DOI: https://doi.org/10.36094/sc.v92.2026.AI_Based_Smart_Agriculture_for_Sustainable_Rural.Dasgupta.120

ARTICLE

Sci. and Cult. 92 (1–2) : 120-122 (2026)

AI-BASED SMART AGRICULTURE FOR SUSTAINABLE RURAL DEVELOPMENT IN INDIA: A SOCIO-TECHNOLOGICAL STUDY

PARTHA PRATIM DASGUPTA*

Smart agriculture is a concept of using AI to make farming more efficient and less prone to failure. This research paper talks about the efforts done in the study of crop yield prediction by applying machine learning. The research specifically employed a huge Indian dataset in the course of the experimental work. This dataset contained over 50,000 data instances with soil, climate, and fertilizer characteristics. Incorporating posing of great potential the incorporated data consists of yield, soil, fertilizer and weather-related information. To achieve this goal, three machine learning models are employed namely Linear Regression, Support Vector Regression, and Random Forest Regressor. These models were used to predict the crop yield in kg per hectare and afterward, their performances are compared via MAE, RMSE, and R^2. The chosen algorithm of Random Forest performs the best among others, showing the least error and the highest R^2. In addition, the paper covers the social and economic impacts and how AI-based farming is helping rural sustainability. According to the research, AI can significantly benefit farmers, the city planner, and the policy implementer, thus supporting the long-term development of rural India.

Introduction

Agriculture is the main source of income for rural India, with many households depending on it. However, farmers encounter challenges such as income instability due to climate change, poor soil and water conditions, price fluctuations, and low profitability from outdated methods. To combat these issues, there is a demand for advanced technological solutions like artificial intelligence (AI) and machine learning (ML), which can enhance crop yield predictions and hazard assessments, leading to better resource management and food security through historical data analysis.

This paper focus on using machine learning technology as an agricultural extension agent for forecasting crop yields. Key data features include various parameters such as soil quality, weather conditions, and fertilizer use, with a particular emphasis on soil study. The dataset spans from 1997 to 2012 and the project also considers the social impacts of integrating technology into rural development. The primary objectives: building machine learning models for crop production prediction, comparing different models using real-world data, assessing outcomes using graphs and tables, and talking about how AI-driven agriculture supports rural sustainability are some of the study's main goals.

Literature Review

AI is currently playing a significant role in the new generation of farming. A study by Kaur et al. underlines the implementation of AI in farm automation and decision support- while Kumari et al. sing AI's key role in climate-adaptable agriculture and building resilience, etc[1,2]. Ensemble models such as Random Forest and Gradient Boosting are increasingly preferred in yield-prediction research over traditional methods like Support and Linear Regression Vector Machines due to their ability to manage

* Assistant Professor, Department of Computational Sciences, Brainware University, Kolkata, India
e-mail: ppdasgupta91@gmail.com

non-linearity[3-6]. Not only Traditional AI but also recent advancements of machine learning and image processing are being employed, for example through aerial photography and field robots[7,8]. The literature is full of instances where AI is taken hand in hand with agricultural sustainability and the key purpose is always the same: that - is ensure that the yield is optimal through the use of the right amount of input and at the same time resource management is the key factor to be observed[9-11]. The debate about whether AI through precision converting crops is of benefit or not continues to rage, one side maintains that the approach is not ecological as not only are the fertilizers used harmful to the environment..., on the other hand, it is only the peculiar cases of very low income and the numerous drawbacks of such technologies that would oppose the fact, under these rather than under others can it do the most good.

Proposed Socio-Technological Ai Model

Data acquisition: yield, soil condition, fertilizer usage, and weather data.

Preprocessing: data cleaning, handling missing values, and normalizing.

Feature engineering: agronomy and climatology indicators.

Models: Linear Regression, SVR, Random Forest.

Evaluation: Mean Absolute Error (MAE), Mean Square Error (MSE), R-squared.

Output: a suggestion of input, income and sustainability.

Dataset Description And Methodology

Dataset: 50,765 Indian records, 20 features

Target: Yield_kg_per_ha

Key inputs: Area, N, P, K

Preprocessing: removing empty rows, imputing the mean value

Train-Test Sets: 80% -training, 20% -testing

Models: LR, SVR (RBF), Random Forest

Mathematical Formulation and Performance Metrics

The performance of prediction models is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). MAE and RMSE evaluate average prediction errors, with RMSE emphasizing larger variances. R^2 indicates the model's overall fit to the data and the proportion of observed data variation explained by the model. Superior model performance is characterized by lower MAE and RMSE values, alongside a higher R^2.

Results and Graphical Analysis

Model Performance Comparison: The analysis of regression models highlights the performance of three approaches. Linear Regression shows a high R^2 value of 0.9997, indicating near-perfect correlation. Support Vector Regression underperformed with an R^2 of 0.2088, suggesting poor learning capability for large-scale data. The Random Forest Regressor demonstrated the performance with the lowest MAE of 0.77 kg/ha, an RMSE of 35.01 kg/ha, and an R^2 of 0.9987, confirming its reliability as the chosen model for predictions.

Table 1: Comparison of Model Performance.

Model	MAE (kg/ha)	RMSE (kg/ha)	R^2
Linear Regression	8.29	16.63	0.9997
Support Vector Regression	566.34	855.38	0.2088
Random Forest Regressor	0.77	35.01	0.9987

Actual vs Predicted Yield Analysis: The plot in Fig. 1 a strong correlation between actual values and Random Forest model estimates, with data points closely aligning to the line. At higher yield levels, points show more dispersion, while lower yields exhibit a linear pattern. Outliers do not suggest overfitting, and the model maintains high accuracy across both yield classes without data distortion.

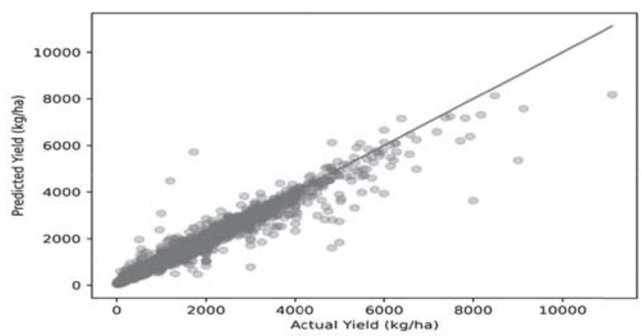

Fig. 1: Actual and Predicted Yield Analysis

Feature Importance Analysis: In Fig. 2 the analysis emphasizes that the most significant input variables affecting yield are cultivated area, total nitrogen input, relative humidity, year, and rainfall. In contrast, COM_code and DIS_code are of lesser importance, while phosphorus and potassium have a minor but noticeable effect. Factors like soil pH, wind speed, solar radiation, and temperature exert negligible influence. Overall, land availability, nutrient inputs, and rainfall conditions are identified as the primary determinants of yield, showcasing a model without target leakage.

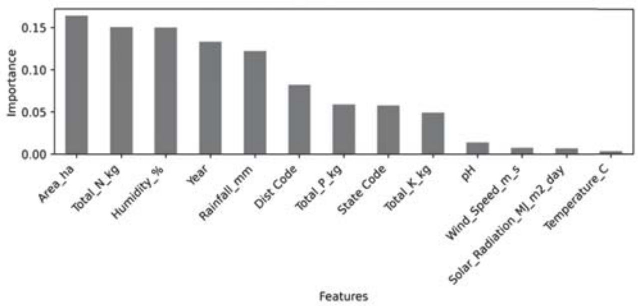

Fig. 2: Feature importance of Random Forest

Residual Analysis: In Fig. 3 the Random Forest model's residual distribution is symmetric around zero with no systematic biases. There are a few severe outliers, but 95% of the residuals fall within ±1500 kg/ha. Approximately 95% of the residuals lie within ±1500 kg/ha, with some severe outliers present. The absence of a funnel structure indicates no heteroscedasticity, suggesting the model is appropriately fit, neither under-fitted nor over-fitted.

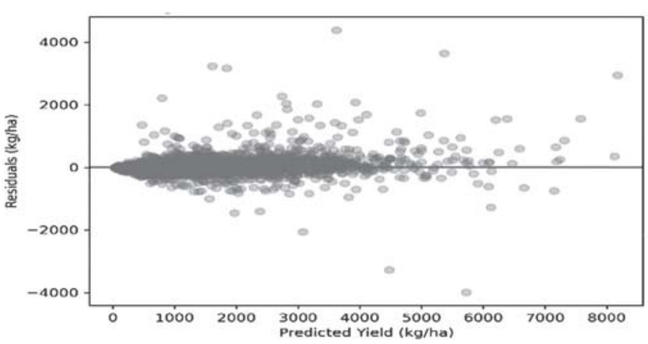

Fig. 3: Residual distribution (Random Forest)

Sustainability and Socio-Economic Impact Discussion

The results of the study have an important impact on sustainable agriculture and rural development.

Environmental Impact: A precise yield prediction would allow the farmers to use the fertilizer and water sparingly hence giving them more yield. This reduces the stress on the soil and water and makes the farming climate friendly.

Economic Impact: When farmers have reliable forecasts, they will not be as financially risky, nor distressed in sales, thus access to credit and insurance will be more improved.

Social and Policy Relevance: AI-based systems for the cases of rural digital inclusion, food security planning, and policy interventions that are timely and location-based.

Model Suitability: Random Forest is the most suitable because of its capability to handle both linear and non-linear relations, noise resistance, and thus preventing overfitting, which is not the case with Linear Regression and SVR.

Conclusion and Future Scope

Here I utilizing machine learning to predict crop yields in India found that the Random Forest model outperformed others, highlighting AI's potential in planning and rural development. However, challenges such as a lack of primary data and real-time sensors were noted. Future improvements could include using diverse data sources like satellite imagery and real-time sensor data, along with deep learning models, to provide better support for farmers through mobile applications. ❐

References

1. G. Kaur, B. Singh, A.K. Angrish and S.K. Bansal, *The Management Accountant Journal*, **57(6)**, 54–57, (2022). https://doi.org/10.33516/maj.v57i6.54-57p
2. K. Kumari, A. Mirzakhani Nafchi, S. Mirzaee, and A. Abdalla, *AgriEngineering*, **7(3)**, 89, (2025). https://doi.org/10.3390/agriengineering7030089
3. N. Alex, C.C. Sobin and J. Ali, *Wireless Pers Commun* **129**, 2345–2385 (2023). https://doi.org/10.1007/s11277-023-10234-5
4. M. Dhanapriya, *Asian Journal of Environment & Ecology*, **24** (6): 141–158, 2025. https://doi.org/10.9734/ajee/2025/v24i6732.
5. Z. Ali, A. Muhammad, N. Lee, M. Waqar and S.W. Lee, *Sustainability*, **17(5)**, 2281, (2025). https://doi.org/10.3390/su17052281
6. Sumit Kurre; Payal Chandrakar; Sudhanshu S. Dadsena. *International Journal of Innovative Science and Research Technology (IJISRT)*, **10**, 5, 2036-2038, (2025) https://doi.org/10.38124/ijisrt/25may1858
7. E. Mohamed, A. Belal, S. Abd-Elmabod, M. El-Shirbeny, A. Gad and M. Zahran, *The Egyptian Journal of Remote Sensing and Space Science.* (2021). https://doi.org/10.1016/j.ejrs.2021.08.007.
8. A. Kaur, *International Journal of Research and Scientific Innovation*, (2025), https://doi.org/10.51244/ijrsi.2025.12050040.
9. M. Batistatos, T. De Cola, M. Kourtis, V. Apostolopoulou, G. Xilouris and N. Sagias, *Agriculture*. (2025). https://doi.org/10.3390/agriculture15080904.
10. H. Shahab, M. Iqbal, A. Sohaib, F. Khan, and M. Waqas, *Comput. Electron. Agric.*, **220**, 108851, (2024). https://doi.org/10.1016/j.compag.2024.108851.
11. F. Shaikh, M. Memon, N. Mahoto, S. Zeadally and J. Nebhen, *IEEE Micro*, **42**, 17-24, (2022). https://doi.org/10.1109/mm.2021.3121279.