International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/07-19
ISSN: 2319-4847

Survey on Crop Prediction using Machine Learning

Ms. P. Indumathy, Hajira Moularan Syed Masood, Hitha Poddar* and Indirani J

Department of Computer Science and Engineering, Easwari Engineering College, Chennai, TN, India

* Corresponding Author

Email: hithapoddar@gmail.com

Article Info

Received: 06-01-2025 ***Revised: 10 -02-2025*** ***Accepted: 20-02-2025*** ***Published:07/03/2025***

Abstract

Agriculture, a vital sector for sustaining global food security, faces significant challenges due to the growing population and environmental changes. Traditional practices often struggle with inefficiencies and adaptability issues. This survey examines recent advancements in crop prediction through the integration of machine learning (ML) and other technologies. It highlights how ML algorithms analyze diverse datasets, including satellite imagery and soil data, to enhance prediction accuracy and optimize resource use. The review underscores the impact of these technologies on improving decision-making for farmers, stabilizing food prices, and promoting agricultural sustainability.

KEYWORDS: Crop yield prediction, Machine learning algorithms, Process based model, Statistical models, Remote sensing, Neural networks.

I. INTRODUCTION

Agriculture is the backbone of human civilization. It is a source of food, fiber and resources necessary for life. It plays an important role in the world economy. Support livelihoods and drive rural development. Agriculture covers the cultivation of crops, livestock, and natural resource management. All of these contribute to food security, nutrition, and overall economic stability. In India, Agriculture is the main source of income and employment.

Therefore, it is an important driving force for poverty alleviation and rural development. In addition to its economic importance, Agriculture also has cultural importance in India. As agricultural practices evolve, the need for innovation in crop management, resource optimization, and yield prediction becomes paramount. Agriculture has traditionally relied on manual labor, historical data, and experience-based decision-making. However, with the world's population projected to reach nearly 10 billion by 2050, global food production must increase by up to 70% to meet demand. Traditional agricultural methods are often inefficient, wasteful, and vulnerable to external shocks, such as unpredictable weather patterns, market volatility, and resource scarcity. The impact of climate change, fluctuating market conditions, and unpredictable weather patterns further intensify the challenges faced by farmers and policymakers alike. Agriculture as a topic of research is not only timely but essential. It offers opportunities to address these challenges through advanced technological interventions, which can help optimize resource use, reduce risks, and enhance productivity.

One of the most critical areas of focus within agriculture today is crop prediction, which has emerged as a powerful tool for improving decision-making processes in farming and agricultural policy. Machine learning models can process and analyze diverse datasets, such as satellite imagery, soil data, and market trends, providing farmers with real-time insights about crop health, growth patterns, and potential yield outcomes. By leveraging these predictive tools, farmers can make more informed decisions about planting, irrigation, fertilization, and harvesting, which in turn maximizes efficiency and reduces waste. Crop yield prediction models not only aid in resource optimization but also mitigate the risks posed by environmental factors like droughts, floods, and pests, which can have devastating effects on agricultural productivity. Predictive analytics enable proactive interventions, allowing farmers to adapt their strategies to minimize losses and optimize crop performance.

Machine learning (ML) revolutionizes agriculture by integrating with traditional models and remote sensing technologies to enhance decision-making and efficiency. Process-based crop models, which simulate physiological processes, are improved with ML to adapt to real-time data.Crop simulation models are key components to test the advances in agricultural technology and to predict crop responses to present and future climate forcing. These models are being used widely to estimate the crop production potential, transfer Agro-technologies, assist strategic decisions, and forecast real-time yield[1]. Statistical models benefit from ML's ability to handle complex, non-linear relationships for more accurate predictions.

Remote sensing technologies, like satellite and drone imagery, provide real-time data on crop conditions, which ML algorithms analyze to detect issues and optimize resources. Remote sensing provides valuable, non-intrusive data on crop conditions through satellite imagery, capturing information on plant health, moisture levels, and overall growth. This data can be used to generate time-series predictions of vegetative indices, which serve as critical indicators of crop development. Hybrid models, which integrate ML with traditional approaches, enhance prediction accuracy. Hybrid model that combines SVM, LSTM, and RNN techniques has demonstrated significant effectiveness in crop yield prediction, achieving high precision and recall rates[2].While neural networks excel in processing and learning from large datasets and complex patterns, such as detecting crop diseases from imagery. Overall, ML enhances agricultural practices by offering precise, actionable insights and improving management and resource optimization. The integration of machine learning, remote sensing, and predictive modeling in crop prediction has the potential to revolutionize agriculture by providing valuable insights to key stakeholders such as farmers, traders, policymakers, and agricultural planners. Accurate crop prediction enables farmers to make informed decisions about planting, irrigation, fertilization, and harvesting, leading to optimized resource use and reduced waste. By providing farmers with tools to better manage crops, anticipate pest infestations, and time market sales, they can make more informed decisions that directly impact their livelihoods. For traders and policymakers, crop yield predictions offer a clearer picture of market conditions, allowing for better planning in terms of supply chain management, storage, and distribution. Traders benefit from accurate forecasts on crop production and prices, aiding in more strategic trading decisions. Policymakers gain data-driven insights to develop policies that bolster food security and income stability. By forecasting both crop yields and market prices, this approach helps stabilize farmer incomes and mitigate the risks associated with fluctuating market conditions and environmental challenges. Additionally, Consumers enjoy increased availability of produce at more stable prices.

II. LITERATURE SURVEY

When reviewing the literature , the articles involving crop yield prediction were studied. They are described in detail below.

A. *Process Based Crop Model*

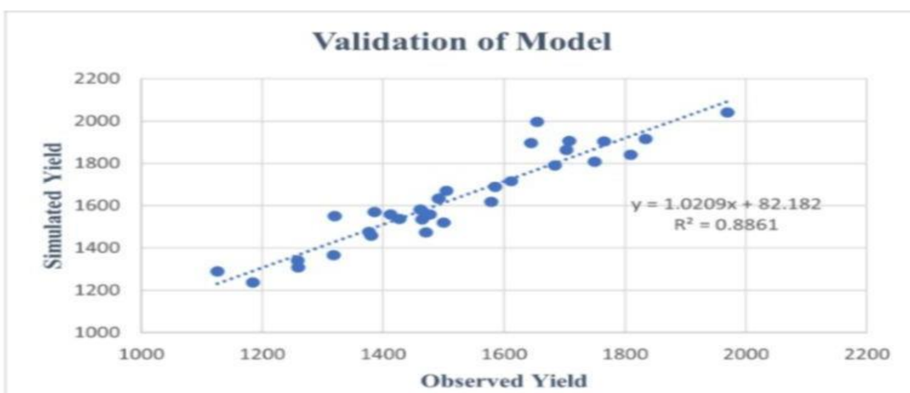
In order to understand the interactions between soil, water, and atmospheric conditions to predict crop growth and yield, process based crop models are used. They integrate various environmental factors to ensure the crop responses under diverse scenarios. PBCMs account for critical soil properties such as pH, moisture, and nutrient levels (NPK), which significantly influence plant performance[3]. Models like BODIUM capture soil function dynamics and complex correlation between soil properties, thereby responding to soil management practices and climate variations, thus enhancing predictive accuracy[4]. The models also serve useful in distinguishing between different types of water flow, thereby incorporating hydrological processes to simulate water availability and its impact on crops[5]. Models such as DSSAT aids in agricultural research by accurately simulating crop growth and yield. DSSAT integrates machine learning and empirical evapotranspiration models for accurate water management, enhancing agricultural decision-making by optimizing irrigation practices based on crop-specific water requirements. Models such as APSIM, the Agricultural Production Systems Simulator, has evolved over 30 years as a vital tool in farming systems research, aiding in interpreting results amidst climate variability and change. In Mali, APSIM has been utilized to analyze sorghum production under varying drought scenarios, revealing critical insights into yield stability across different rainfall zones[7]. Some PBCMs are being used to address water logging and aeration stresses. However, the challenges remain in accurately modeling these conditions. PBCMs play a crucial role in assessing the impacts of extreme weather events. Thus, they are useful to identify adaptation strategies for agricultural systems under climate change. The integration of modular frameworks in these models allows for flexibility in model complexity by tailoring predictions to specific research needs while managing uncertainty[6]. A range of investigations point out the strong performance of both direct and curved strategies, where curved approaches like K-Nearest

International journal of application or innovation in engineering and management
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/07-19
ISSN: 2319-4847

Neighbors (KNN) and Although PBCMs have advanced significantly, challenges still persist in capturing all environmental interactions comprehensively, especially under extreme conditions. Further research and data collection is required in order to refine these models and enhance their predictive capabilities. The DSSAT (Decision Support System for Agrotechnology Transfer) is used for spatial yield estimation of chickpea in Vidisha and Nagaur, India, showing high accuracy in simulating crop yields. [11]

Table 1Observed and simulated yield for Vidisha monitoring sites

Vidisha				
Latitude	Longitude	Observed Yield (kg/ha)	Simulated Yield (kg/ha)	Deviation (%)
Variety I - JG 16				
23.51063	78.06736	1464	1536	4.9
23.71673	77.86417	1655	1996	19.6
23.72553	77.83993	1995	2011	0.8
23.68868	78.00718	1765	1904	7.9
23.70251	77.99177	1318	1366	3.6
24.06780	77.91728	1427	1537	7.7
23.63285	77.84866	1185	1238	4.5
Variety II - JAKI 9218				
23.71113	77.90449	1750	1809	3.4
23.48960	78.05553	1834	1916	4.5
23.47606	78.03633	1812	1922	6.1
23.70830	77.93528	1969	2041	3.7
24.02634	77.86156	1703	1863	9.4
23.39478	78.01640	1684	1790	6.3
Variety III - RVG 202				
23.67509	77.97400	1492	1633	9.5
23.73018	77.86488	1585	1689	6.6
23.58353	77.98783	1612	1715	6.4
23.53913	78.02460	1505	1670	11.0
24.00687	78.14264	1320	1550	17.4
24.06878	77.74187	1386	1570	13.3
23.54101	77.98305	1198	1260	5.2

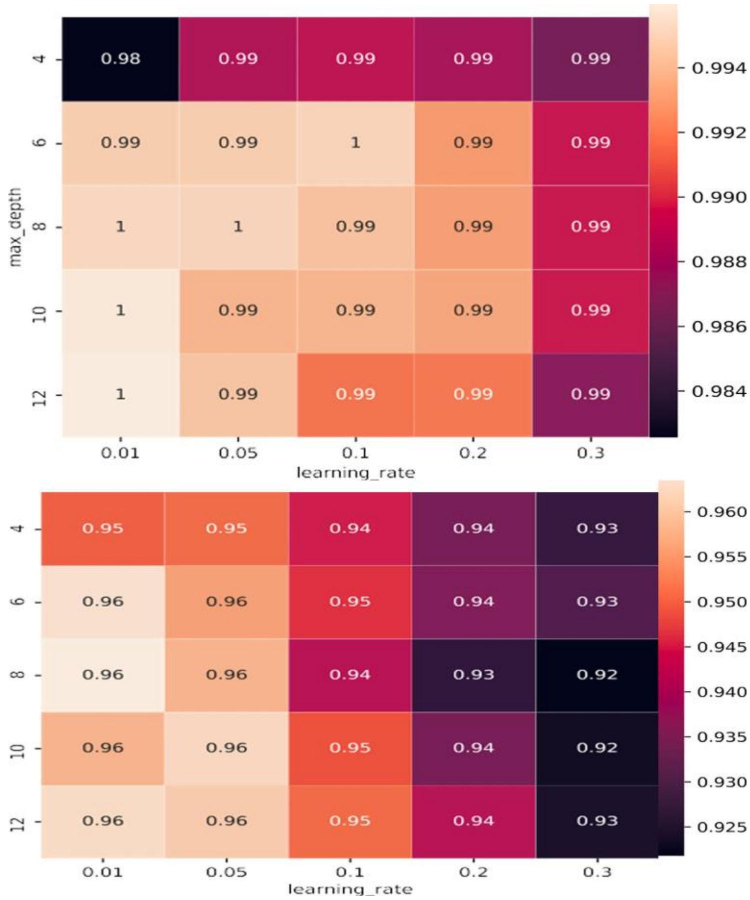


B. Statistical Model

Using Statistical Models to enhance agricultural productivity and sustainability is a critical area of research that leverages machine learning techniques. The effectiveness of different regression models relies on diverse input variables, including weather data and soil conditions. Regression models such as linear regression, decision trees, and ensemble methods have proved to be of significance. Random Forest Regression model achieved a determination coefficient ($r^2 = 0.94$) in predicting crop yields across 196 countries, utilizing parameters like pesticides, rainfall, and temperature[8]. Stacking of ensemble methods have been found to outperform individual models. R-squared value of 98.92% has been achieved by using a stacking approach with Linear Regressor, Decision Tree, and AdaBoost Regressor[9]. Multivariate regression proves to be a powerful tool for predicting crop yields, integrating various environmental and agricultural factors. Multiple Linear Regression is employed in estimating yields for crops like rice and wheat. It demonstrated strong correlations ($R = 0.96$) with historical data, while also addressing climate variables and pollutants. Effective yield prediction relies on comprehensive data preprocessing, including feature selection and engineering, as well as the integration of historical weather data and soil conditions. The integration of AI-based monitoring systems and machine learning techniques has shown promise in enhancing yield predictions and managing agricultural risks, particularly in disease detection[10]. However, challenges such as variable interdependence, data availability and data quality remain critical in agricultural forecasting.

C. Machine Learning Methods

Machine learning methodologies have surfaced as critical assets for predicting agricultural productivity, thus rendering considering help to farmers and the agricultural economy. processes of predictive models are crucial for cultivating trust among diverse stakeholders. Scholarly investigations Support Vector Regression (SVR) achieve better precision than their direct counterparts [13]. Sophisticated models, including Random Forests and Long Short-Term Memory (LSTM) networks, have also demonstrated considerable potential, especially in the analysis of intricate datasets that encompass historical yield data, meteorological conditions, and soil classifications [14]. Furthermore, ensemble techniques, particularly Categorical Boosting (CatBoost), Light Gradient-Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost), have been recognized for their high accuracy in yield forecasting, with CatBoost achieving a remarkable accuracy rate of 99.123% [12]. These advancements in ML not only improve yield forecasting but also facilitate strategic agricultural planning and risk management [15].



Hyperparameters for XG Boost and CatBoost algorithms

D. *Remote Sensing*

Remote sensing has emerged as an indispensable instrument for the forecasting of agricultural yield, harnessing sophisticated technologies such as machine learning and deep learning methodologies. This paradigm significantly bolsters agricultural decision-making processes and food security initiatives by furnishing timely and precise yield forecasts. The integration of Remote Sensing and Machine Learning is vital for the examination of crop vitality and the forecasting of yields. Empirical research indicates that deep learning frameworks, notably Convolutional Neural Networks (CNNs), can attain an accuracy rate of approximately 90% in yield predictions [16]. The deployment of various vegetation indices (VIs) derived from Unmanned Aerial Vehicle (UAV) imagery, such as the Normalized Difference Vegetation Index (NDVI) and the Green NDVI (GNDVI), has shown significant promise in yield prediction for

International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/07-19
ISSN: 2319-4847

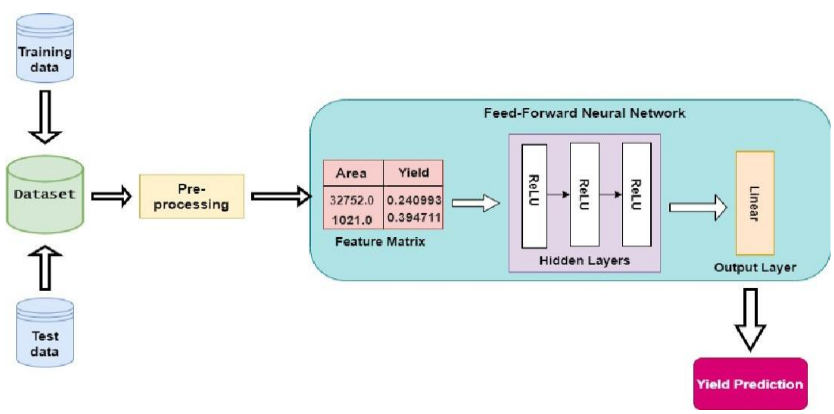
soybeans, with the accuracy of predictive models varying according to growth phases [17].Mechanistic Models and Ground Truth Data An innovative framework that employs solar-induced chlorophyll fluorescence (SIF) has been introduced, showcasing enhanced performance in geographical regions characterized by limited ground truth data, such as India [18]. This underscores the versatility of remote sensing methodologies across diverse agricultural environments. Understanding the decision-making underscore the necessity for explainable artificial intelligence in the context of crop yield forecasting to elucidate critical growth phases and enhance the reliability of predictive models [19]. Although remote sensing considerably augments the accuracy of crop yield forecasts, challenges persist, particularly in locales where ground data is scant. Subsequent research endeavors should prioritize the integration of heterogeneous data sources and the enhancement of model transparency to propel advancements in this domain.

E. Hybrid Models

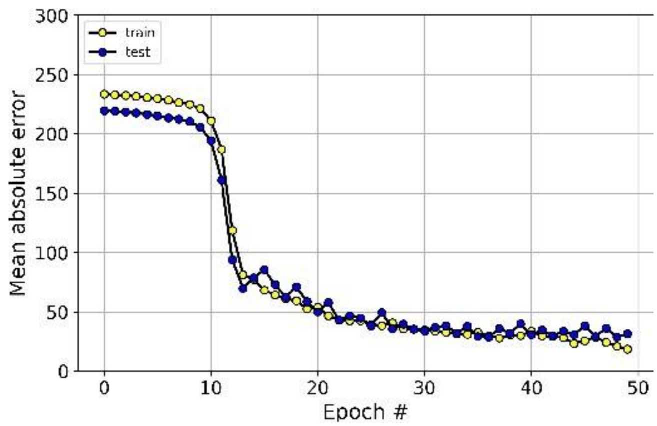
Hybrid machine learning models are an advanced approach that combines the strengths of multiple algorithms to tackle the complexity of tasks like crop yield prediction. These models typically integrate different techniques, such as decision trees, neural networks, or optimization algorithms, to improve accuracy and robustness. The hybrid model [20] combines Decision Tree (DT), XGBoost, and Random Forest (RF) and successfully predicts crop yields with an accuracy of 98.6%. The fusion of these methods allows the model to handle the variability in large agricultural datasets, making it highly adaptable.[21] employ a combination of artificial neural networks (ANN) with optimization algorithms like Gray Wolf Optimizer (GWO), revealing that such hybrid models can outperform standalone algorithms by enhancing learning and reducing prediction error. Hybrid systems often combine elements from both data-driven models and time-series models, as demonstrated in [22] incorporated exogenous variables like irrigation in their ARIMAX-LSTM model. This blend allows the model to account for external factors, further refining predictions. Hybrid models hold the potential to offer more accurate and scalable predictions by leveraging the unique advantages of multiple algorithms. They also come with challenges such as increased computational complexity and a need for careful feature selection to avoid overfitting. The strength of hybrid models lies in their ability to combine various techniques.

F. Neural Networks

Neural networks, particularly deep learning models, have significantly advanced crop yield prediction by leveraging their ability to model complex, non-linear relationships within extensive datasets. A notable approach involves Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units, which address the vanishing gradient problem and enhance time-series forecasting. [23] utilized an RNN-LSTM model to predict wheat yields in northern India, demonstrating superior accuracy compared to traditional methods like Random Forest and Multivariate Linear Regression. Convolutional Neural Networks (CNNs) have been applied to analyze spatial and temporal features, providing robust predictions based on historical data. [24] emphasizes how CNNs, by classifying data and performing regression analysis, can effectively capture relationships between predictor and target variables, thereby improving yield forecasts. Neural network methodologies enhance predictive accuracy and support better decision-making in agriculture. Feed Forward Neural Networks (FFNNs), as explored in [25] further contribute to crop yield forecasting by modeling various influencing factors in a more straightforward architecture compared to RNNs and CNNs. Feed Forward Neural Networks are effective for predicting crop yields by processing input features related to soil type, weather conditions, and crop attributes, thus offering another layer of forecasting capability. These models excel at processing complex and large datasets.



Sequential Model Architecture for predicting crop yield



Mean absolute error of the state Andhra Pradesh

CONCLUSION

Crop yield prediction is essential for advancing agricultural productivity and sustainability, addressing challenges posed by climate change, resource variability, and market fluctuations. This survey highlights the effectiveness of various approaches, including process-based crop models, statistical and machine learning methods, remote sensing, and hybrid models. Each methodology offers unique benefits, from detailed environmental simulations to accurate predictions using advanced algorithms. However, challenges remain, such as integrating diverse data sources and improving model accuracy in data-sparse regions and making this system more real time by identifying the underlying trend in crop prediction. Our review underscores the importance of these technologies and suggests that integrating and enhancing these approaches can further refine predictions and support better decision-making in agriculture.

REFERENCES

[1]. Bannayan M, Crout NM. A stochastic modelling approach for real-time forecasting of winter wheat yield. Field Crops Research. 1999 Jun 1;62(1):85-95.

[2]. E., Revathi. (2023). An Innovative Hybrid Approach for Predicting Crop Yields across Diverse States and Crops. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11):138-148. doi: 10.17762/ijritcc.v11i11.9117

[3]. APSIM-based modeling approach to understand sorghum production environments in Mali- Madina Diancoumba,Jan Kholova,Myriam Adam,Mahamoudou Famanta,Benoît Clerget,Pierre C. S. Traore,Eva Weltzien,Michel Vacksmann,Greg McLean,Graeme L

Hammer,Erik van Oosterom,Vincent Vadez

[4]. Bhagyalakshmi, S. K. Suman and K. Murugan, "Corona based clustering with mixed routing and data aggregation to avoid energy hole problem in wireless sensor network," *2012 Fourth International Conference on Advanced Computing (ICoAC)*, Chennai, India, 2012, pp. 1-8, doi: 10.1109/ICoAC.2012.6416860.

[5]. S. K. Suman, L. Bhagyalakshmi and K. Murugan, "Non cooperative power control game for wireless ad hoc networks," *2012 Fourth International Conference on Advanced Computing (ICoAC)*, Chennai, India, 2012, pp. 1-6, doi: 10.1109/ICoAC.2012.6416822.

[6]. Suman S. K., Porselvi, S., Bhagyalakshmi L., and Kumar, D., “Game Theoretical Approach for Improving Throughput Capacity in Wireless Ad Hoc Networks”,in proceedings of International Conference on Recent Trends in Information Technology (ICRTIT 2014, MIT Chennai), 10-12 April 2014. 10.1109/ICRTIT.2014.6996152

[7]. S. K. Suman, B. Rajalakshmi, I. Khan, V. Alekhya, S. Lakhanpal and A. A. Ali, "Spatial Modulation Techniques for Improved ISAC Throughputs," *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*, Raigarh, India, 2024, pp. 1-5, doi: 10.1109/OTCON60325.2024.10688297.

[8]. Bhavani, B.G., Singh, S.N., Suman,S.K., Bhat, N., Sasirekha, D., Bharath, K. R., “ECG Dimensionality Reduction using PCA and Feature Abstraction Expending ICA Built with Power Spectral Estimation”, in proceeding of 9th International Conference on Signal Processing and Communication (ICSC 2023), 21-23 December 2023. 10.1109/ICSC60394.2023.10441071

[9]. G. R, S. K. Suman, K. Navaneetha, D. Rajeshswari, N. K. K and P. S. Wani, "The Ethical Implications and Societal Impact of Widespread WSN Deployments," *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, Indore, India, 2024, pp. 1-6, doi: 10.1109/ACROSET62108.2024.10743950.

[10]. M. V. Reddy, K. V. Shahnaz, P. Narayana, S. K. Suman, A. Asha and A. Balamurali, "An Effective Path Convergence Approach Founded on Recurrent Ant Colony Optimization (RECACO) in Mobile Ad Hoc Networks," *2023 9th International Conference on Signal Processing and Communication (ICSC)*, NOIDA, India, 2023, pp. 221-226, doi: 10.1109/ICSC60394.2023.10441147.

[11]. Narra, Naga, Nitin., R, Sai, Srikar., Pranav, Kumar, Dileep., D., Deva, Hema. (2024). 4.

Maximizing Crop Yield: Crop Yield Prediction using Advanced ML Algorithms. doi: 10.1109/accai61061.2024.10601735

[12]. J, Kavipriya., G., Vadivu. (2024). 3. Exploring Crop Yield Prediction with Remote Sensing Imagery and AI. doi: 10.1109/accai61061.2024.10601854

[13]. Sadia, Alam, Shammi., Yanbo, Huang., Gary, Feng., Haile, Tewolde., Xin, Zhang., Johnie, Jenkins., Mark, W., Shankle. (2024). 4. Application of UAV Multispectral Imaging to Monitor Soybean Growth with Yield Prediction through Machine Learning. Agronomy, doi: 10.3390/agronomy14040672

[14]. Oz, Kira., J., Wen., Jimei, Han., Andy, McDonald., Christopher, B., Barrett., Ariel, Ortiz-Bobea., Yanyan, Liu., Liangzhi, You., Nathan, Mueller., Ying, Sun. (2024). 5. A scalable crop yield estimation framework based on remote sensing of solar-induced chlorophyll fluorescence (SIF). Environmental Research Letters, doi: 10.1088/1748-9326/ad3142

[15]. Hiba, Najjar., Miro, Miranda., Marlon, Nuske., Ribana, Roscher., Andreas, Dengel. (2024). 1. Explainability of Sub-Field Level Crop Yield Prediction using Remote Sensing. doi: 10.48550/arxiv.2407.08274

[16]. Manasa, Chitradurga, Manjunath., Blessed, Prince, Palayyan. (2023). An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model. doi: 10.18280/ria.370428

[17]. Saeed, Nosratabadi., Imre, Felde., Károly, Széll., Sina, Ardabili., Bertalan, Beszedes., Amir, Mosavi. (2020). Hybrid Machine Learning Models for Crop Yield Prediction.arXiv: Neural and Evolutionary Computing,

[18]. Pramit, Pandit., Atish, Sagar., Bikramjeet, Ghose., Prithwiraj, Dey., Moumita, Paul., Saeed, AlQadhi., Javed, Mallick., Hussein, Almohamad., Hazam, Ghassan, Abdo. (2023). Hybrid time series models with exogenous variable for improved yield forecasting of major Rabi crops in India. Dental science reports, 13 doi: 10.1038/s41598-023-49544-w

[19]. Bali, Nishu & Singla, Anshu. (2021). Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India. Applied Artificial Intelligence. 35. 1-25. 10.1080/08839514.2021.1976091.

[20]. Ritik, Bohra., Shrunkhal, More., Shwetal, Kamat., Rishabh, Pathak., Prof., Manisha, Shitole. (2024). Crop Yield Prediction using CNN. International Journal of Advanced Research in Science, Communication and Technology, doi: 10.48175/ijarsct-18234

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/07-19
ISSN: 2319-4847

[21]. Srinivasa, Reddy, Seelam., Charitha, Sai, Thivari., Parabattina, Bhagath., Narendra, Kumar, G., K, Sumanth. (2024). Neural Harvest : Advancements in Crop Yield Forecasting. doi: 10.1109/icict60155.2024.10544942