DOI 10.5281/twist.10049652#348

TWIST, 2024, Vol. 19, Issue 4, pp. 395-402

TWIST

Journal homepage: www.twistjournal.net

Machine Learning Techniques for Crop Yield Prediction:

A Survey on Techniques, Applications, and Future Directions

Nagaveni B. Nimbal*
[1]- Research Scholar, AMC Engineering College, Bengaluru,
Affiliated to Visvesvaraya Technological University,
Belagavi-590018, Karnataka, India

[2]- Associate Professor, Department of CSE, K S School of Engineering & Management,
Affiliated to Visvesvaraya Technological University, Belagavi-590018, Karnataka, India
[**Corresponding author*]

Mareeswari V.
Professor and Head, Department of CSE, AMC Engineering College, Bengaluru, Affiliated to
Visvesvaraya Technological University, Belagavi-590018, Karnataka, India

Abstract
Accurate crop yield prediction is crucial for global food security and effective agricultural management. This paper reviews the latest machine learning (ML) techniques used in crop yield prediction, focusing on applications, challenges, and future prospects. Traditional methods based on historical data often fail to capture the complexities of environmental and management factors. Recent ML advancements have improved accuracy by using high-dimensional data like satellite imagery and weather patterns. Key challenges include data quality, model interpretability, generalizability, and computational demands. Solutions such as data sharing, interpretable models, transfer learning, and hybrid approaches are proposed to advance crop yield prediction and promote sustainable agriculture.

Keywords
Crop yield prediction, Machine learning, Deep learning, Sustainable agriculture, Precision agriculture

INTRODUCTION

Agriculture plays a vital role in the global economy, with crop production being a key driver of food security and economic growth. According to the Food and Agriculture Organization (FAO), the global population is expected to reach 9.7 billion by 2050, necessitating a 70% increase in food production to meet the growing demand [1]. However, predicting crop yields accurately remains a significant challenge due to the complex interplay of various factors such as weather conditions, soil health, and management practices.

Crop yield prediction estimates expected agricultural produce considering various factors. Accurate predictions are crucial for decision-making at multiple levels. Traditionally, crop yield prediction relied on empirical models based on historical data and expert knowledge. These methods often lacked the ability to capture the intricate relationships between variables and adapt to changing environmental conditions. In recent years, the rapid advancements in machine learning (ML) techniques have opened up new possibilities for more accurate and robust crop yield prediction. ML has emerged as a powerful tool, leveraging vast data to capture complex interactions and provide timely predictions. Challenges include handling high-dimensional, non-linear relationships in agricultural data. ML techniques can automatically learn representations, but often require large labeled datasets. Performance can be sensitive to data quality and representativeness.

Image processing plays a crucial role across various domains, including applications like facial recognition [2], medical imaging [3], and autonomous vehicles [4], by enabling the extraction and analysis of meaningful information from visual data. Furthermore, image processing has become an integral part of modern agricultural practices, enabling the extraction of valuable information from satellite imagery and UAV-captured data. By processing high-resolution images, ML models can identify patterns and features that are not easily visible through traditional methods, thus enhancing the accuracy of crop yield predictions. This approach allows for the real-time monitoring of crops, aiding in early detection of potential issues and optimizing resource allocation.

Accurate yield prediction has far-reaching implications for agriculture and beyond. It enables resource optimization, informs policy decisions, and supports climate change adaptation. Developing robust ML models can have spillover effects in related domains like precision agriculture and supply chain management.

The primary objective of this survey is to provide a comprehensive review of the state-of-the-art machine learning techniques used in crop yield prediction. We aim to highlight the recent advances and future directions in this field, focusing on the most effective ML algorithms for different crops and regions. By synthesizing the findings from various studies, we intend to provide valuable insights for researchers, practitioners, and policymakers working towards improving agricultural productivity and sustainability.

LITERATURE REVIEW

The block diagram exhibits the outline of all approaches and factors of the categories of ML based crop yield prediction.

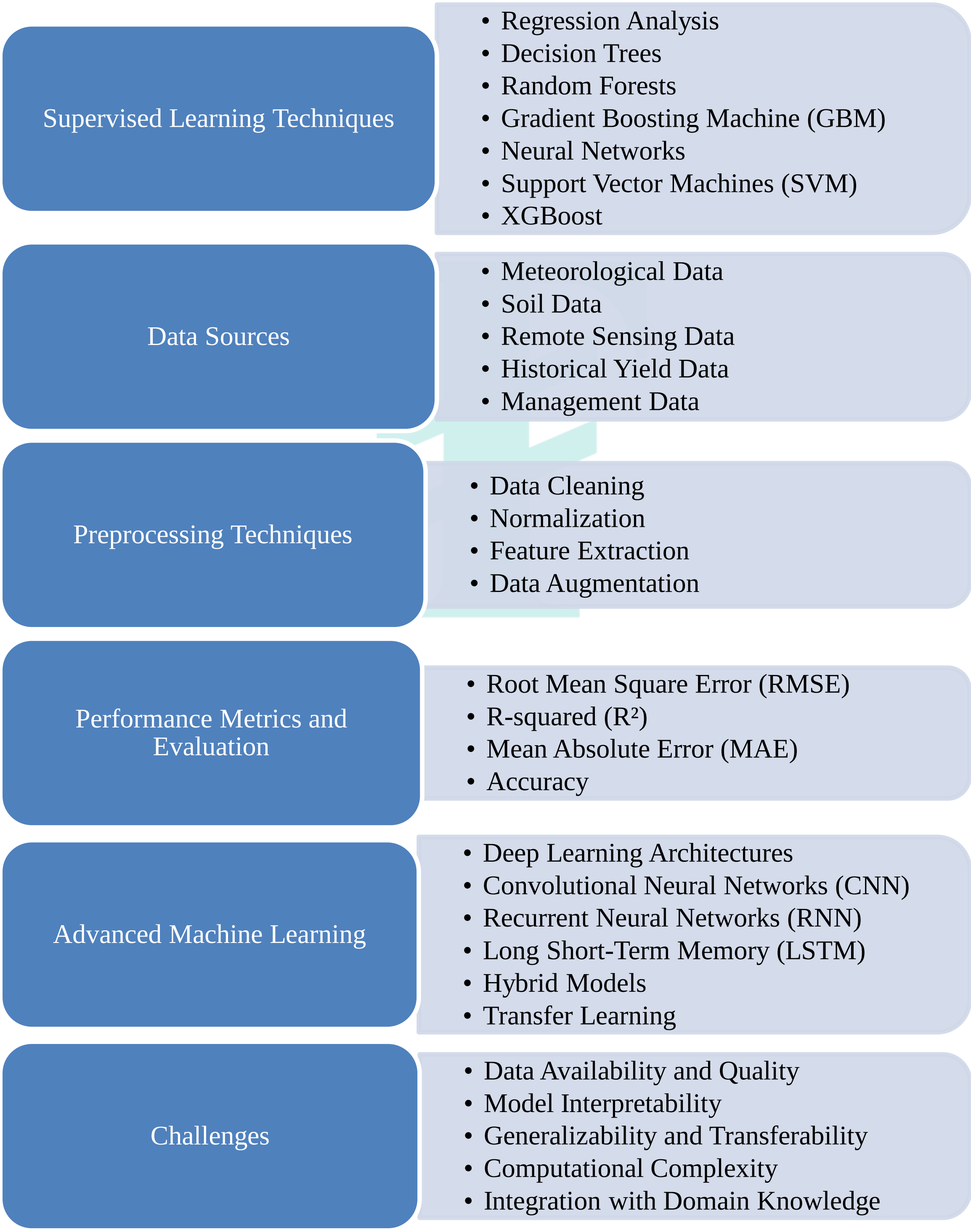

Fig. 1 Overview of Approaches and Factors in Crop Yield Prediction Using ML

Numerous studies have been conducted to predict crop yield using various machine learning and data mining techniques applied to time-series and satellite imagery datasets. Despite achieving different practical outcomes, prediction approaches continue to pose significant challenges. Some of the recent research papers in this area are reviewed as follows:

The work proposed in [1] highlights several key trends and challenges shaping the future of food and agriculture. It notes that global population growth, income increases in low- and middle-income countries, and structural shifts in economies are driving changes in agricultural production and consumption patterns. This is leading to increased pressure on natural resources and raising concerns about food security, poverty, and sustainability. Additionally, the text discusses how climate change is disproportionately affecting food-insecure regions and how satisfying increased agricultural demands could exacerbate issues like greenhouse gas emissions and deforestation. Overall, the information provided paints a complex picture of the multifaceted challenges facing the future of food systems worldwide.

The work [5] explores the application of machine learning models to predict the yield of corn hybrids, addressing the challenges of evaluating all possible hybrid combinations in breeding programs. By leveraging various algorithms, including XGBoost, the study demonstrates that accurate yield predictions can be made for untested hybrids, thereby identifying high-yielding hybrids efficiently. This approach highlights the potential of machine learning in optimizing crop production, reducing both time and costs associated with traditional field evaluations.

The work [6] investigates the enhancement of rice yield predictions through machine learning techniques, emphasizing the significance of integrating phenological, climatic, and geographical data. The research reveals that machine learning methods, particularly Support Vector Machine (SVM), outperform traditional regression models by incorporating phenological variables, which significantly influence carbon allocation and yield accuracy. The findings suggest that a combined approach using diverse data sources can markedly improve prediction precision, underscoring the importance of detailed phenological and climatic factors in agricultural forecasting.

The work proposed in [7] explores the use of machine learning to enhance wheat yield predictions by integrating multi-source data, including climate, remote sensing, and soil information. It evaluates different time windows within the wheat growth period to determine their impact on prediction accuracy, revealing that Support Vector Machine (SVM), Gaussian Process Regression (GPR), and Random Forest (RF) models provide high accuracy, with RF showing the best performance. The research underscores the importance of time window selection and regional variability in yield prediction, suggesting that the developed framework and Google Earth Engine (GEE) platform can be broadly applied to improve yield forecasting for various crops globally.

The work proposed in [8] investigates maize yield prediction in China by integrating diverse data sources, including optical, fluorescence, thermal satellite data, and environmental variables. It demonstrates that ML and DL methods, particularly RF and XGBoost, outperform traditional regression techniques, with Solar-Induced Chlorophyll Fluorescence (SIF) and thermal metrics proving valuable. The research emphasizes the benefits of combining multi-spectral satellite data and environmental factors for more accurate and efficient yield predictions, highlighting the potential for these methodologies to be applied globally.

The work [9] explores the application of Long Short-Term Memory (LSTM) models for estimating county-level corn yields in the U.S. Corn Belt by integrating diverse data sources, including crop phenology, meteorology, and remote sensing. The research highlights the LSTM model's superior performance in capturing yield variations and managing extreme weather events compared to traditional methods like LASSO regression and RF. By accounting for complex, nonlinear relationships and cumulative effects, the LSTM model offers a robust framework for understanding and predicting crop yields under varying climate conditions, demonstrating its potential for global agricultural yield forecasting.

The work [10] evaluates the CNN-LSTM-Attention model's effectiveness in predicting maize, rice, and soybean yields in Northeast China, contrasting it with traditional models like RF, XGBoost, and CNN. The research demonstrates that the advanced deep-learning approach, integrating Convolutional Neural Networks, Long Short-Term Memory, and an attention mechanism, significantly enhances prediction accuracy by capturing complex spatial and temporal variations. The findings underscore the value of sophisticated models and vegetation indices, such as kNDVI, in improving agricultural predictions and optimizing food production strategies.

The work [11] develops a rice yield prediction model using the Least Squares Support Vector Machine (LSSVM) optimized by Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Grey Wolf Optimization (GWO) to enhance accuracy. The research highlights that incorporating multi-source data, including meteorological and phenotypic factors, improves model performance, with GWO-LSSVM achieving the highest accuracy. The study underscores the effectiveness of hybrid optimization algorithms in refining model predictions, suggesting that the GWO-LSSVM approach is particularly suitable for predicting rice yields in China.

The work [12] explores the use of machine learning for predicting Kharif season rice yields at the district level in India, utilizing a range of models including CatBoost, LightGBM, and Extremely Randomized Trees. The research, based on 20 years of climate, satellite, and yield data, achieved high accuracy with out-of-sample R^2 values up to 0.82. It also developed an interactive dashboard to visualize predictions and assess model performance across districts. The study highlights the potential for integrating machine learning into agricultural early warning systems, offering a benchmark for predictive accuracy and enhancing decision-making in crop management.

The work [13] investigates how integrating vegetation indices into wheat yield prediction models enhances accuracy. By employing a CNN to capture spatial features and a LSTM network to process temporal data, combined with a fully connected network (FCN) for final predictions, the research demonstrates that incorporating satellite-derived vegetation indices significantly improves yield forecasting. The approach underscores the value of combining multiple data sources for more precise agricultural predictions.

The work [14] evaluates the effectiveness of various drought indices in predicting rain-fed barley yields using the random forest algorithm. By comparing the standardized precipitation evapotranspiration index (SPEI), modified SPEI (M-SPEI), reconnaissance drought index (RDI), modified RDI (M-RDI), standardized precipitation index (SPI), and modified SPI (M-SPI), the research finds that the SPEI and M-SPEI indices generally provided the most accurate predictions across different stations. The study highlights the utility of these indices in enhancing agricultural drought assessment and yield forecasting.

The work [15] introduces a hybrid deep learning model that combines classification and regression techniques for rice yield prediction using drone-collected environmental data. By employing feature selection methods like Pearson correlation coefficients, SHAP, and RFECV, the model efficiently identifies key predictors and reduces training time. The proposed approach outperforms traditional deep learning models, achieving a root-mean-square error of 344.56 and an R-squared value of 0.64, with high accuracy in distinguishing yield levels. The study also highlights the potential for enhancing model performance through data augmentation techniques like generative adversarial networks.

The work [16] focuses on developing a 3D CNN model for predicting soybean yield at the plot level using UAV-based RGB imagery collected across multiple time points. The research evaluates the performance of 3D variants of VGG, ResNet, and DenseNet architectures, finding that DenseNet provides the best balance between accuracy and model complexity. It demonstrates that while higher spatiotemporal resolutions do not necessarily improve predictions, UAV imagery alone can sufficiently inform accurate yield estimates, highlighting the potential for integrating UAV technology into crop breeding and precision agriculture.

The work [17] explores various multivariate models, including ANN, PCA-ANN, and LASSO, for predicting soybean yield across eight districts in Uttarakhand, emphasizing the influence of weather conditions on crop development. By analyzing historical time series data and using both weighted and unweighted weather indices, the research identifies the PCA-ANN model as the most accurate for yield prediction. This finding underscores the importance of integrating advanced modeling techniques in agricultural forecasting to enhance decision-making processes related to crop management and policy formulation.

The work [18] leverages machine learning techniques to predict the impact of climate change on the yields of five staple crops in Ethiopia from 2021 to 2050. By integrating climate projections from multiple Global Climate Models (GCMs) and various soil indicators, the research identifies key climate variables influencing crop production. The study demonstrates the effectiveness of ensemble learning methods in improving prediction accuracy and provides insights into future crop yield trends under different climate scenarios, highlighting implications for agricultural planning and policy formulation in Ethiopia.

The work [19] explores the impact of climate change on maize yields in Ethiopia's southern central rift valley using the DSSAT model, focusing on adaptation strategies such as altered planting dates, nitrogen application, tillage, and mulching. The research finds that combining nitrogen use with mulching and adjusted planting dates offers the most promising adaptation strategy, effectively mitigating yield losses under varying climate scenarios. The study highlights the importance of integrating multiple agronomic practices to enhance maize production resilience in the face of climate change.

Data Sources and Preprocessing

Common data types include meteorological, soil, remote sensing, historical yield, and management data. Sources range from ground-based stations to satellites and agricultural surveys. Before feeding the raw data into ML models, several preprocessing steps are necessary to ensure data quality, consistency, and compatibility. Preprocessing steps include data cleaning, normalization, feature extraction and selection, and data augmentation. Ensuring data quality is crucial for the success of ML-based crop yield prediction models. However, several challenges need to be addressed: Challenges include data accuracy, consistency, sparsity, variability, integration, and privacy concerns.

Machine Learning Techniques for Crop Yield Prediction

Supervised learning-based machine learning techniques for crop yield prediction involve training models on labeled datasets where input features like soil properties, weather conditions, and crop type are associated with known yield outcomes. Techniques such as regression analysis, decision trees, random forests, and neural networks analyze these patterns to forecast future yields. By leveraging historical data, these models can accurately predict crop performance under varying conditions, enabling farmers to optimize agricultural practices and resource allocation. This approach enhances decision-making, supports precision farming, and can lead to improved productivity and sustainability in agriculture by anticipating and mitigating potential yield-reducing factors.

Table 1 presents a comparison of the performance of different supervised learning algorithms in terms of Root Mean Square Error (RMSE) on a common benchmark dataset for crop yield prediction (e.g., the Syngenta Crop Challenge dataset [20]). Lower RMSE values indicate better model performance. Among the listed algorithms, XGBoost achieves the lowest RMSE (0.0509), suggesting it has the highest accuracy in predicting crop yields. Both the Gradient Boosting Machine (GBM) and Neural Network models follow closely with an RMSE of 0.0520. Random Forest and Adaboost show slightly higher RMSE values of 0.0529 and 0.0578, respectively, while the Decision Tree model has the highest RMSE (0.0621), indicating the least accurate predictions among the compared models.

Table 1 Performance Comparison of Supervised Learning based ML Models [3]

Algorithm	RMSE
Decision Tree	0.0621
Adaboost	0.0578
GBM	0.0520
Random Forest	0.0529
Neural Network	0.0520
XGBoost	0.0509

Evaluating the performance of crop yield prediction models is crucial for assessing their reliability and usefulness in real-world applications. Common metrics include Accuracy, R-squared, Root Mean Square Error, and Mean Absolute Error.

Rice Yield Prediction

Supervised learning-based machine learning techniques for rice yield prediction involve training models on labeled datasets where input features such as soil properties, weather conditions, and rice variety are associated with known yield outcomes. Techniques like BPNN, RF, and SVM analyze these patterns to forecast future yields. By leveraging historical data, these models can accurately predict rice performance under varying conditions, enabling farmers to optimize agricultural practices and resource allocation. This approach enhances decision-making, supports precision farming, and can lead to improved productivity and sustainability in rice cultivation by anticipating and mitigating potential yield-reducing factors.

Table 2 presents a comparison of the performance of different machine learning algorithms in terms of RMSE and R^2 on a common benchmark dataset for rice yield prediction. Lower RMSE values indicate better model performance, while higher R^2 values indicate better explanatory power of the model. Among the listed algorithms, SVM achieves the lowest RMSE (737) and the highest R^2 (0.33), suggesting it has the highest accuracy and explanatory power in predicting rice yields. RF follows with an RMSE of 744 and an R^2 of 0.31. The BPNN model shows the highest RMSE (800) and the lowest R^2 (0.24), indicating the least accurate predictions and explanatory power among the compared models.

Table 2 Performance Comparison of ML Models for Rice Yield Prediction [6]

Algorithm	RMSE	R^2
BPNN	800	0.24
RF	744	0.31
SVM	737	0.33

Wheat Yield Prediction

Supervised learning-based machine learning techniques for wheat yield prediction involve training models on labeled datasets where input features such as soil properties, weather conditions, and wheat variety are associated with known yield outcomes. Techniques like Gaussian Process Regression (GPR), RF, and SVM analyze these patterns to forecast future yields. By leveraging historical data, these models can accurately predict wheat performance under varying conditions, enabling farmers to optimize agricultural practices and resource allocation. This approach enhances decision-making, supports precision farming, and can lead to improved productivity and sustainability in wheat cultivation by anticipating and mitigating potential yield-reducing factors.

Table 3 presents a comparison of the performance of different machine learning algorithms in terms of R^2 on a common benchmark dataset for wheat yield prediction. Higher R^2 values indicate better explanatory power of the model. Among the listed algorithms, RF achieves the highest R^2 (0.81), suggesting it has the best explanatory power in predicting wheat yields. GPR follows with an R^2 of 0.79. The SVM model shows the lowest R^2 (0.77), indicating slightly less explanatory power among the compared models.

Table 3 Performance Comparison of ML Models for Wheat Yield Prediction [7]

Algorithm	R^2
GPR	0.79
RF	0.81
SVM	0.77

Maize Yield Prediction

Supervised learning-based machine learning techniques for maize yield prediction involve training models on labeled datasets where input features such as soil properties, weather conditions, and maize variety are associated with known yield outcomes. Techniques like RF, XGBoost, and LASSO regression analyze these patterns to forecast future yields. By leveraging historical data, these models can accurately predict maize performance under varying conditions, enabling farmers to optimize agricultural practices and resource allocation. This approach enhances decision-making, supports precision farming, and can lead to improved productivity and sustainability in maize cultivation by anticipating and mitigating potential yield-reducing factors.

Table 4 presents a comparison of the performance of different machine learning algorithms in terms of R^2 on a common benchmark dataset for maize yield prediction. Higher R^2 values indicate better explanatory power of the model. Among the listed algorithms, XGBoost achieves the highest R^2 (0.77), suggesting it has the best explanatory power in predicting maize yields. RF follows with an R^2 of 0.75. The LASSO regression model shows the lowest R^2 (0.39), indicating significantly less explanatory power among the compared models.

Table 4 Performance Comparison of ML Models for Maize Yield Prediction [8]

Algorithm	R^2
RF	0.75
XGBoost	0.77
LASSO	0.39

OBSERVATIONS AND FINDINGS

The comprehensive literature review highlights significant progress in using ML techniques for crop yield prediction, revealing both advancements and ongoing challenges. A key observation is the increased accuracy of predictions due to ML's ability to process high-dimensional and unstructured data, such as satellite imagery and weather sequences. Models like XGBoost and Random Forest have consistently outperformed traditional regression methods in predicting yields across various crops, including maize, rice, and wheat. For instance, integrating diverse data sources, such as optical, fluorescence, and thermal satellite data, has improved prediction accuracy for maize, while incorporating phenological, climatic, and geographical data has enhanced rice yield predictions with Support Vector Machines.

Despite these advancements, challenges persist, particularly regarding data availability and quality. There is a scarcity of high-quality, consistent data on crop yields, weather, soil, and management practices, which limits the generalizability and transferability of ML models to new regions or crops. To address these issues, researchers advocate for collaborative data sharing and the standardization of data collection methods, such as remote sensing and crowdsourcing. Additionally, model interpretability and transparency remain significant hurdles, as the complexity of ML models can make it difficult for users to understand and trust the predictions. Developing interpretable and explainable models, such as decision trees and rule-based systems, is crucial for enhancing user confidence and facilitating the integration of ML predictions into practical agricultural decision-making.

The review also emphasizes the computational complexity and resource requirements of ML models, which can be prohibitive for small-scale farmers or regions with limited access to advanced technology. Solutions such as cloud computing, distributed learning frameworks, and the development of efficient algorithms are proposed to overcome these barriers. Furthermore, integrating domain knowledge with ML techniques is essential, as hybrid models that combine ML with crop growth models and expert systems can enhance prediction accuracy while incorporating biophysical and socioeconomic constraints. These interdisciplinary collaborations are vital for developing robust, actionable insights that drive sustainable agricultural practices and improve global food security. In summary, while ML techniques have substantially advanced crop yield prediction, addressing challenges related to data quality, model interpretability, and computational demands is essential for realizing their full potential in agricultural applications.

Table 3 summarizes the main challenges and limitations of machine learning for crop yield prediction, along with potential solutions and research directions.

Table 3 Main challenges and limitations of machine learning for crop yield prediction

Challenges	Description	Potential Solutions
Data availability and quality	Lack of high-quality and consistent data on crop yields, weather, soil, and management practices	- Collaborative data sharing and standardization- Low-cost and high-resolution data collection methods (e.g., remote sensing, crowdsourcing)
Model interpretability and transparency	Difficulty in interpreting and explaining complex machine learning models	- Development of interpretable and explainable models (e.g., decision trees, rule-based systems) Visualization and communication of model results
Generalizability and transferability	Limited transferability of models to new regions, crops, or management practices	- Transfer learning and domain adaptation techniques -Incorporation of biophysical and socioeconomic constraints into models
Computational complexity and resource requirements	High computational and resource requirements for training and deploying complex models	- Development of efficient and scalable models and algorithms - Cloud computing and distributed learning frameworks
Integration with domain knowledge	Lack of integration with biophysical and socioeconomic processes that influence crop yields	- Hybrid modeling approaches that combine machine learning with crop growth models and expert systems - Interdisciplinary collaboration between machine learning experts and domain specialists

POTENTIAL SOLUTIONS AND FUTURE DIRECTIONS

To address the challenges and limitations of machine learning for crop yield prediction, several potential solutions and future research directions have been proposed:

- The integration of Internet of Things (IoT) sensors and real-time data analytics can enable the collection and processing of high-resolution and high-frequency data on crop growth, weather, and management practices
- The development of advanced deep learning techniques, such as attention mechanisms, graph neural networks, and generative models, can enhance the ability of machine learning models to capture complex spatial and temporal dependencies in crop yield data.
- The use of blockchain technology can enable secure and transparent data sharing and traceability in agricultural supply chains

CONCLUSION

This study comprehensively reviewed state-of-the-art ML techniques for crop yield prediction. ML models, especially deep learning, show promising results across crops and regions. Performance depends on various factors including data quality and model architecture. Hybrid approaches combining ML with crop growth models can provide more accurate predictions but increase complexity. Challenges include data availability, model interpretability, generalizability, computational requirements, and domain knowledge integration. Potential solutions involve data sharing, low-cost collection methods, interpretable models, transfer learning, hybrid approaches, and interdisciplinary collaboration. The key is to develop machine learning models that are not only accurate and reliable, but also relevant and actionable for the end-users. This survey will be helpful in stimulating further research, innovation, and collaboration in this area, contributing to the development of more productive, sustainable, and resilient agricultural systems around the world. This will support in several practical implications including improved agricultural planning and risk management, enhanced food security and sustainability, and empowerment of small-scale farmers.

ACKNOWLEDGEMENT

I would like to thank (Dr. Mareeswari V) for the guidance and support provided. I am also grateful to my colleagues for their feedback and to my family for their encouragement throughout this research.

DECLARATION OF CONFLICT

The authors report there are no competing interests to declare.

REFERENCES

1. FAO (2017). The future of food and agriculture – Trends and challenges. Rome: Food and Agriculture Organization of the United Nations. Retrieved from http://www.fao.org/3/i6583e/i6583e.pdf
2. Mareeswari, V., Patil, S. S., Lingraj, & Upadhyaya, P. (2020). A Novel Approach to Identify Facial Expression Using CNN. *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE*, 323-331.
3. Nasir, N., Afreen, N., Patel, R., Kaur, S., & Sameer, M. (2021). A Transfer Learning Approach for Diabetic Retinopathy and Diabetic Macular Edema Severity Grading. *Rev. d'Intelligence Artif.*, *35*(6), 497-502.
4. Chen, C., Wang, C., Liu, B., He, C., Cong, L., & Wan, S. (2023). Edge intelligence empowered vehicle detection and image segmentation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13023-13034.
5. Sarijaloo, F. B., Porta, M., Taslimi, B., & Pardalos, P. M. (2021). Yield performance estimation of corn hybrids using machine learning algorithms. *Artificial Intelligence in Agriculture*, *5*, 82-89.
6. Y. Guo, Y. Fu, F. Hao, X. Zhang, W. Wu, X. Jin, C. R. Bryant, and J. Senthilnath, ''Integrated phenology and climate in rice yields prediction using machine learning methods,'' Ecol. Indicators, vol. 120, Jan. 2021, Art. no. 106935.
7. Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sensing*, *12*(2), 236.
8. 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 machine learning approaches. *Remote Sensing*, *12*(1), 21.
9. Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., Huang, J., & Li, Z. (2020). A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. Global Change Biology, 26(3), 1754-1766. https://doi.org/10.1111/gcb.14885
10. Lu J, Li J, Fu H, Tang X, Liu Z, Chen H, Sun Y, Ning X. Deep Learning for Multi-Source Data-Driven Crop Yield Prediction in Northeast China. *Agriculture*. 2024; 14(6):794. https://doi.org/10.3390/agriculture14060794
11. Zhao, L., Qing, S., Wang, F. *et al.* Prediction of Rice Yield Based on Multi-Source Data and Hybrid LSSVM Algorithms in China. *Int. J. Plant Prod.* **17**, 693–713 (2023). https://doi.org/10.1007/s42106-023-00266-z
12. Djavan De Clercq., Adam Mahdi., (2024). Feasibility of machine learning-based rice yield prediction in India at the district level. arXiv preprint arXiv:2403.07967v1.
13. Patel, P., Shah, M., Raval, M.S., Chaudhary, S., Parmar, H. (2023). Impact of Vegetation Indices on Wheat Yield Prediction Using Spatio-Temporal Modeling. In: Chaudhary, S., Biradar, C.M., Divakaran, S., Raval, M.S. (eds) Digital Ecosystem for Innovation in Agriculture. Studies in Big Data, vol 121. Springer, Singapore.
14. Zarei, A.R., Mahmoudi, M.R. & Moghimi, M.M. Determining the most appropriate drought index using the random forest algorithm with an emphasis on agricultural drought. *Nat Hazards* **115**, 923–946 (2023). https://doi.org/10.1007/s11069-022-05579-2

15. Chang C-H, Lin J, Chang J-W, Huang Y-S, Lai M-H, Chang Y-J. Hybrid Deep Neural Networks with Multi-Tasking for Rice Yield Prediction Using Remote Sensing Data. *Agriculture*. 2024; 14(4):513. https://doi.org/10.3390/agriculture14040513
16. Bhadra, S., Sagan, V., Skobalski, J. *et al.* End-to-end 3D CNN for plot-scale soybean yield prediction using multitemporal UAV-based RGB images. *Precision Agric* **25**, 834–864 (2024)
17. Khan, Y., Kumar, V., Setiya, P. *et al.* Forecasting soybean yield: a comparative study of neural networks, principal component analysis and penalized regression models using weather variables. *Theor Appl Climatol* **155**, 2937–2952 (2024). https://doi.org/10.1007/s00704-023-04767-9
18. Xu N, Li F D, Zhang Q Y, Ai Z P, Leng P F, Shu W, Tian C, Li Z, Chen G, Qiao Y F. (2024). Crop yield prediction in Ethiopia based on machine learning under future climate scenarios. Chinese Journal of Eco-Agriculture, 32(3): 490−504. DOI: 10.12357/cjea.20230257
19. Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., & Asseng, S. (2023). Exploring adaptation responses of maize to climate change scenarios in South Africa using projections of 20 coupled ensemble climate models. Nature Scientific Reports, 23(39795). Available at: https://www.nature.com/articles/s41598-023-39795-y.pdf
20. Syngenta Crop Challenge (2021). Syngenta Crop Challenge in Analytics. Retrieved from https://www.ideaconnection.com/syngenta-crop-challenge/