meteorological series, and satellite indicators enables early-season forecasts that outperform conventional statistical baselines and even some official forecasts. County-level yield prediction in the U.S. Midwest has shown that XGBoost models using hundreds of environmental features can provide reliable maize forecasts several months before harvest, improving on models based only on basic weather or historical yields. Reviews of precision agriculture emphasize that such predictive systems contribute to resource optimization, risk management, and food-security planning by linking sensing technologies, big data platforms, and advanced analytics into operational decision support tools.

Despite these advances, several limitations constrain the reliability and transferability of current soil-climate yield models. Studies comparing algorithms against simple baselines show that, under realistic forecasting setups using ordered train-test splits, ML models sometimes offer only modest gains over farm-level average yields, especially when weather forecast errors are ignored. Systematic reviews also highlight persistent challenges with obtaining high-quality, harmonized datasets on soil nutrients, management, and high-resolution yields, which can limit model generalization across regions and seasons. In addition, many models are trained and validated under random data partitioning, leading to over-optimistic performance estimates for true out-of-sample prediction. Future research directions point toward hybrid, transferable, and explainable frameworks. Hybrid models that couple process-based crop simulators with ML or deep learning have improved accuracy and reduced uncertainty in semi-arid maize systems, particularly when fusing remote sensing, climate, and soil information. Domain adaptation and transfer-learning approaches, including partial adversarial networks, are beginning to address domain shifts between ecological zones and could substantially improve cross-regional maize yield prediction. Reviews stress the need for standardized data protocols, interpretable architectures (e.g., SHAP- or XAI-enhanced models), and scalable, crop-agnostic pipelines so that soil nutrient and climate-based yield prediction can be robustly embedded in precision agriculture and sustainability strategies.

Acknowledgments

We would like to thank the anonymous reviewers for their detailed review of the draft. Their specific feedback helped us correct the logical loopholes in our arguments.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdel-Salam M., Kumar N., and Mahajan S., 2024, A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning, Neural Computing and Applications, 36: 20723-20750.
https://doi.org/10.1007/s00521-024-10226-x

Aghighi H., Azadbakht M., Ashourloo D., Shahrabi H., and Radiom S., 2018, Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11: 4563-4577.
https://doi.org/10.1109/jstars.2018.2823361

Ahmed Z., Krupnik T., Timsina J., Islam S., Hossain K., Kurishi A., Emran S., Harun-Ar-Rashid M., McDonald A., and Gathala M., 2024, Prediction of spatial heterogeneity in nutrient-limited sub-tropical maize yield: implications for precision management in the eastern indo-gangetic plains, Artificial Intelligence in Agriculture, 12: 1-15.
https://doi.org/10.1016/j.aiia.2024.08.001

Archontoulis S., Castellano M., Licht M., Nichols V., Baum M., Huber I., Martinez-Feria R., Puntel L., Ordóñez R., Iqbal J., Wright E., Dietzel R., Helmers M., Vanloocke A., Liebman M., Hatfield J., Herzmann D., Córdova S., Edmonds P., Togliatti K., Kessler A., Danalatos G., Pasley H., Pederson C., and Lamkey K., 2020, Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt, Crop Science, 60: 721-738.
https://doi.org/10.1002/csc2.20039

Asamoah E., Heuvelink G., Chairi I., Bindraban P., and Logah V., 2024, Random forest machine learning for maize yield and agronomic efficiency prediction in Ghana, Heliyon, 10: e37065.
https://doi.org/10.1016/j.heliyon.2024.e37065

Bischl B., Binder M., Lang M., Pielok T., Richter J., Coors S., Thomas J., Ullmann T., Becker M., Boulesteix A., Deng D., and Lindauer M., 2021, Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(2): e1484.
https://doi.org/10.1002/widm.1484