

# Crop Yield Prediction using Machine Learning

Akash Tony<sup>1</sup>, Dona Lalkumar<sup>2</sup>, Luke Saint<sup>3</sup> and Mohammed Hazim A S<sup>4</sup>

<sup>1-4</sup>Department of Computer Science and Engineering, Sahridaya College of Engineering and Technology Thrissur, India  
Email: akashtony171@gmail.com, donalal1710@gmail.com, lukesjc70@gmail.com, mohammedhazim092@gmail.com

**Abstract**—As the world faces growing challenges in ensuring food security, predicting crop yields accurately becomes increasingly crucial. Traditionally, this prediction relied on methods limited by their inability to handle complex interactions between various factors like climate, soil, and farming practices. This is where Machine Learning (ML) emerges as a game-changer. By harnessing the power of data-driven insights, ML offers highly accurate predictions surpassing traditional methods. This paper specifically dives into the potential and challenges of utilizing ML for crop yield prediction, highlighting its strengths and areas for further development. While traditional approaches often fall short in capturing the intricate relationships impacting yield, the proposed method leverages satellite imagery to extract valuable features and deliver precise predictions on granular levels. Ultimately, this paves the way for farmers to make informed decisions, optimize resource allocation, maximize production, and contribute significantly to global food security.

**Index Terms**— Crop yield prediction, Satellite imagery, Machine Learning, Regression, Deep learning and Convolutional Neural Network.

## I. INTRODUCTION

Traditionally, farmers relied on methods prone to inaccuracy, often failing to capture the complex interplay between various factors like climate, soil, and their own crop management practices. Imagine a world where farmers can confidently anticipate their harvest, allowing them to optimize resource allocation, make informed decisions about planting and irrigation, and ultimately maximize their production. Informed decisions about planting and irrigation, and ultimately maximize their production. This is the power that Machine Learning promises to bring to agriculture. Machine learning algorithms, with their ability to learn from vast and complex datasets like satellite imagery, can identify patterns and relationships that escape the human eye. This knowledge translated into highly accurate crop yield predictions that traditional methods simply cannot match. The need for such precise predictions is more profound than ever. With a rapidly evolving agricultural landscape, ensuring food security hinges on understanding and optimizing crop yields. This holistic approach allows for the extraction of meaningful features from diverse data sources, ultimately leading to highly accurate predictions. This study not only empowers farmers but also informs policy makers in identifying food insecure regions and planning crucial food storage and transport logistics.

## II. LITERATURE REVIEW

### A. Yield Prediction with Deep Learning and Remote Sensing

Deep learning's potential for predicting crop yield, leveraging remote sensing data, sparks hope for optimized

agricultural practices. This review highlights a [1] crucial research gap in specific deep learning methodologies and the impact of vegetation indices and environmental factors on yield. Analyzing studies from 2012 to 2022, it identifies Long Short-Term Memory (LSTM) [2] and Convolutional Neural Networks (CNN) as prominent deep learning approaches, with satellite remote sensing, particularly MODIS data, reigning supreme. While vegetation indices dominate feature selection, their effectiveness varies across methods. Despite promising [3] results, challenges persist: improving model accuracy, ensuring practical application for agriculturalists, and tackling the "black box" nature of deep learning models. This review underscores the potential of this technology while urging further research to navigate these key hurdles and revolutionize crop yield prediction for a more food-secure future. [4]

#### *B. Research on Rice Yield Prediction Model Using Deep Learning*

Deep learning, particularly its prowess in image recognition and processing, offers promising solutions. This paper explores combining remote sensing images [5] with deep learning to emulate the key factors influencing rice yield. Rice yield depends on four key factors like number of ears per area, number of grains per ear, seed set rate, and grain weight. In this article, [6] they select characteristic variables that best represent rice yield based on simulations of these four factors. They then construct a regression model using these selected features to estimate the actual yield of a rice plot. [7] Imagine you have numerical data and need to predict missing values based on existing patterns. Regression steps in to tackle this problem by creating a mathematical equation, called a "régression équation," that connects inputs to expected outputs. To assess how well this equation fits the data, we use a measure called "goodness-of-fit" (R<sup>2</sup>). [8] The higher the R<sup>2</sup> (closer to 1), the better the equation fits the data and the more accurate its predictions. [9] We compare various linear and non-linear regression models, analyzing their prediction performance and significance. The ultimate goal is to develop a method for accurate rice yield estimation, contributing to the broader aim of food security. [10]

#### *C. Soybean Yield Prediction Using Deep CNN model*

Climate changes have a significant impact on agricultural crops in India. This has created a need for tools that empower farmers to make informed decisions. [11] This project addresses this need by developing a deep learning model for preharvest soybean yield prediction. By analyzing data containing soil conditions, weather patterns, and historical yield performance, the model aims to assist farmers in optimizing their agricultural practices. [12] The research proposes a convolutional neural network (CNN) as the deep learning framework for this task. They compared the model's performance against other popular methods like artificial neural networks (ANNs) and long short-term memory (LSTM) networks. [13] The CNN achieved superior results, the mean squared error, the root mean squared error, the mean absolute error, and the R-squared value (R<sup>2</sup>), demonstrating its effectiveness in predicting soybean yield. [14] These metrics highlight the outperformance of the CNN model, suggesting its potential as a valuable tool for Indian farmers. [15]

#### *D. Crop Yield Prediction Using Machine Learning*

Currently, the agricultural sector grapples with a complex challenge of ensuring food security for a growing population while facing constraints on land, water, and resources. [16] This research dives into the promising world of ML, exploring how its ability to analyze vast datasets including real-time data from sensors that can empower farmers with knowledge. [17] By understanding factors influencing crop growth, farmers can make informed decisions regarding planting, watering, and harvesting, leading to optimized yields and minimized waste across the entire production cycle. [18] This research meticulously examines the current state of ML in agriculture, comparing and analyzing fifteen different algorithms to identify the most effective for various agricultural applications. Furthermore, a novel feature combination scheme enhanced algorithm is introduced, boasting superior performance in data analysis. [19] The results are nothing short of impressive. The proposed model achieves a remarkable classification accuracy of 99.59 percent using the Bayes Net algorithm, with similarly high accuracy achieved using Naive Bayes Classifier and Hoeffding Tree algorithms. These figures translate into tangible benefits for farmers and consumers. [20].

### III. PROPOSED SYSTEM ARCHITECTURE

The proposed system aims to revolutionize crop yield prediction in agriculture, especially corn and soybean by leveraging Machine learning techniques and satellite imagery data. The system architecture is designed to effectively integrate heterogeneous data sources, including satellite imagery, crop yield data, and production information, to capture the complex interactions.

References	Year	Methodology	Properties	Description
[1] Crop Prediction Model Using Machine Learning Algorithms	2023	Using satellite imagery bands proposes a histogram dimensionality reduction technique to combat over fitting.	Compares multiple machine learning algorithms to identify the best performer for a specific scenario.	In this article the Satellite imagery data, historical yield data and climate data are collected and trained. Support Vector Regression (SVR), Random Forest (RF), and Multilayer Perceptron (MLP) Algorithms are used. Evaluation is done by using R-squared, mean squared error (MSE), and root mean squared error (RMSE) method.
[2] A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing	2022	Explores the application of deep learning and remote sensing for crop yield prediction.	Provides a comprehensive overview of current research trends in the field. Identifies knowledge gaps and suggests future research directions.	This paper includes Searching various datasets and focuses on deep learning methods, and uses remote sensing data. Extract information on methodology, accuracy, crops, features, and other relevant details.
[3] A Deep Learning Model For Soybean Yield Prediction	2023	Explores different modelling approaches to relate crop yields to environmental factors like topography and soil characteristics.	Utilizes multiple feature groups. High accuracy. Can optimize crop management through accurate yield prediction.	This paper investigates the feasibility of predicting soybean yields using a deep learning approach. The dataset consisted of 25,345 records encompassing various features. A Convolutional neural Network (CNN) model was selected and trained. Data preprocessing, model selection, and loss function definition were performed. Ultimately, the trained model achieved high accuracy, demonstrating its strong potential for accurate soybean yield prediction.
[4] Research on 2022 Rice Yield Prediction Model Based on Deep Learning	2022.	Investigates the use of deep learning and remote sensing images to estimate rice yield. Various regression models are compared and evaluated for their prediction accuracy and significance.	Utilizes remote sensing images and other numerical data for yield estimation. Aims to achieve high accuracy. Non-linear relationships	The study involves remote sensing images and potentially other numerical data, focusing on variables extracted from simulations of four key yield-influencing factors. It employs feature selection techniques to identify rice yield and develops both linear and non-linear regression models to estimate yield. R-squared is used as a goodness-of-fit measure.

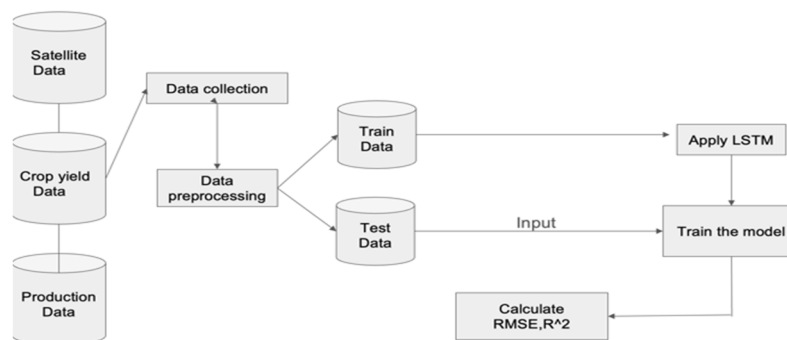


Fig.1. Architecture Diagram of the Proposed System

Influencing crop growth. By leveraging the capabilities of Machine learning models, the system aims to extract meaningful features from raw satellite imagery production information, to capture the complex interactions influencing crop growth. By leveraging the capabilities of Machine learning models, the system aims to extract meaningful features from raw satellite imagery and other environmental data, enabling more accurate and robust crop yield predictions. At the core of the proposed system is a Machine learning architecture optimized for learning from large- scale, high-dimensional data sets. The architecture consists of LSTM model for spatial feature extraction from satellite imagery and recurrent layers for capturing temporal dynamics in crop

growth patterns. This architecture allows the system to effectively capture the intricate relationships between environmental variables and crop yields, leading to improved prediction accuracy and reliability. One of the key innovations of the proposed system is its ability to handle the inherent challenges associated with working with satellite imagery data, such as variability in cloud cover, atmospheric conditions, and sensor noise. To address these challenges, the system incorporates advanced pre-processing techniques, including image normalization, data augmentation, and feature extraction, to enhance the quality and consistency of the input data. Additionally, the system employs transfer learning and domain adaptation strategies to adapt pre-trained machine learning models to the specific characteristics of agricultural regions, thereby improving generalization and robustness. To evaluate the performance of the proposed system, comprehensive experimentation and validation are conducted using real-world data sets collected from agricultural regions across the US. The system is trained and tested on historical crop yield data paired with corresponding satellite imagery and environmental variables, allowing for rigorous evaluation of its predictive capabilities. Performance metrics such as accuracy, R-squared method, root mean square error (RMSE) are used to assess the system's predictive accuracy and reliability across different regions and crop types.

#### IV. RESULT

The experimental results demonstrate the effectiveness of the proposed system in accurately predicting crop yields in US agriculture. The system achieves significant improvements in prediction accuracy compared to traditional methods, highlighting the potential of deep learning techniques to revolutionize crop yield prediction in the US. Moreover, the system's ability to provide timely and reliable predictions enables farmers, policymakers, and other stakeholders to make informed decisions and take proactive measures to mitigate risks and maximize agricultural productivity. The system leverages a comprehensive approach for robust performance evaluation. It undergoes training and testing with historical data sets featuring yield data, satellite imagery, and environmental variables. This multifaceted approach empowers a thorough assessment of its predictive capabilities across various regions and crop types. A suite of metrics including accuracy, R-squared, and RMSE is employed to provide a rigorous evaluation of the system's generalizability and reliability.

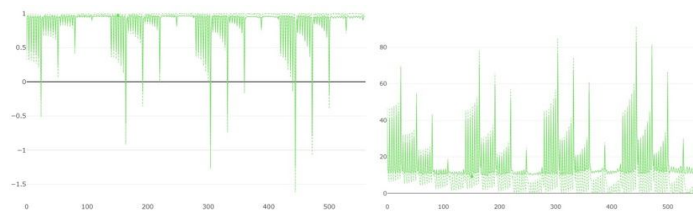


Fig. 2: R-Squared Method

Fig. 3: RMSE Training 1

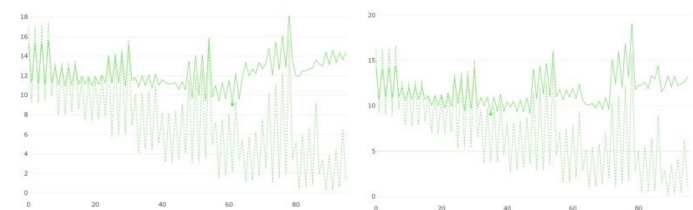


Fig. 4: RMSE Training 2

Fig. 5: RMSE Training 3

Overall, the proposed system represents a significant advancement in the field of crop yield prediction in agriculture. By harnessing the power of machine learning and satellite imagery data, the system offers a scalable and cost-effective solution for addressing the complex challenges faced by farmers and policymakers. With further research and development, the proposed system has the potential to drive transformative change in agricultural practices, enhance food security, and promote sustainable development across the world.

#### V. CONCLUSION

In conclusion, this conference paper explores the ability to accurately predict crop yields for ensuring food security, optimizing resource allocation, and mitigating risks for farmers. Machine learning, coupled with the vast information gleaned from satellite imagery, is revolutionizing this field, offering unprecedented insights and

tools for yield prediction. Machine learning and satellite imagery have ushered a transformative era for crop yield prediction. This transformative approach offers unprecedented accuracy, extracting valuable insights from diverse factors like weather, soil quality, and historical yields. Farmers and stakeholders empowered by these insights can make informed decisions. By addressing the remaining challenges and harnessing the vast potential of these technologies, we can build a more food-secure and sustainable future for agriculture.

#### ACKNOWLEDGMENT

This conference paper is a result of collaborative efforts of several individuals. Firstly, we deeply appreciate the organizers of this conference for offering a platform to showcase our research on “Crop Yield Prediction Using Machine Learning.” Special thanks to colleagues, especially Ms. Anusree k, Assistant Professor, Computer Science and Engineering Department, Sahrdaya College of Engineering and Technology. Her invaluable support and guidance were instrumental in shaping this paper. Furthermore, we extend our sincere thanks to the broader research community and conference participants whose diverse perspectives and constructive criticism greatly refined our ideas. Their contribution to this work is invaluable. Lastly, we cannot forget the unwavering support of our families and friends who have been with us throughout this journey. Their encouragement and belief in our work have sustained us throughout the process. This work embodies the collaborative essence of research, driven by a collective dedication to expand our knowledge within this field.

#### REFERENCES

- [1] P. Muruganantham and N. Islam, “A systematic literature review on crop yield prediction with deep learning and remote sensing,” *Remote Sensing*, vol. 14, no. 9, p. 1990, 2022.
- [2] L. Zheng, Z. Qiao, X. He, X. Song, Y. Wang, and Y. Xiang, “A deep learning framework for crop yield prediction from sentinel-2 satellite imagery,” *Remote Sensing*, vol. 14, no. 1, p. 226, 2022.
- [3] L. Ghazaryan, B. Blasch, V. Nenov, and P. Maragos, “Yield prediction using time-series satellite imagery and recurrent neural networks,” *Remote Sensing*, vol. 13, no. 7, p. 1398, 2021.
- [4] M. Croci, G. Impollonia, M. Meroni, and S. Amaducci, “Dynamic maize yield predictions using machine learning on multi- source data,” *Remote Sensing*, vol. 15, 12 2022.
- [5] X. Han, F. Liu, X. He, and F. Ling, “Research on rice yield prediction model based on deep learning,” *Computer Systems Science and Engineering*, vol. 43, no. 4, pp. 1267–1276, 2022.
- [6] T. H. Santos, C. H. do Amaral, J. P. Molin, and Z. Yun, “Predicting crop yield in the Brazilian cerrado using convolutional neural networks,” *Remote Sensing*, vol. 13, no. 13, p. 2649, 2021.
- [7] Z. Chu and J. Yu, “An end-to-end model for rice yield prediction using deep learning fusion,” *Computers and Electronics in Agriculture*, vol. 174, p. 105471, 2020
- [8] van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* 2020, 177, 105709.
- [9] Crop Prediction Model Using Machine Learning Algorithms by E. Elbasi, C. Zaki, A. E. Topcu, W. Abdelbaki, A. I. Zreikat, E. Cina, A. Shdefat, and L. Saker. (*Applied Sciences*, 2023, Volume 13, Pages 9288)
- [10] Crop yield prediction using machine learning: A systematic literature review by T. van Klompenburg, A. Kassahun, and C. Catal. (*Computers and Electronics in Agriculture*, 2020, Volume 177, Pages 105709)
- [11] Crop yield prediction using deep learning: A comprehensive review by A. M. Raza et al. (*Computers and Electronics in Agriculture*, 2021, Volume 190, Pages 106401)
- [12] Deep learning-based crop yield prediction: A survey by M. Usman et al. (*Remote Sensing*, 2021, Volume 13, Issue 17, Pages 3460)
- [13] Crop yield prediction using deep learning: A review by H. Liu et al. (*Field Crops Research*, 2020, Volume 255, Pages 107860)
- [14] Crop yield prediction with deep convolutional neural networks by D. Khan et al. (arXiv preprint arXiv:1702.05742, 2017)
- [15] Convolutional neural networks for corn yield prediction using UAV-based remote sensing imagery by T. Wu et al. (*Remote Sensing*, 2020, Volume 12, Issue 7, Pages 1098)
- [16] A comprehensive review of machine learning for agricultural yield prediction by A. Fountas et al. (*Computers and Electronics in Agriculture*, 2021, Volume 180, Pages 105888)
- [17] Crop yield prediction using machine learning techniques: A review” by M. S. Islam et al. (*Computers and Electronics in Agriculture*, 2021, Volume 189, Pages 106389)
- [18] A review of machine learning applications in agriculture by A. K. Kamilaris et al. (*Computers and Electronics in Agriculture*, 2017, Volume 142, Pages 283-297)
- [19] Machine learning models for yield prediction: A review by S. Li et al. (*Computers and Electronics in Agriculture*, 2021, Volume 189, Pages 106347)

[20] A comprehensive review of machine learning techniques for crop yield prediction by M. R. I. Shaon et al. (Computers and Electronics in Agriculture, 2021, Volume 188, Pages 106361).