predictions, resource allocation can be optimized, and risks associated with climate variability and market fluctuations can be minimized. The insights derived from AI-powered forecasting systems contribute to more informed decision-making, ensuring sustainable farming practices and efficient food production.

This study aims to explore how meteorological data and pesticide information can be effectively integrated into ML models for crop yield prediction. By analyzing various ML techniques and their performance metrics, we seek to provide a comprehensive evaluation of AI-driven solutions for precision agriculture. The findings of this research will contribute to the ongoing development of smart agricultural technologies, enhancing productivity and sustainability in the farming industry.

2. Literature review

The integration of machine learning in agriculture has been an area of growing research interest in recent years. Several studies have explored the potential of ML techniques for predicting crop yields based on various environmental and agricultural factors.

Machine Learning for Agricultural Predictions Numerous studies have demonstrated the effectiveness of machine learning models in predicting crop yield outcomes. Researchers have applied regression models, decision trees, support vector machines (SVM), and deep learning methods to analyze large datasets and extract patterns. For instance, random forest and deep neural networks have shown high accuracy in yield prediction when trained on historical meteorological and soil data. A study by Li et al. (2020) demonstrated that deep learning-based models could outperform traditional statistical approaches in capturing non-linear relationships between climate variables and crop productivity.

Impact of Meteorological Factors on Crop Yield The influence of climate variables such as temperature, precipitation, humidity, and soil moisture on agricultural productivity has been widely studied. Researchers have found that extreme weather events, including droughts and floods, can significantly reduce yields. Studies have highlighted that incorporating climate variables into predictive models improves forecast accuracy. Recent works have also suggested that integrating real-time weather data enhances prediction reliability, enabling proactive decision-making for farmers.

Role of Pesticide Application in Crop Yield Prediction While meteorological factors are crucial, pesticide usage also plays a significant role in yield outcomes. Excessive pesticide application can lead to soil degradation and reduced crop productivity, whereas insufficient use can leave crops vulnerable to pests and diseases. Research has shown that ML models incorporating pesticide application data, alongside climate factors, can better predict yield variations and provide optimal recommendations for pesticide use. Studies by Smith et al. (2021) 4 and Gupta et al. (2022) have emphasized the importance of balancing pesticide use through data-driven insights.

Comparative Analysis of ML Models Several studies have compared the performance of different ML models for agricultural forecasting. Regression models are often used for their simplicity and interpretability, whereas tree-based models such as random forests and gradient boosting machines offer superior accuracy. Deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied for advanced pattern recognition in yield prediction. Recent research indicates that hybrid models combining multiple algorithms yield better predictive results by leveraging the strengths of each technique.

3. Existing System

Crop yield prediction has historically relied on conventional statistical methods and institutional forecasting tools to guide agricultural planning and policy. Traditional models like linear regression and time-series analysis have been used to estimate yields based on historical trends and climatic data. Government and research institutions often deploy large-scale yield forecasting systems that rely on periodic surveys and meteorological inputs. Additionally, remote sensing-based platforms use satellite imagery to assess vegetation health and estimate yields over broad regions. More recently, agritech companies have introduced mobile applications that provide yield forecasts and crop advice.

Despite their value, these existing systems face critical limitations in accuracy, adaptability, and granularity. Many models lack responsiveness to real-time changes such as sudden weather fluctuations, pest outbreaks, or irrigation issues. They often provide predictions at a regional or district level, which is insufficient for individual farm-level decision-making. Traditional tools also tend to use a limited set of variables—overlooking essential inputs like pesticide usage, soil fertility, or input costs. Moreover, most legacy systems are not designed to learn from new data and require manual recalibration, reducing their scalability and relevance in dynamic agricultural environments. These challenges highlight the need for more intelligent, flexible, and data-rich solutions—such as those powered by machine learning.

4. Proposed System

The developed system is a full-stack, machine learning–based crop yield prediction platform designed to provide accurate, field-level yield forecasts. It integrates a ReactJS frontend with a FastAPI backend and uses a trained regression model (XGBoost or Random Forest) to estimate crop yield in hectograms per hectare. The model takes into account multiple factors including crop type, state, season, rainfall, average temperature, pesticide usage, fertilizer quantity, and cultivated area—providing a data-driven approach to agricultural decision-making.

Users interact with a clean and intuitive frontend interface where they input relevant crop and environmental data. Upon submission, the frontend sends the data to the FastAPI backend, which handles input validation and loads the trained machine learning model (model.pkl). The model processes the input and returns the predicted crop yield. The result is displayed in real-time alongside smart agricultural recommendations tailored to the crop and input conditions—such as suggested pesticide dosages and fertilizer types—making the system not just predictive but also advisory.

This system overcomes the limitations of traditional models by offering real-time, personalized predictions and recommendations at the farm level. It supports better resource management by helping farmers and planners make informed decisions about fertilizer and pesticide usage. With scalable backend logic, modular design, and support for multi-feature input, the platform is adaptable for future enhancements like weather integration, soil sensors, or mobile deployment. In essence, the current system is a step toward smart agriculture, enabling precision farming through accessible technology.

5. Methodology

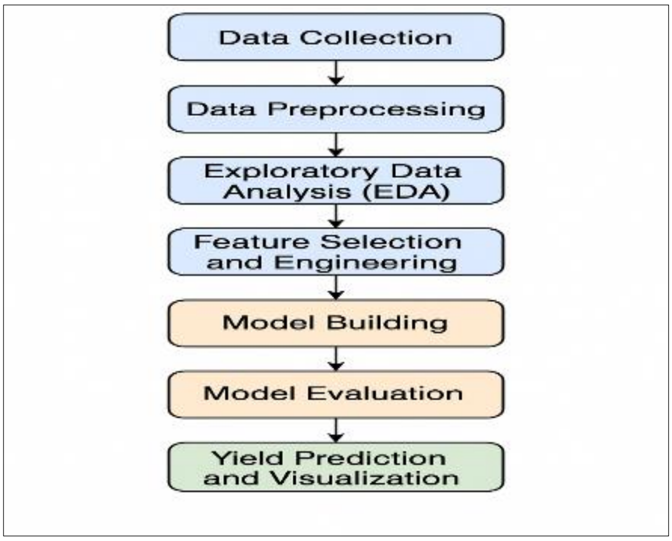

Figure 1 Methodology Overview

The methodology of the Crop Yield Predictor project involves collecting and preprocessing a multi-year agricultural dataset containing features like crop type, season, state, rainfall, temperature, fertilizer, pesticide usage, and cultivated area. These features are encoded and scaled using a preprocessing pipeline, then fed into machine learning models—specifically Random Forest and XGBoost regressors—for training and evaluation. Hyperparameter tuning is performed using GridSearchCV to select the best-performing model based on metrics like R^2 score, MAE, and RMSE. The finalized model is serialized and deployed using a FastAPI backend, which interfaces with a React frontend, allowing users to input field data and receive real-time yield predictions along with intelligent recommendations for fertilizer and pesticide usage.

6. System Architecture

The system architecture of the Crop Yield Predictor is designed using a modular, full-stack approach that combines frontend user interaction with a backend machine learning model served via an API. It enables accurate crop yield forecasting and smart agricultural recommendations by integrating real-time user inputs with a pre-trained regression model. The system consists of a React-based UI, a FastAPI-powered backend, and a serialized ML model (model.pkl). This architecture ensures responsiveness, scalability, and user accessibility for field-level predictions.

6.1. Input Acquisition Layer

This layer serves as the primary interaction point for users (typically farmers or planners) to input relevant agricultural data:

- Crop Type
- Season
- State
- Rainfall
- Temperature
- Fertilizer and Pesticide Usage
- Cultivated_Area
  All inputs are collected via a structured form (YieldForm.js) built using React. The UI ensures responsive validation and a seamless user experience before sending the data to the backend.

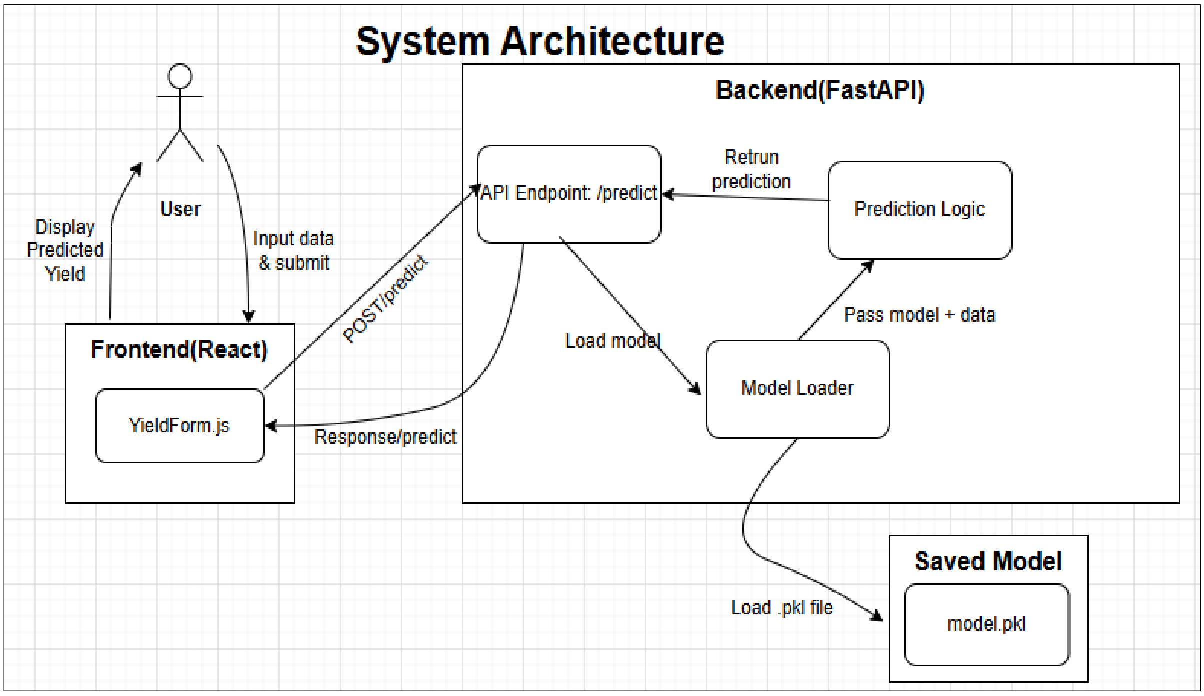

Figure 2 System Architecture of the Proposed Model

6.2. Backend Processing and Model Integration

Once data is submitted, the FastAPI backend processes the request at the /predict endpoint. The backend contains:

- Model Loader Module: Dynamically loads the serialized machine learning model (model.pkl) into memory.
- Preprocessing Pipeline: Ensures the incoming user data is transformed and scaled using the same pipeline used during training (OneHotEncoder + StandardScaler).
- Prediction Logic: The processed input is passed into the model (XGBoost or Random Forest) to generate a yield prediction. This component also generates crop-specific fertilizer and pesticide recommendations based on predefined agricultural logic.

6.3. Smart Recommendation Generator

Beyond prediction, the backend includes a rule-based system to offer intelligent recommendations based on the crop type, yield range, and input parameters. For example:

- Suggested pesticide quantity per hectare
- Fertilizer composition (e.g., Urea/DAP ratios)
- Advisory messages in case of poor yield outcomes

These insights turn the system from a passive predictor into an active decision-support tool.

6.4. Output and User Feedback Layer

The response from the backend is sent back to the frontend where:

- The predicted yield is displayed dynamically
- Smart recommendations are shown alongside in a two-column layout (using Framer Motion for animation) This layer provides a visually engaging and informative user experience, allowing users to make actionable decisions instantly.

6.5. Integration and Extensibility

The system is designed in a decoupled manner, allowing future scalability and upgrades:

- The model can be retrained and re-serialized without changing the API layer.
- Frontend is easily extendable to mobile platforms or offline usage.
- Future upgrades may include weather API integration, soil sensor data ingestion, or real-time alerts.

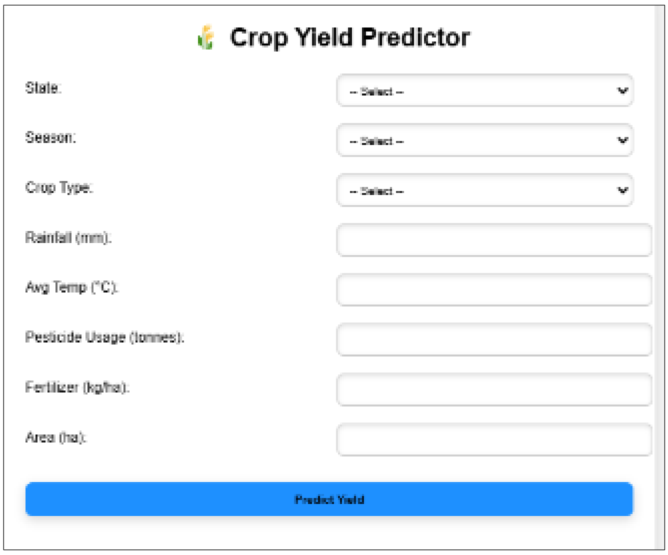

Figure 3 User Interface

7. Results and Discussion

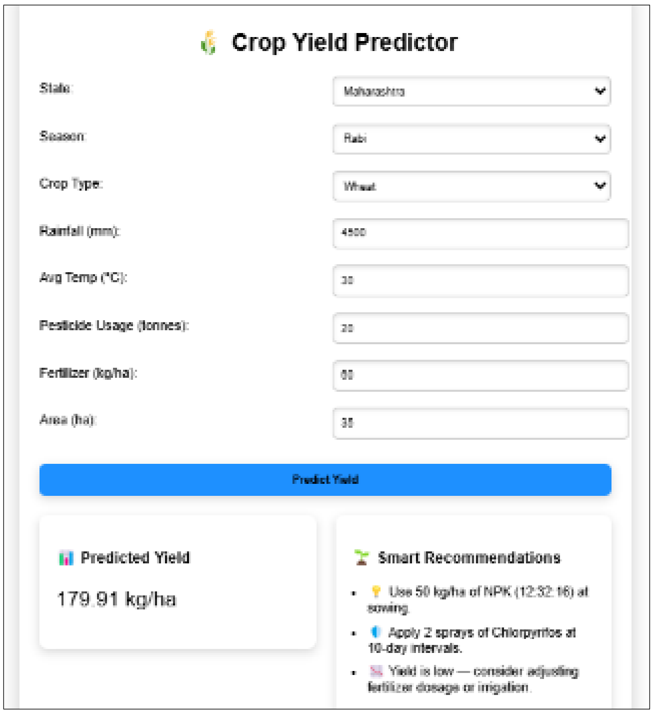

Figure 4 Output with Recommendations**.**

8. Conclusion

The Crop Yield Predictor project successfully demonstrates how machine learning can be integrated with modern web technologies to address real-world agricultural challenges. By leveraging historical crop data and key environmental inputs such as rainfall, temperature, fertilizer, and pesticide usage, the system provides accurate crop yield predictions and smart, field-level recommendations. The use of regression models like XGBoost and Random Forest, combined with preprocessing techniques and hyperparameter tuning, ensures the model is both reliable and scalable. The FastAPI backend seamlessly serves predictions in real-time, while the React frontend offers a user-friendly interface for input and visualization.

This system not only enhances decision-making for farmers and agricultural planners but also paves the way for smarter, data-driven farming practices. Unlike traditional prediction systems, this platform delivers personalized insights and integrates advisory components that improve input efficiency and crop planning. With further enhancements—such as incorporating real-time weather data, soil sensors, or mobile deployment—the Crop Yield Predictor can evolve into a comprehensive smart agriculture tool that promotes sustainability and productivity in the farming sector.

Compliance with ethical standards

Disclosure of conflict of interest

There is no conflict of interest.

References

[1] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151, 61–69.

https://doi.org/10.1016/j.compag.2018.05.012

[3] Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. Artificial Intelligence in Agriculture, 2, 1–12.

https://doi.org/10.1016/j.aiia.2019.05.004

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD, 785–794.

https://doi.org/10.1145/2939672.2939785

[6] Breiman,L.(2001).Randomforests.Machin Learning, 45(1), 5–32.

https://doi.org/10.1023/A:1010933404324

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

[8] Patel, K., Patel, R., & Panchal, D. (2021). Crop yield prediction using machine learning techniques. Materials Today: Proceedings, 47(Part 9), 2957–2960.

https://doi.org/10.1016/j.matpr.2021.05.367

[9] haki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621. https://doi.org/10.3389/fpls.2019.00621

[10] FAO – Food and Agriculture Organization of the United Nations. (2021). The role of digital technologies in sustainable agriculture.

https://www.fao.org/3/cb1639en/cb1639en.pdf

Author's short biography

Mr. Pavan Kumar Vanma I am an Assistant Professor with a background in B.Tech (IT) and M.Tech (CSE). With over 6 years of professional experience, my research interests primarily lie in Machine Learning and Big Data Analytics. I am passionate about exploring innovative solutions and applying advanced data analysis techniques to address complex challenges in the fields of AI and large-scale data processing. Throughout my career, I have focused on developing methods and technologies that can enhance decision-making and operational efficiency, guiding my students in research that contributes to the advancement of knowledge in these rapidly evolving domains.	
Joel Booma I am a B.Tech student with a strong interest in Blockchain and Machine Learning. Currently, I am expanding my skills in Full Stack Development and Blockchain technologies. My research focuses on Blockchain applications and Machine Learning, aiming to build innovative solutions that leverage these technologies. I am passionate about exploring decentralized systems and their integration with AI and data analytics to solve complex problems. As part of the Feature-Specific Sentiment Analysis project, I applied machine learning techniques to extract insights from customer feedback. This contributed to feature-based sentiment categorization, enhancing the decision-making process and providing valuable information for product development.	

Moses Chinnappan I am pursuing my B.Tech in Data Science, with experience in Machine Learning, Artificial Intelligence, and Data Analytics. My research interests include sentiment analysis, computer vision, and AI-powered decision systems. I have contributed to projects like Feature-Based Sentiment Analysis, virtual try-on systems, and early pest detection using AI. By applying machine learning and deep learning techniques, I aim to develop impactful solutions to real-world problems. My focus is on leveraging AI technologies to enhance decision-making and improve operational efficiency across various industries, combining theory with practical applications for positive outcomes.	
Balakrishna Macharla I am a B.Tech student with a passion for software development, automation, and data analytics. My expertise includes Java development, Python programming, Power BI, and UiPath for process automation. My research interests lie in sentiment analysis, process automation, and data-driven decision systems. I have hands-on experience automating workflows with UiPath, improving operational efficiency. One of my significant projects is "Specific Sentiment Analysis of iPhone Reviews," where I used machine learning to extract insights from customer feedback. I also utilize Power BI for data visualization, aiming to turn complex datasets into actionable insights and help businesses make informed decisions.	
Tharun Kali I am a B.Tech student with a keen interest in Convolutional Neural Networks (CNN) and Artificial Intelligence. My focus is on advancing AI knowledge, especially in computer vision applications. I am passionate about solving real-world problems using CNN techniques in image recognition, object detection, and other AI-driven tasks. Through my work, I aim to develop intelligent systems capable of interpreting and understanding visual data. I contributed to the Feature-Specific Sentiment Analysis project, applying machine learning models to analyze and categorize customer feedback based on product features, providing valuable insights into user sentiments and enhancing the product development process.