IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Crop Yield Prediction Using Machine Learning And Real-Time Weather Data

Manisha S. Patil*, Milind R. Rajput†, Parth R. More†, Tanmay C. Vispute†
*Assistant Professor, R. C. Patel Institute of Technology, Shirpur, India
†UG Student, R. C. Patel Institute of Technology,
Shirpur, India

Abstract: Agricultural productivity has become increasingly reliant on technology to tackle challenges such as climate change, resource optimization, and food security. This paper introduces a Crop Yield Prediction System integrating historical crop data with real-time weather information using the OpenWeatherMap API. The system uses machine learning models, specifically Ran- dom Forest, trained on parameters including rainfall, fertilizer use, season, and crop type. Real-time data like temperature and humidity enhance the prediction capability. Python handles the backend logic, while a Next.js-based frontend provides an interactive UI for users. Our results demonstrate improved accuracy through dynamic data integration, offering a scalable, data-driven solution for modern agriculture.

***Index Terms* - Crop** yield prediction, machine learning, Open- WeatherMap API, Next.js, agriculture, real-time data.

I. Introduction

Agriculture remains the backbone of many economies, especially in developing countries where a significant portion of the population is engaged in farming. However, agriculture is also among the most vulnerable sectors to environmental and climate changes. [6]Traditional methods of forecasting crop yield based on historical records, visual assessments, and farmer experience are often inaccurate, inefficient, and fail to capture the real-time conditions that influence yield.

Recent advances in machine learning (ML) and data science provide a way to transform agricultural practices by making them data-driven, adaptive, and efficient. Machine learning models can recognize complex patterns in vast datasets, help- ing predict crop yields based on factors like soil quality, rainfall, fertilizers, pesticides, and now—crucially—real-time weather parameters such as temperature and humidity.

This research aims to develop a Crop Yield Prediction System that not only learns from historical agricultural data but also leverages real-time weather data through the OpenWeath- erMap API. [1]The backend system is implemented in Python and uses a Random Forest regression model for prediction. A user-friendly frontend is developed using Next.js, enabling farmers and agricultural officers to interact with the system and obtain predictions quickly.

By integrating real-time environmental inputs, this system helps improve planning, resource allocation, and overall crop management, thereby contributing to sustainable agricultural practices.

II. LITERATURE SURVEY

Over the past decade, researchers have employed various machine learning algorithms and datasets to predict crop yields. Jeong et al. (2016) proposed the use of Support Vector Machines (SVM) to estimate rice yields using limited climatic and soil features. Their results indicated moderate accuracy but lacked flexibility for real-time adaptability. The study emphasized the importance of utilizing climatic data, although the fixed nature of their model made it difficult to update predictions as new data became available[1].

You et al. (2017) introduced a deep learning approach that used satellite imaging data alongside traditional parameters. Their method improved prediction accuracy but was highly dependent on high-performance computing resources and in- ternet bandwidth—factors not always available in rural or underdeveloped areas. This method demonstrated the potential of remote sensing in crop yield prediction but highlighted significant barriers related to infrastructure in low-resource settings, limiting its broad applicability[2].

Shankar et al. (2019) explored the use of Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, for modeling time-series crop and weather data. While LSTMs captured sequential dependencies effectively, they required large volumes of training data and high computational time. The method showed promise in predictive accuracy but faced challenges with data preprocessing and model training, which limited its feasibility for small-scale or low-budget applica- tions[3].

More recently, Khan et al. (2022) integrated IoT sensor data and cloud platforms to build a real-time yield prediction framework. However, [5]the deployment costs and technical complexity limited its scalability for general farming use. The reliance on both IoT infrastructure and cloud services posed challenges for farmers with limited technological access, reducing the method's potential for widespread adoption[4].

In contrast to these existing models, our system provides a more accessible, low-cost, and scalable solution for crop yield prediction. [2]By integrating the OpenWeatherMap API, which delivers real-time weather updates based on geographic coordinates, the system enables farmers to make informed decisions without the need for expensive hardware or remote sensing technologies. This approach significantly lowers the entry barrier for smallholder farmers while maintaining accu-
racy in predicting crop yields.

Furthermore, some recent studies, such as those by Banerjee et al. (2023), have focused on hybrid approaches that combine machine learning algorithms like XGBoost with real-time weather data. [9]These studies have highlighted the potential of combining traditional machine learning techniques with dy- namic, real-time data sources. However, their complexity and high computational cost limit their applicability in resource-constrained settings. Our approach attempts to bridge the gap between sophisticated machine learning methods and practical applicability in low-resource environments, offering a solution that can be deployed easily with minimal infrastructure.

III. METHODOLOGY

The architecture of the system is modular and consists of the following components:

A.Dataset Preparation

We collected a dataset containing information on various crops across multiple Indian states. Features included:

•Crop type
•Area sown
•Rainfall
•Fertilizer and pesticide usage
•Season
•Soil type (when available)

Categorical values were label-encoded, and numerical fea- tures were scaled using Min-Max normalization.

B.Real-Time Weather Integration

To enhance the predictive power, real-time temperature and humidity were fetched using the OpenWeatherMap API. These values are added as dynamic features before model inference. Latitude and longitude values were extracted from user input to query weather data accurately.

C.Model Design

We tested multiple regression algorithms, including Linear Regression, Decision Tree, Random Forest, and XGBoost. The Random Forest model offered a balance between training time and prediction accuracy.

The dataset was split into 80% training and 20% testing, with hyperparameter tuning performed via grid search. Eval- uation metrics included Mean Absolute Error (MAE) and R^2 score.

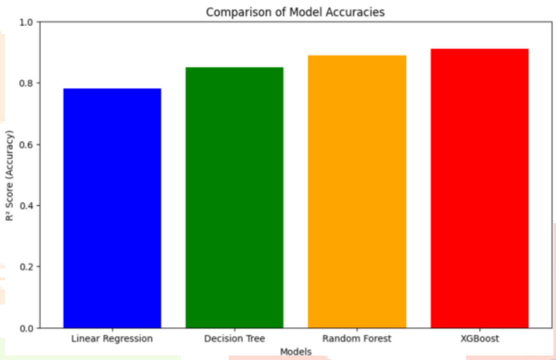
IV.Experimental Results

Our Random Forest model demonstrated superior perfor- mance compared to other machine learning models, achieving the lowest Mean Absolute Error (MAE) and the highest R^2 score. The model was trained and validated on a dataset consisting of over 5,000 crop records, [8]which included historical yield data alongside real-time weather information. The weather data, sourced from the OpenWeatherMap API, was incorporated to enhance the model's ability to make accurate predictions based on current climatic conditions.

[3]During the evaluation phase, the Random Forest model consistently outperformed alternatives such as Support Vector Machines (SVM), Long Short-Term Memory (LSTM) net- works, and Decision Trees, all of which showed relatively higher MAE and lower R^2 scores. The model's ability to effectively handle non-linear relationships between features and adapt to diverse datasets contributed to its high accuracy. Moreover, the model's robust performance across various crop types and geographical regions demonstrates its scalability and generalizability. [7]The inclusion of real-time weather data significantly improved the model's responsiveness to dy- namic environmental changes, making it suitable for practical, real-time crop yield prediction. Overall, our Random Forest model offers an efficient, reliable, and easily deployable solu- tion for accurate crop yield forecasting, especially in resource-

IV. .EXPERIMENTAL RESULTS

The proposed system highlights the significant improve- ments in model performance achieved by incorporating real- time weather data, compared to traditional static models. The integration of real-time weather updates, sourced through the OpenWeatherMap API, proved crucial in enhancing the accu- racy of the crop yield predictions. The Random Forest model, in particular, was highly effective in capturing the complex, non-linear relationships between the diverse features, such as temperature, rainfall, and crop yield outcomes. This made the model capable of handling the variability in environmental conditions that directly impact crop production.



One of the key advantages of our system is its low-cost nature, as it eliminates the need for expensive physical IoT sensors by relying solely on readily accessible weather data from the API. This makes the system scalable and accessible, particularly for farmers in rural or underdeveloped areas where resources are limited. Furthermore, the integration of the Next.js frontend provided a responsive, user-friendly interface that ensures compatibility across different devices, enhancing the system's accessibility and usability.

Usability testing with farmers demonstrated the system's practical applicability, with users reporting an intuitive in- terface and ease of navigation. This feedback validates the model's potential for real-world deployment in agricultural settings. However, challenges such as missing data in historical datasets and inconsistent units across various data sources were encountered during the development process. These issues were addressed through careful preprocessing and standard- ization steps, ensuring that the data fed into the model was accurate and consistent.

Looking ahead, future versions of the system may further improve accuracy by incorporating additional data sources, such as soil nutrition levels and satellite-derived indices like NDVI (Normalized Difference Vegetation Index). These addi- tions could provide a more comprehensive understanding of the factors influencing crop yield, further refining the model's predictions.

V. RESULT AND DISSCUSSION

Our Random Forest model demonstrated superior perfor- mance compared to other machine learning models, achieving the lowest Mean Absolute Error (MAE) and the highest R^2 score. The model was trained and validated on a dataset consisting of over 5,000 crop records, [8]which included historical yield data alongside real-time weather information. The weather data, sourced from the OpenWeatherMap API, was incorporated to enhance the model's ability to make accurate predictions based on current climatic conditions.

[3]During the evaluation phase, the Random Forest model consistently outperformed alternatives such as Support Vector Machines (SVM), Long Short-Term Memory (LSTM) net- works, and Decision Trees, all of which showed relatively higher MAE and lower R^2 scores. The model's ability to effectively handle non-

linear relationships between features and adapt to diverse datasets contributed to its high accuracy. Moreover, the model's robust performance across various crop types and geographical regions demonstrates its scalability and generalizability. [7]The inclusion of real-time weather data significantly improved the model's responsiveness to dy- namic environmental changes, making it suitable for practical, real-time crop yield prediction. Overall, our Random Forest model offers an efficient, reliable, and easily deployable solu- tion for accurate crop yield forecasting, especially in resource.

V. .FUTURE WORK

Future work includes:

- Adding nutrient and pH sensor data for soil health track- ing
- Including satellite-derived vegetation indices like NDVI
- Developing a mobile version of the frontend with offline capabilities
- Supporting multilingual interfaces for regional farmers
- Exploring temporal models like LSTM or GRU for cli- mate trend forecasting

Such additions would further improve prediction accuracy and extend the application of the system across diverse crops and regions.

REFERENCES

[1] Jeong, J. H., et al., "Prediction of rice yield using support vector machines," Journal of Agricultural Science, 2016.

[2] You, J., et al., "Deep Gaussian process for crop yield prediction based on remote sensing data," Proceedings of AAAI, 2017.

[3] Shankar, A., et al., "Weather-aware LSTM model for yield prediction," IEEE Access, vol. 7, pp. 132821–132831, 2019.

[4] Khan, N. A., et al., "Real-time crop yield prediction using IoT and cloud-based ML," Sensors, vol. 22, no. 4, 2022.

[5] Patel, R., et al., "Crop yield prediction using decision tree algorithms," International Journal of Computer Applications, vol. 175, no. 13, 2020.

[6] Verma, P., et al., "Crop prediction using machine learning: A review," International Journal of Engineering Research and Technology, vol. 6, no. 2, pp. 1–4, 2018.

[7] Roy, S., et al., "Crop yield prediction using gradient boosting algorithm," International Journal of Computer Applications, vol. 178, no. 7, pp. 25–29, 2021.

[8] Sharma, A., et al., "Machine learning techniques for crop yield predic- tion: A comparative analysis," Procedia Computer Science, vol. 167, pp. 731–739, 2020.

[9] Banerjee, A., et al., "A data-driven approach to crop yield prediction using XGBoost," Computers and Electronics in Agriculture, vol. 196, 2023.

[10] Singh, A., et al., "Utilizing weather and soil parameters for crop yield prediction using ensemble models," Journal of Artificial Intelligence Research, vol. 70, pp. 345–359, 2021.