ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

AI Based Crop Yield Prediction

Shital Sapkal[1]; Priya Taware[2]; Dr. J.S. Rangole[3]; Harshal Borate[4]

[1]Electronics and Telecommunication Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology Baramati, Maharashtra
[2]Electronics and Telecommunication Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology Baramati, Maharashtra
[3]Electronics and Telecommunication Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology Baramati, Maharashtra
[4]Electronics and Telecommunication Vidya Pratishthan's
Kamalnayan Bajaj Institute of Engineering and Technology Baramati, Maharashtra

Publication Date: 2025/05/16

Abstract: Agricultural productivity has a key role in ensuring food availability, promoting economic resilience, and supporting the long-term well-being of rural communities. This work focuses on harnessing machine learning to better crop production prediction by assessing critical agricultural characteristics like precipitation, temperature, humidity, pH, and nutritional levels of the soil. The expanding population and associated growth in food demand provide considerable challenges to agricultural productivity, demanding creative technology solutions to optimize land management and boost crop yield.The proposed approach leverages machine learning models trained on agricultural datasets, including a Kaggle dataset for crop recommendations. Numerous classification methods are used, including K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Extra Tree Classifier (ETC), and boost predicted accuracy. By incorporating machine learning into farming methods, the system intends to improve judgment in relation to irrigation, fertilization, and crop choice. This research emphasizes the potential of AI-driven solutions to solve important concerns in agriculture, including soil degradation, water shortages, and insect control. The results illustrate the ability of machine learning to boost agricultural efficiency, decrease risks, enhance food security, and encourage sustainable farming methods.

Keywords: *Crop Yield Prediction, Artificial Intelligence, Machine Learning, Agriculture, Review.*

How to Cite: Shital Sapkal; Priya Taware; Dr. J.S. Rangole; Harshal Borate (2025). AI Based Crop Yield Prediction. *International Journal of Innovative Science and Research Technology*, 10(5), 211-218.
https://doi.org/10.38124/ijisrt/25may536

I. INTRODUCTION

Agriculture has a significant role in food production, making it one of the important areas of social concern. However, despite technical breakthroughs, many nations still battle with food shortages owing to increased population growth, climate change, and diminishing soil quality. These problems need new solutions to ensure continuous agricultural expansion and food security. Strengthening food production requires contemporary methodologies that focus on precise crop yield estimation, land appraisal, and crop preservation to support national food policies. Accurate agricultural output assessment is vital for policymakers to analyze export and import demands and increase food security. Various factors impact agricultural yield, including soil conditions, weather patterns, fertilizer use, and seed quality. While classic crop modelling and forecasting approaches have yielded adequate results, researchers are increasingly turning to machine learning (ML) and deep learning (DL) methods to boost prediction accuracy. Machine learning, known for its adaptability across numerous fields, is employed here to anticipate crop categorization based on soil and meteorological characteristics. The prediction model includes major soil properties, including Nitrogen, Phosphorus, Potassium, and pH levels, plus environmental parameters like temperature, humidity, and rainfall.

II. PROBLEM STATEMENT

The agricultural sector is a vital component of food security and economic stability, particularly in nations like India, where a major section of the population depends on farming for their living. Yet, the industry encounters various obstacles, including irregular climate shifts, declining soil quality, degradation, and natural disasters, which significantly effect agricultural yield. These concerns, along with the expanding global population and climate change, heighten uncertainties in agricultural production, leading to financial misery among farmers and decreased interest in

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

agriculture. Despite the availability of government subsidies and traditional crop forecasting methods, accurate and reliable predicting crop output still stands as a difficulty owing to the intricate interaction of different elements, including factors like soil quality and climatic conditions unpredictability, fertilizer usage, and seed quality. Traditional techniques sometimes fall short in addressing these nuanced linkages, which effect productivity and resource optimization.

III. LITERATURE REVIEW

Recent studies have emphasized the relevance involving machine learning techniques and artificial intelligence in enhancing agricultural production prediction and tackling the difficulties posed by climate variability and food security. The scientific article "Crop Yield Prediction using Machine Learning: A Systematic Literature Review" shows how different research apply varied characteristics based on study focus, crop type, geographical location, and data availability, showing the diversity of methodologies in yield forecasting.

- The article "A Comprehensive Review of Crop Yield Prediction using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction" further emphasizes that dataset accessibility and study objectives influence feature selection, contributing to disparities in crop types, locations, and modelling intensity.
- According to the publication "Artificial Intelligence (AI) For Crop Yield Forecasting," rising global population and increased extreme weather events induced by climate change demand accurate crop yield estimates to boost food security. The study highlights several approaches used for forecasting, including crop models, remote sensing, and sophisticated machine learning algorithms, which have proved usefulness both at the community level and in real-world agricultural contexts.
- Meanwhile, the journal publication "Deep Learning for Crop Yield Prediction: A Systematic Literature Review" underlines the complexity of utilizing deep learning algorithms in agricultural productivity prediction, noting that unique hurdles and potential solutions exist for each case study.
- Additionally, the book "Effects of Modern AI Trends on Smart Farming to Boost Crop Yield" gives a detailed analysis of recent breakthroughs in AI-driven smart farming technology. It emphasizes advances such as soil monitoring, climate analysis, irrigation optimization, the use of unmanned aerial vehicles (UAVs), insect and disease control, and weed management, all of which lead to better agricultural yields and sustainable farming methods.

IV. BLOCK DIAGRAM

The block diagram demonstrates the structure and workflow of a machine learning-driven crop yield estimation system that integrates several data sources and predictive algorithms to boost agricultural output.

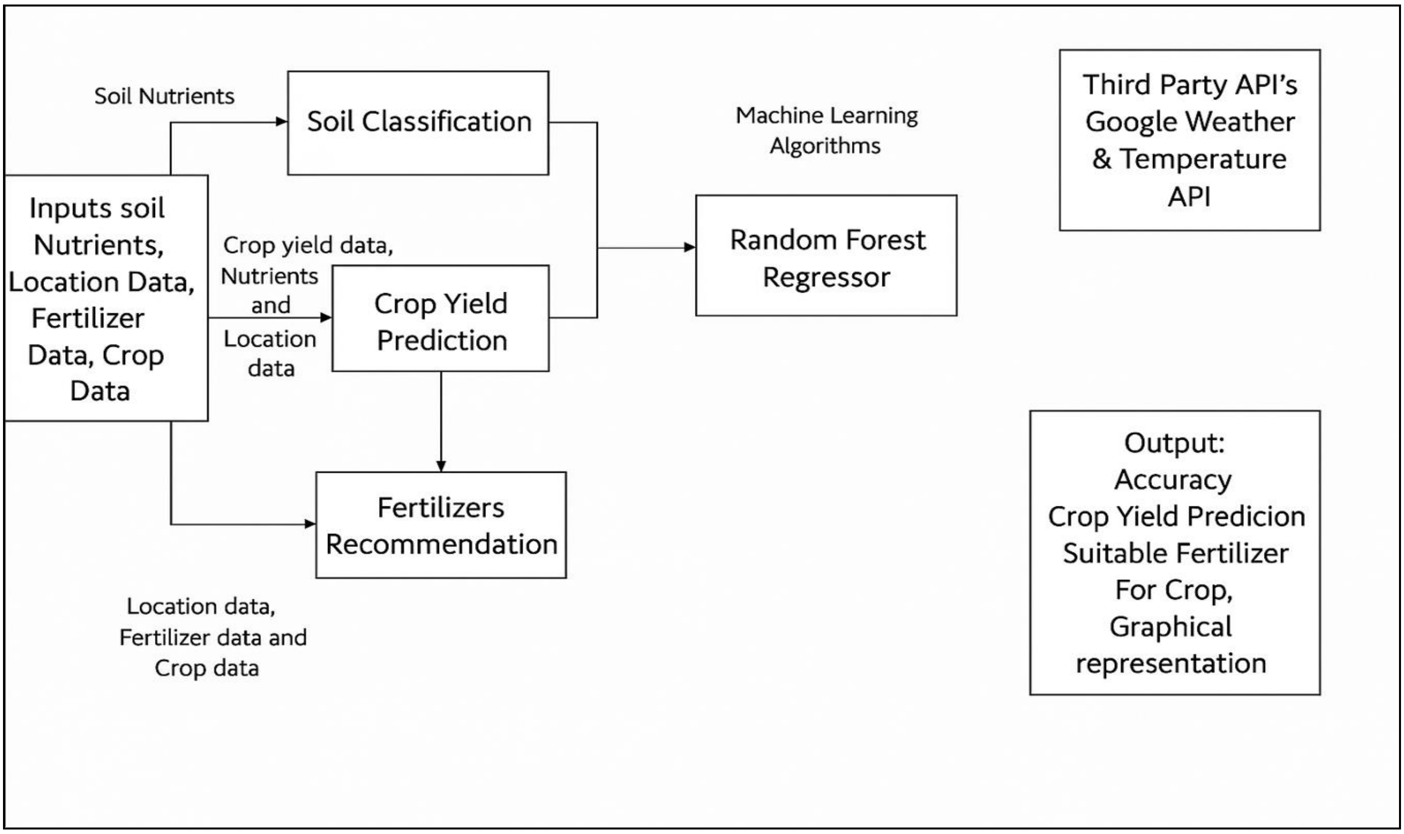

Fig 1 Block Diagram of System

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

The system takes incoming data, conducts categorization and prediction tasks, and outputs vital insights for farmers and agricultural stakeholders. The components and their roles are outlined below:

➢ *Inputs*

In this system, the input block captures vital agricultural data, including soil nutrients, location data, fertilizer consumption information, and crop data. This information is important for developing accurate crop production projections, since it offers the context needed to analyze soil health, meteorological conditions, and fertilizer influence.

➢ *Soil Classification*

This module processes soil nutrient data to categorize soil based on its fertility, type, and nutrient profile. Soil categorization aids in determining soil suitability for different crops and informs suitable fertilizer recommendations.

➢ *Machine Learning Alogorithms*

The system incorporates machine learning methods, notably logistic regression, for both soil categorization and crop production prediction. Logistic regression is used owing to its effectiveness across both binary and multiclass classification tasks issues and its interpretability in forecasting crop results and soil types. We forecasted results using a range of machine learning approaches to guarantee maximum efficiency and subsequently evaluated their performance. In our research, several machine learning algorithms were employed to predict agricultural yield. To further enhance model performance, hyperparameter tuning was implemented using optimization strategies. Additionally, a comparative analysis was conducted among these models, which include:

- Random Forest Regressor (RFR)
- Linear Regression
- Logistic Regression

➢ *Crop Yield Prediction*

Grounded in crop yield data statistics, soil nutrients, and geographical information, this module calculates the predicted yield for a given crop. The forecast uses numerous parameters to increase accuracy, helping farmers optimize resource usage and maximize yield.

➢ *Fertilizer Recommendation*

This component proposes acceptable fertilizers based on the soil categorization results, fertilizer data, crop type, and location. It tries to optimize production while limiting the environmental impact of over-fertilization.

➢ *Third-Party APIs (Google Weather & Temperature API)*

The system incorporates third-party APIs to access real-time weather and temperature data, which are crucial elements impacting crop growth and production. These APIs give up-to-date meteorological data to increase forecast accuracy.

➢ *Output*

The system's final output comprises many crucial elements:

- Prediction Accuracy: Evaluation of the model's precision in forecasting crop yield.
- Crop Yield Forecast: The anticipated yield for a specific crop, determined by the provided input parameters.
- Fertilizer Recommendation: Suggestions for appropriate fertilizers suited to the soil and crop.
- Graphical depiction: A visual depiction of the forecasts and suggestions to assist simple comprehension by users.

This system provides an integrated approach to modernized farming practices, combining real-time data and machine learning techniques and multi-source inputs in order to enhance decision-making, boost crop yields, and promote sustainable agriculture.

V. FLOW CHART

This flowchart provides a systematic procedure for calculating crop production in an agricultural environment, with a focus on soil moisture management and irrigation planning. The process starts with input parameters that comprise numerous aspects crucial to agricultural growth, including crop type, water availability, soil characteristics, climatic data, and other field-related inputs. These parameters constitute the basis for the iterative simulation model. Initially, the variable t t, denoting time is set to $t = 1$ t=1, signalling the beginning of the crop growth phase.

In each iteration, the model calculates critical variables such as the daily reference evapotranspiration (ET 0 0), prospective evapotranspiration (ET), soil water depletion factor, root zone depth, effective rainfall, and actual evapotranspiration. These variables combined assist define the crop's water demands and the prevailing soil moisture conditions. The model then verifies whether the day under consideration corresponds to an irrigation event. If it is an irrigation event, the depth of irrigation is assessed and added to the soil moisture balance calculation. If not, the depth of irrigation is set to zero, and the model proceeds with the natural water balance evaluation based on rainfall and evapotranspiration.

Next, the daily soil moisture balance is updated, reflecting the combined impacts of water intake (through rainfall or irrigation) and water loss (via evapotranspiration). The model then determines whether the crop growth season has completed. If the crop season is still underway, the time variable t t is increased by one day, and the loop restarts with a fresh set of computations for the following day. This iterative approach continues till the crop time is ended.

Once the crop cycle concludes, the model does a cumulative calculation of the total depth of irrigation administered during the whole growing season. This aggregated irrigation depth, coupled with the recorded soil moisture balance and evapotranspiration data, is then utilized to determine the actual crop yield. This technique blends

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

irrigation management zones and soil types and agricultural water dynamics, which are crucial to estimating crop production under varied water availability situations. This flowchart presents a structured framework for constructing AI-based agricultural yield prediction models, where real-time data inputs Key methods for machine learning might be utilized to boost the precision and flexibility of the yield predictions. The flowchart describes a systematic approach to irrigation management and crop production estimation based on field data. It starts with inputting factors such as crop type, water availability, soil qualities, and climate conditions. The procedure initializes with time (t) set to day 1, signifying the beginning of the crop growing phase. For each day, essential factors such potential evapotranspiration (ET_0), soil water depletion, root zone depth, effective rainfall, and real evapotranspiration are determined. The system then examines whether irrigation is required for that day. If irrigation is scheduled, the depth of water to be applied is estimated; otherwise, it is set to zero.

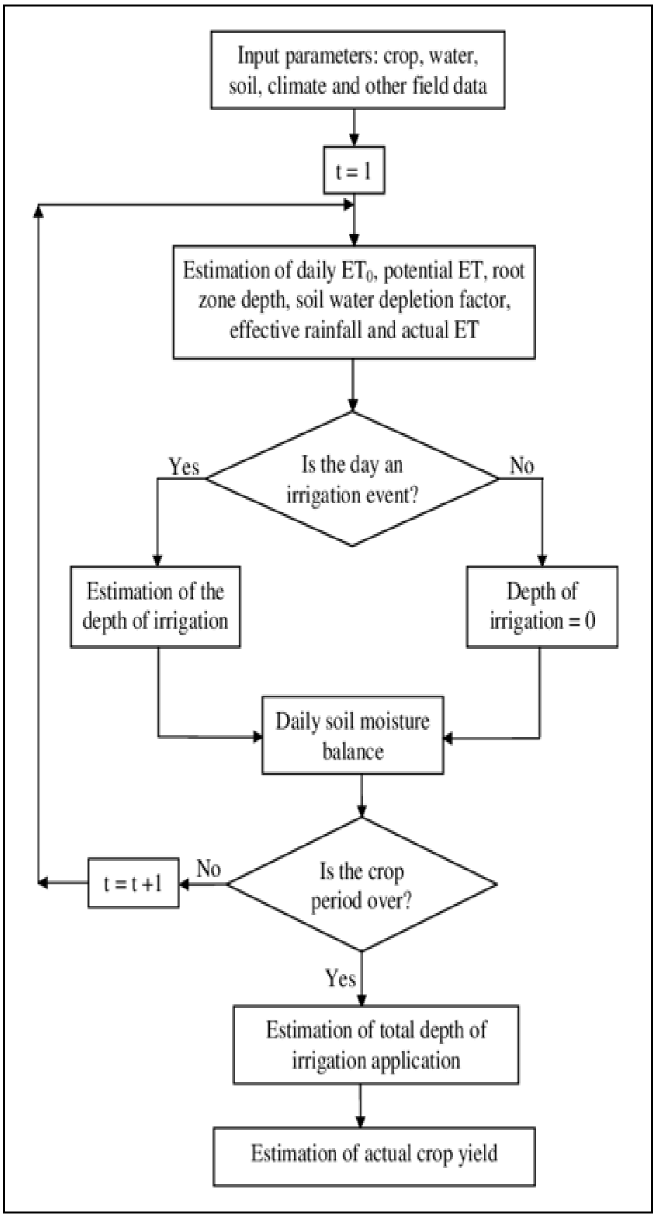

Fig 2 Flow Chart

VI. METHODOLOGY

➢ *Data Preprocessing:*

Data preprocessing is an essential step to ensure that the dataset is clean, consistent, and well-suited for machine learning modeling. The initial phase involved collecting and analysing a crop yield dataset, which contained both numerical and categorical variables.

- *To Prepare the Data:*

✓ **Missing values** were identified and removed to maintain data integrity.
✓ The **target variable**, *Yield*, was separated from the input features.
✓ The dataset included a categorical feature, **Season**, which could not be directly used by most machine learning models. To address this, **one-hot encoding** was applied, converting the categorical variable into a binary format that preserves its information.
✓ **Standardization** was applied to the numerical features to normalize the data. This step ensures that every feature makes an equal contribution to the model's learning process by standardizing them to a uniform scale without distorting differences in their ranges.

```
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
from sklearn.model_selection import train_test_split
import joblib


df = pd.read_csv("crop_dataset.csv")


df['Season'] = df['Season'].str.strip()


df = df[['Fertilizer', 'Pesticide', 'Annual_Rainfall', 'Area', 'Season', 'Yield']]


df = df[df['Season'].isin(['Summer', 'Kharif', 'Rabi', 'Whole Year'])]


df = pd.get_dummies(df, columns=['Season'])

X = df.drop('Yield', axis=1)
y = df['Yield']
```

➢ *Model Development:*

The model was developed using a **technique for supervised machine learning intended for regression issues.**

- *A Random Forest Regressor was chosen due to its:*

✓ Stability against noise and outliers.
✓ Capability to manage non-linear dependencies.
✓ Resistance to **overfitting** by combining predictions from multiple decision trees.

➢ *The Random Forest Algorithm:*

- Constructs numerous decision trees during training.
- Averages their outputs to produce a more accurate and stable prediction.
- Handles complex agricultural data where variables like rainfall, temperature, and soil properties interact in non-linear ways.

➢ *A Pipeline Structure was used to:*

- Seamlessly integrate **data preprocessing** (e.g., scaling and encoding).
- Ensure consistent and repeatable transformation during both training and testing phases.
- Improve overall workflow organization and lower the data risk leakage.

```
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

➢ *Model Training and Evaluation:*

The dataset was separated into training and testing subsets in order to train and evaluate the model. The model was fitted using the training set, and its performance on unseen data was assessed using the testing set.

- MSE - Standard regression metrics, such as Mean Squared Error (MSE), which measures the average squared difference between expected and actual yield values, were used to evaluate the model's performance. The error in the target variable's original units is provided by the Root Mean Squared Error (RMSE), which is the square root of MSE.
- RMSE - The R-squared (R^2) score is used to calculate the percentage of the dependent variable's variance that can be accounted for by the independent variables.
- R-Squared - The model's ability to accurately anticipate outcomes and capture underlying patterns was demonstrated by its strong R2 score. The accuracy of the model and its potential for real-world application in agricultural settings were further validated by the comparatively low RMSE.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

VII. RESULTS

To examine the success of our AI-driven crop production forecast system, we designed a user-friendly graphical interface that facilitates easy entry of agricultural information. As seen in Figure 1, the interface contains clearly labelled areas for inputting amounts of fertilizer (in kilograms), pesticide (in kilograms), rainfall (in millimetres), and the area under cultivation (in hectares). These inputs reflect the most significant elements impacting crop output and are critical for developing accurate agronomic projections. In addition, the interface allows users to pick the proper planting season—such as Rabi, Kharif, or Zaid—through a dropdown menu, ensuring that seasonal variance in crop cycles is accounted for during the forecast process. The design focuses ease of use, making it accessible to people with minimum technical skills, such as farmers or field officers, but still being powerful enough for agricultural researchers and analysts. Each input field is checked to prevent mistakes in data entry, hence boosting the dependability of the model's output. The season selection function is extremely essential, as various crops are farmed throughout different times of the year, and the environmental and soil conditions change substantially between these seasons. Integrating this seasonal context helps the model to produce more context-aware predictions that coincide with actual agricultural cycles.

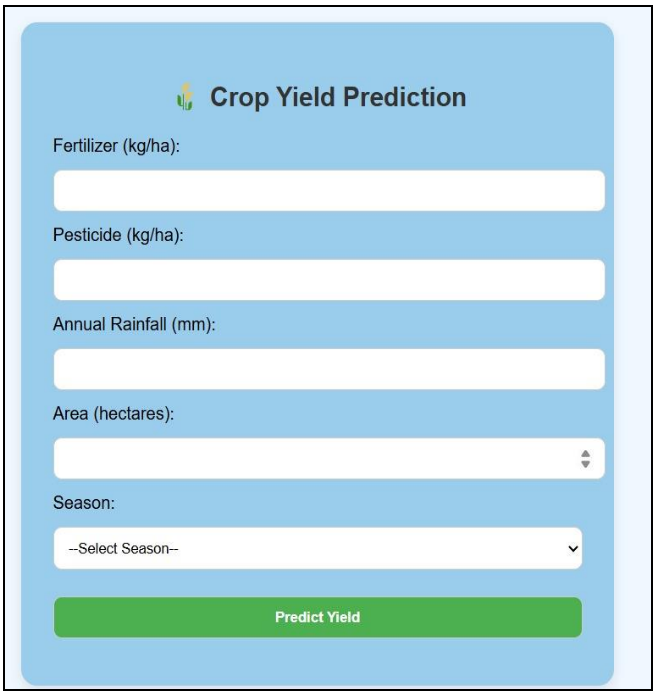

Fig 3 User Interface

In order to confirm the model's functioning, we tested it using a realistic set of agricultural data inputs. As illustrated in Figure 2, the form was filled with the following example values: 31310.93 kg of fertilizer, 101.99 kg of pesticide, 1,556.1 mm of rainfall, and 329 hectares of cultivated land. The season picked was "Summer," which is traditionally defined by winter crops cultivated between October and March. Upon submission, the model processes these inputs through the trained machine learning pipeline and delivers the anticipated crop production. The processing is immediate and does not require any extra user participation, making the tool viable for real-world application.

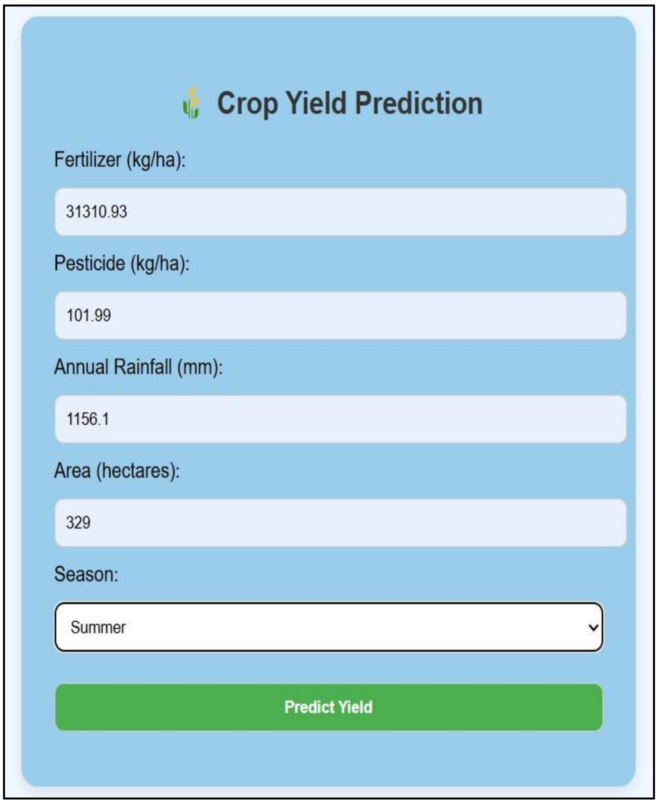

Fig 4 Data Inputs

The result of this test run is illustrated in Figure 3, where the model properly predicted a crop yield of 2.0829 quintals. The yield output is given below the form with a visual highlight for greater user comprehension, containing a crop symbol and color-coded text for emphasis. After prediction, the input fields are immediately removed, enabling for a fresh set of inputs to be inputted seamlessly. This predicted performance indicates the model's practical application in agricultural planning and advising systems. Farmers and agricultural officials may apply this application to make data-driven choices, optimize resource consumption, and even boost production by modifying input factors before implementation. The results confirm that our algorithm can effectively analyze huge and complicated information to create meaningful and practical agricultural forecasts.

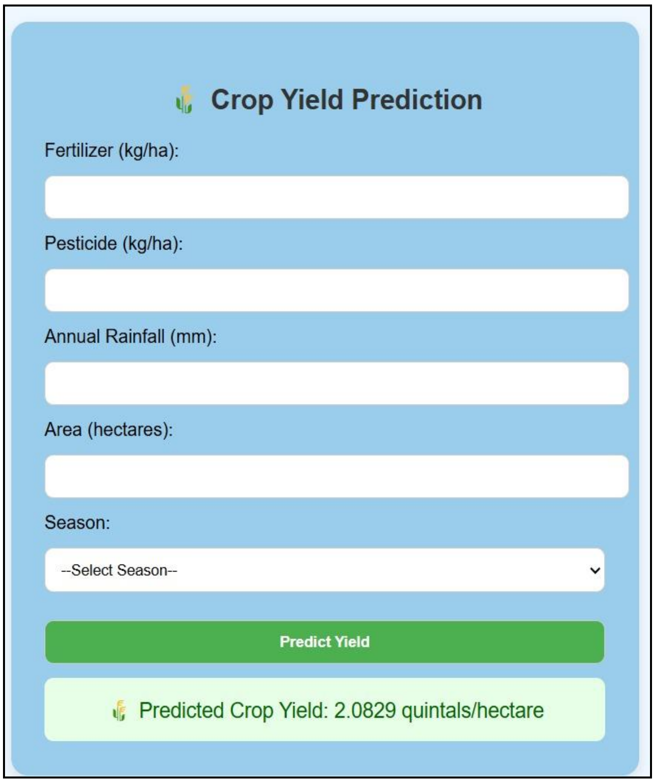

Fig 5 Predicted Crop Yield

VIII. CONCLUSION

The AI-Based Crop Yield Prediction system effectively utilizes machine learning techniques and instantaneous agricultural data to improve judgment in farming. By analyzing key factors like soil nutrients, weather conditions, and geographical aspects, the model offers accurate yield predictions and fertilizer recommendations. This approach not only boosts crop productivity but also optimizes resource usage, minimizes environmental impact, and promotes sustainable farming practices. The integration of algorithms such as the Random Forest Regressor has demonstrated promising accuracy in yield prediction across various seasons and conditions. This system provides farmers with actionable insights, supporting food security and agricultural resilience amidst challenges like population growth and climate change. Future improvements, such as the incorporation of more dynamic datasets, IoT integration, or deployment on cloud platforms, could further enhance the solution's scalability, precision, and accessibility.

REFERENCES

[1]. V. Yadav, A. Sinha, M. R. Gupta, and S. Patel, ‘‘Ethical considerations in deploying ML models in agriculture,’’ AI & Soc., vol. 38, pp. 85–101, May 2023.

[2]. X. Jiang, Y. Sun, and F. Zhao, ‘‘Decision tree-based regression models for predicting crop yield using remote sensing and climate data,’’ Agricult. Syst., vol. 211, Jul. 2023, Art. no. 103277.

[3]. R. Sharma, P. K. Tiwari, and K. Gupta, ‘‘Temporal crop yield prediction using LSTM networks,’’ Agricult. Forest Meteorol., vol. 318, Mar. 2024, Art. no. 108924.

[4]. H. Patel, K. Shah, and A. Mehta, ‘‘Impact of soil properties on crop yield: A comprehensive analysis,’’ Soil Sci. Soc. Amer. J., vol. 87, pp. 45–58, Jul. 2023.

[5]. A. Banerjee, S. Ray, and M. Sarkar, ‘‘Machine learning in precision agriculture: A review of techniques and applications,’’ Comput. Electron. Agricult., vol. 199, Oct. 2023, Art. no. 107175.

[6]. D. Gomez, J. E. Garcia, and S. Lopez, ‘‘ML models for pest and disease detection in crops: Current status and future directions,’’ Comput. Electron. Agricult., vol. 202, Jan. 2024, Art. no. 107340.

[7]. M. U. Hassan, S. Ahmed, and A. Khan, ‘‘A machine learning-based framework for crop selection and yield optimization for smallholder farmers,’’ Agricult. Syst., vol. 209, Nov. 2023, Art. no. 103204.

[8]. R. Silva, J. Rodrigues, and A. C. Antunes, ‘‘Cloud-based platforms for agricultural data analysis: A review,’’ Comput. Electron. Agricult., vol. 200, p. 107287, Sep. 2023

[9]. Divyanshu Tirkey, Kshitiz Kumar Singh, and Shrivishal Tripathi, "Performance analysis of AI-based solutions for crop disease identification, detection, and classification," pp. 2772–3775, April 2023

[10]. A Systematic Review of Current Trends in Artificial Intelligence for Smart Farming to Enhance Crop Yield, by M. H. Widianto, M. I. Ardimansyah, H. I. Pohan, and D. R. Hermanus, Journal of Robotics and Control (JRC), Vol. 3, No. 3, DOI: 10.18196/jrc.v3i3.13760, pp. 269-278, May 2022.

[11]. "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction," IEEE Comprehensive Review, April 2021, by Mamunur Rashid, Bifta Sama Bar, Yusri Yusup, Mohamad Anuar Kamaruddin, and Nuzhat Khan

[12]. Computers & Electrical Engineering, vol. 92, p. 107104, 2021; G. Idoje, T. Dagiuklas, and M. Iqbal, "Survey for smart farming technologies: Challenges and issues"

[13]. In their article "Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey," published in March 2021 in the Archives of Computational Methods in Engineering, N. Bali and A. Singla provide evidence of this.

[14]. "A thorough review on the crop prediction algorithms," Materials Today: Proceedings, March 2021, doi: 10.1016/j.matpr.2021.01.549, C. G. Anupama and C. Lakshmi.

[15]. "Crop Yield Forecasting using Data Mining," Global Transitions Proceedings, August 2021, doi:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may536

10.1016/j.gltp.2021.08.008, P. Kamath, P. Patil, S. S., Sushma, and S. S.

[16]. The article "An Efficient Crop Yield Prediction Using Machine Learning" by M. A. Manivasagam, P. Sumalatha, A. Likitha, V. Pravallika, K. V. Satish, and S. Sreeram was published in the International Journal of Research in Engineering, Science, and Management in March 2022. It was accessed on December 19, 2022. [Online]. https://journals.resaim.com/ijresm/article/view/1862 is accessible.

[17]. In the International Journal of Reconfigurable and Embedded Systems (IJRES), vol. 11, no. 1, pp. 25–33, March 2022, M. H. Widianto, A. Ramadhan, A. Trisetyarso, and E. Abdurachman, "Energy saving on IoT using LoRa: a systematic literature review," doi: 10.11591/ijres.v11.i1.pp25-33.

[18]. "Crop Yield Prediction Techniques Using Machine Learning Algorithms," 2022 8th International Conference on Smart Structures and Systems (ICSSS), April 2022, doi: 10.1109/icsss54381.2022.9782246, M. Gupta, B. V. Santhosh Krishna, B. Kavyashree, H. R. Narapureddy, N. Surapaneni, and K. Varma.

[19]. Khaki, H. Pham, and L. Wang, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning," Scientific Reports, vol. 11, no. 1, p. 11132, May 2021, doi: 10.1038/s41598-021-89779-z.

[20]. "An Ensemble Algorithm for Crop Yield Prediction," Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), February 2021, doi: 10.1109/icicv50876.2021.9388479, M. Keerthana, K. J. M. Meghana, S. Pravallika, and M. Kavitha.

[21]. "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," IEEE Access, vol. 9, pp. 4843–4873, 2021, doi: 10.1109/ACCESS.2020.3048415, A. Sharma, A. Jain, P. Gupta, and V. Chowdary.

[22]. L. Yu and colleagues, “Dynalytic Hierarchy Process and Delphi-Based Comprehensive Assessment of Soil Moisture Sensing Technology Applications,” Agriculture, vol. 11, no. 11, p. 1116, Nov. 2021, doi: 10.3390/agriculture11111116.

[23]. "Smart Home Security Analysis System Based on The Internet of Things," by R. Yu, X. Zhang, and M. Zhang, in IEEE 2nd International Conference on Big Data, Artificial Intelligence, and Internet of Things Engineering (ICBAIE), March 2021, pp. 596–599. 10.1109/ICBAIE52039.2021.9389849 is the doi.

[24]. Pathfinding Augmented Reality for Fire Early Warning IoT Escape Purpose, International Journal of Engineering Trends and Technology, vol. 69, no. 7, pp. 190–197, 2021, doi: 10.14445/22315381/IJETT V69I7P226; M. H. Widianto, T. E. Suherman, and J. Chiedi.