International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 18s,2025
https://theaspd.com/index.php

Advances In Crop Yield Prediction Through Machine Learning And Deep Learning: Insights Into Yield Estimation

[1]Dipika Laxman Tidke, [2]Bhagwant B. Handge, [3]Madhuri V. Malode, [4]Vishakha A. Gaidhani, [5]Pratibha V. Kashid,
[1,5] Assistant Professor, Computer Engineering Department, MVPS's KBT College of Engineering, Nashik, India
[2]Lecturer, Computer Technology, MVPS's, RSM Polytechnic, Nashik, India
[3]Assistant Professor, Department of MBA, SVIT, Nashik, Maharashtra, India
[4]Associate Professor, Department of Information Technology, SVIT, Nashik, Maharashtra, India
Corresponding author email: [1]dipikat86@rediffmail.com

Abstract: *Crop yield prediction methods are becoming increasingly accurate and scalable because of the growing demand for sustainable agriculture. We identify various Machine Learning (ML) and Deep Learning (DL) techniques in yield prediction through a systematic review, which significantly contributes to state-of-the-art research for modern agriculture. It identifies and examines essential data sources such as remote sensing images, climate variables or meteorological information, soil health indicators, and sensor metrics that are key to improving prediction. In particular, it explores challenges with these methodologies relating to varying data quality, model size scalability, and a lack of interpretability in complex DL models. The future trajectory of these trends is then discussed, highlighting the increased utilization of hybrid models that capture the strengths and informational value of multiple techniques alongside further integration of multimodal learning approaches to take advantage of complementary data sources. These developments show promise to enhance the accuracy of yield predictions while overcoming regional as well as environmental discrepancies. The paper ends with some recommendations for future work, including the need to model the region; the use of explainable AI (XAI) techniques as integration between common and domain/ML approaches; and scalable frameworks to support real-time agricultural prediction. The present study synthesizes recent achievements and points to key challenges that can foster AI-driven agricultural practices, sustainability of agriculture, and meeting global food security demands.*
Keywords: *Crop Yield Prediction, ML, DL, Agricultural Forecasting, Remote Sensing, Multimodal Learning*

1. INTRODUCTION

As the foundation of global food security, agriculture is vital to our survival, and rice, serving billions worldwide is one of the most important staple crops (Food and Agriculture Organization, 2021). Moreover, predicting crop yield such as rice is crucial to sustainable agricultural practices because it stimulates solutions for mitigating risks associated with climate change that results in food insecurity and the never-ending hunger pang of a growing population. Wart et al. (2013) highlight the limitations of current approaches of predicting crop yields through historical trends or practical observations to address the complexity of contemporary food production systems. The advent of ML and DL has transformed the field, enabling experts to perform “instant predictions” according to the authors’ reliance not on complex calculations but on measured estimates provided by various datasets, including climatic parameters, soil characteristics, or satellite imagery. This review paper provides the reader with a brief outline of recent breakthroughs in ML and DL methods of rice yield prediction. In this paper, the research assesses the authors’ interest in methods, the pros, and cons and examples of using such models, and concludes the positioning on multimodal utilization options, ranging from addressing the lack of heterogeneous data sources to enhance the broad applicability of AI models to limitedly interpretable model adjustments. Therefore, this paper aims to bridge a gap between theory and practice, as no comprehensive overview on AI in the crop yield prediction domain has been conducted.

Major Contributions of the paper are as follows:

1. A comprehensive review of state-of-the-art ML and DL methods used for yield prediction
2. Identification of key data sources and their role in enhancing predictive performance
3. Analysis of challenges such as data quality, model scalability, and explainability
4. Discussion of emerging trends, including hybrid models and multimodal learning approaches

5. Recommendations for future research directions in AI-powered agricultural forecasting.

The paper is organized as follows: Section 2 provides a detailed review of traditional and AI-based techniques for crop yield prediction. Section 3 discusses the significance of data sources and their integration into predictive models. Section 4 evaluates the challenges and limitations of ML and DL methods in this domain. Section 5 explores emerging trends and innovative approaches for yield estimation. Section 6 concludes the paper with key findings, recommendations, and future research opportunities.

2. LITERATURE REVIEW

Sharifi A (2021) discusses the application of ML algorithms with satellite imagery to predict crop yields. The new model emphasizes the integration of data on vegetation indices; weather and soil properties to improve predictiveness. The results make remote sensing a responsive and scalable approach applicable to agricultural forecasting. It also touches upon the challenges of data quality and algorithm scalability.

K. P. Murphy, (2023) Probabilistic ML: Advanced topics It covers deeper topics about probabilistic ML such as Bayesian methods, Gaussian processes, and deep generative models. It guides us in using probabilistic methods for uncertainty quantification and decision-making, which is important such as in agriculture, where knowing when predictions can be trusted is essential.

Posch, K. et al. (2021) Variable selection with nearest neighbor Gaussian processes We present a new approach for variable selection in Gaussian processes based on nearest-neighbor techniques and examine its performances compared to existing approaches. This approach minimizes computing costs making it applicable to huge datasets found in agricultural forecasts. The datasets from the validation experiment fully support each method and show a potential to be applied in crop yield modeling.

Sparapani, R. et al. (2021) present Nonparametric Bayes via Bayesian additive regression trees. Bayesian Analysis. This article is on Bayesian Additive Regression Trees (BART) and their implementation in the BART R package. The model is capable of handling non-linear relationships and this property was applied in a case study for crop yield predictions, throughs that the authors pointed out [4]. It is then proposed that BART provides a fast option for non-parametric ML in agriculture.

Elbasi, E. et al. (2023) Predictive model of the crop using supervised Machine Learning algorithms. This work proposes a model for the estimation of crops based on ML algorithms like decision trees, support vector-machines and nn. A systems and method for improved forecasting of specific crop types and their potential yields through use of climate, soil and remote sensing data and information. These results demonstrate the flexibility of the model in diverse agro-climatic scenarios.

Zhao, H. et al. (2019) DL based crop classification in heterogeneous areas: case of trough, early crop assessment based on three dl models and spectral information using sentinel-1a imagery time series. Three DL models (VGG16, InceptionV3 and ResNet50) were evaluated in this paper for early season crop classification using Sentinel-1A satellite observations. It demonstrates the importance of time: when time information is considered, early characterizations of plants could be much more accurate and would provide inputs to predict yield.

Ilyas, Q. M. et al. (2023) AI-Based Remote Sensing Techniques for Autonomous Crop Yield Estimation. For the book chapter, we specifically write about an AI and remote sensing framework for crop yield estimation. It relies on high-resolution satellite imagery and ML methods for improved spatial and temporal prediction. Source Title: Utilising AI for Water Security: Reinventing the Opportunities and Challenges. This study serves to spotlight significant obstacles such as the expense of calculations and data availability.

Hu, T. et al. (2023) Research on crop yield -box estimation of climate change effects on crop yield with XAI and interpretable ML We provided summary of the importance of XAI in crop yield prediction because of the black box nature of (lack of explainability) in the machine learning models. It discusses the dangers of relying on non-interpretable models for consequential decision-making and calls it a dangerous temptation when it comes to climate change. They suggest hybrid interpretable methods to maintain both the accuracy and interpretability.

Aslan, M. F. et al. (2024) A comprehensive survey on artificial intelligence methods in crop yield estimation using Sentinel-2 data This survey focuses on artificial intelligence (AI) methods that have taken advantage of the Sentinel-2 satellite data for crop yield estimation. The categorization is proposed according to conventional ML, DL, and hybrid techniques. It was found that prediction improvement using spectral indices and multimodal data contribute significantly.

Sharma, P. et al. (2023) ML-Based Agriculture Yields Prediction Applying Regression & DL In this research, regression-based ML models and DL approaches are compared concerning predicting agricultural yields. The authors highlight the better performance of DL models in predicting yield by capturing complex relationships between soil properties and climatic factors.

Patil, P. et al. (2023) ML-Based Crop Selection and Yield Prediction This paper presents a framework based on machine learning (ML) that can be employed for crop selection and yield prediction. Using past data, climate, and soil characteristics, the model tells you which are the most appropriate crops for a location and if so at what expected yield.

Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. In this paper Talaat et al (2023), the MATLAB-based algorithms involved in crop yield prediction; using IoT-based sensors and climate data are discussed. Incorporating real-time data from the environment, the algorithm can provide more accurate predictions by tackling one of the challenges that many algorithms face because agricultural systems are highly dynamic and governed by climate variability.

M. Sathya & PGnanasekaran (2023) MLR-LSTM based Paddy Yield Prediction in Delta Region of Tamilnadu. To predict the paddy yield in Tamil Nadu, India, we implement a hybrid MLR–LSTM model in this study. Statistical and DL methods can be combined to perform better than stateless or standalone models – having specific successes for time-series studies.

Reddy, D.J. & Kumar M.R (2021) ML Algorithm Based Crop Yield Prediction Aspect of Machine Learning ML algorithms such as Random Forest RF and Gradient Boosting GB have been successfully used in predictive modeling. Using these datasets, the authors assess the models based on historical crop and climate data, finding that ML approaches consistently provide improvements in accuracy over conventional predictive methods.

Gosh, R. (2022) Crop Yield and Environmental Feature in Agriculture land: A Survey on Deep Learning Architecture. DNN-based approaches to crop yield prediction using high-dimensional, multivariate datasets have been demonstrated in this paper. As this study shows, the complexity and nonlinear relationships DNNs can handle make them a good candidate to be utilized for Agriculture applications.

Waseem, M. et al. (2024) Predicting Crop Yield and Disease Detection in Agroecosystems Based on Machine Learning Techniques Chapter three covers the AI frameworks for crop yield prediction and disease detection. To enhance prediction precision and sustainability, it focuses on the amalgamation of ML & DL techniques with agroecosystem data. It handles the challenges of scalability and data heterogeneity.

Kishore, P. et al. PradeepJotheri, K and Tapparel, Sophie and Validiv, Alesya (2024) Improving Crop Yield Prediction using Ensemble Learning Methods on Multi-Modal Agricultural Data. It explores ensemble learning techniques to leverage multi-modal agricultural data, including satellite images and weather and soil measurements by fusing various ML models. The results show higher accuracy and stability over methods that use one model.

Crop yield prediction is very crucial for smart agriculture. In this paper, we introduce a smart agriculture system based on the cloud that implements hybrid DL algorithms. The comparison model is short and long-range memory LSTM is part of the framework to analyze remote sensing and IoT data from mobile devices for crop yield prediction. It highlights the ability to scale out and, process data in real time.

Anbananthen, K. S. M. et al. Deep and Hybrid Machine Learning Decision Support System for Predicting Crop Yield: An Intelligent Approach (2021) This paper presents a hybrid machine learning-based intelligent decision support system for yield prediction. It integrates linear regression, decision trees, and neural networks into a single user-friendly non-linear regression system with greater accuracy and flexibility for varying agricultural datasets.

Mallika, T. et al. (2023) DeepNeuralNet: A New Approach to Predicting Crop Production. They introduce a new deep neural network paradigm at the framework level, one that captures both spatial and temporal dependencies in data, to improve crop yield prediction. Combining high-resolution satellite imagery with climate datasets, this method shows higher predictive performance.

Article 10 Al-Adhaileh, M. H. & Aldhyani, T. H. H. (2022) An Artificial Intelligence Framework for Modelling and Predicting the Crop Yield for Improving Food Security in Saudi Arabia In this paper, we propose a framework of AI specific to the agricultural context in Saudi Arabia for crop yield prediction that incorporates machine learning (ML) models trained with remote sensing and Internet-of-Things data. It helps to offer decision-makers and the agricultural sector access to information on food security challenges from the perspectives of crop production management.
Ramzan, S. et al. And Then (2024) An Efficient IoT Based Crop Prediction System By Using ML and EL. In the existing study, an IoT-based system is presented that integrates ML and EL techniques to predict crops. This approach focuses on integrated real-time acquisition and processing of relevant data that allows for real-time in-field adaptive accurate yield predictions over diverse agro-climatic conditions.
Mahankale, N. et al. AI-driven geospatial analysis of weather–crop yield relationships in satellite agrometeorology (2023) The present study investigates the applications of AI in spatial analysis of crop yield as influenced by weather parameters. Satellite-derived agrometeorological data are integrated with ML algorithms to increase the spatial detail and precision of yield forecasts.
Olofintuyi, S. S. et al. An ensemble deep learning approach for predicting the yield of cocoa — VOSviewer (2023) The authors provide an ensemble DL model for predicting cocoa yield. The integrated LSTM-CNN-fully connected layers performance model sequentially captures both temporal and spatial patterns in agriculture data and achieves better accuracy than individual models.
Sun, J. et al. (2020) This paper presents a county-level corn yield prediction model via satellite imagery and environmental data based on a multilevel DL network. It achieves multi-scale representational capabilities through a hierarchical approach.
Freitas Cunha, R. L. & Silva, B. (2020) The use of remote sensing data and DL models for crop yield estimation is widely studied in this respect. A hybrid framework that combines CNNs with multi-spectral satellite data to predict yields at high spatial resolution It highlights the high scalability and flexibility of DL to capture complex patterns in large-scale agricultural forecasting.
Cunha, R. L. et al. (2018) A scalable machine learning system for predicting agriculture yield before planting season We built a scalable machine learning system for pre-season yield prediction which we propose in this work. Our approach uses regression algorithms trained on weather and soil data to provide a budget-friendly tool for early agricultural decision-making.
Perich, G. et al. (2023) The prediction of pixel-based yield using spectral indices and neural networks from Sentinel-2 This study proposes a pixel-based mapping method based on the Sentinel-2 data. The presented study uses neural networks for field-level yield prediction based on spectral indices, and the results are very accurate in different agriculture scenarios.
Ma, Y., & Zhang, Z. (2022). A Bayesian domain adversarial neural network-based method for corn yield prediction. In this paper, a Bayesian domain adversarial neural network (BDANN) is developed to flexibly adapt across multiple agricultural domains for corn yield prediction. The proposed method improves generalization under different environmental conditions and regions.
Divakar, M. S. et al. (2022) present the Forecasting Crop Yield using a DL-based ensemble model Combining LSTM & CNN with gradient Boosting A framework for time-series and remote-sensing data-driven crop yield forecasting.
Zhang, L. et al. (2021) Deep learning-based smallholder maize yield prediction under the use of satellite-derived climatic and vegetation indices To this end, in this study, we investigate the incorporation of climatic and vegetation indices developed from satellite data into a DL framework for predicting maize yield. This shows that DL models are a feasible solution to enhance yield prediction for smallholder farms.
Ma, Y. et al. (2021) A bayesian convolutional neural network model for corn biomass and uncertainty prediction as a function of remotely sensed covariates. For corn yield estimate and uncertainty quantification, the authors propose a Bayesian neural network. Regression When localisation on-farm level is the objective, it merges remote-sensing layers with local applications for Precision Agriculture.
Tian, H. et al. (2021) Attention mechanism based deep learning model for winter wheat yield prediction using remotely sensed indices in the Guanzhong Plain of PR China. In this study, we propose a deep learning model with attention mechanism for wheat yield estimation using remotely sensed indices. These

attention layers are beneficial to the model so that it is interpretable and more accurate in predicting yield.

Zhang, L. et al. (2019) ML approaches for dynamic prediction of county-level maize yield in China integrating optical, fluorescence and thermal satellite data and environmental data The research explores a novel approach to dynamically simulate county-level maize yield by integrating multi-source (optical, fluorescence, thermal) satellite image and environmental data into machine learning (ML) models. Our results demonstrate the importance of diverse data in building more robust yield prediction models.

Di, Y. et al. (2022) A DL and optimization-based framework for winter wheat yield prediction Given the situation presented above, in this article, we propose a new framework which combines the DL models and Bayesian optimization to effectively improve the accuracy of winter wheat yield predictions. It utilizes, but does not bypass, multi-temporal data to achieve better performance than a pure traditional framework.

Ariza-Sentís, M. et al. (2023) Predicting spinach (Spinacia oleracea) seed yield from 2D UAV images using DL We used UAV imagery and DL models to design an approach to predicting the seed yield of spinach in this research. The author has presented a UAV-based algorithm for accurate yield estimation with 2D images and studied the possibility of practical field-level application of this technology in agriculture.

Apolo-Apolo, O. E. et al. A UAV based cloud computing framework for generating an apple yield map Orchard: (2020) This paper provides a cloud-based facility for apple yield estimation from UAV images and DL. These are real-time yield maps, one of the best illustrations of cloud benefits in precision ag on an epic scale.

Priyatikanto, R. et al. (2023) Domain adaptation enhances generalization and transferability of maize yield prediction models using machine learning. The goal of this study is to extend the generalization ability of the ML-based maize yield prediction models by using domain adaptation. This upscaling knowledge allows to its use in various climate conditions by a direct use of model trained in one site.

Barbosa, B. D. S. et al. (2021) Estimation of coffee yield by using a UAV with feature selection and DL, the authors propose a UAV (Unmanned Aerial Vehicle) methodology to estimate coffee yield, which combine feature selection methods with DL (Deep Learning) models. This leads us to the conclusion that the synergy between Remote Sensing and AI is of paramount importance for the research on yield prediction.

Livieris, I. E. et al. A neural network model with multiple inputs for the cotton production quantity prediction: a case study (2020) Fan et al. propose a multiple-input neural network to predict cotton production that integrates heterogeneous agricultural datasets [10]. The model is also quite accurate and adaptable, reinforcing its promise for experience with other crop categories.

Bai, D. et al. Soybean yield estimation parameters under lodging conditions based on unmanned aerial vehicle-derived RGB information. This study aims to explore the use of UAV-based RGB imagery for soybean yield estimation in severe lodging conditions. The results show that UAVs have the potential to provide important yield parameters with high accuracy, especially under less favorable conditions.

Abbaszadeh, P. et al. Bayesian multi-modeling of deep neural nets to make crop yield predictions using probability (2022).

A multi-model framework for crop yield prediction combining deep neural networks: a Bayesian approach This model, which considers uncertainty, provides reliable predictions in diverse environmental settings.

Terliksiz, A. S. & Altýlar, D. T.(2019) Deep Neural Network for Crop Yield Prediction: A Case Study on the Soybean Yield in Lauderdale County (Alabama - USA), Lecture Notes in Computer Science Springer International Publishing AG, 173-178. 234 Deep learning applied to satellite and environmental data for soybean yield prediction: A case study As a result, the results point out how DNNs outperform traditional models in identifying complex patterns in agriculture.

Qiao, M. et al. Liu, Z., Lin, F.(2021). 3D-CNN multi-kernel GP for crop yield prediction with hierarchical feature use. It fuses 3D CNN and multi-kernel GP for leveraging hierarchical characteristics in the prediction of crop yield. Our hybrid model achieves better performance by integrating spatial and temporal information.

Tanabe, R. et al. (2023) A Convolutional Neural Network approach on winter wheat yield prediction using multitemporal multispectral images from UAV. CNNs are also employed to process UAV multispectral data to predict winter wheat yield in this research. This study presents the fusion of UAV and DL models to deliver reliable and scalable yield estimates.

Qiao, M. et al. (2021) Crop yield prediction based on multi-spectral temporal observation in remotely sensed imagery with recurrent 3D convolutional neural network (R3D-CNN) A similar study as above that employs recurrent 3D CNN for multi-spectral and multi-temporal remotely-sensed data while predicting crop yield. It is this way of capturing spatial and temporal dependencies which led to high accuracy in large-scale agriculture prediction.

MacEachern, C. B. et al. (2023) DL CNN for detecting Fruit Maturity Stage and Yield Estimation in wild blueberries Wild blueberry fruit maturity stage detection and yield estimation was performed using convolution neural networks (CNN) in this study. The results indicate that DL can significantly improve the accuracy of yield estimation and automatize maturity discrimination.

Lee, S. et al. (2019)Deep learning based self-predictable crop yield platform (SCYP) using diseases for kharif crops . SCYP frame-work uses the DL technology and crop disease identification along with yield forecasting for crops. Dual role: This system can serve dual roles to provide crop yield and health information simultaneously and lead to better decision-making in agriculture Commissioning of first PCT Unfortunately protocols work without any commissioning.

3. Significance of Data Sources and Integration into Predictive Models

Data sources of the crop yield prediction are also critical to why the model's accuracy and robustness is important. The recent advances in ML and DL models allow us to solve the challenging agricultural problems that are dominated by massive amount of high-resolution multi-modal data. The next section maps data sources together with their importance with regard to combining them with predictive constructs:

1. Remote Sensing Imagery

Remote sensing from satellite and UAV is an important source of information about crop health, soil texture, vegetation indices (e.g., NDVI) and climatic conditions. These sources provide spatial and temporal resolution to improve the model. By inputting multi-spectral and hyper-spectral data in CNNs and recurrent models, temporal patterns in crop growth and anomalies can be detected more efficiently which can result in more accurate predictions.

2. Weather Data

Temperature, rain, and moisture are the main elements that effect on plant growth. Weather variables record short- and long-term climatic impacts on yield. Integrating weather data to time-series models (e.g., LSTMs) or hybrid methods can assist in predicting yields dynamically under changing environmental conditions.

3. Soil Health Metrics

Ph, nutrient and moisture determine the soil productivity in an immediate way. Utilizing such parameters presents the basis for estimating the potential growth of individual crops.

Outcomes Combining soil data with ML model such as Random Forest or gradient boosting improves the model’s performance while predicting outcomes at certain regions or farming practices.

4. Historical Yield Data

Historical records provide an image of a myriad of forces influencing yield over the years which can be used for predictions. The historical information, if used into the supervised learning models, presents a ground-truth that can be used to train and validate the ML/DL models.

5. IoT and Sensor Data

Internet of things (IoT) devices capture real-time metrics such as soil humidity, temperature and nutrient level, thereby providing detailed and continuous sensor reading. IoT data can be integrated into predictive models to facilitate dynamic, real-time yield prediction, given cloud-based frameworks especially.

6. Geospatial and Topographic Data

Topographic features such as altitude, gradient, and slope orientation impact water flow and sunlight radiation, in turn exerting effects on crop growth. Geospatial data augment spatially explicit models, such as those delivered by GIS-based predictive methodologies and give a sense for region-specific yield responses.

Challenges in Data Integration

Despite its benefits, data integration poses challenges such as:

- Data heterogeneity across sources.
- Scalability issues with large datasets.
- The need for preprocessing and harmonization of formats and resolutions.

To operate at their optimal potential in agriculture, integrated predictive models need to address these challenges. New methodologies such as federated learning or more advanced fusion schemes are promising to address these limitations. By utilizing different sources of information in an optimal manner, predictive models can provide a novel drive for changing the way agriculture is being conducted towards a more secure and sustainable global food system.

4. Challenges and Limitations of ML and DL Methods in Crop Yield Prediction

There are a number of limitations and challenges when using ML and DL techniques in predicting crop yield, despite the remarkable progress being made in the ML and DL technologies regarding the crop yield prediction problems, which mainly is based on the complexity of agricultural system. These problems affect the robustness, the scalability, and the real--world application of these models.

1. Data Challenges

- **Heterogeneity of Data**: Data in agriculture stem from a variety of sources like satellite imagery, weather stations and Iot devices, which are usually in different formats, resolutions and temporal frequencies. This very variety of cotton causes difficulties for the preprocessing of the data and the training of models.
- **Data Quality and Noise**: The model may suffer from errors or missing data, as well as other noises from environmental conditions. For example, error-prone or low-quality weather station measurements and low-resolution space-borne images may cause limited prediction accuracy.
- **Limited Access to High-Quality Data**: There exists large regions and developing countries where no detailed information is available, which limits the overall applicability of the model.

2. Model Complexity

- **Overfitting**: DL models tend to overfit to the training data in case the training data is small, that may capture noise patterns instead of useful ones. This is difficult to achieve especially in very small sample data such as agricultural datasets.
- **Lack of Explainability**: DL models have a reputation of being “black boxes”. The lack of interpretability of these models also creates a difficulty for stakeholders — farmers and policymakers, who cannot trust predictions and do not understand why they were driven by these models.
- **High Computational Costs**: Due to the nature of the classical ML and DL algorithms, which is required to train these complex models, the computational cost is high, and they tend to be impractical to deploy in resource limited systems.

3. Scalability and Generalization

- **Region-Specific Models**: Models developed in one area of a region may not be valid in other areas with different agriculture, climate and soil management systems. The non-existence of any widely applicable models restricts the degree to which solutions are scalable.
- **Dynamic Environmental Factors**: Agricultural systems are vulnerable to dynamic events such as changes in the weather, pests and disease. The requirement of so many manual adjustments makes static models ill-equipped to account for these dynamic conditions and creates difficulties in performing real-time applications of these models.

4. Integration Challenges

- **Combining Multimodal Data**: Representation of mixed data (e.g., in a single stream or from different modalities, e.g., remote sensing images and IoT data) demands sophisticated fusion techniques. If not well integrated, this may cause model performance degradation.
- **Interoperability Issues**: Systems used by various organizations for storing and gathering data may not be interoperable with each other, resulting in non-uniform data integration and model deployment.

5. Implementation Barriers

- **Skill and Knowledge Gaps**: Most of the small holders lack knowledge to interpret and implement ML and DL models which hinders their adoption.

- **Infrastructure Constraints**: The computational infrastructure including the Internet connectivity are usually not available in rural settings where Agriculture is more practiced and thus not suitable for deploying sophisticated ML/DL solutions.
- **Cost of Deployment**: The cost of purchasing IoT devices, high-resolution satellite data, and computing power can be expensive for small farmers.

6. Validation and Evaluation

- **Lack of Standardized Metrics**: Different research papers employ different model evaluation metrics like accuracy, RMSE, and F1 score and so on, making one model difficult to compare with that other model.
- **Limited Field Validation**: Numerous models are tested in simulation or based on historic data rather than on real-world conditions, resulting in different performance in wild.

7. Ethical and Regulatory Issues

- **Privacy Concerns**: The use of sensitive data (e.g., location of farms) by sensors raise a number of privacy issues that can be addressed via secure data practices.
- **Regulatory Hurdles**: AI and digital technologies, from economics' standpoint, may require changes in legislation, especially in terms data sharing and data privacy in countries having stringent laws.

Strategies to Overcome Challenges

1. **Data Augmentation**: Using synthetic data generation techniques, such as GANs, to overcome data scarcity issues.
2. **XAI**: Developing interpretable models to enhance trust and usability among stakeholders.
3. **Transfer Learning**: Leveraging pre-trained models to improve scalability and generalization across regions.
4. **Federated Learning**: Enabling collaborative model training without sharing raw data, addressing privacy and data scarcity concerns.
5. **Policy Support**: Governments and organizations should support the development and deployment of ML/DL systems in agriculture through subsidies and infrastructure improvements.

By addressing these challenges and limitations, ML and DL methods can become more accessible, reliable, and impactful, ultimately transforming agriculture into a more sustainable and resilient sector.

5. Emerging Trends and Innovative Approaches for Yield Estimation

Research in the area of crop yield prediction develops quickly by the progression of ML, DL, and data collection technologies. New trends and novel methodologies are bringing solutions to the current bottlenecks, increasing model performance and opening new ways for sustainable agriculture. Here are some of the most important developments:

1. Hybrid and Multimodal Models

Integrating ML and DL to take advantage of strengths of each other. CNN-LSTM-based hybrid models for spatial-temporal processing. Multimodal satellite, weather, soil and IoT (Internet of Things) sensor joint models. Improved treatment of complex relationships between heterogeneous data sources. Enhanced accuracy and generalizability.

2. Transfer Learning

Using pre-trained models developed in one region or crop type and adapting them to different conditions. Pre-trained CNNs on large datasets for remote sensing imagery applied to specific crops or regions. Reduces computational costs and training time. Addresses data scarcity issues in underrepresented regions.

3. XAI

Developing interpretable models to provide actionable insights rather than just predictions. Feature attribution methods like SHAP (Shapley Additive Explanations) to understand the contribution of each input variable. Builds trust among stakeholders like farmers and policymakers. Facilitates regulatory compliance and ethical AI use.

4. Integration of IoT and Edge Computing

Real-time data collection and processing using IoT devices and edge computing. IoT sensors for monitoring soil moisture, temperature, and crop growth combined with on-site edge devices for immediate processing. Reduces latency in decision-making. Enables real-time yield estimation in resource-constrained environments.

5. Domain Adaptation and Federated Learning

Techniques to enhance model adaptability and privacy. Domain adaptation for training models in one region and applying them to others. Federated learning for collaborative model training without sharing raw data. Improves scalability and applicability across diverse agro-climatic conditions. Ensures data privacy and security.

6. Advanced Data Fusion Techniques

Combining data from multiple sources to create comprehensive datasets for yield estimation. Fusion of multispectral, hyperspectral, and thermal satellite imagery with weather and soil data. Captures multi-dimensional aspects of crop growth. Provides a holistic view of factors influencing yield.

7. Bayesian Neural Networks and Uncertainty Quantification

Incorporating uncertainty quantification into DL models to improve reliability. Bayesian Neural Networks (BNNs) for estimating yield with confidence intervals. Helps in decision-making under uncertainty, particularly in volatile agricultural conditions. Improves model robustness.

8. Generative Models and Synthetic Data

Using Generative Adversarial Networks (GANs) and other techniques to generate synthetic data. GANs for augmenting datasets with additional training samples.

Addresses data scarcity issues, particularly for rare crop types or regions.

Enhances model training and validation.

9. Attention Mechanisms and Transformers

Leveraging attention-based models for improved feature representation. Vision Transformers (ViTs) for analyzing high-resolution remote sensing imagery. Attention-enhanced LSTMs for identifying critical temporal patterns. Improves model interpretability and focus on relevant features. Handles large-scale data more efficiently.

10. Crop-Specific and Climate-Aware Models

Developing models tailored to specific crops and climate conditions. Climate-aware models integrating historical weather patterns and future climate projections for yield forecasting. Improves regional and crop-specific accuracy. Provides insights into the impact of climate change on agriculture.

11. Cloud-Based and Scalable Architectures

Using cloud platforms for deploying scalable yield estimation systems. Cloud-based frameworks integrating real-time satellite data for large-scale applications. Supports real-time analytics and decision-making. Provides accessibility to farmers and agricultural agencies.

12. Automation and Smart Farming

Incorporating autonomous systems for yield estimation and farm management. Autonomous UAVs collecting data on crop health and yields. Reduces labor-intensive data collection processes. Increases operational efficiency.

Emerging trends in ML and DL for yield estimation demonstrate a promising shift toward precision agriculture. Continued innovations in data integration, model interpretability, and real-time analytics are expected to bridge the gap between theoretical advancements and practical applications, ensuring sustainable and efficient agricultural systems globally.

6. CONCLUSION

The integration of ML and DL techniques into crop yield prediction has revolutionized traditional agricultural practices, in terms of accuracy, scalability, and transferability. In this paper, we have showcased the spectrum of developments, focusing on the importance of a variety of data sources such as remote sensing imagery, weather data, soil health indicators, and IoT sensor data. These data-based methods have enabled researchers to construct novel predictive models that capture complex spatial and temporal dynamics that characterize the many dimensions of modern agriculture.

The major contributions of this paper emphasize the revolutionary power of hybrid, multimodal methods that exploit the best features of a range of ML/DL approaches and perform better. Attention mechanisms, transfer learning and domain adaptation have appeared as the trends with potential to improve generalization performance of models and make them scalable in different agro-nomic conditions. Additionally, the inclusion of XAI in agricultural predictions may contribute to the establish trust and usability with stakeholders.

Despite these advancements, Data heterogeneity, high computational overhead, model interpretability and lack of generalizability to under- representing regions are all major stopping points to relying on these systems on a large scale. These roadblocks will need to be overcome by a partnership among researchers, policymakers and technology developers.

Future studies should develop geographically-tailored models by considering local environmental factors, and towards the ability of hyper-scalability, accomplish the model adaptation through transfer learning and federated learning methodologies. Furthermore, predictive frameworks can be enhanced for their stochastics and Bayesian optimization added to increase the reliability of model outputs, particularly for dynamic agricultural systems. Infrastructure investments, like IoT networks and cloud-based solutions, will help to enable decision-making in real-time and make these technologies more feasible to implement in low-resource settings.

With the constraints of food security and climate change facing the global community, the continuous development of ML and DL for crop yield prediction is of great significance in sustainable agriculture practices. This field may significantly contribute to global food sustainability and resilience through adoption of emerging trends and addressing current limitations.

REFERENCES:

1. FAO. 2021. *World Food and Agriculture – Statistical Yearbook 2021.* Rome.
2. J. van Wart, K. C. Kersebaum, S. Peng, M. Milner, and K. G. Cassman, "Estimating crop yield potential at regional to national scales," *Field Crops Research*, vol. 143, pp. 34-43, 2013.
3. Sharifi, A. (2021). Yield prediction with ML algorithms and satellite images. *Journal of the Science of Food and Agriculture*, *101*(3), 891-896.
4. Murphy, K. P. (2023). *Probabilistic ML: Advanced topics*. MIT Press.
5. Posch, K., Arbeiter, M., Truden, C., Pleschberger, M., & Pilz, J. (2021). Variable selection using nearest neighbor Gaussian processes. *arXiv preprint arXiv:2103.14315*.
6. Sparapani, R., Spanbauer, C., & McCulloch, R. (2021). Nonparametric ML and efficient computation with Bayesian additive regression trees: The BART R package. *Journal of Statistical Software*, *97*, 1-66.
7. Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop prediction model using ML algorithms. *Applied Sciences*, *13*(16), 9288.
8. Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., & Feng, M. (2019). Evaluation of three DL models for early crop classification using sentinel-1A imagery time series—A case study in Zhanjiang, China. *Remote Sensing*, *11*(22), 2673.
9. Ilyas QM, Ahmad M, Mehmood A. Automated Estimation of Crop Yield Using Artificial Intelligence and Remote Sensing Technologies. *Bioengineering*. 2023; 10(2):125. https://doi.org/10.3390/bioengineering10020125
10. T. Hu, X. Zhang, G. Bohrer, Y. Liu, Y. Zhou, J. Martin, Y. Li, and K. Zhao, "Crop yield prediction via explainable AI and interpretable ML: Dangers of black box models for evaluating climate change impacts on crop yield," *Agricultural and Forest Meteorology*, vol. 336, p. 109458, 2023. doi: 10.1016/j.agrformet.2023.109458.
11. Aslan MF, Sabanci K, Aslan B. Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey. *Sustainability*. 2024; 16(18):8277. https://doi.org/10.3390/su16188277
12. P. Sharma, P. Dadheech, N. Aneja and S. Aneja, "Predicting Agriculture Yields Based on ML Using Regression and DL," in *IEEE Access*, vol. 11, pp. 111255-111264, 2023, doi: 10.1109/ACCESS.2023.3321861.
13. Patil P, Athavale P, Bothara M, Tambolkar S, More A. Crop Selection and Yield Prediction using ML Approach. Curr Agri Res 2023; 11(3). doi: http://dx.doi.org/10.12944/CARJ.11.3.26
14. Talaat, F.M. Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. *Neural Comput & Applic* **35**, 17281–17292 (2023). https://doi.org/10.1007/s00521-023-08619-5
15. Sathya P & Gnanasekaran P. (2023) Paddy Yield Prediction in Tamilnadu Delta Region Using MLR-LSTM Model. *Applied Artificial Intelligence* 37:1.
16. D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using ML Algorithm," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2021, pp. 1466-1470, doi: 10.1109/ICICCS51141.2021.9432236.
17. Khaki S and Wang L (2019) Crop Yield Prediction Using Deep Neural Networks. *Front. Plant Sci.* 10:621. doi: 10.3389/fpls.2019.00621
18. Waseem, M., Raza, A., & Malik, A. (2024). AI-Driven Crop Yield Prediction and Disease Detection in Agroecosystems. In B. Singh, C. Kaunert, K. Vig, & S. Dutta (Eds.), *Maintaining a Sustainable World in the Nexus of Environmental Science and AI* (pp. 229-258). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-6336-2.ch009

19. Kishore, P. ., Kumar, R. ., Negi, H. S. ., Sharma, K. ., Srivastava, A. P. ., Kumar, N. ., & Shrivastava, A. . (2024). Enhancing Crop Yield Prediction Using Ensemble Learning Techniques on Multi-Modal Agricultural Data. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(20s), 262–268.
20. Sharma A. K, Rathore A. S. Design and Implementation of a Cloud-Based Smart Agriculture System for Crop Yield Prediction using a Hybrid DL Algorithm. Curr Agri Res 2024; 12(2). doi: http://dx.doi.org/10.12944/CARJ.12.2.17
21. Anbananthen KSM, Subbiah S, Chelliah D, Sivakumar P, Somasundaram V, Velshankar KH, Khan MKAA. An intelligent decision support system for crop yield prediction using hybrid ML algorithms. F1000Res. 2021 Nov 11;10:1143. doi: 10.12688/f1000research.73009.1
22. Mallika, T. ., Kavila, S. D. ., Pydi, B. ., & Rajesh, B. . (2023). A Novel Method for Prediction of Crop Yield Using Deep Neural Networks. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(4), 308–315.
23. Al-Adhaileh MH, Aldhyani THH. 2022. Artificial intelligence framework for modeling and predicting crop yield to enhance food security in Saudi Arabia. *PeerJ Computer Science* 8:e1104 https://doi.org/10.7717/peerj-cs.1104
24. S. Ramzan, Y.Y. Ghadi, H. Aljuaid, A. Mahmood, and B. Ali, "An Ingenious IoT Based Crop Prediction System Using ML and EL," *Comput. Mater. Contin.*, vol. 79, no. 1, pp. 183-199, 2024. https://doi.org/10.32604/cmc.2024.047603
25. N. Mahankale, S. Gore, D. Jadhav, G. S. P. S. Dhindsa, P. Kulkarni and K. G. Kulkarni, "AI-based spatial analysis of crop yield and its relationship with weather variables using satellite agrometeorology," *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Mumbai, India, 2023, pp. 1-7, doi: 10.1109/ICACTA58201.2023.10392944.
26. Olofintuyi, S. S., Olajubu, E. A., & Olanike, D. (2023). An ensemble DL approach for predicting cocoa yield. *Heliyon*, *9*(4).
27. Sun, J., Lai, Z., Di, L., Sun, Z., Tao, J., & Shen, Y. (2020). Multilevel DL network for county-level corn yield estimation in the us corn belt. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5048-5060.
28. de Freitas Cunha, R. L., & Silva, B. (2020, March). Estimating crop yields with remote sensing and DL. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)* (pp. 273-278). IEEE.
29. Cunha, R. L., Silva, B., & Netto, M. A. (2018, October). A scalable ML system for pre-season agriculture yield forecast. In *2018 IEEE 14th International Conference on e-science (e-Science)* (pp. 423-430). IEEE.
30. Perich, G., Turkoglu, M. O., Graf, L. V., Wegner, J. D., Aasen, H., Walter, A., & Liebisch, F. (2023). Pixel-based yield mapping and prediction from Sentinel-2 using spectral indices and neural networks. *Field Crops Research*, *292*, 108824.
31. Ma, Y., & Zhang, Z. (2022). A Bayesian domain adversarial neural network for corn yield prediction. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5.
32. Divakar, M. S., Elayidom, M. S., & Rajesh, R. (2022). Forecasting crop yield with DL-based ensemble model. *Materials Today: Proceedings*, *58*, 256-259.
33. Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., & Li, S. (2021). Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using DL. *Agricultural and Forest Meteorology*, *311*, 108666.
34. Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, *259*, 112408.
35. Tian, H., Wang, P., Tansey, K., Han, D., Zhang, J., Zhang, S., & Li, H. (2021). A DL framework under attention mechanism for wheat yield estimation using remotely sensed indices in the Guanzhong Plain, PR China. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102375.
36. Zhang, L., Zhang, Z., Luo, Y., Cao, J., & Tao, F. (2019). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using ML approaches. *Remote Sensing*, *12*(1), 21.
37. Di, Y., Gao, M., Feng, F., Li, Q., & Zhang, H. (2022). A new framework for winter wheat yield prediction integrating DL and Bayesian optimization. *Agronomy*, *12*(12), 3194.
38. Ariza-Sentís, M., Valente, J., Kooistra, L., Kramer, H., & Mücher, S. (2023). Estimation of spinach (Spinacia oleracea) seed yield with 2D UAV data and DL. *Smart Agricultural Technology*, *3*, 100129.
39. Apolo-Apolo, O. E., Pérez-Ruiz, M., Martinez-Guanter, J., & Valente, J. (2020). A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a DL technique. *Frontiers in plant science*, *11*, 1086.
40. Priyatikanto, R., Lu, Y., Dash, J., & Sheffield, J. (2023). Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation. *Agricultural and Forest Meteorology*, *341*, 109652.
41. Barbosa, B. D. S., Costa, L., Ampatzidis, Y., Vijayakumar, V., & dos Santos, L. M. (2021). UAV-based coffee yield prediction utilizing feature selection and DL. *Smart Agricultural Technology*, *1*, 100010.
42. Livieris, I. E., Dafnis, S. D., Papadopoulos, G. K., & Kalivas, D. P. (2020). A multiple-input neural network model for predicting cotton production quantity: A case study. *Algorithms*, *13*(11), 273.
43. Bai, D., Li, D., Zhao, C., Wang, Z., Shao, M., Guo, B., ... & Jin, X. (2022). Estimation of soybean yield parameters under lodging conditions using RGB information from unmanned aerial vehicles. *Frontiers in Plant Science*, *13*, 1012293.
44. Abbaszadeh, P., Gavahi, K., Alipour, A., Deb, P., & Moradkhani, H. (2022). Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction. *Agricultural and Forest Meteorology*, *314*, 108773.
45. Terliksiz, A. S., & Altýlar, D. T. (2019, July). Use of deep neural networks for crop yield prediction: A case study of soybean yield in Lauderdale County, Alabama, USA. In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp. 1-4). IEEE.
46. Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Tian, Z., & Guo, H. (2021). Exploiting hierarchical features for crop yield prediction based on 3-D convolutional neural networks and multikernel Gaussian process. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 4476-4489.

47. Tanabe, R., Matsui, T., & Tanaka, T. S. (2023). Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery. *Field Crops Research*, *291*, 108786.
48. Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Zhang, L., & Tian, Z. (2021). Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102436.
49. MacEachern, C. B., Esau, T. J., Schumann, A. W., Hennessy, P. J., & Zaman, Q. U. (2023). Detection of fruit maturity stage and yield estimation in wild blueberry using DL convolutional neural networks. *Smart Agricultural Technology*, *3*, 100099.
50. Lee, S., Jeong, Y., Son, S., & Lee, B. (2019). A self-predictable crop yield platform (SCYP) based on crop diseases using DL. *Sustainability*, *11*(13), 3637.