National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
Journal of Advanced Research in Embedded System
Volume 13,Issue 1&2, 2026

Review Article

Crop Yield Prediction: A Comprehensive Review of Machine Learning and Deep Learning Approaches

Mansi Gera[1], Vipul Sharma[2]

[1] Assistant Professor, Department of Computer Science, I.K. Gujral Punjab Technical University (IKGPTU), Kapurthala, India
[2] Assistant Professor, Department of Computer Science, IKGPTU Amritsar Campus, India
DOI: https://doi.org/10.24321/2395.3802.202604

I N F O

Corresponding Author:
Mansi Gera, Department of Computer Science, I.K. Gujral Punjab Technical University (IKGPTU), Kapurthala, India
E-mail Id:
mansi.gera91@gmail.com
Orcid Id:
https://orcid.org/0000-0001-5158-0920
How to cite this article:
Gera M, Sharma V. Crop Yield Prediction: A Comprehensive Review of Machine Learning and Deep Learning Approaches . *J Adv Res Embed Sys*. 2026;13(1&2); 13-17

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-29

A B S T R A C T

Predicting crop yields accurately is essential for farm management, policymaking, and food security. With an emphasis on applications in India, this paper examines current developments in machine learning (ML) and deep learning (DL) techniques for crop yield estimation worldwide. The current study examines different obstacles (data scarcity, model transferability, interpretability), input variables (weather, soil, satellite indices), model types (regression, tree ensembles, neural networks, hybrid architectures), and future possibilities (transformers, multimodal fusion, IoT). In addition, this paper points out gaps and suggests recommended practices for further study.

Keywords: Crop Yield Prediction, Machine Learning, Deep Learning, Ndvi, Weather Data, Soil Parameters, Punjab Agriculture

Introduction

In order to allocate inputs, forecast prices, mitigate risks, and ensure food security, governments and farmers benefit from estimating crop output prior to harvest. Although traditional statistical and mechanistic crop models have been around for a while, their scalability is restricted, and they frequently require extensive agronomic data. ML and DL techniques have become more popular in yield prediction as satellite imaging, IoT sensors, and processing power have increased. These techniques have potential for enhancing conventional forecasting in India, a country with a high degree of agricultural diversification and data heterogeneity. This study examines current research on ML and DL techniques for yield prediction from 2015 to 2025, including viewpoints from India and throughout the world. This review synthesises crop yield prediction research using ML and DL methods from 2015 to 2025, emphasising the Indian agricultural context. Unlike earlier reviews[1,6,23] it highlights recent developments in transformer-based architectures, multimodal data fusion (RS, soil, weather), and explainable AI. The paper also identifies key gaps in data quality, model transferability, and reproducibility, providing guidelines for future research.

Input Variables and Datasets

Weather and Climate Variables

Temperature, precipitation, humidity, solar radiation, and growing degree days are standard inputs. Many studies incorporate seasonal aggregates or monthly time series.

Soil and Management Features

Soil organic carbon, nitrogen, phosphorus, pH, fertiliser application, pesticide use, irrigation status, field practices. These features help explain yield beyond just weather.

Remote Sensing Indices

Vegetation indices such as NDVI, EVI, SAVI, and LAI derived from satellite data (MODIS, Landsat, and Sentinel) are widely used to represent crop health. Remote sensing can capture spatial variability not captured in station data. For example, the review "Crop yield prediction using machine learning: A systematic literature review" uses remote sensing variables as common features.[1]

In Crop Yield Prediction Using Multi-Sensor Remote Sensing (Review), the authors emphasise combining multiple sensor data sources to enhance yield models.

Datasets & Geographic Scope

Public datasets include those based on MODIS, Sentinel, and regional agricultural statistics. In India, state-level datasets combining district production, rainfall, and soil are common. The review Advancements in remote sensing-based crop yield modelling in India survey such Indian-specific efforts.[22] The selection of these input variables is supported by prior studies linking environmental, soil, and remote sensing indicators to crop productivity. Weather variables such as temperature, rainfall, and humidity strongly influence plant growth stages and yield formation.[1,4,11] Soil attributes like nitrogen, phosphorus, and organic carbon directly affect nutrient availability and productivity.[21,23] Remote sensing indices (NDVI, EVI, LAI) provide spatially continuous information reflecting crop vigour and biomass.[6,12,24] Integrating these variables improves predictive performance across spatial and temporal scales.[5,10,19] In addition to MODIS and Sentinel data, several benchmark datasets have been introduced recently: the SustainBench Crop Yield Dataset.[18] USDA-NASS datasets for maize and soybean yield prediction[5,10] and India AgroData and PRADHAN datasets for Indian crops.[21,23]

These datasets enhance cross-region model comparison and reproducibility in crop yield prediction research.

Methodological Review

Traditional and Statistical Methods

Earlier approaches include linear regression, ridge, lasso, and time-series models (e.g., ARIMA). These handle baseline relationships but struggle with non-linear interactions and high dimensionality.

Machine Learning Models

Random Forest (RF) and Gradient Boosting Machines (GBM/ XGBoost / LightGBM) are among the most used models, due to robustness to nonlinearity, feature interactions, and missing values.[1,9] Support Vector Regression (SVR), k-Nearest Neighbours, and Bayesian methods are sometimes used for smaller or mid-sized datasets. Some hybrid models use ensemble stacking of RF, GBM, and linear models.

Deep Learning Models

Owing to the capacity to model temporal and spatial patterns:

- **LSTM / RNN:** For modelling sequential weather or NDVI time series. Studies show LSTM outperforms classical methods when long sequences are available.[6]
- **CNN / ConvLSTM / CNN–RNN hybrids:** Combine spatial (image) features from remote sensing with temporal modelling. Khaki et al. proposed a CNN–RNN framework showing strong performance on U.S. corn/ soy datasets.[3,10]
- **Transformer/Attention-based Models:** Emerging recently; some works apply ViT/attention to satellite imagery for yield estimation and denoising (e.g., "Quartile Clean Image" + ViT approaches).[24]
- **Hybrid / Feature-Selection Models:** Frameworks combining feature selection with ML (e.g., hybrid feature selection + optimised SVR) have been proposed for improved robustness.[16]

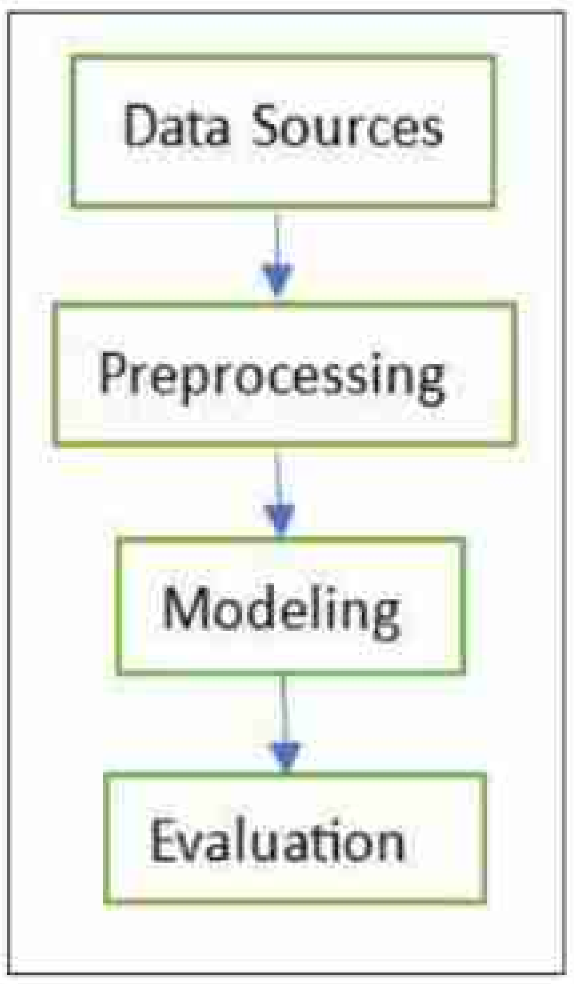

Figure 1.Framework of Crop Yield Predic-tion Using ML/DL

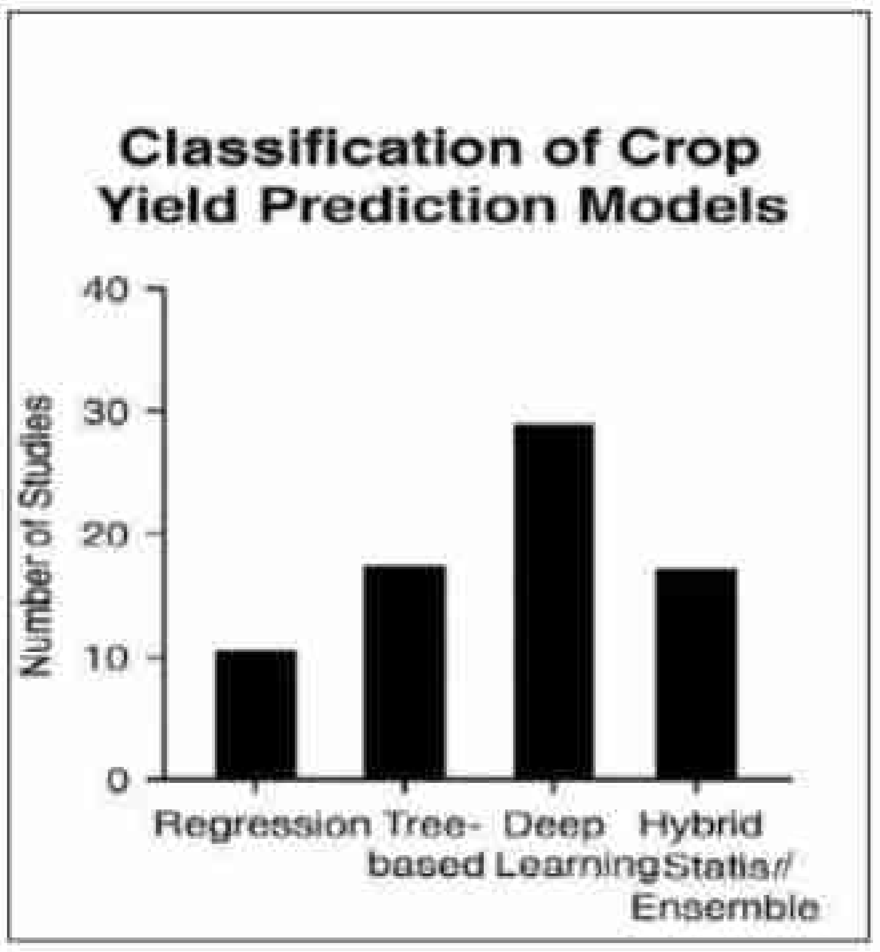

Figure 2.Classification of Crop Yield Predic-tion Models

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202604

Review Papers & Surveys

Several systematic reviews provide overviews of methods, inputs, and evaluation practices; notable reviews include van Klompenburg et al.[1] and Muruganantham et al.[6] Additional comprehensive reviews, such as Jabed et al. (2024),[2] Ilyas et al. (2023),[13] and Dey et al. (2025),[23] further analyse recent ML/DL frameworks. Compared with these, the present study emphasises the integration of Indian datasets, transformer-based architectures, and benchmark reproducibility, offering a broader synthesis of 2015–2025 literature.

Comparative Insights and Trends

Performance Comparisons

Multiple comparative studies show that tree ensembles (RF, GBM) outperform basic linear models on heterogeneous agricultural data; deep learning can outperform ML where ample data and temporal depth exist.[1,5,9,16]

Data & Training Challenges

- **Data scarcity and imbalance:** Many regions have limited yield records and missing values.[21]
- **Scale mismatch:** Field-plot vs district-aggregated data create scaling issues.
- **Generalisation/spatial transfer:** Models trained on one region may not generalise across climates and soils.[19]
- **Interpretability:** DL models are often black boxes; use of explainability (SHAP, permutation importance) is increasing but not universal.
- **Temporal mismatch:** Remote sensing cadence, growing-season windows and cloud/noise issues complicate modelling.

Indian Context

Indian studies commonly use district-level production, rainfall, soil, and NDVI; rice and wheat are frequent targets.[22,27] Here, combining management inputs (fertiliser, pesticide) with climatic variables improves interpretability and policy relevance.

Emerging Directions

Multimodal fusion (RS + weather + soil + management).[18]

Attention & transformer-based models for spatio-temporal dependencies.[24]

Federated/transfer learning to share knowledge across regions with privacy constraints.[5,19]

Edge/IoT integration and sensor-based real-time prediction.

Benchmarking & reproducibility: community benchmarks (e.g., SustainBench) are improving comparability.[18]

Recent works explore AI-driven ensemble and hybrid architectures for real-time yield estimation.[28,31,32] Transformer-based spatial–temporal fusion models have shown superior performance on heterogeneous agricultural datasets.[24,33]

Challenges, Gaps, and Best Practices

Key Challenges

Data quality, missingness, and inconsistent reporting; model generalisation and domain transfer, Explainability and trust for stakeholders, Lack of standardised benchmarks and reproducibility.

Best Practices

Use spatial-temporal cross-validation (grouped/time-split). Employ explainability (SHAP, PDP), Conduct ablation/sensitivity analyses, Release code and data for reproducibility. Consider hybrid/ensemble approaches for robustness.

Future Directions

Transformer and attention models for long-range spatio-temporal dependencies. Multi-scale modelling (field → district → region). Federated/transfer learning across crops and regions. Integration with IoT and near-real-time sensors (soil moisture, drones). Crop-specific explainable AI and decision support for extension services. Climate scenario modelling—predictive models under future climates. Community benchmarks and open datasets for fair comparisons.

Conclusion

This review summarises advances in ML and DL for crop yield prediction, with emphasis on global trends and India-specific use cases. Tree-based ensembles remain robust baselines; DL and hybrid models are promising when data permit. Addressing data quality, transferability, and explainability will be crucial for operational adoption. Future work should emphasise multimodal fusion, ethical deployment, and reproducibility.

References

1. T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review—2010 to 2020," Computers and Electronics in Agriculture, vol. 177, p. 105709, 2020. doi:10.1016/j.compag.2020.105709. https://www.sciencedirect.com/science/article/pii/S0168169920302301
2. M. A. Jabed, M. R. Hasan, and A. S. Islam, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches," Heliyon, vol. 10, p. e40836, 2024. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11667600/
3. S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN–RNN framework for crop yield prediction," arXiv:1911.09045,

2019. https://arxiv.org/abs/1911.09045

4. J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian Process for crop yield prediction based on remote sensing data," in Proc. AAAI Conf. on Artificial Intelligence, 2017, pp. 4559–4565. https://cs.stanford.edu/people/jiaxuan/files/Jiaxuan_AAAI17.pdf
5. A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in Proc. ACM COMPASS, 2018. doi:10.1145/3209811.3212707. https://dl.acm.org/doi/10.1145/3209811.3212707
6. P. Muruganantham, S. Ramesh, and S. R. Santhi, "A systematic literature review on crop yield prediction using deep learning," Remote Sensing, vol. 14, no. 9, p. 1990, 2022. doi:10.3390/rs14091990.https://www.mdpi.com/2072-4292/14/9/1990
7. M. Engen, O. Smith, and D. Adams, "Farm-scale crop yield prediction from multi-temporal satellite imagery," Agronomy, vol. 11, no. 12, p. 2576, 2021. doi:10.3390/agronomy11122576. https://www.mdpi.com/2073-4395/11/12/2576
8. Q. You, X. Li, J. Jia, D. Lobell, and S. Ermon, "A deep learning framework for crop yield prediction from remote sensing data," International Journal of Remote Sensing (related AAAI/Stanford work), 2017.https://cs.stanford.edu/people/jiaxuan/files/Jiaxuan_AAAI17.pdf
9. O. Ennaji, M. Benbrahim, and H. El Allati, "Gradient boosting for yield prediction of elite maize hybrid," PLOS ONE, vol. 19, no. 3, 2024. doi:10.1371/journal.pone.0315493. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0315493
10. S. Khaki, H. Pham, and L. Wang, "Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning," arXiv:2012.03129, 2020. https://arxiv.org/abs/2012.03129
11. D. Lobell, "Remote sensing of crop yields: current status and prospects" (representative works on RS & yields). Representative reading: D. Lobell et al., Nat. Climate Change, various publications. https://www.nature.com/search?author=David+Lobell
12. A. Mateo-Sanchis, N. Brisco, and M. I. Chuvieco, "Synergistic integration of optical and microwave satellite observations for improved crop monitoring and yield estimation,"Remote Sensing of Environment, 2019. https://www.sciencedirect.com/science/article/pii/S0034425719304791
13. S. Ilyas, N. Aslam, and M. Z. Ilyas, "Automated estimation of crop yield using artificial intelligence techniques," Sensors, vol. 23, no. 12, p. 5178, 2023. doi:10.3390/s23125178. https://www.mdpi.com/1424-8220/23/12/5178
14. M. Thiruvengadam and S. R. Suresh, "Machine learning models for crop yield prediction: A comparative study," Journal of Agricultural Informatics, vol. 12, no. 2, pp. 45–60, 2022.
15. Y. Zhu, L. Zhang, and H. Li, "A deep learning crop model for adaptive yield estimation in large regions," Agricultural Systems, vol. 192, p. 103172, 2022. https://www.sciencedirect.com/science/article/pii/S0308521X22001043
16. M. S. Qiao et al., "SSTNN: A spatial–spectral–temporal neural network for crop yield prediction," ISPRS Journal of Photogrammetry and Remote Sensing, 2021. https://www.sciencedirect.com/science/article/pii/S0303243421001434
17. A. Joshi, R. Singh, and V. Thakur, "Deep-Transfer-Learning Strategies for Crop Yield Prediction," Remote Sensing, vol. 16, no. 24, p. 4804, 2024. doi:10.3390/rs16244804. https://www.mdpi.com/2072-4292/16/24/4804
18. SustainBench Group, "SustainBench — Crop Yield Prediction dataset & benchmarks," SustainLab / Berkeley AI Research.
19. M. Shahhosseini, G. Hu, S. V. Archontoulis, and I. Huber, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," arXiv:2008.04060, 2020. https://arxiv.org/abs/2008.04060
20. S. Sabini, "Understanding satellite-imagery-based crop yield predictions," Stanford CS231n report (tutorial/survey), 2017. https://cs231n.stanford.edu/reports/2017/pdfs/555.pdf
21. D. Gangopadhyay, S. Pradhan, and R. Srivastava, "Applications of machine learning for crop yield prediction in India: A review," Agricultural Reviews, 2021. (Available through agricultural journal archives)
22. "Advancements in remote sensing based crop yield modelling in India" — Indian-focused RS+ML survey (journal article).
23. S. A. Dey, R. K. Verma, and P. K. Singh, "Precision agriculture for improving crop yield predictions: a literature review," Frontiers in Agronomy, 2025. (Use Frontiers portal to fetch DOI)
24. A. Thakkar, "Enhancing crop yield estimation from remote sensing data: Preprocessing and ViT approaches," SN Applied Sciences, 2024. doi:10.1007/s42452-024-06329-8. https://link.springer.com/article/10.1007/s42452-024-06329-8
25. "Extreme Gradient Boosting for yield estimation compared with Deep Learning," Computers and Electronics in Agriculture, 2022. https://www.sciencedirect.com/science/article/abs/pii/S0168169922006548
26. H. Zhong, X. Li, D. Lobell, S. Ermon, and M. L. Brandeau, "Hierarchical modeling of seed

variety yields and decision making for future planting plans," arXiv:1711.05809, 2017. https://arxiv.org/abs/1711.05809
27. "Feasibility of machine learning-based rice yield prediction in India at the district level using climate reanalysis data," arXiv / journal (2024).
28. A. Ikram et al., "A fuzzy-optimized hybrid ensemble model for yield prediction," Frontiers in Plant Science, 2025. https://www.frontiersin.org/journals/plant-science
29. M. Ilyas, F. Ahmed, and Q. M. Ilyas, "Automated estimation of crop yield using artificial intelligence," Sensors / Computers & Electronics in Agriculture, 2023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9952812/
30. "A systematic review of the use of deep learning in satellite imagery for agriculture," arXiv (2022).
31. https://iasri.res.in/agridata/24data/Home.HTML
32. Luis Jácome Galarza, Miguel Realpe, Marlon Santiago Viñán-Ludeña, "AgriTransformer: A Transformer-Based Model with Attention Mechanisms for Enhanced Multimodal Crop Yield Prediction",2023.
33. Satyveer Singh Meena, agini Naikoti," EMERGING FRONTIERS OF COMMERCE, MANAGEMENT AND INFORMATION TECHNOLOGY",2024.