Agriculture Yield Prediction Using Edge Computing and Machine Learning

Prof. Karishma Misal[1], Aditya Nimankar[2], Prajakta Nakhate[3] , Divyank Khobragade[4] , Chetna Nagpure[5]
[1]*Professor, Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India*
[2,3,4,5]*Students, Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India*

Abstract— When it comes to international trade and food security, agriculture is vital. Yield prediction has changed drastically since the introduction of powerful computational methods. In this study, we survey all the new developments in agricultural output prediction that have made use of edge computing and machine learning (ML). The effectiveness of various ML techniques in yield forecasting is evaluated in this research. These algorithms include regression models, deep learning approaches, ensemble methods, and support vector machines. Also covered is the role of edge computing in improving precision agriculture decision-making through real-time data processing with reduced latency. It also delves into the integration of smart farming based on the Internet of Things, monitoring using drones, and decision support systems driven by artificial intelligence. Data integration, computational limitations, and scalability are still problems, despite the fact that machine learning and edge computing provide good solutions. Improving the quality of datasets, investigating federated learning methods, and optimizing ML models for edge devices are all areas that might use more investigation in the future. The purpose of this review is to offer helpful information on how to improve agricultural sustainability, food security, and productivity by using modern technology.

I. INTRODUCTION

Millions rely on agriculture for their food, jobs, and economic security, making it one of the most vital industries in the world. Optimizing agricultural production, however, has emerged as a major obstacle in the face of a rising world population and accompanying increase in food consumption. When it comes to farming, the old ways of doing things are mostly dependent on guesswork and past mistakes, which doesn't necessarily lead to the best use of resources. In order to effectively manage a farm, it is essential to precisely predict crop yield. This allows for better planning, allocation of resources,and mitigation of risks. Predictions in agriculture have been radically altered by the advent of ML and edge computing. Machine learning algorithms take into account variables like weather, soil fertility, crop type, and geographical impacts to produce highly accurate agricultural output forecasts using both historical and real-time data. At the same time, edge computing improves real-time processing by decreasing the dependency on centralized cloud infrastructure and moving computational capacity closer to data sources, such as sensors in the Internet of Things (IoT) and drones.

A number of advantages, such as increased efficiency, decreased waste, and better precision farming, accrue from the combination of ML with edge computing in the agricultural sector. These tools can help agronomists, legislators, and farmers by evaluating massive information and environmental variables in real time. Full potential, however, can only be achieved by overcoming obstacles like inconsistent data gathering,high computational requirements, and implementation costs. Predicting agricultural yields using current ML methods and edge computing applications is the goal of this article. Various prediction models are examined, their performance is compared, and the function of real-time data processing in improving agricultural productivity is addressed. Emphasizing the necessity for ongoing developments in smart agricultural technologies, it also addresses present problems and prospective future research initiatives in the industry.

II. LITERATURE SURVEY

[1] The authors Yewle et al. propose RicEns-Net, an intelligent deep ensemble model that integrates diverse data types—such as Sentinel satellite imagery, synthetic aperture radar, and meteorological data—for accurate crop yield prediction. The study effectively addresses the curse of dimensionality by reducing 100+ features to a more meaningful set of 15 across five modalities. The ensemble approach outperforms traditional methods by leveraging strengths of various machine learning algorithms. The model achieved a mean absolute error (MAE) of just 341 kg/Ha, making it significantly more accurate than previous models. This contributes to more reliable forecasting and informed decision-making in agriculture. However, challenges remain, such as the computational complexity of processing high-dimensional multimodal data and ensuring timely data availability. *Yewle, A. D., Mirzayeva, L., & Karakuş, O. (2025). Multi-modal Data Fusion and Deep Ensemble Learning for Accurate Crop Yield Prediction. arXiv preprint arXiv:2502.06062.*

[2] The authors Kamangir et al. introduce CMAViT, a multimodal vision transformer that integrates climate data, farm management practices, and remote sensing imagery to enhance grape yield prediction. Tested over 2,200 hectares with eight grape varieties, CMAViT achieved an R^2 of 0.84 and a MAPE of 8.22%, outperforming baseline models like UNet-ConvLSTM. Ablation studies confirmed that removing key modalities like management data significantly reduced model performance, proving the effectiveness of data integration. The model is especially valuable for fine-grained yield mapping and decision support in precision viticulture. Despite its success, the transformer model's dependency on large datasets and significant computational power could limit its adoption among small-scale farmers or low-resource regions. *Kamangir, H., Sams, B. S., Dokoozlian, N., Sanchez, L., & Earles, J. M. (2024). CMAViT: Integrating Climate, Management, and Remote Sensing Data for Crop Yield Estimation with Multimodal Vision Transformers. arXiv preprint arXiv:2411.16989.*

[3] The authors Kalmani et al. present a deep learning model that combines CNN-LSTM architecture with attention layers and skip connections to predict yields for crops like wheat and rice. This hybrid model captures both spatial and temporal dependencies and demonstrates superior performance over classical regressors such as Decision Tree and Random Forest. The inclusion of attention mechanisms enhances interpretability by focusing on the most influential features, while skip connections ensure better training convergence. With an RMSE of 0.017 and R^2 score of 0.967, the model sets a benchmark for accuracy. A notable advantage of this model is its ability to dynamically learn from seasonal patterns and environmental shifts. However, one downside is its training complexity, which may hinder real-time applications without substantial computational resources. *Kalmani, V. H., Dharwadkar, N. V., & Thapa, V. (2024). Crop Yield Prediction using Deep Learning Algorithm based on CNN-LSTM with Attention Layer and Skip Connection. Indian Journal of Agricultural Research.*

[4] The authors Lin et al. propose MMST-ViT, a multi-modal spatial-temporal vision transformer that factors in climate change impacts for crop yield prediction. This model integrates remote sensing data and climatic variables to analyze spatial-temporal crop dynamics. The use of transformer-based architecture allows the model to focus attention on relevant data points, thereby improving accuracy and generalization across seasons. MMST-ViT was tested on large-scale datasets and showed marked improvement over existing CNN-LSTM models. Its consideration of climate variability adds substantial real-world relevance. While the model demonstrates state-of-the-art performance, its dependence on large labeled datasets and high-performance computing infrastructure can be a barrier for implementation in developing regions. *Lin, F., Crawford, S., Guillot, K., Zhang, Y., Chen, Y., Yuan, X., & Tzeng, N.-F. (2023). MMST-ViT: Climate Change-Aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer. arXiv preprint arXiv:2309.09067.*

[5] The authors Manjunath and Palayyan developed a hybrid machine learning framework for predicting crop yields by combining Decision Tree, XGBoost, and Random Forest models. Their system was trained on open-source data from Kaggle and achieved a prediction accuracy of 98.6%, surpassing individual

model performances. The study also introduces a GUI-based tool named 'Crop Yield Predictor,' enabling easy access to predictive insights for non-technical stakeholders. This significantly improves accessibility and practical application. The hybrid nature of the model ensures robustness and better generalization. A notable advantage is the tool's real-world usability in helping farmers make informed choices. However, the paper does not elaborate on the impact of region-specific variables, which could affect model generalizability across different agricultural zones. *Manjunath, M. C., & Palayyan, B. P. (2023). An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model. Revue d'Intelligence Artificielle, 37(4), 491–498.*

[6] The authors Wang et al. propose a deep learning framework that merges Convolutional Neural Networks (CNN), Graph Attention Networks (GAT), and Long Short-Term Memory (LSTM) units to predict crop yield using spatial and temporal information. Applied to a soybean dataset across 1,115 U.S. counties, their model improved RMSE by 5% and R^2 by 6% over baseline methods. CNNs extract image-based features, GATs handle geospatial relations, and LSTMs manage historical data sequences. This triple-layered approach captures rich patterns across space and time. The study's main contribution lies in integrating topological information via GATs—rare in agriculture-focused DL. However, limitations include computational demands and a potential need for extensive preprocessing of geographic and temporal data. *Wang, L., Chen, Z., Liu, W., & Huang, H. (2024). A Temporal–Geospatial Deep Learning Framework for Crop Yield Prediction. Electronics, 13(21), 4273.*

[7] The authors Yuan et al. conducted a systematic literature review focusing on crop yield prediction using UAV remote sensing and machine learning. Reviewing 76 studies, they identified wheat, corn, rice, and soybeans as the most studied crops. Their analysis reveals that multispectral and RGB imaging dominate data sources, with Random Forest and CNNs frequently used for modeling. The review highlights both the growing importance of UAVs for real-time data collection and the benefits of multimodal fusion. A major strength of this study is its broad coverage, which offers a panoramic view of recent progress in this sub-field. However, it also points out major challenges such as dataset inconsistency, limited model generalization, and the high cost of UAV deployment. *Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain Crop Yield Prediction Using Machine Learning Based on UAV Remote Sensing: A Systematic Literature Review. Drones, 8(10), 559.*

[8] The author Pandya, P., & Gontia, N. K. explains the application of early crop yield prediction in drought-prone regions using a combination of meteorological drought indices, satellite-based indices, and machine learning techniques. Their study focuses on the Saurashtra region of Gujarat, India, using a 33-year dataset (1986–2019) to build prediction models for cotton and groundnut. They implement and compare Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Random Forest (RF), identifying RF as the most effective, with R^2 values between 0.77 and 0.92. The study emphasizes that indices like the Standardized Precipitation Evapotranspiration Index and Vegetation Condition Index are strongly correlated with crop yields. The positive impact of this research lies in its ability to provide accurate and timely drought monitoring, aiding farmers and policymakers in decision-making. However, a potential downside could be reliance on historical data and static models, which might not generalize well to unpredictable future climate changes. *Pandya, P., & Gontia, N. K. (2023). Early Crop Yield Prediction for Agricultural Drought Monitoring Using Drought Indices, Remote Sensing, and Machine Learning Techniques. Journal of Water and Climate Change, 14(12), 4729–4746.*

[9]The authors Bansal et al. explore crop yield prediction for winter wheat using machine learning models across multiple, varied datasets. Think of this study as trying to bake the same cake using ingredients from different kitchens—the data is diverse and inconsistent, but the goal is a reliable, consistent result. By comparing and validating models like Random Forest, XGBoost, and Support Vector Machines across datasets with differing formats and sources, they address the challenge of model generalizability. Their research demonstrates that certain models maintain accuracy even when applied to unfamiliar data environments, making them robust tools for real-world use. A key strength is the emphasis on

adaptability, making it suitable for practical deployment in countries with varying data quality. However, a downside is that results can still vary based on the dataset's quality and feature availability. *Bansal, Y., Lillis, D., & Kechadi, M. T. (2023). Winter wheat crop yield prediction on multiple heterogeneous datasets using machine learning. arXiv preprint arXiv:2306.11946.*

[10] The authors Nikhil et al. investigate machine learning-based crop yield prediction in South India, comparing models like Linear Regression, Decision Tree, Random Forest, and XGBoost. This study is like a talent show for algorithms, testing which model performs best under the specific environmental and agricultural conditions of the region. They found that Random Forest and XGBoost consistently outperformed others in predicting yields of crops like rice and maize. The model's strength lies in its ability to manage non-linear data and capture hidden patterns, like how an experienced farmer can "read" the signs of the land. The research provides actionable insights for regional policymakers and farmers. One limitation is the reliance on historical data, which may not reflect future climate changes or sudden shifts in agricultural practices. *Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in South India: Performance analysis of various models. Computers, 13(6), 137.*

III. MACHINE LEARNING IN AGRICULTURE YIELD PREDICTION

The agricultural sector has made extensive use of machine learning approaches for yield prediction based on both historical and real-time data. A few ML methods that are frequently employed are:

Classification Systems:
Linear regression is a statistical method for modeling the link between agricultural yield and variables including soil quality, temperature, and rainfall. One non-linear model that may accurately predict features based on their importance is decision tree regression, which uses datasets to divide them into smaller groupings. Random Forest Regression is an ensemble method that uses a number of decision trees together to lessen the effects of overfitting and increase accuracy.

Methods for Deep Learning:
Soil moisture and vegetation index patterns can be discovered using Convolutional Neural Networks (CNNs) applied to satellite data. One kind of RNN that can handle temporal dependencies in weather and agricultural growth data is the Long Short-Term Memory (LSTM) network.

Learning Techniques for Ensembles:
To improve the accuracy of yield predictions, XGBoost, Bagging Regressor, and Random Forest all combine numerous models.

Machines for Supporting Vectors (SVM):
A supervised learning technique that works well with minimal datasets and is used for prediction models based on classification in agricultural yield estimation. Research shows that when it comes to predicting agricultural yields, Random Forest and CNN outperform other models, with accuracy levels as high as 98.96%. To improve forecast accuracy, these models use hyperparameter tuning and feature selection methods.

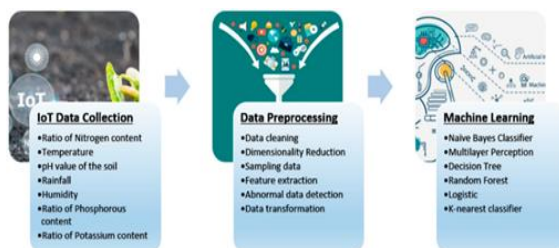
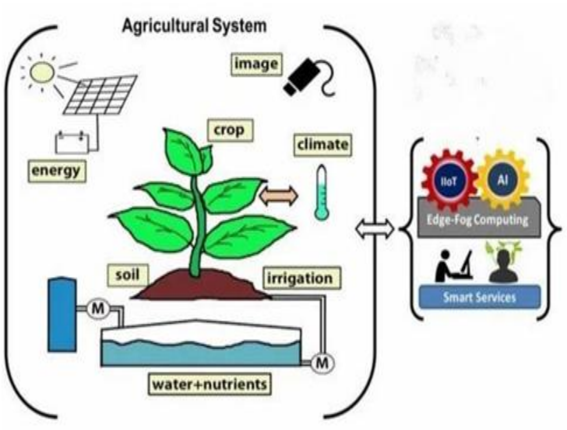


Fig. IoT and machine learning-based crop analysis and prediction process.

IV. EDGE COMPUTING IN AGRICULTURE

By moving processing of agricultural data closer to its point of origin, edge computing reduces latency and dependence on cloud computing, allowing for real-time processing. Here are a few important uses:
In Internet of Things (IoT) smart farming, sensors measure soil temperature, humidity, and wetness; this data is analyzed on edge devices to provide immediate insights, eliminating the need for constant internet connection. The use of drones and remote sensing technology to take high-resolution pictures and

process them on edge devices for the purpose of disease identification, insect infestation monitoring, and predicting crop yields.



Automated Decision Support Systems: Utilizing sensor data and insights driven by machine learning, edge devices powered by artificial intelligence offer real-time recommendations for watering, fertilizing, and pest control. Automated sowing, harvesting, and water management are just a few examples of how precision agriculture technologies are improving farming efficiency.

V. COMPARATIVE ANALYSIS OF TECHNIQUES

Various machine learning and edge computing models have been tested for their ability to forecast yield using metrics including accuracy, MAE, and RMSE. According to research:

The best accuracy (98.96%) and lowest MAE (1.97), respectively, are achieved by Random Forest.

When it comes to deep learning approaches, CNN is superior to LSTM, and its lower test loss of 0.00060 makes it a good fit for predictions based on satellite images.

With edge computing, data transmission times to centralised cloud servers are decreased, enhancing real- time prediction capabilities while also reducing bandwidth consumption.

Because of its interpretability, Decision Trees are better suited to real-world agricultural applications that call for explanations of model projections.

VI. INTEGRATION OF EDGE COMPUTING, SENSORS WITH MACHINE LEARNING MODEL

The integration of edge computing, IoT sensors, and machine learning has revolutionized precision agriculture by enabling real-time data analysis and decision-making. Traditional cloud-based approaches often suffer from high latency, bandwidth limitations, and dependency on stable internet connectivity, which poses challenges in remote farming regions. To address these issues, edge computing allows data to be processed locally on edge devices, eliminating the need for continuous cloud access while ensuring low latency and efficient decision-making.

IoT sensors play a crucial role in collecting real-time data on soil moisture, temperature, humidity, and nutrient levels. Additionally, aerial imagery captured by drones or mobile devices enhances crop health monitoring. These data streams are processed on-site by edge devices such as Raspberry Pi, or microcontrollers, reducing the need for centralized cloud processing. The localized processing ensures that farmers receive immediate insights into critical agricultural parameters, facilitating timely interventions for irrigation, fertilization, and pest control.

Machine learning models are deployed on edge devices to analyze sensor data and predict crop yield. Convolutional Neural Networks (CNNs) are widely used for image-based crop health analysis, while decision trees and random forests help in classifying and predicting yield based on historical and real-time data. Support Vector Machines (SVM) assist in detecting crop diseases and soil conditions. By running these AI models directly on edge devices, the system provides real-time recommendations, helping farmers optimize their resources efficiently.

The integration of these technologies offers several advantages, including reduced latency, lower dependency on cloud infrastructure, enhanced security by keeping data localized, and improved sustainability by minimizing resource wastage. However, challenges such as hardware limitations of edge devices, the need for lightweight AI models, and scalability issues still need to be addressed. Future advancements in federated learning, 5G-enabled edge AI, and autonomous farming solutions can further enhance the efficiency and adaptability of this technology in modern agriculture. The combination of edge computing, IoT sensors, and AI-driven analytics marks a significant step toward smart and sustainable farming, providing farmers with real-time, data-driven insights to improve productivity and crop management.

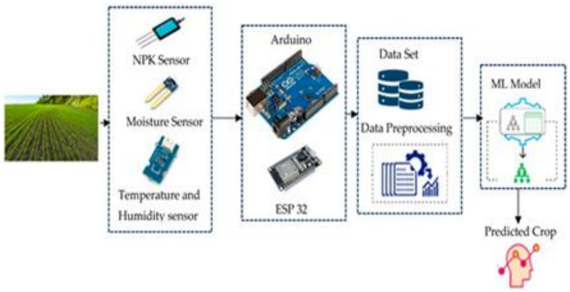


Fig. Integration of Edge Computing, Sensors with Machine Learning Model.

VII. DIFFERENCE BETWEEN PREVIOUS STUDIES AND OUR STUDY

Previous investigations into agricultural yield forecasting have predominantly depended on cloud-based computing for data processing, which presents challenges such as significant latency, reliance on stable internet connectivity, and heightened operational expenditures. These models frequently necessitate that farmers transfer substantial datasets to centralized servers, resulting in delays in obtaining actionable insights. Furthermore, conventional methodologies primarily utilize remote sensing data or static datasets, rather than employing real-time sensor-based analytics.

In contrast, our research amalgamates edge computing, Internet of Things (IoT) sensors, and machine learning to facilitate on-site, real-time data processing. Unlike models that rely on cloud infrastructure, our methodology processes sensor data locally through edge devices such as Raspberry Pi and Jetson Nano, thereby ensuring minimal latency and continuous operational capacity even in remote regions characterized by inadequate internet connectivity. Additionally, prior precision farming and deliberates on the function of edge computing in the optimization of decision-making studies predominantly emphasize singular data source inputs such as satellite imagery. our proposed framework synergistically amalgamates sensor-derived metrics, encompassing soil moisture, temperature, humidity, and nutrient concentrations, with aerial imagery and artificial intelligence-enhanced analytical techniques to enable superior precision and adaptability in yield forecasting.

Moreover, while earlier studies primarily employed basic machine learning techniques, our investigation integrates advanced deep learning frameworks, notably Convolutional Neural Networks (CNNs) for image- based assessments of crop vitality and Decision Trees for accurate yield predictions. This integrated approach significantly enhances both precision and the efficacy of decision-making processes. Furthermore, the safeguarding of data confidentiality and security is substantially reinforced within our framework, as sensitive agricultural information is confined to localized edge networks, thus reducing vulnerability to cyber threats in comparison to cloud-centric models.

VIII. CONFLICTS OF INTEREST

A conflict of interest in the domain of agricultural yield forecasting through machine learning methodologies may emerge when a researcher or entity engaged in the development of such predictive models possesses a financial interest in a particular agricultural input (such as fertilizers, seeds, or pesticides), thereby potentially biasing the model to favor forecasts that serve to enhance the profitability of their product, even if such bias compromises the overall precision for the broader agricultural community. Key areas where conflicts could occur: Data selection: Should a researcher selectively utilize data that exclusively supports the product they are advocating, it may distort the model's predictions towards endorsing that particular product, notwithstanding the possibility that alternative options could be more advantageous under specific circumstances. Feature engineering: The process of determining which variables to incorporate into the model may be manipulated to accentuate the advantages of a designated product, while concurrently minimizing the significance of other variables that could affect yield outcomes.

Model evaluation metrics: The selection of particular metrics for assessing model efficacy that align with preferred outcomes may result in the neglect of other critical dimensions of predictive accuracy.

Funding sources: In instances where research is financed by an enterprise with a significant stake in certain agricultural commodities, there may be an inclination to modify the research outcomes to bolster the interests of those products.

Examples of potential conflicts: A seed corporation that is engaged in the development of a yield prediction model may emphasize its proprietary seed varieties in the output, potentially minimizing the performance evaluation of competing seed options. A fertilizer corporation could devise a model that advocates for excessive fertilizer application rates, ostensibly to enhance their sales, notwithstanding potential long-term detrimental effects on soil health. A research institution financed by a pesticide manufacturer may concentrate its yield prediction investigations on the efficacy of the sponsor's pesticides, potentially overlooking the advantages conferred by integrated pest management methodologies.

How to mitigate conflicts of interest: Transparency: Unambiguously revealing any prospective conflicts of interest in academic publications and presentations pertaining to research. Independent data sources: Employing publicly accessible datasets or data from a variety of sources to guarantee that the model remains impartial and is not skewed towards particular products. Peer review process: A comprehensive and rigorous peer review process can assist in the detection of potential biases inherent in research methodologies and resultant conclusions. Ethical guidelines: Complying with established ethical frameworks for research endeavors in the fields of agriculture and machine learning.

IX. CONCLUSION

The amalgamation of edge computing, Internet of Things (IoT) sensors, and machine learning algorithms possesses the capacity to revolutionize precision agriculture by furnishing real-time, data-informed insights for agronomists. In contrast to conventional cloud-centric frameworks, our methodology capitalizes on localized data processing via edge devices, thereby diminishing latency, reliance on internet connectivity, and operational expenditures. Through the implementation of artificial intelligence models such as Convolutional Neural Networks (CNNs) and Decision Trees, our framework provides precise yield forecasts, optimized irrigation plans, and proactive pest identification, culminating in enhanced agricultural productivity and sustainability.

The research underscores the merits of edge artificial intelligence in the agricultural domain, encompassing expedited decision-making, improved resource allocation, and augmented data confidentiality. Nonetheless, obstacles such as hardware constraints, model refinement, and scalability warrant further investigation. Prospective developments in 5G technology, federated learning, and energy-efficient edge computing will be instrumental in enhancing the efficiency and accessibility of smart farming practices.

REFERENCE

[1] Yewle, A. D., Mirzayeva, L., & Karakuş, O. (2025). Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. *arXiv preprint arXiv:2502.06062*. https://arxiv.org/abs/2502.06062

[2] Kamangir, H., Sams, B. S., Dokoozlian, N., Sanchez, L., & Earles, J. M. (2024). CMAViT: Integrating climate, management, and remote sensing data for crop yield estimation with multimodal vision transformers. *arXiv preprint arXiv:2411.16989*. https://arxiv.org/abs/2411.16989

[3] Kalmani, V. H., Dharwadkar, N. V., & Thapa, V. (2024). Crop yield prediction using deep learning algorithm based on CNN-LSTM with attention layer and skip connection. *Indian Journal of Agricultural Research*. https://arccjournals.com/journal/indian-journal-of-agricultural-research/A-6300

[4] Lin, F., Crawford, S., Guillot, K., Zhang, Y., Chen, Y., Yuan, X., ... & Tzeng, N.-F. (2023). MMST-ViT: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer. *arXiv preprint arXiv:2309.09067*. https://arxiv.org/abs/2309.09067

[5] Manjunath, M. C., & Palayyan, B. P. (2023). An efficient crop yield prediction framework using hybrid machine learning model. *Revue d'Intelligence Artificielle*, *37*(4), 491–498. https://iieta.org/journals/ria/paper/10.18280/ria.370428

[6] Wang, L., Chen, Z., Liu, W., & Huang, H. (2024). A temporal–geospatial deep learning framework for crop yield prediction. *Electronics*, *13*(21),

4273. https://www.mdpi.com/2079-9292/13/21/4273

[7] Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review. *Drones*, *8*(10), 559. https://www.mdpi.com/2504-446X/8/10/559

[8] Pandya, P., & Gontia, N. K. (2023). Early crop yield prediction for agricultural drought monitoring using drought indices, remote sensing, and machine learning techniques. *Journal of Water and Climate Change*, *14*(12), 4729–4746. https://iwaponline.com/jwcc/article/14/12/4729/99202

[9] Bansal, Y., Lillis, D., & Kechadi, M. T. (2023). Winter wheat crop yield prediction on multiple heterogeneous datasets using machine learning. *arXiv preprint arXiv:2306.11946*. https://arxiv.org/abs/2306.11946

[10] Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in South India: Performance analysis of various models. *Computers*, *13*(6), 137. https://www.mdpi.com/2073-431X/13/6/13