



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Machine Learning Methods for Crop Yield Prediction and Climate Change Assessment in Agriculture

Dr.S.Maruthuperumal

Associate Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

Pujari Shivaram, R.Nithin Kumar, R.Praveen Reddy, P.Koushik Reddy

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: Agriculture contributes a significant amount to the economy of India due to the dependence on human beings for their survival. The main obstacle to food security is population expansion leading to rising demand for food. Farmers must produce more on the same land to boost the supply. Through crop yield prediction, technology can assist farmers in producing more. This paper's primary goal is to predict crop yield utilizing the variables of rainfall, crop, meteorological conditions, area, production, and yield that have posed a serious threat to the long-term viability of agriculture. Crop yield prediction is a decision-support tool that uses machine learning and deep learning that can be used to make decisions about which crops to produce and what to do in the crop's growing season. It can decide which crops to produce and what to do in the crop's growing season. Regardless of the distracting environment, machine learning and deep learning algorithms are utilized in crop selection to reduce agricultural yield output losses. To estimate the agricultural yield, machine learning techniques: decision tree, random forest, and XGBoost regression; deep learning techniques- convolutional neural network and long-short term memory network have been used. Accuracy, root mean square error, mean square error, mean absolute error, standard deviation, and losses are compared. Other machine learning and deep learning methods fall short compared to the random forest and convolutional neural network. The random forest has a maximum accuracy of 98.96%, mean absolute error of 1.97, root mean square error of 2.45, and standard deviation of 1.23. The convolutional neural network has been evaluated with a minimum loss of 0.00060. Consequently, a model is developed that, compared to other algorithms, predicts the yield quite well. The findings are then analyzed using the root mean square error metric to understand better how the model's errors compare to those of the other methods.

I. INTRODUCTION

Agriculture is vital for the development of the world. We, humans, benefit from agriculture one way or the other, which has made agriculture a key area of study. Farmers will always need information to refer to, most especially when growing crops that are not common in their land or culture .

The average farmer has access to crude sources of information such as TV, radio, newspapers, fellow farmers, government agricultural agencies, farm supply, and traders. There is, therefore, a need for a system that allows farmers access to relevant information. Machine learning is among the trending technologies; hence, there exist several technologies and systems that run on a machine learning framework.

Project Overview

In recent times, several machine learning systems in agriculture have been tested and created. Research of several machine learning algorithms' effectiveness in agriculture and other application domains has also been conducted and this is because machine learning is a very effective tool for efficient use of resources, prediction, and management, which are needed in agriculture. Machine learning is the ability of an electrical processing system to acquire knowledge and apply that knowledge.

As we all know that agriculture depends largely on the nature of soil and the climatic conditions and many a times, we face unpredictable changes in climate like, non-seasonal rainfall or heat waves or fluctuations in humidity levels, etc. and all such events cause a great loss to our farmers and farming, because of which they are not able to utilize their agricultural land to its fullest.

So to solve all such problems, I have build a Machine Learning Model by the virtue of which we can help farmers, optimize the agricultural production, because this predictive model will help them understand that for a particular soil & given climatic condition, which crop will be best suitable for the harvest.

II. LITEARTURE SURVEY

I .CROP YIELD PRODUCTION USING RANDOM FOREST ALGORITHM

Priya et al.(2020):

There was concerned with the use of the random forest algorithm generate predictions for crop yield and improvement. They used random forest algorithm was used for yield production using dataset with four features or parameters. A training set as used to train the algorithm rules which were then applied to the remaining datasets. The results showed that we can attain accurate crop yield prediction using random forest algorithm achieves a largest number of crop yield models. It is suitable for massive crop yield prediction in agricultural planning.

MING ET AL.(2020):

This work involved classification of land cover based on image and remote sensing. Random forest is an efficient classification algorithm and performs effectively without using special selected features.

JEONG ET AL.(2020):

This work aimed at examining the performance efficiency of the random forest algorithm in crop yield prediction for the wheat crop, potato crop, and maize crop. The work showed that. The random forest algorithm is far more effective in crop yield prediction

II .CROP YIELD PRODUCTION USING ARTIFICIAL NEURAL NETWORKS LIAKOS ET(2021):

This work involved a research into the use of machine learning agricultural production systems using Artificial Neural Networks. The work showed that machine learning models have been used in several agriculture related areas. Mainly crop production and aiding management decision making processes.

MITRA ET AL.(2021):

This work focused on simulating and predicting crop yield for effective crop management and adequate results. The artificial neural network was Effective for simulation and prediction.

III .ADVANCED AGRICULTURE PRODUCTION USING NAÏVE BAYES, MULTILAYER PERCEPTRON, SUPPORT VECTOR MACHINE NITZE ET(2022):

They compared the effectiveness of several machine learning algorithms: Support Vector Machines, Multilayer Perceptron, Support Vector Machine used for classification of different crops Image processing is also done so the SVM is better than ANN.

III. METHODOLOGY

Requirements: The search has become more intense and concentrated on the software's requirements at this time. To comprehend the nature of the programs to be developed, the software engineer must first comprehend the software's information domain, which includes the required functionalities, user interface, and so on. The customer must be informed about the second activity, which must be recorded and presented.

Design: This step is used to transform the above criteria as a representation in the form of "blueprint" software before coding begins. The design must be able to meet the criteria laid out in the previous.

Implementation: The design was converted into a machine-readable format in order for it to be interpreted by a computer in some circumstances, i.e., through the coding process into a programming language. This was the stage in which the programmer will put the technical design phase into action.

Verification: It, like anything else constructed, must first be put to the test. The same may be said for software. To ensure that the application is error-free, all functions must be checked, and the results must closely comply to the previously specified requirements.

Maintenance: Software maintenance, including development, is essential since the software that is being generated is not always exactly like that. It may still have minor faults that were not identified previously when it runs, or it may require additional capabilities that were not previously available in the software.

Useful factors: The Waterfall model has its advantages like it is simple to use. Additionally, while using the model all the system requirements can be defined as a whole, explicitly and at the start the product can run without many issues. It is economic to make changes in the early stages of the project when there are problems with system requirements then when the problems which arise in later stages.

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the of the proposed system. The non- functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

- The system should provide better accuracy.
- The system should have simple interface for users to use.
- To perform efficiently in short amount of time

Reliability : The reliability of the product will be dependent on the accuracy of the dataset of purchase, how much stock was purchased, high and low value range as well as opening and closing figures. Also, the stock data used in the training would determine the reliability of the software.

Security: The user will only be able to access the website using his login details and will not be able to access the computations happening at the back end.

Maintainability: The maintenance of the product would require training of the software by recent data so that their commendations re up to date.

MODELS:

A machine learning model is defined as a mathematical representation of the output of the training process. Machine learning is the study of different algorithms that can improve automatically through experience & old data and build the model. A machine learning model is similar to computer software designed to recognize patterns or behaviors based on previous experience or data. The learning algorithm discovers patterns within the training data, and it outputs an ML model which captures these patterns and makes predictions on new data. In Our Project we compared different machine learning models like oLinear Regression

- XGBOOST
- ADA BOOST
- Decision Tree
- Random Forest

Train and Test:

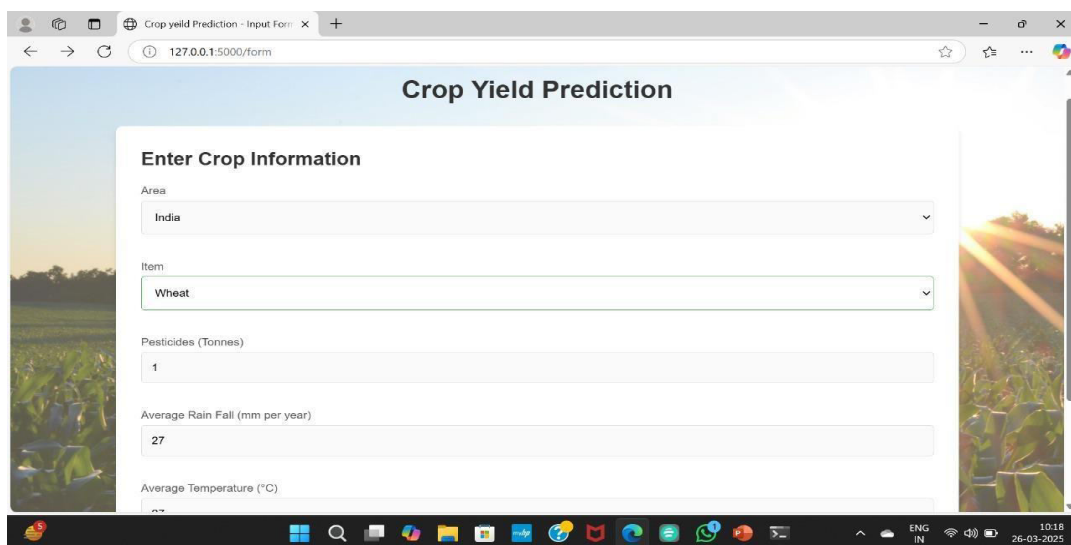
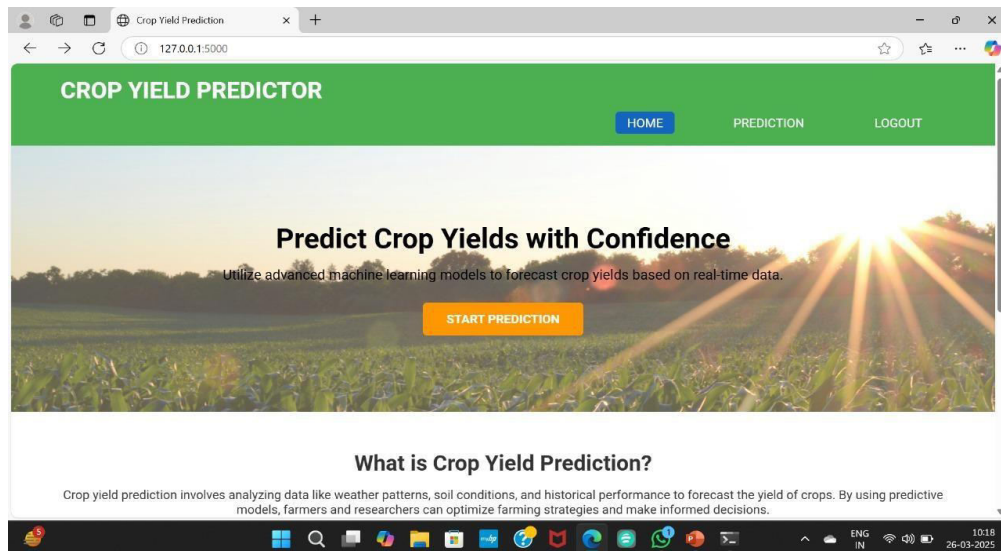
Machine Learning is one of the booming technologies across the world that enables computers/machines to turn a huge amount of data into predictions. However, these predictions highly depend on the quality of the data, and if we are not using the right data for our model, then it will not generate the expected result. In machine learning projects, we generally divide the original dataset into training data and test data. We train our model over a subset of the original dataset, i.e., the training dataset, and then evaluate whether it can generalize well to the new or unseen dataset or test set. Therefore,

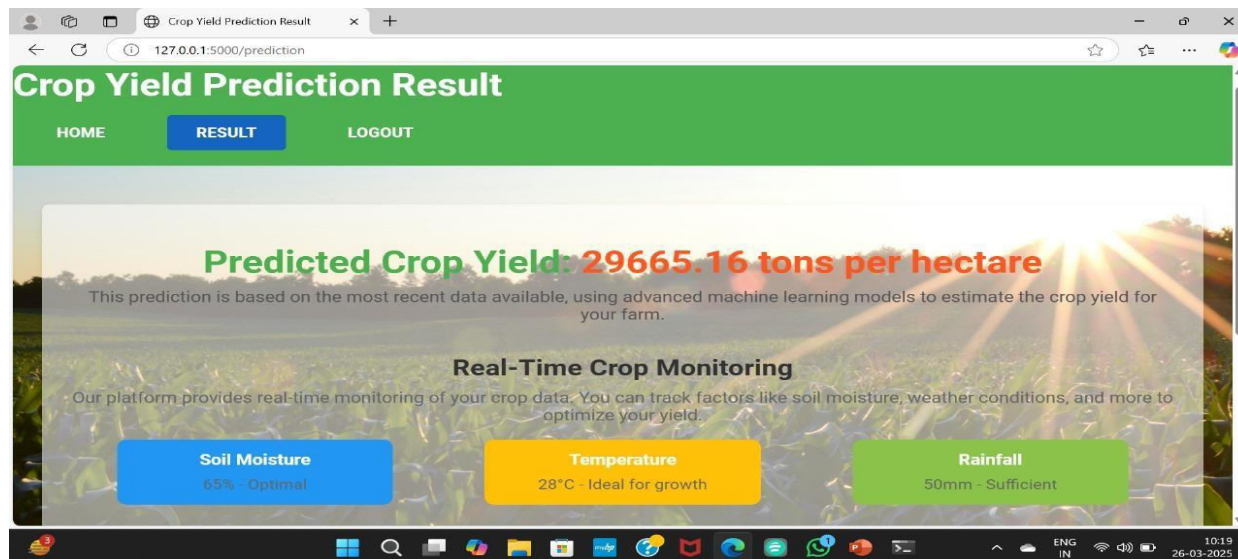
train and test datasets are the two key concepts of machine learning, where the training dataset is used to fit the model, and the test dataset is used to evaluate the model.

IV. RESULTS AND DISCUSSION

In machine learning, testing is mainly used to validate raw data and check the ML model's performance. The main objectives of testing machine learning models are:

- Quality Assurance
- Detect bugs and flaws
- Once your machine learning model is built (with your training data), you need unseen data test your model. This data is called testing data, and you can use it to evaluate the performance and progress of your algorithms' training and adjust or optimize it for improved results. Testing data has two main criteria. It should:
 - Represent the actual dataset
 - Be large enough to generate meaningful predictions





V. CONCLUSION

crop yield prediction was effectively carried out using a combination of machine learning and deep learning techniques. By analyzing critical factors such as rainfall, temperature, crop type, and area of cultivation, the proposed models demonstrated strong predictive capabilities. Among the algorithms tested, the Random Forest and Convolutional Neural Network models outperformed others in terms of accuracy and minimal error, making them highly suitable for real-world agricultural applications. These results emphasize the potential of intelligent data-driven systems to support farmers in making informed decisions, optimizing crop production, and contributing to sustainable agricultural practices. With further enhancements and real-time data integration, such predictive models can play a vital role in ensuring food security and increasing farm productivity across diverse agro-climatic zones.

The future scope of this crop yield prediction project is vast and promising. With advancements in technology, the integration of real-time data from IoT devices such as soil sensors, weather stations, and satellite imagery can significantly enhance the accuracy and responsiveness of yield predictions. The system can be expanded to support a wider range of crops and geographical regions, making it inclusive and scalable. Incorporating reinforcement learning can help the model adapt to changing environmental conditions and farming practices over time. Additionally, a user-friendly mobile or web application can be developed to provide personalized recommendations and alerts to farmers, enabling better decision-making. Collaboration with agricultural departments and local governments can further help in large-scale deployment and policy formulation for farming.

REFERENCES

1. Review article by Indian journal of agricultural sciences
2. F. Amador, M.J. Sumpsi, C. Romero A non-interactive methodology to assess farmers' utility functions: an application to large farms in Andalusia Spain. *European Review of Agricultural Economics*, 25 (1998), pp. 92-109 View Record in Scopus Google Scholar
3. J. Berbel, A. Rodriguez An MCDM approach to production analysis: an application to irrigated farms in southern Spain *European Journal of Operational Research*, 107 (1998), pp. 108-118 Article Download PDF View Record in Scopus Google Scholar
4. J. Gomez-Limon, L. Riesgo Irrigation water pricing: Differential impacts on irrigated farms *Agricultural Economics*, 31 (1) (2004), pp. 47-66 Article Download PDF View Record in Scopus Google Scholar
5. B. Manos, T. Bournaris, M. Kamruzzaman, A. Begum, J. Papathanasiou The regional impact of irrigation water pricing in Greece under alternative scenarios of European policy: a multicriteria analysis *Regional Studies*, 40 (9) (2006), pp. 1055-1068 View PDF Cross Ref View Record in Scopus Google Scholar .

6. IBEF. (2023). Agriculture in India: Industry Overview, Market Size, Role in Development. Accessed: Jan. 20, 2023. India. [Online]. Available: <https://www.india.gov.in/>
7. H.S.Gill, G.Murugesan, B.S.Khehra, G.S.Sajja, G.Gupta, and A. Bhatt, "Fruit recognition from images using deep learning applications," *Multimedia Tools Appl.*, vol. 81, no. 23, pp. 33269–33290, Sep. 2022.
8. S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers Plant Sci.*, vol. 10, p. 621, May 2019.
9. E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environ., Develop. Sustainability*, vol. 22, no. 6, pp. 5687–5708, Aug. 2020.
10. Vemulapalli, G., & Pulivarthy, P. (2025). Integrating Green Infrastructure With AI-Driven Dynamic Workload Optimization: Focus on Network and Chip Design. In *Integrating Blue-Green Infrastructure Into Urban Development* (pp. 397-422). IGI Global Scientific Publishing.
11. P. S. Maya Gopal and R. Bhargavi, "Optimum feature subset for optimizing crop yield prediction using filter and wrapper approaches," *Appl. Eng. Agricult.*, vol. 35, no. 1, pp. 9–14, 2019.
12. P. S. M. Gopal and R. Bhargavi, "Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms," *Appl. Artif. Intell.*, vol. 33, no. 7, pp. 621–642, Jun. 2019.
13. The Food and Agriculture Organization (FAO). (2023). FAOSTAT. Accessed: Jan. 20, 2023. [Online]. Available: https://www.fao.org/faostat/en/#rankings/countries%20_by_com%20modity_exports
14. F. Raimundo, A. Glória, and P. Sebastião, "Prediction of weather forecast for smart agriculture supported by machine learning," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, May 2021, pp. 160–164.
- A. Suruliandi, G. Mariammal, and S. P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Math. Comput. Model. Dyn. Syst.*, vol. 27, no. 1, pp. 117–140, Jan. 2021.
15. P. Mohan and K. K. Patil, "Crop production rate estimation using parallel layer regression with deep belief network," in *Proc. Int. Conf. Electr., Electron., Commun., Comput., Optim. Techn. (ICEECCOT)*, Dec. 2017, pp. 168–173



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