

Crop Prediction Based on Characteristics of the Agriculture Environment using Machine Learning

Prof.H.D. Gunjal¹, Pratiksha Vijaykumar Sole², Sakshi Rangnath Sonawane³,

Pratiksha Sanjay Supekar⁴

Asst. Professor, Department of Computer Engineering, Vishwabharati Academy's College of Engineering, India¹

Student, Department of Computer Engineering, Vishwabharati Academy's College of Engineering, India^{2,3,4}

Abstract: Agriculture plays a vital role in the economy and food production. Selecting the right crop based on soil and environmental conditions is an important challenge for farmers. Traditional crop selection methods often depend on experience and may not provide accurate recommendations under changing environmental conditions. This paper presents a machine learning based crop recommendation system that combines Convolutional Neural Network (CNN) based soil classification and Random Forest based crop prediction. The system analyzes soil images along with agricultural parameters such as Nitrogen, Phosphorus, Potassium, pH, carbon, moisture, temperature, humidity, and rainfall. The CNN model identifies the soil type, while the Random Forest model recommends the most suitable crop for cultivation. The proposed system helps farmers make informed decisions, improve productivity, and reduce agricultural risks by providing accurate crop recommendations.

KEYWORDS: Agriculture, CNN, Random Forest, Soil Classification, Crop Recommendation, Machine Learning.

I. INTRODUCTION

Agriculture is one of the most important sectors contributing to economic growth and food security. Farmers often face difficulties in selecting suitable crops due to variations in soil properties and environmental conditions. Factors such as soil type, pH value, temperature, humidity, rainfall, and nutrient levels significantly influence crop growth and productivity.

With the advancement of Machine Learning (ML) and Artificial Intelligence (AI), intelligent systems can assist farmers in making better agricultural decisions. Machine learning algorithms are capable of analyzing large datasets, identifying hidden patterns, and generating accurate predictions based on historical and real-time data.

This project proposes an intelligent crop recommendation system that combines image-based soil classification and environmental parameter analysis. A Convolutional Neural Network (CNN) is used to classify soil type from uploaded soil images, while a Random Forest classifier is used to recommend suitable crops based on soil and environmental characteristics. The integration of these techniques provides an efficient and reliable solution for modern agriculture.

The proposed system aims to support farmers by improving crop selection, increasing productivity, and reducing losses caused by unsuitable crop choices. The system can be further enhanced by integrating weather forecasting services and larger agricultural datasets in future developments.

Motivation

- The unpredictability of the environment and the dearth of scientific forecasting techniques cause losses for farmers.
- Conventional crop yield models have few parameters and are not very generalizable.
- Repetitive or unnecessary features are frequently present in feature-rich datasets, which lowers prediction performance.
- Crop selection, water management, and fertilizer use can all be optimized with the application of machine learning techniques, especially ensemble models like Random Forest.

II. LITERATURE REVIEW

ShuaipengFei, Lei Li, ZhiguoHan: A novel ensemble feature selection (EFS) technique was presented in this research to enhance yield prediction using hyperspectral data. The canopy hyperspectral reflectance of 207 wheat cultivars and breeding lines was assessed throughout the flowering, early grain filling (EGF), mid-grain filling (MGF), and late grain filling (LGF) phases after they were cultivated under full and reduced watering regimes, respectively. Then, in order to train deep neural network (DNN) models for yield prediction, 115 vegetation indicators were taken from the hyperspectral reflectance and integrated with four feature selection techniques: mean decrease impurity (MDI), Boruta, FeaLect, and RRelieff.

P Sathya& P Gnanasekaran: A major use of machine learning is crop production prediction, and feature selection is essential to this process. Crop characteristics, soil, and weather are crucial for making an accurate paddy yield prediction. The prediction model is thus made simpler by using feature selection approaches to find pertinent and non-redundant attributes from a bigger dataset. An ensemble feature selection technique is presented in this work, which combines several attribute subsets based on information shared between attributes and between attributes and the class to choose an optimal subset of attributes. K-nearest neighbor, Random Forest, Support Vector Machine, Naive Bayes, and bagging are the five classification algorithms used to test the suggested ensemble approach.

AnujMehla, Sukhvinder Singh Deora, SandeepDalal: The Principal Component Analysis (PCA) technique, feature extraction algorithms, and machine learning models are used in this study to determine the main variables influencing crop productivity in India. Five decades' worth of data from the Indian Departments of Agriculture, Statistics, and Meteorology are examined to give farmers and policymakers insightful information. Twenty factors were examined in this study to ascertain how they affected crop yield in the Indian economy. To determine the crucial elements, three feature selection algorithms-forward feature selection, backward feature selection, and recursive feature elimination-were employed. Three ML models were then used to measure the accuracy of the feature selection algorithms after these three algorithms were used to choose the most crucial aspects from the twenty factors that were chosen: Multiple Linear Regression, XGBoost, and Random Forest.

Mahmoud Abdel-salam, Neeraj Kumar, Shubham Mahajan: The intricate relationships between different environmental elements and crop development, however, are frequently difficult for current methods to quantify, resulting in less-than-ideal forecasts. Knowing which feature is most crucial is therefore essential when using Support Vector Regressor to estimate crop yield. Furthermore, great accuracy might not always be obtained by manually adjusting SVR hyperparameters. In this study, we present a unique paradigm that tackles these issues in agricultural production prediction. To improve prediction accuracy effectively, our framework combines an optimized SVR model with a novel hybrid feature selection technique. The three stages of the suggested framework are the prediction, hybrid feature selection, and preprocessing stages. In order to create a smaller dataset, K-means clustering is applied in conjunction with the correlation-based filter after data normalization during the preprocessing step. A unique hybrid FMIG-RFE feature selection method is then suggested for the hybrid feature selection phase. Lastly, the prediction phase presents ICOA, an enhanced version of the Crayfish Optimization Algorithm (COA), which is used to optimize the SVR model's hyperparameters and achieve higher prediction accuracy in addition to the innovative hybrid feature selection method. The performance of the suggested framework is tested and evaluated through a number of experiments.

YashrajPatil, Rani Fathima, SridheviSundarajan, P. SrideviPonmalar: In order to ascertain how feature selection and model complexity impact the predictive performance of several regression models for agricultural production prediction, this study was conducted. On a wheat production dataset, two experiments were carried out: one with a smaller number of features and one with a larger set. The research employed models of Linear Regression, Random Forest, and Gradient Boosting. Among the performance metrics were R-squared (R^2), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). With an MSE of 4,316, RMSE of 65.7, and R^2 of 0.98, linear regression excelled over complex models in Experiment 1. Experiment 2 demonstrated significant improvement with Linear Regression, reducing the MSE to 1,757 and increasing R^2 to 0.99. In predicting crop yield, these results cast doubt on the general preference for intricate models.

Ping Jiang, GuizhenNiu, Guangming Li: The model of deep learning Soil moisture is frequently predicted using LSTM; nevertheless, the forecasts produced by 12 LSTM alone are not very accurate. In this paper, we suggest a

strategy to address this problem by predicting soil moisture by combining meteorological variables with correlation 13 of soil at various depths. 14 It is established by confirming the autocorrelation and cross-correlation of soil at various depths that the properties of soil moisture may be thought of as a smooth time series, and that soil moisture can be predicted using meteorological data. 15 Subsequently, soil moisture was predicted using the LSTM model based on meteorological data, and the difference between the actual and anticipated values was found to be considerable 16 and rebuilding the model increased the prediction's accuracy. We chose six monitoring points in unirrigated (rainfed) wheat fields in the designated regions of the Yellow 18 and Huaihai Seas to collect moisture data in order to validate the 17 proposed model. Accurate moisture prediction values were obtained, demonstrating the model's validity.

III. METHODOLOGY

Step 1: Data Collection

Collect agricultural data such as:

- Temperature
- Humidity
- Rainfall
- Soil Type
- pH Value
- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)
- Carbon
- Moisture
- Season
- Taluka

Step 2: Data Preprocessing

Prepare the dataset by:

- Removing missing values
- Removing duplicate records
- Converting categorical data into numerical form
- Normalizing the dataset

Step 3: Feature Selection

Relevant agricultural parameters such as Nitrogen, Phosphorus, Potassium, pH, temperature, humidity and rainfall are selected and prepared for model training.

Step 4: CNN based soil classification

The uploaded soil image is processed using a Convolutional Neural Network (CNN). The CNN extracts important visual features from the soil image and predicts the soil type. The CNN model classifies the uploaded soil image into Black Soil, Cinder Soil, Laterite Soil, Peat Soil, or Yellow Soil categories.

Step 5: Random Forest Model Training

The agricultural dataset containing soil and environmental parameters is used to train a Random Forest classifier. Multiple decision trees are generated and combined to improve prediction accuracy.

Step 6: Crop Recommendation

Based on the predicted soil type and user-provided parameters such as Nitrogen, Phosphorus, Potassium, pH, carbon, moisture, temperature, humidity, and rainfall, the Random Forest model recommends the most suitable crop. The system also provides fertilizer recommendations based on the predicted crop and soil characteristics.

Step 7: Result Evaluation

The performance of the model is evaluated using accuracy, precision, recall, and confusion matrix.

IV. EXISTING SYSTEM

The current crop prediction systems primarily rely on farmers' expertise and conventional statistical models to predict yields. Single machine learning techniques like Naïve Bayes and Decision Trees are also used in some methods. But because these algorithms struggle to manage unbalanced datasets, overlook redundant features, and are unable to adjust effectively to climate fluctuation, they frequently produce inaccurate results. This limits the scalability and dependability of their forecasts.

V. PROPOSED SYSTEM

The proposed system is an intelligent crop recommendation platform that utilizes Machine Learning techniques to assist farmers in selecting suitable crops. The system consists of two major components: CNN based soil classification and Random Forest based crop recommendation.

Initially, the user uploads a soil image to the system. The image is processed using a Convolutional Neural Network (CNN), which identifies the soil type based on visual characteristics and patterns present in the image. After soil classification, the user provides agricultural parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH value, carbon, moisture, temperature, humidity, rainfall, season, and taluka.

These parameters are analyzed using a Random Forest classifier trained on agricultural datasets. The model evaluates the provided inputs and predicts the most suitable crop for cultivation. The system then displays the recommended crop along with relevant information to support agricultural decision-making.

The proposed solution improves accuracy, reduces manual effort, and provides quick recommendations, helping farmers maximize crop yield and resource utilization.

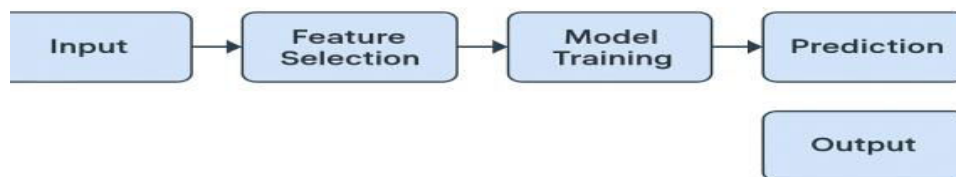


Figure 1. System Architecture

VI. CONCLUSION

The proposed system integrates CNN based soil classification with Random Forest based crop recommendation. By analyzing soil images and agricultural parameters, the system recommends suitable crops for cultivation. The system can assist farmers in making informed decisions, improving productivity, and reducing agricultural risks.

Future work may include integration with weather forecasting APIs, larger agricultural datasets, and enhanced fertilizer recommendation techniques.

VII. RESULT



Fig 01: Home Page

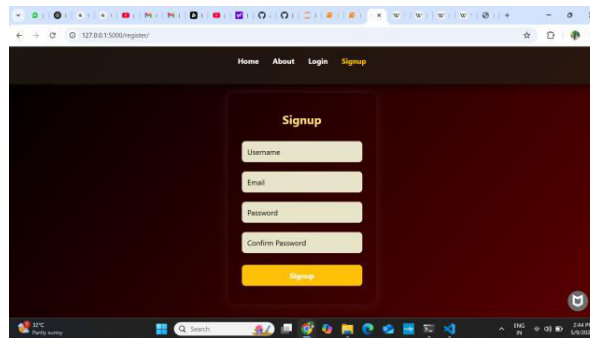


Fig 02: Registration Page

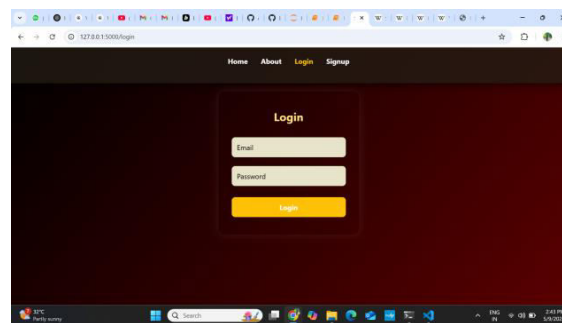


Fig 03: Login Page

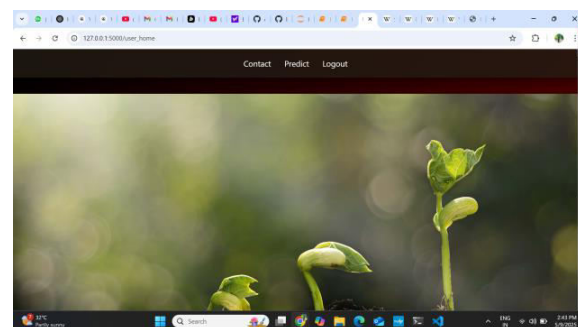


Fig 04: User home Page

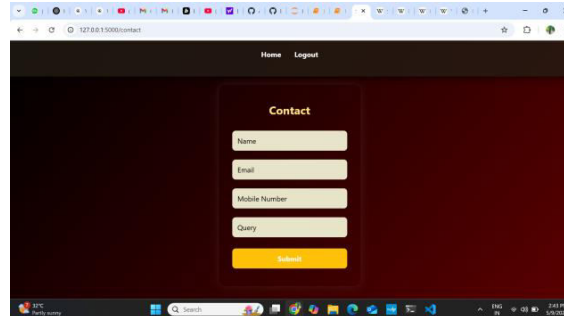


Fig 05: Contact Page

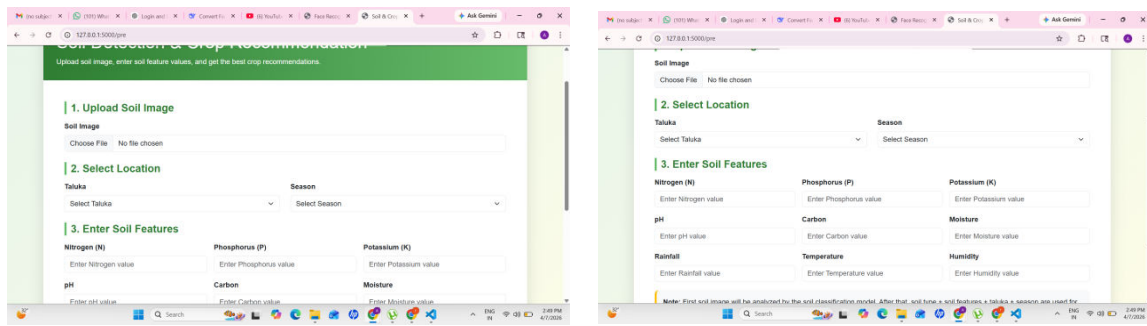


Fig 06: Crop input Page

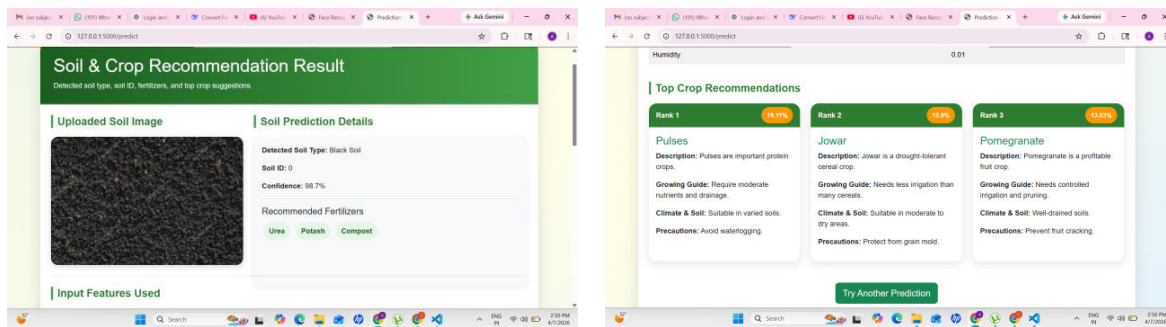


Fig 07: Crop and fertilizer recommendation Output Page

REFERENCES

1. S. Fei, L. Li, and Z. Han, "Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield," *Plant Methods*, vol. 18, no. 119, 2022. doi: 10.1186/s13007-022-00949-0
2. P. Sathya and P. Gnanasekaran, "Ensemble feature selection framework for paddy yield prediction in Cauvery Basin using machine learning classifiers," *Cogent Engineering*, vol. 10, no. 1, p. 2250061, 2023. doi: 10.1080/23311916.2023.2250061
3. A. Mehla, S. S. Deora, and S. Dalal, "Improving crop yield prediction models with optimization-based feature selection and filtering approaches," *Indian Journal of Science and Technology*, vol. 16, no. 47, pp. 4512–4524, 2023.
4. M. Abdel-salam, N. Kumar, and S. Mahajan, "A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning," *Neural Computing and Applications*, vol. 36, pp. 20723–20750, 2024. doi: 10.1007/s00521-024-10226-x

5. Y. Patil, R. Fathima, S. Sundarajan, and P. SrideviPonmalar, “Impact of feature selection on wheat yield prediction using machine learning,” *International Journal of Design & Nature and Ecodynamics*, vol. 19, no. 6, pp. 1909-1917, 2024. doi: 10.18280/ijdne.190607
6. P. Jiang, G. Niu, and G. Li, “Soil moisture prediction based on long short-term memory networks and meteorological data,” *Research Square* (preprint), Sep. 20, 2024. doi: 10.21203/rs.3.rs-4888081/v1.
7. M Gopal P S and B. R, “Selection of important features for optimizing crop yield prediction,” *Int. J. Agricult. Environ. Inf. Syst.*, vol. 10, no. 3, pp. 54–71, Jul. 2019.
8. D. H. Zala and M. B. Chaudhri, “Review on use of BAGGING technique in agriculture crop yield prediction,” *Int. J. Sci. Res. Develop.*, vol. 6, no. 8, pp. 675–677, 2018.
9. A. Bahl et al., “Recursive feature elimination in random forest classification supports nanomaterial grouping,” *NanoImpact*, vol. 15, Mar. 2019, Art. no. 100179.
10. P. S. Maya Gopal and R. Bhargavi, “Feature selection for yield prediction in boruta algorithm,” *Int. J. Pure Appl. Math.*, vol. 118, no. 22, pp. 139–144, 2018.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details