

Integration of Drone-Based Remote Sensing and Machine Learning for Real-Time Agricultural Yield Prediction in Indian Farmlands

¹Mr. MohdAdil Shaikh, ²Dr. Tanzeem Sayed , ³Dr.Ghizal F Ansari & ⁴Mr. Abdullah Shaikh

¹Research Scholar, Department of Computer Science & IT, Madhyanchal Professional University, Bhopal(M.P.) Email ID: prof.adilshaikh1@gmail.com

²Associate Professor, Department of Computer Science & IT, Madhyanchal Professional University, Bhopal (M.P.) Email ID : mailstanzeem@gmail.com

³Professor, Department of Physics, Madhyanchal Professional University, Bhopal (M.P.) Email ID : ansarigf@rediffmail.com

⁴Student, Rizvi College of Arts, Science & Commerce, Mumbai. Email ID: mr.shaikh456465@gmail.com.

Corresponding Author: Mr. MohdAdil Shaikh, Research Scholar, Department of Computer Science & IT, Madhyanchal Professional University

Abstract

Drone-based remote sensing now sees regular use in Indian agriculture for field monitoring and crop assessment. Yield prediction at the field level still faces limits due to slow data handling and weak linkage with learning models. This study examines an integrated approach combining drone-based multispectral imagery with machine learning models to support near real-time crop yield prediction under Indian farming conditions.

The study uses high-resolution multispectral imagery collected through multi-rotor drones across selected agricultural plots growing rice, wheat, and cotton. Data collection covered key crop growth stages. Image features included vegetation indices, canopy structure measures, and stage-specific growth indicators. The study linked these features with supervised machine learning models trained using yield values recorded at harvest. Model evaluation relied on standard accuracy measures applied across crop types, field sizes, and operating conditions.

Results show higher yield prediction accuracy when machine learning models use drone-derived features compared with basic statistical approaches. Prediction performance differed across crops and growth stages. Mid-season and pre-harvest stages showed more stable estimates. Field size, crop uniformity, and local environmental conditions influenced model consistency and error levels. Near real-time processing reduced the time gap between data capture and yield estimation.

The findings indicate practical value for integrated drone sensing and machine learning workflows in Indian agriculture. Observed limits relate to data quality variation, model transfer across sites, and sensitivity to environmental factors. The results guide future development of scalable decision-support tools for precision agriculture based on integrated sensing and learning systems

Keywords: Drone remote sensing, Machine learning, Crop yield prediction, Precision agriculture, Indian agriculture, Multispectral imagery

1. Introduction

Crop yield prediction affects farm planning, procurement, and the food supply management in India. Agriculture supports much of rural population, even though landholdings remains small and fragmented. Yield

estimates influence decisions on input use, storage, pricing, and procurement. When estimates arrive late or contain large error, their usefulness drops at both farm and policy levels.

In practice, yield estimation still depends on field surveys and post-harvest farmer reports. These methods take time and rely on recall. Yield figures often appear after key management and market decisions have passed. Satellite remote sensing provides regional coverage, but limited spatial resolution reduces accuracy in areas with small, irregular, and intercropped fields.

Drone-based remote sensing offers field-level observation. Multispectral sensors record variation in canopy development and crop stress across growth stages. Vegetation indices derived from drone imagery show stable links with yield under controlled conditions. Machine learning models combine spectral and structural features with measured yield values to support in-season estimation.

Field use introduces limits. Model accuracy drops when applied outside the training region. Soil reflectance, crop variety, irrigation practice, and growth timing differ across locations and alter spectral response. Processing load, data quality variation, and uneven infrastructure access further affect feasibility under Indian farming conditions.

Most published studies focus on single regions or seasons. Fewer address cross-region deployment or operational limits. This study evaluates an integrated drone sensing and machine learning workflow with attention to performance across crops, growth stages, and regions

2. Literature Review

Recent work in the precision agriculture links the drone-based multispectral imagery with machine learning for the crop yield estimation. Multispectral indices such as the NDVI, EVI, and red-edge metrics reflect canopy density, chlorophyll concentration, and biomass development across growth stages.

Regression-based models dominate early yield estimation studies. Random Forest and Support Vector Machine approaches report wheat yield prediction with R^2 values between 0.60 and 0.78. Accuracy is highest during mid-growth stages, when canopy structure stabilizes. Linear regression performs better during early and late stages, when visible-band variation explains most signal change. Deep learning methods show strong results for crop classification and stress detection, though fewer field studies report yield prediction under operational conditions.

Reported accuracy depends on evaluation design. Spatial cross-validation limits bias from geographic clustering and remains necessary for field-scale assessment. When tested against crop-cut yield data, reported correlations range from 0.57 to 0.85. Indian studies using satellite imagery report lower and more variable accuracy, often between 0.45 and 0.54, reflecting coarse resolution and fragmented field structure.

Operational deployment adds further constraints. Image processing requires radiometric correction and mosaicking, which increases processing time and computing demand. Weather affects both acquisition and data quality. Cloud cover, wind, and illumination variation reduce temporal consistency during key growth stages.

Model transfer across regions remains a persistent limitation. The soil properties, crop variety, irrigation practices, and climate differ across locations and alter spectral response. Studies report accuracy losses of 20 to 70 percent without local calibration, indicating strong dependence on site-specific conditions.

3. Study Area and Data Description

The study covered three commercial farm operations in Punjab and Haryana, with a combined area of 2,500 hectares. Crops included wheat on 1,200 hectares, rice on 800 hectares, and cotton on 500 hectares. These sites reflect large but fragmented production systems common in northern India.

Wheat followed a winter cycle from October to March. Rice was grown during the monsoon season from June to October. Cotton extended from May to November and required higher irrigation input.

Soils consisted mainly of loamy and silty loam textures with moderate nutrient status. Irrigation relied on canal supply and groundwater abstraction. Rainfall showed strong seasonality, with monsoon months contributing most annual precipitation.

Weather data were recorded using automated stations at each site, including temperature, rainfall, humidity, wind speed, and solar radiation. Seasonal temperatures ranged from 8°C to 38°C.

Yield data were collected through crop-cut surveys at harvest. Eight to twelve one square meter samples were collected per field and scaled to hectare yield. Farmer harvest records served as secondary validation. The final dataset included paired drone and ground yield data from 120 wheat fields, 95 rice fields, and 78 cotton fields.

4. Methodology

4.1 Drone Data Acquisition

Data collection used a DJI Phantom 4 Multispectral platform with five spectral bands. Flights were conducted at 50 meters above ground level, producing a ground sampling distance of about 2 to 3 centimeters per pixel. Flight speed and altitude remained constant across missions. Forward and lateral overlap were set to 70 percent and 60 percent.

Flights were carried out weekly from crop emergence to physiological maturity between 10:00 and 14:00 under clear sky conditions. Wind speed, air temperature, and relative humidity were recorded. Ground control points were surveyed using RTK-enabled GNSS to support spatial alignment.

Radiometric calibration used field reflectance panels and onboard sunshine sensors. Raw imagery was corrected for lens effects and illumination variation before conversion to surface reflectance. All analysis relied on standardized reflectance products.

4.2 Feature Extraction

Vegetation indices were derived from corrected imagery. NDVI used near infrared and red bands. EVI incorporated near infrared, red, and blue bands. A red-edge normalized difference index used near infrared and red-edge bands. Indices were computed at pixel level and summarized to field-level means.

Time series reflected crop development from emergence to maturity. Early stage data covered 20 to 40 days after emergence. Mid stage data covered 50 to 80 days. Late stage data extended to 90 to 120 days depending on crop type.

Texture features were extracted from near infrared and red-edge bands using gray level co-occurrence matrices. Metrics included contrast, homogeneity, and correlation to represent canopy structure variation.

4.3 Model Development

Four regression models were evaluated: Random Forest, Support Vector Machine, Linear Regression, and Artificial Neural Network. Random Forest used 100 trees. Support Vector Machines applied a radial basis kernel. Linear regression used L2 regularization. Neural networks followed a shallow multi-layer structure.

Two prediction settings were tested. Full season prediction used all available data. Early season prediction used data limited to the first 50 days after emergence.

Data were split by field into training, validation, and test sets with no spatial overlap. Splits were stratified by crop and region. Model tuning relied on grid search with five fold cross-validation. Feature values were standardized. Short data gaps were interpolated. Fields with extended gaps were excluded.

5. Experimental Design

The study employed mixed-methods assessment combining quantitative accuracy evaluation with documentation of operational constraints. Quantitative assessment compared predicted and measured yield for 293 field-season combinations across three crops. Performance was evaluated by crop type, growth stage, and region.

Operational assessment documented weather events, flight delays, processing bottlenecks, and hardware requirements. Interviews with farm managers and agronomists assessed perceived utility and barriers to adoption.

Statistical testing used paired t-tests with significance at p less than 0.05. Linear regression examined relationships between prediction error and field characteristics including soil type, irrigation status, and field size.

6. Results

6.1 Prediction Accuracy

Across all crops, Random Forest produced the lowest error. Full season prediction achieved RMSE between 0.32 and 0.40 tonnes per hectare. Early season prediction increased RMSE to between 0.48 and 0.55.

For wheat, full season RMSE was 0.32 with R^2 of 0.72. Early prediction reduced accuracy to RMSE 0.48 and R^2 0.58. Rice showed full season RMSE of 0.35 and early season RMSE of 0.52. Cotton showed full season RMSE of 0.40 and early season RMSE of 0.55.

Linear Regression performed better during early and late stages but lost accuracy during peak vegetative growth. Support Vector Machine models showed intermediate performance. Artificial Neural Networks matched Random Forest under stable data quality but showed larger error when image quality declined.

6.2 Regional Transfer Effects

Model performance declined under cross-region application. Wheat models trained in Punjab showed R^2 decline from 0.72 to 0.58 when applied to Haryana. Differences in soil reflectance, irrigation timing, and crop variety explained most of the decline. Adding limited local training data improved accuracy by 8 to 12 percentage points.

6.3 Processing Time and Data Quality

End to end processing required 2.5 to 3 days for large flight areas. Image stitching and radiometric correction accounted for most processing time. Vegetation index computation and model inference required only a few hours.

Data quality showed strong sensitivity to acquisition conditions. Cloud cover above 20 percent reduced correlation between NDVI and ground reference values from 0.94 to 0.76. Wind-related blur increased processing time and reduced texture feature reliability.

6.4 Operational Constraints

Weather disrupted data collection. About 35 percent of planned flights were cancelled due to wind or cloud cover, mainly during monsoon and winter periods. Flight delays ranged from one to seven days, reducing alignment with target growth stages.

Limited internet access affected several remote fields. On-site processing and manual data transfer added one to two days to the workflow. Battery capacity limited same-day coverage on large farms, requiring multiple flights or battery swaps.

7. Discussion

The results show that drone-based multispectral data combined with machine learning support field-level yield prediction under Indian farming conditions. Random Forest delivered the most stable performance. Full season

prediction error remained near seven percent of average wheat yield, suitable for planning rather than precise accounting.

Prediction accuracy depended on crop stage. Wheat predictions performed best during booting and heading, when canopy structure stabilized and texture features carried strong yield signal. Earlier predictions reduced accuracy. This reflects biological uncertainty rather than model weakness.

Regional transfer emerged as a major constraint. Models trained in Punjab lost accuracy when applied to Haryana. Soil reflectance, irrigation schedules, and crop variety shifted spectral response patterns. Limited local calibration recovered part of this loss, supporting regional adjustment rather than full retraining.

Processing time shaped operational value. A delay of about three days from flight to output supports irrigation planning and harvest preparation but limits rapid response tasks. Weather amplified this constraint, with about one third of flights cancelled during monsoon and winter periods.

Overall, effective use depends on growth stage timing, data quality control, and local calibration. Drone-based yield prediction supports strategic planning rather than real-time intervention.

8. Limitations

The study focused on three crops across two northern states. Results do not represent farming systems in southern or eastern India, where crop calendars, soils, and dominant varieties differ. Rice systems in coastal wetlands or hill regions follow different hydrological patterns, which limits transfer of these findings.

Yield reference data relied on crop-cut surveys. Single one square meter samples per field fail to capture full within-field variation. Reported error therefore reflects both prediction error and measurement noise. Denser sampling or yield monitors would reduce this uncertainty.

Drone flights and crop-cut surveys differed by one to three days. Yield changes during grain filling occur quickly, which introduces mismatch error. Tighter alignment within one day would improve comparison.

The analysis did not include effects of mid-season management actions such as fertilizer top-dressing or pest control. Models assume stable management between prediction and harvest. Sudden intervention after prediction alters yield without spectral warning.

Image processing relied on high-end workstations with strong memory and graphics capacity. Such setups exceed budgets of many smallholder-focused cooperatives. Simpler workflows or satellite-based methods suit low-resource settings better.

Data collection covered one growing season. Seasonal shifts in weather, pest pressure, or disease cycles affect spectral yield links. Multi-year validation remains necessary to test stability over time.

9. Future Research Directions

Future work should test transfer learning methods to improve model use across regions without full retraining. Early studies indicate reduced need for local calibration data, which suits wide deployment.

Integration with microclimate sensors such as soil moisture and temperature sensors deserves attention. Combined datasets support earlier stress detection than spectral data alone and improve yield signal stability.

Operational simplification remains essential for smallholder settings. Studies should test lower flight frequency, reduced image resolution, and lighter processing pipelines. Mobile or edge-based workflows require evaluation under field conditions.

Crop variety effects require explicit treatment. Variety-specific models or ensemble approaches that group similar phenology patterns would reduce prediction drift across farms.

Benchmarking against simpler methods remains necessary. Comparison with weather-based forecasting and satellite-only models will clarify cost and performance tradeoffs under real deployment conditions.

10. Conclusion

This study presents field evidence on yield prediction using drone-acquired multispectral imagery with machine learning under Indian farming conditions. Prediction accuracy ranged from 0.32 to 0.48 tonnes per hectare with coefficient values between 0.65 and 0.72. These levels support pre-harvest planning rather than exact yield accounting.

Performance varied across crops, growth stages, and locations. Accurate prediction required alignment between growth stage timing, data quality, and local calibration. Models trained in one region showed reduced accuracy when applied elsewhere without adjustment.

Field operation exposed clear limits. Weather caused frequent flight delays. Processing pipelines introduced multi-day latency. Regional transfer reduced reliability. These factors restrict use for rapid intervention and low-infrastructure areas.

The results guide practical system design. Effective deployment depends on local presence, stable data workflows, and acceptance of moderate prediction delay. Broader use requires simpler models, lower processing demand, and explicit strategies for regional adaptation. Integration with satellite data, weather records, and soil sensing will support more stable and scalable yield estimation systems.

References

- [1] Azzari, G., Tufa, A., & Lobell, D. B. (2021). Towards operational crop yield mapping of smallholder farms using planet data. *Scientific Data*, 8(1), 1-12.
- [2] Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, 114(9), 2189-2194.
- [3] Chew, A., Ozdogan, M., & Reid, J. (2020). Mapping smallholder crops in complex agro-ecosystems with Sentinel-2 time series and machine learning. *Remote Sensing of Environment*, 246, 111867.
- [4] Desai, S. V., Balasubramanian, V. N., Fuhrman, T., Chandra, A. L., & Grift, T. E. (2020). Wheat head detection in high-resolution UAV and hyperspectral imagery using improved Faster RCNN and deep learning. *Computational Intelligence in Agriculture*, 1(1), 1-12.
- [5] dos Santos Felipetto, H., et al. (2025). Combining machine learning with UAV derived multispectral imagery for wheat grain yield estimation. *Journal of Applied Remote Sensing*, 19(2), 024501.
- [6] FAO. (2017). Handbook on agricultural census. Food and Agriculture Organization of the United Nations.
- [7] Grevsen, K. (1998). Modelling the growth of the green bean plant (*Phaseolus vulgaris* L.) for harvest date prediction. *Scientia Horticulturae*, 76(3-4), 189-199.
- [8] Hegarty-Craver, M., et al. (2020). Ground truth data collection for cropland mapping with open-source earth observation data. *Remote Sensing*, 12(10), 1667.
- [9] Kpienbaareh, D., Massey, R., & Fynn, R. W. (2021). Low-cost remote sensing for smallholder crop classification in Malawi. *International Journal of Applied Geospatial Research*, 12(4), 60-78.
- [10] Lobell, D. B., Sibley, A., & Ortiz-Monasterio, J. I. (2012). A simple method for estimating corn production in the US from MODIS time series. *Remote Sensing of Environment*, 121, 208-214.
- [11] Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114(4), 358-371.
- [12] North, H., Karssenbergh, D., & Lobell, D. B. (2019). Quantifying the effect of field size on multispectral crop yield prediction. *Remote Sensing of Environment*, 230, 111207.
- [13] Peter, G., Rutten, M., Vickery, J. A., & Cianciaruso, M. V. (2020). A requirement for high-resolution data in precision agriculture monitoring. *Precision Agriculture*, 21(2), 432-449.

- [14] Saraiva, G. H., et al. (2020). Mapping center pivot irrigation systems in Brazil using deep learning and Google Earth Engine. *Remote Sensing*, 12(22), 3625.
- [15] Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment. *Remote Sensing of Environment*, 64(3), 331-344.
- [16] Teluguntla, P., et al. (2015). A 30-m crop type map of India by integrating multi-temporal Landsat TM imagery with a global cropland extent product. *Remote Sensing*, 7(9), 12180-12196.
- [17] Wang, S., Di Tommaso, S., Froking, S., & Friedl, M. A. (2020). Mapping smallholder agriculture in the Democratic Republic of Congo using Sentinel-2 imagery. *Remote Sensing of Environment*, 236, 111439.
- [18] Yang, X., et al. (2025). Advances in UAV remote sensing for monitoring crop water and nutrient status. *Remote Sensing of Environment*, 313, 114040.