*2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*
*Karnataka, India. Dec 29-31, 2023*

2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) | 979-8-3503-1912-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMARTGENCON60755.2023.10442746

# Enhancing Crop Yield Prediction in India: A Comparative Analysis of Machine Learning Models

Kothakonda Chandhar
*Asst Professor, School of Computer Science & Artificial Intelligence*
*SR University*
Warangal, Telangana, India.
chandu19024@gmail.com.

Kamalakar Ramineni
*Asst Professor, School of Computer Science & Artificial Intelligence*
*SR University*
Warangal, Telangana, India

Erroju Ramakrishna
*Asst Professor, School of Computer Science & Artificial Intelligence*
*SR University*
Warangal, Telangana, India.

Dr T Venkata Ramana
*Associate Professor, Department of AIML, CVR Engineering College,*
Telangana, India.

Achi Sandeep
*Asst Professor, Computer Science & Engineering*
*Sri Indu College of Engineering & Technology, Ibrahimpatnam*,
Rangareddy, India

Karmakonda Kalyan
*Asst Professor, Computer Science & Engineering,*
*Jyothismathi Institute of Technology and Science,*
Karimnagar, Telangana, India.

***Abstract*— Farming is very important for India's economy. Lots of people in India depend on farming to make a living. But farming has become harder because the weather, climate, and environment keep changing. These changes make it tough to grow crops well. We proposed a mechanism called crop yield prediction. It helps us predict how much crops will grow and helps farmers decide what to plant. We use things like where the farm is, how much rain it gets, how hot or cold it is, and what kind of soil is there to make these predictions. These things help us figure out how much crop we can expect. In this proposal we tried out some ma-chine learning models like the Gradient Boosting Regressor, Random Forest Regressor, Support Vector Regressor, and Decision Tree Regressor. We wanted to see which one is the best at predicting crop yields. It turns out the Decision Tree Regressor did the best job and was really accurate with the da-ta we had. This method helps farmers because it means we can use this model to make better predictions about how much food we can grow, which helps the farming community in India grow more food in a good way.**

***Keywords*— *Gradient Boosting Regressor, Random Forest Regressor, Support Vector Regressor, Decision Tree Regressor***

## I. Introduction

Population growth and farming are intricately connected factors that shape the world's future. As the global population continues to expand, the demand for food production grows exponentially. This places immense pressure on agriculture to meet the increasing food requirements. Sustainable farming practices are crucial to balance this equation. In future, more people will want to buy food from farms. So, we need to improve the fields and grow more crops. But, as the Earth warms, bad weather often destroys crops. One problem is that sometimes the soil is not good enough, the climate changes a lot, floods, insufficient water in the land and other problems that destroy the crops. This is causing loss to the farmers who have lost their crops.

Crop yield prediction help to solve these problems. It's like guessing how much crop a farm will make in the future. With this information, farmers can plan better. For example, if we know there might be bad weather or not enough water, we can prepare in advance. We can use special methods to make the soil better and protect the crops from bad weather. This way, even if there are problems, they won't be as bad, and farmers won't lose as many crops. It's like being ready for a storm so it doesn't cause too much damage.

Accurate crop yield prediction plays a crucial role in this scenario by helping farmers plan and optimize their agricultural practices. By predicting how much crop they can expect to harvest, farmers can make informed decisions about planting, resource allocation, and crop selection. This, in turn, enables them to maximize their yield, ensuring a stable food supply.

## II. Literature Survey

Tseng [1] employed IoT devices in smart agriculture to monitor and predict crop yields. Traditional agriculture models often relied on extensive data analysis to fore-cast crop outcomes, which were influenced by various weather conditions. Tseng's novel approach integrated IoT sensors to comprehensively monitor agricultural conditions, including atmospheric or barometric pressure, humidity, moisture levels, temperature, and soil salinity. This IoT-driven big data analysis aimed to better understand farming practices and environmental dynamics. One notable feature of the model was its application of 3D cluster analysis to explore the relationships between environmental factors and farmer-provided guidelines. However, the model dis-played anomalous patterns when confronted with potential risks associated with air humidity, soil moisture content, and temperature variations.

P. M. Gopal and R. Bhargavi [2] tackle the intricate task of crop yield prediction within agriculture by investigating the synergy between two widely-used prediction methodologies: Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN). They introduce an inventive hybrid approach that initializes ANN's input layer weights and bias using MLR's coefficients and bias, leading to enhanced pre-diction precision. This unique hybrid MLR-ANN model is deployed specifically for predicting paddy crop yield, employing a Feed Forward ANN with Back Propagation training. Unlike the conventional ANN, which initializes weights and bias randomly, this hybrid model capitalizes on MLR's parameters for its initialization process. The article conducts a comparative analysis of the hybrid model against ANN, MLR, Support Vector Regression (SVR), k-Nearest Neighbor (KNN), and Random Forest (RF)

models, employing diverse performance metrics. The findings underscore the superior performance of the proposed hybrid MLR-ANN model, underscoring its potential to elevate crop yield prediction accuracy while maintaining computational efficiency.

P. Sivanandhini and J. Prakash [3] center their research on crop yield prediction models that utilize Support Vector Regression, K-Nearest Neighbor, and Decision Tree Regression. They evaluate the performance of these models using assessment metrics like Root Mean Square Error (RMSE) and Mean Squared Error (MSE) by comparing their predictions to actual values. The results indicate that Support Vec-tor Regression demonstrates superior predictive accuracy when compared to the other two models. The study also suggests future research directions, advocating for the incorporation of Artificial Neural Networks (ANN) in crop yield prediction. This extended approach would encompass additional factors such as temperature, soil quality, humidity, and more, with the goal of bolstering the stability and accuracy of crop yield prediction models. This literature review offers valuable insights into the current landscape of crop yield prediction models and provides a glimpse into potential advancements in the field.

O Bazrafshan, El-Shafie [4] underscores the importance of accurately estimating crop yield, with a particular focus on tomato yield (TY), due to its economic significance and health benefits. Accurate TY prediction can boost farmers' incomes and contribute to the economic well-being of countries where tomatoes are cultivated for export. The study introduces a novel approach by employing a Bayesian Model Averaging (BMA) model in conjunction with multiple Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multi-Layer Perceptron (MLP) models for prediction. This innovation showcases the potential of ensemble models in forecasting various variables, providing not only predictions but also uncertainty estimates for input and model parameters. Additionally, the study introduces a new hybrid Genetic Test (GT) for selecting optimal predictors, which can be a valuable tool for forecasting meteorological variables in future studies. Lastly, the paper highlights the integration of robust optimization algorithms to enhance the precision of soft computing models, offering a comprehensive approach to improving crop yield prediction. This literature review underscores the multifaceted nature of the research, combining novel modeling techniques, uncertainty estimation, predictor selection, and optimization strategies to advance the field of crop yield prediction, particularly for tomatoes.

Bondre and Mahagonkar [5] used smart computer techniques to guess how much crop will grow and recommend the right kind of fertilizer. Guessing how much crop will grow is a big challenge in farming, but they made a computer program that could do it. They tested how well their program worked for figuring out how much crop can be produced on a farm. One good thing about their program is that it used old data to guess how much crop will grow in the future. They also used special computer tools like random forest and SVM to help suggest the right fertilizer for each type of crop. However, they didn't include a method for a smart watering system to make the crops grow better.

## III. METHODOLOGY

### A. Data Collection

Collecting data was really important to understand Indian crop patterns. We carefully gathered a lot of information, concentrating only on Indian crops from a big dataset. This careful work made sure that our study focused on the specific features and difficulties of Indian farming. The data set collected from Kaggle Website. about only 8 different crop types. We've made a pie chart that shows in Fig 1. how often each crop was observed.

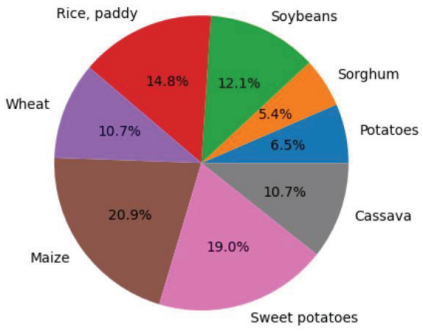


Fig. 1. Distribution of observed crops in India

### B. Data Preprocessing

Data preprocessing is a crucial step for predicting crop yields. It starts with gathering historical data on crop yields, weather, soil, and related info from reliable sources. Then, we carefully check the data for problems like missing numbers or strange values. Fixing these issues is vital to have good data for analysis.

Dealing with missing data and odd values is really important because they can mess up our predictions. We use methods like guessing missing numbers or removing bad data to handle these problems. After that, we create new or adjust existing data to help the computer make better predictions. For crop yields, this might mean combining weather data, calculating plant health, or changing types of crops into numbers.

To make sure our data works well for the computer, we make sure all the numbers are on the same scale. Then, we split the data into groups for training, checking, and final testing. We also pay attention to time if the data has dates. If there are too many numbers to handle, we might use tricks to make it easier for the computer.

### C. Features Extraction

Feature extraction for crop yield prediction involves selecting and transforming relevant data into meaningful variables. This process identifies key factors impacting yields, such as weather data, soil attributes, and crop-specific indicators. Techniques include aggregating weather variables over time, calculating vegetation indices from remote sensing data, and encoding categorical factors. The resulting features provide essential information for predictive models, helping to capture and understand the complex relationships between variables. Effective feature extraction enhances the accuracy of crop yield predictions, benefiting agricultural decision-making and productivity.

### D. Machine Leaning models

*a)* Gradient Boosting Regressor: Using the Gradient Boosting Regressor model for crop yield prediction is like hav-ing a smart tool to guess how much a crop will grow in a specific place and weather. This tool is good at making guesses with numbers, which is perfect for predicting crop yields. Here's how it works: First, we collect old data about crop growth, weather, soil, and other things we need. This data helps teach the smart tool how to make predictions. The smart tool is like a detective. It looks at all the data and tries to figure out the best way to make predictions. It does this by combining lots of smaller guesses (decision trees) into one big guess. Each small guess tries to fix the mistakes of the previous one, like teamwork. After lots of practice, the smart tool becomes really good at finding patterns in the data. For example, it learns how hot or rainy weather affects crop growth. Now, we can give the smart tool new data, like today's weather and soil conditions, and it uses what it learned to guess how much the crop will yield. The great thing about this smart tool is that it can handle tricky situations and give us good predictions, even if the data is not perfect. Farmers and people who plan farming can use this tool to decide when to plant, how to use re-sources, and manage risks. And the more we use it with new data, the better it gets at predicting crop yields, making it super useful for modern farming.

*b) Random Forest Regressor:* The Random Forest Regressor is a machine learning model used in crop yield prediction. It works by combining multiple decision trees to predict crop yields based on various input features like weather data, soil proper-ties, and crop characteristics. Each decision tree in the forest independently makes predictions, and the final prediction is an average or weighted combination of these individual tree predictions. This ensemble approach improves prediction accuracy, handles complex interactions among features, and minimizes overfitting. Random Forest Regressor is particularly effective for crop yield prediction because it can capture the nonlinear and intricate relationships between factors, aiding farmers in making informed decisions for optimal crop management.

*c) Random Forest Regressor (SVR):* Support Vector Regression (SVR) helps predict how many crops a farmer might get. To do this, it uses information like past crop yields, weather, soil details, and more. It's like using the past to guess what might happen with the crops in the future. First, we collect all this information and make sure it's clean and organized. We fix any missing or strange data. Then, we create new numbers that help the computer understand things better, like calculating the total rain for the season or how hot it's been. Next, we split our data into three groups: one to teach the computer, one to check how good it's doing, and one to test it. We also pick the best way for SVR to learn from the data. Once it's learned, we use it to make predictions about future crop yields. This helps farmers make smarter choices about when to plant, water, and harvest their crops. We keep checking and updating our predictions as we get more data to make sure they stay accurate over time.

*d) Decision Tree Regressor:* A decision tree is like a tool that helps us make choices and predictions. It looks at facts and data about something and then gives us an answer or predict. This tool can be used to solve two main types of problems: one is about guessing numbers, and the other is about putting things into categories. The idea is to make a set of rules based on what we already know, and then use those rules to guess the answer for something new. It's like having a map where we start at the top, follow the directions, and at the end, we get our answer. When we're trying to guess numbers, it's called "regression," and when we're sorting things into categories, it's called "classification." So, deci-sion trees help us make smart guesses by following a set of simple rules, and they can be used for different types of problems.

### E. Models Training and Evaluation

We train the chosen model using the training dataset. This involves feeding the model with historical data on crop yields and relevant environmental factors. Additionally, we fine-tune the model's hyperparameters to ensure it performs at its best. This iterative process helps optimize the model's ability to make accurate crop yield pre-dictions.

After training the model, we evaluate its performance to assess how well it can predict crop yields. We use various metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) on a separate testing dataset. These metrics provide insights into the model's accuracy and effectiveness. To ensure we have the best model, we repeat the evaluation process with different models and select the one that performs the best, ensuring reliable and precise crop yield predictions.

## IV. RESULTS

In the context of crop yield prediction, our study evaluated the performance of several ma-chine learning models, including the Gradient Boosting Regressor, Random Forest Regressor, Support Vector Regressor, and Decision Tree Regressor. Among these models, our analysis revealed that the Decision Tree Regressor exhibited the highest accuracy rate when applied to the provided dataset. The results of this assessment are visually depicted in the Fig.2 and are further detailed in the accompanying Table 1. These findings underscore the efficacy of the Decision Tree Regressor as a promising tool for accurate crop yield prediction, offering valuable insights for agricultural planning and resource management.

TABLE I. PERFORMANCE COMPARISON AMONG MACHINE LEARNING MODELS.

| S. No | Machine Learning Models Tested | Accuracy (%) |
|---|---|---|
| 1 | Gradient Boosting Regressor | 89 |
| 2 | Random Forest Regressor | 68 |
| 3 | Support Vector Regressor | 20 |
| 4 | Decision Tree Regressor | 96 |

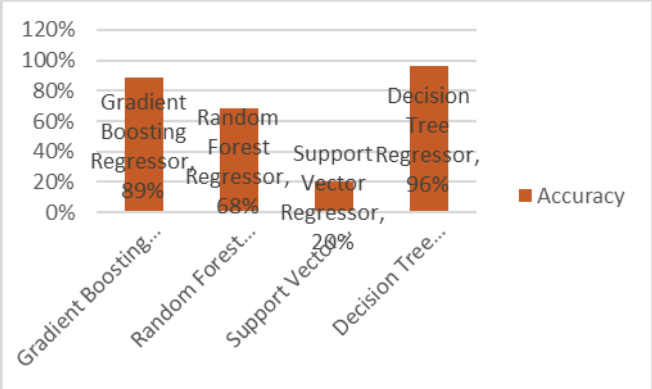


Fig. 2. Performance of Machine learning models

## V. CONCLUSION

In our investigation, we tried out different computer methods to guess how many crops farmers might get. We tested things like Gradient Boosting, Random Forest, Support Vector, and Decision Tree. We found that Decision Tree was the best at making accurate predictions about crops. This could help make forthcoming crop predictions even better. We also realized that it's important to make it easy for farmers to use this prediction method. So, we made a simple phone and computer app that farmers can use. This app will keep helping farmers, even if they aren't computer experts, to know more about their crops. Our goal is to make advanced computer methods easy to use, so farmers can make good decisions about their crops. This will help farming get better, and it will be good for farming communities in the future.

## REFERENCES

[1] F. H. Tseng, H. H. Cho, and H. T. Wu: "Applying big data for intelligent agriculture-based crop selection analysis," IEEE Access, vol. 7, pp. 116965-116974, 2019.

[2] P. M. Gopal, and R. Bhargavi: "A novel approach for efficient crop yield prediction," Computers and Electronics in Agriculture, vol. 165, pp. 104968, 2019.

[3] P. Sivanandhini, and J. Prakash: "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.

[4] Ommolbanin Bazrafshan, Mohammad Ehteram, Sarmad Dashti Latif, Yuk Feng Huang, Fang Yenn Teo, Ali Najah Ahmed, Ahmed El-Shafie, predicting crop yields using a new robust Bayesian averaging model based on multiple hybrid ANFIS and MLP models, Ain Shams Engineering Journal, Volume 13, Issue 5, 2022.

[5] D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371-376, 2019.

[6] Kalichkin, V & Alsova, O & Maksimovich, Kirill. (2021). Application of the decision tree method for predicting the yield of spring wheat. IOP Conference Series: Earth and Environmental Science. 839. 032042. 10.1088/1755-1315/839/3/032042.

[7] Badenko V L, Garmanov V V and Ivanov D A 2015 Prospects for the use of dynamic models of agroecosystems in the tasks of medium and long-term planning of agricultural production and land management Russian Agricultural Sciences 1-2 72-76.

[8] Lutsenko E V, Loiko V I and Velikanova L O 2008 Forecasting and Decision Making in Crop Production Using Artificial Intelligence Technologies.

[9] T. Senthil Kumar, "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm," Journal of Artificial Intelligence, vol. 2, no. 03, pp. 185-193, 2020.

[10] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," Ictact journal on soft computing, vol. 11, no. 01, 2020.

[11] P. Tiwari, and P. Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[12] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Predicition of Crop Yeild Using SVM Approch with the Facility of E-MART System" Easychair 2020.

[13] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.

[14] R. Kumar, M. Singh, P. Kumar, and J. Singh, "Crop selection method to maximize crop yield rate using machine learning technique," in 2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM). IEEE, 2015, pp. 138–145.