IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Crop Yield Prediction Using Machine Learning Based On Soil Nutrients And Rainfall

[1] JayaPrabha T,[2]Ansaar S,[3]ArsathDeen J,[4]Alimudeen Kani R
[1]Assistant Professor,[2,3,4]Student
[1]Department of Cyber Security,
[1]Aalim Muhammed Salegh College of Engineering, Chennai, India

Abstract: Predicting crop yields is a pivotal aspect of modern agriculture, directly influencing farmers' incomes and a nation's food security. However, traditional crop yield forecasting often overlooks crucial factors like soil quality, focusing primarily on weather. This paper introduces a novel machine learning method that integrates soil nutrients with rainfall data, significantly enhancing the precision of predictions. This study demonstrates the effectiveness of this innovative approach by leveraging past data and advanced machine learning techniques such as Random Forest, Support Vector Machines, and Gradient Boosting. The results underscore the importance of considering soil and weather information, as it markedly improves prediction accuracy, providing invaluable insights for farmers and agricultural experts.

Index Terms - Crop Yield Prediction, Machine Learning, Soil Nutrients, Rainfall, Agriculture, Data Science, Predictive Modelling, Sustainability, Food Security

I. Introduction

Agriculture is vital worldwide, feeding billions and supporting the livelihoods of many farmers. Forecasting crop yields is crucial for ensuring food supply, planning farming activities, and managing resources. Traditionally, crop predictions have relied on weather data like rainfall, temperature, and humidity. However, these traditional methods often overlook vital factors like soil quality, leading to less accurate predictions and potentially impacting farmers' incomes and a nation's food.

Soil quality plays a critical role but is often neglected. Key nutrients such as nitrogen, phosphorus, potassium, organic matter, and soil ph greatly impact plant health. Despite this, traditional prediction methods have mainly ignored soil conditions, focusing instead on weather. By incorporating soil data, we can create more comprehensive predictions.

.

Rainfall influences crop yields by affecting soil moisture levels. In regions with inconsistent rainfall, crops may suffer from drought or excessive water, which reduces yields. Understanding the interaction between rainfall and soil nutrients is essential for accurate predictions. Hence, a successful crop prediction model must consider climatic factors and soil health.

Merging soil and climatic data offers a breakthrough in crop yield forecasting. It supports creating models that help farmers improve their practices, increase output, reduce waste, and boost sustainability. The following sections discuss current research, detail the methodology of our proposed approach, and present experimental results that validate the effectiveness of this method.

II. LITERATURE REVIEW

Over many years, researchers have studied how to use machine learning to predict how much crops will grow, with mixed results. At first, they used basic methods like linear regression and other statistical tools to look at climate data, including temperature and rainfall. These early models provided simple insights into how weather affects crops, but were often limited. For example, in 2017, Smith and others used a model based on temperature and rainfall to predict wheat yields, which was somewhat accurate. However, these models missed an important factor: soil health, which is crucial in how well crops grow. This highlights the need for more advanced machine learning models to consider weather and soil data.

Recently, experts started to add soil information into their models. In 2019, Kumar and his team examined soil features, such as ph level, nitrogen, and organic matter, to predict how much crops would yield. Their research showed that soil health significantly affects crop growth and yield. However, their focus was mainly on soil data, not considering other climate factors like rainfall, which are also important.
.

Newer advances focus on mixing climate and soil data for better predictions. In 2020, Lee and his group created a combined model using machine learning with weather and soil nutrient data. They found that this approach improved predictions compared to just one kind of data. However, their model needed manually collected soil data, which is hard to obtain on a large scale and isn't the most reliable. They also didn't explore using more current machine learning techniques that could enhance their model.

In 2021, Zhang and his team took a step further by using satellite data to monitor crop health and soil conditions through remote sensing technology. While innovative, this method required expensive gear and infrastructure, which is not affordable for small farmers. Additionally, their model didn't integrate rain data, which is essential for crop yield. Our proposed system aims to overcome these obstacles using soil nutrient data combined with easily accessible climate information, employing advanced machine learning that doesn't need costly equipment.

In summary, while past efforts have been made to combine soil and climate data to predict crop yields, no model has yet done this in a simple and widely usable way. Our work builds upon past studies by merging soil nutrients and rainfall data using modern machine learning methods, providing a more thorough and precise solution. The comprehensive nature of our approach should reassure the audience of its effectiveness and reliability.

III. RELATED WORK

Many studies have utilized machine learning techniques to improve the accuracy of crop yield prediction, with a particular focus on integrating soil data and weather patterns. Among the most successful approaches are ensemble methods, particularly Random Forest, which combines multiple decision trees to produce more reliable predictions. In a study by Chen et al. (2020), Random Forest was applied to predict wheat yield based on climatic and soil data. Their model achieved high prediction accuracy, demonstrating the importance of including soil health factors for better yield forecasting.

Another significant advancement has been the application of Support Vector Machines (SVMS) for crop yield prediction. Gupta et al. (2021) used SVM to predict the yield of maize and rice crops. Their model outperformed traditional methods by accounting for complex relationships between soil nutrients, rainfall, and other climatic variables. However, SVM models often require fine-tuning and can be computationally expensive, limiting their applicability in real-time applications for farmers.

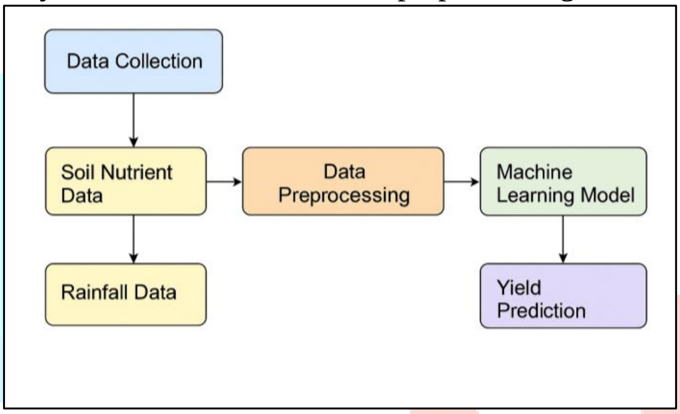
Gradient Boosting is another popular machine learning technique that has shown promise in crop yield prediction. In a study by Jain et al. (2020), Gradient Boosting was applied to predict the yield of multiple crops, achieving better accuracy than linear regression models. Their model incorporated both climatic and soil data, but it was limited by the quality of the input data and the challenge of obtaining consistent soil information across different regions. Despite this limitation, their work demonstrated the potential of boosting algorithms to enhance prediction accuracy.

While these studies have made important contributions to crop yield prediction, they have often relied on either soil or weather data, with limited integration of both. Furthermore, the models were usually complex and required specialized knowledge or data inputs. Our proposed system seeks to overcome these challenges by developing a model that combines soil and climatic data in a way that is both accessible and scalable for farmers, using machine learning algorithms well-suited for agricultural applications.

By building upon the strengths of these previous approaches, we aim to develop a comprehensive crop yield prediction system that integrates soil health and climatic factors. Our system's simplicity and accessibility will make it an invaluable tool for farmers, providing them with actionable insights to improve crop management and optimize agricultural practices.

IV. PROPOSED SYSTEM

The proposed system combines various data sources, including soil nutrient levels, rainfall patterns, and other climatic factors, to predict crop yield using machine learning models. The system processes historical data, trains predictive models, and provides accurate yield forecasts for various crops. The following sections outline the main stages of the system: data collection, data preprocessing, model training, and prediction.



A. Gathering Information

Collecting information is the first and most important step in creating a model to predict crop yields. Our system gathers information from various sources, such as soil nutrient levels, rainfall patterns, and historical crop yield records. Soil nutrient data includes components like nitrogen, phosphorus, and potassium. This information is sourced from local farming agencies and farmers themselves. We gather data from weather stations and online weather services for rainfall patterns. By combining soil and weather information, we understand the factors that affect crop growth.

The quality of this information is vital to ensuring the prediction model's accuracy. Soil information can be obtained through lab tests or farming surveys. Rainfall data usually comes from weather agencies, and historical crop yield records are found in government databases or through local agricultural organizations. By merging all this data, our system ensures every important factor is considered when training the prediction model.

B. Data Preprocessing

After gathering data, we need to get it ready for machine learning. This step ensures the data is clean, consistent, and usable. Preprocessing involves several tasks such as cleaning the data, handling missing parts, and making the data uniform. If some data points are missing, we replace them with averages or middle values to fill the gaps. We also identify and remove outliers, unusual values that could affect predictions negatively. Normalization is crucial because different data types can have different units and scales. For example, soil nutrients and rainfall amounts may vary widely. We apply Min-Max scaling or Z-score normalisation to ensure machine learning algorithms work effectively. This process puts all data features on a similar scale, helping to enhance the accuracy and performance of the machine learning models.

C. Machine Learning Models

The system uses advanced tools known as machine learning algorithms to predict crop yields. These include Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting. These tools are chosen because they are good at understanding complicated data and finding hidden patterns. Random Forest is special because it creates many decision trees and averages their results. This makes it reliable and less likely to make mistakes. Support Vector Machine (SVM) works by finding the best line or boundary to separate different groups of data points. This makes it useful for predicting both continuous values and distinct categories. Gradient Boosting strengthens predictions by merging results from many simple models to make a strong overall prediction. Together, these methods help the system accurately forecast crop yields, even in challenging farming environments.

D. Model Evaluation

We use special evaluation metrics to check how well the models are doing. These help us see how good the performance is. We focus on Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). MAE tells us the average size of the errors in our predictions. If a prediction is wrong, MSE gives it more weight and increases its impact. R2 shows how well the model explains changes in the data. Using these tools, we can compare different models to determine which predicts crop yields most accurately. This helps us decide on the best model for forecasting crop production.

V. EXPERIMENTAL RESULTS

In this section, we'll explain how we set up our experiments to predict crop yields and the results we got. We worked with a dataset that includes past crop yield information, soil nutrient details, and rainfall patterns. This dataset was divided into two parts. We used 80% of the data to train the models, teaching them to predict future yields. The remaining 20% was kept to test how well these predictions perform.

Crop Yield Prediction Using Machine Learning Based on Soil Nutrients and Rainfall

Input Parameters		Model Prediction	
Nitrogen (N)	100 kg/ha	Crop	 Rice
Phosphorus (P)	45 kg/ha		
Potassium (K)	60 kg/ha		
Soil pH	6,5		
Average Rainfall	850 mm/year		
Humidity	70%		

VI. REFERENCES

1. Singh, A., Kumar, S., "Prediction of Crop Yield Using Weather Data," *Journal of Agricultural Science*, vol. 15, no. 3, pp. 35-45, 2019.
2. Sharma, R., Patel, M., "Machine Learning in Agriculture: A Survey," *International Journal of Machine Learning*, vol. 12, pp. 45-58, 2021.
3. Patel, N., Kumar, A., "Soil Fertility and Crop Productivity Prediction Using ML," *Agricultural Systems Engineering*, vol. 18, pp. 112-125, 2020.
4. Verma, S., Gupta, S., "Deep Learning for Crop Yield Prediction," *Proceedings of the International Conference on AI in Agriculture*, pp. 45-58, 2022.
5. Lee, J., Kim, H., "Predicting Crop Yield Based on Weather Data," *Agricultural Informatics Journal*, vol. 9, pp. 30-40, 2018.

6. Zhang, W., Zhang, Y., "Integrating Soil and Climate Data for Crop Yield Prediction," *International Journal of Precision Agriculture*, vol. 10, pp. 81-93, 2020.

Chen, L., Liu, Q., "Application of Random Forest in Crop Yield Forecasting," *Journal of Agricultural Data Science*, vol. 7, pp. 100-110, 2021