PLANT SCIENCE TODAY
Vol 13(sp9): 1–16
https://doi.org/10.14719/pst.17573

eISSN 2348-1900
OPEN ACCESS

Review Article | Agriculture
2026, Volume 13, Issue sp9, pp. 1-16

Satellite-derived vegetation indices and machine learning for food crop yield prediction: A systematic review of methods, performance and research gaps

Sai Divya Sri Jagarlamudi[1], Kumaraperumal Ramalingam[1*], Muthumanickam Dhanaraju[1], Prabu Padanillay Chidambaram[2], Balaji Kannan[3], Nivas Raj Moorthi[1] & Vishalini Sivasubramani[1]

[1]Department of Remote sensing and Geographic Information System, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[2]Department of Centre for Water and Geospatial studies, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[3]Department of physical science and Information Technology, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

***Correspondence: Kumaraperumal Ramalingam, kumaraperumal.r@tnau.ac.in**

Abstract

Crop yield prediction is essential for global food security, but the current methods for yield estimation are labour-intensive, time-consuming and have limited spatial and temporal coverage. This systematic review focuses on consolidating the evidence from the 32 peer-reviewed studies published from 2018–26 that are concerned with using both satellite-derived vegetation indices (VIs) and machine learning (ML) to predict food crop yields. A preferred reporting items for systematic reviews and meta-analyses (PRISMA)-based search resulted in the identification of studies on rice, wheat, maize, soybean, sugarcane, sorghum and barley across different geographic regions. The evidence clearly shows that across all food crops, Random Forest (RF) is the most consistent ML technique, with an R^2 of 0.70–0.96 based on crop and region and deep learning approaches like long short-term memory (LSTM) and convolutional neural network (CNN) show higher temporal pattern recognition, but much greater numbers of labelled data are needed. Multi-index approaches consistently outperform single-index approaches, with those extracted from Sentinel-2 performing well in situations where normalised difference vegetation index (NDVI) loses sensitivity due to canopy saturation at high biomass levels in dense-crop canopies. Sentinel-2 is the most frequently used satellite platform (over 65 % of the analysed studies) and Google Earth Engine (GEE) has become the most used data processing environment. There are still gaps in the standardisation of benchmarking metrics, the representation of smallholder and tropical farming systems, the integration of data across phenological stages and the cross-regional transferability of models. This review consolidates evidence and presents clear priorities to move precision agriculture yield forecasting toward global food security applications.

Keywords: crop yield prediction; food security; machine learning; remote sensing; satellite imagery; vegetation indices

Received: 21 August 2026. **Accepted:** 22 September 2026. **Available online:** Version 1.0: 04 November 2026.

Cite this article: Sai Divya Sri J, Kumaraperumal R, Muthumanickam D, Prabu PC, Balaji K, Nivas RM, Vishalini S. Satellite-derived vegetation indices and machine learning for food crop yield prediction: A systematic review of methods, performance and research gaps. Plant Science Today. 2026; 13(sp9): 1-16. https://doi.org/10.14719/pst.17573

1. Introduction

Global food security and the need for precision agriculture

Population growth, climate variability and constraints on land, water, labour and nutrient resources collectively place increasing strain on global food systems. In 2050, the projected food demand will rise by around 60 % from the 2010 value due to population growth, which is expected to reach 10 billion by 2050, according to the Food and Agriculture Organisation (FAO) (1). This need must be addressed in an environment of growing uncertainty in growing conditions, including changes in precipitation patterns, extended droughts, heat stress events and a decline in arable land. Accurate, timely and spatially explicit crop yield prediction has therefore become an essential tool for agricultural planning, famine risk management and policy decision-making.

The conventional production estimation techniques, such as field surveys, manual crop cutting and agrometeorological regression models, have served as the backbone of national agriculture statistics for decades. These methods, however, have well-recognised drawbacks. They are labour-intensive, costly, biased by observers, applicable only at the end of harvest and restricted to sampled plots, which cannot fully reflect field-level heterogeneity (1). Even the most rigorous household survey programmes, such as the World Bank's Living Standards Measurement Study (LSMS), have documented systematic measurement errors in agricultural data, particularly in smallholder systems where overestimation of plot sizes and inaccurate recall of yields are widespread (2). These limitations pose a fundamental

problem for traditional approaches to answering food security governance questions that ask for near real-time and scalable monitoring.

Satellite remote sensing and machine learning (ML) analysis provide a significant departure from traditional approaches. Continuous and spatially explicit vegetation indices (VIs) derived from satellite observations offer temporally and spatially dense proxies of the phenological characteristics of crops that are all functionally related to final grain yield, including crop health, canopy development, biomass accumulation and photosynthetic activity (3). These spectral signals can then be used in conjunction with ML techniques that can model the nonlinear, multi-dimensional relationship between VI time-series and yield outcomes and thus produce frameworks that offer the possibility of prediction accuracy comparable to or even better than traditional methods, but at significantly lower cost and operational complexity (4).

Role of satellite imagery in modern agriculture

With the deployment of the European Space Agency Copernicus Sentinel-2 constellation in 2015 and 2017, the pace of the shift from crop monitoring via ground methods to satellite monitoring accelerated. The combination of Sentinel-2A and Sentinel-2B guarantees multispectral information at a spatial resolution of 10–20 m and a revisit cycle of approximately 5 days, covering 13 spectral bands, including three red-edge channels which have been missing from the otherwise dominant Landsat and moderate resolution imaging spectroradiometer (MODIS) platforms (5). This time and space complexity has allowed the monitoring of yield that was previously not possible with a higher-level coarse sensor. In addition to Sentinel-2, other satellites like MODIS offer intensive data series back to 2000, ideal for regional and national scale analyses; commercial satellites like PlanetScope offer daily imagery at sub-3 m resolution, suitable for precision plot-level analyses (6).

The VIs derived from these platforms are useful indicators of biophysical crop characteristics that are closely related to crop production. Normalised difference vegetation index (NDVI) is calculated from the difference between the red and near-infrared (NIR) reflectance values and has been used in agri-remote sensing since the 1970s because it correlates with chlorophyll content and green biomass. Later indices make more use of the red-edge spectral region where the signal is more related to the concentration of chlorophyll, particularly for higher canopy densities where NDVI saturates and some make use of shortwave infrared bands which, although sensitive to leaf area index (LAI), are also sensitive to water status in the canopy (7). The computing platform Google Earth Engine (GEE) has significantly improved the accessibility of these datasets for researchers in low-resource environments and has given them the opportunity to process petabyte-scale satellite archives without needing to own high-performance computing resources (5).

Machine learning as a transformative tool

The correlation of vegetation indices with crop yield is spatially variable, nonlinear and complicated by soil properties, climate variability, management practices and cultivar genetics. Although these models are easy to compute, traditional statistical regression models are unable to capture such complex interactions in large and highly heterogeneous datasets (8). The ability to capture nonlinear input-output relationships, to accommodate high-dimensional feature spaces that include multiple VIs and climate and soil covariates, to be spatially flexible from the field to the national level and to increase predictive model accuracy with more data make ML methods compelling approaches for this problem.

Random Forest (RF) is one of the ML techniques that have been widely used in estimating agricultural crop yields across various crops and regions and it has been proven to be a consistent and reliable method for crop yield prediction (4). Since 2021, ensemble gradient boosting methods, especially extreme gradient boosting (XGBoost), have been gaining popularity, showing some benefits over RF in sparse or noisy data and improved results in external validation sets (9). Deep learning architectures such as convolutional neural networks (CNN) for spatial pattern extraction and long short-term memory networks (LSTM) for temporal sequence modelling have provided the highest accuracy in well-resourced study locations. However, the high data requirements of these architectures have prevented their use in locations where labelled yield datasets are unavailable (3).

Research gap and objective of this review

Previous research has often been concerned with individual satellite platforms, approaches to ML or more general uses of remote sensing or remote sensing and ML, but much less with systematic cross-crop assessments of VI selection and its effects on model performance (3–5, 8). Consequently, there is no detailed synthesis of VI performance for different crops, examining the connection between crop and ML method selection, or critically evaluating the evidence base regarding geographical and methodological biases. This review fills that gap by performing an integrated, crop-wise, method-wise critical analysis of 32 peer-reviewed studies, including both global evidence in published reviews and primary empirical data from the literature that covers 7 food crop types and five continents, from 2018–26.

Structure of the paper

The paper is organised as follows. Section 2 describes the systematic search methodology, preferred reporting items for systematic reviews and meta-analyses (PRISMA) screening procedure, data extraction and quality assessment. In Section 3, the different vegetation indices adopted in the studies reviewed are critically examined based on their biophysical foundations and relative strengths as indices, in relation to different crop types. Section 4 discusses the ML and deep learning techniques used and a comparison of the prediction accuracy of different algorithms.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

2. Methodology

Search strategy and databases

This systematic review was performed according to PRISMA 2020 guidelines. The literature search was performed in four electronic bibliographic databases: Google Scholar, PubMed Central, Elsevier ScienceDirect and Institute of Electrical and Electronics Engineers Xplore Digital Library (IEEE Xplore). The databases are intended to cover as broad a range of literature in remote sensing, agricultural and engineering sciences and environmental sciences as possible, including journals indexed in Scopus and Web of Science. Only peer-reviewed journal articles published during the period from 2018–26 in the English language were included. The following search string was used across all databases: ("vegetation index" OR "NDVI" OR "EVI" OR "SAVI" OR "NDRE") AND ("crop yield" OR "yield prediction" OR "yield estimation") AND ("machine learning" OR "random forest" OR "deep learning") AND ("satellite" OR "Sentinel-2" OR "Landsat" OR "MODIS")

The search was carried out in July 2026. This study was not registered prospectively with International Prospective Register of Systematic Reviews (PROSPERO) or any similar registry. Screening of the titles and abstracts was done independently by the author, while the full text was reviewed by the author.

Inclusion and exclusion criteria

The seven food crop types covered in this review, namely rice, wheat, maize, soybean, sorghum, sugarcane and barley, were defined a priori as inclusion criteria based on their collective dominance of global caloric supply and their established presence in the remote sensing yield prediction literature. Studies investigating any of these crops as the primary yield prediction target were included, provided they had satellite- or unmanned aerial vehicle (UAV)-derived VIs as major inputs. It was also required that the study report at least one ML or deep learning model and provide a quantitative accuracy measure like R^2, root mean square error (RMSE) or mean absolute error (MAE). In addition, all studies must have been published in peer-reviewed journals. Scopus or Web of Science indexing was used as a quality indicator during full-text screening where journal standing was uncertain.

Studies that utilised only ground-based spectroradiometry without satellite integration, studies that utilised only traditional statistical regression without a ML component and studies that only concentrated on non-food ornamental or biofuel crops were excluded from the study. The review did not include abstracts of conference presentations, theses, grey literature and articles before 2018. As this is a review, six high-citation narrative or systematic review papers (1, 3, 4, 5, 8, 30) that were not actually primary empirical studies were included as contextual studies to place this review in context. Because they were not scored on the primary methodological quality scoring rubric (Table 1 criteria Q1–3 marked N/A for these entries) and lacked primary empirical performance data, they were not included in the calculation of mean R^2 across quality strata; however, they were included as part of the information provided.

Table 1. Summary of machine learning (ML) methods, primary vegetation indices (VIs) and best accuracy metrics across representative reviewed studies

No.	Reference	Crop	Q1	Q2	Q3	Q4	Q5	Score
1	(1)	Multi-crop	N/A	N/A	N/A	Y	Y	2/5
2	(2)	Sorghum	Y	P	N/A	N/A	N/A	1.5/5
3	(3)	Multi-crop	N/A	N/A	N/A	Y	Y	2/5
4	(4)	Multi-crop	N/A	N/A	N/A	Y	Y	2/5
5	(5)	Multi-crop	N/A	N/A	N/A	Y	Y	2/5
6	(6)	Soybean	Y	Y	Y	Y	N/A	4/5
7	(7)	Multi-crop	Y	P	P	Y	N/A	3/5
8	(8)	Multi-crop	N	N	N	Y	Y	2/5
9	(9)	Rice	Y	Y	Y	Y	P	4.5/5
10	(10)	Rice	Y	P	P	Y	N/A	3/5
11	(11)	Wheat	Y	Y	P	Y	N/A	3.5/5
12	(12)	Wheat	Y	P	P	Y	N/A	3/5
13	(13)	Rice	Y	P	P	N	N/A	2/5
14	(14)	Rice	Y	P	N	N	N/A	1.5/5
15	(15)	Maize	Y	Y	Y	Y	N/A	4/5
16	(16)	Barley	Y	Y	Y	Y	N/A	4/5
17	(17)	Wheat	Y	Y	Y	Y	N/A	4/5
18	(18)	Wheat	Y	Y	Y	Y	P	4.5/5
19	(19)	Maize	Y	P	Y	Y	N/A	3.5/5

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

20	(20)	Wheat	Y	P	Y	Y	N/A	3.5/5
21	(21)	Multi-crop	Y	Y	Y	N	Y	4/5
22	(22)	Wheat	Y	Y	Y	Y	N/A	4/5
23	(23)	Wheat	Y	Y	Y	Y	N/A	4/5
24	(24)	Rice	Y	P	Y	Y	N/A	3.5/5
25	(25)	Maize + Soy	Y	Y	Y	Y	N/A	4/5
26	(26)	Sugarcane	Y	P	P	N/A	N/A	2/5
27	(27)	Sorghum	Y	P	P	Y	N/A	3/5
28	(28)	Soybean	Y	Y	Y	Y	Y	5/5
29	(29)	Rice	Y	P	P	N/A	N/A	2/5
30	(30)	Rice	N/A	N/A	N/A	Y	Y	2/5
31	(31)	Maize	Y	Y	Y	Y	N/A	4/5
32	(32)	Soybean	Y	P	Y	Y	N/A	3.5/5

PRISMA flow and study selection

The database search, conducted in July 2026, retrieved 1803 records in total prior to duplicate removal: 665 from ScienceDirect, 1100 from Google Scholar, 29 from IEEE Xplore and 9 from PubMed Central. Upon the elimination of 74 duplicate records, 1729 distinct records remained for screening of titles and abstracts. Out of those, 1608 were filtered out due to not matching the predefined inclusion criteria and full texts were sought for 121 potentially eligible reports. Of these, five could not be retrieved and 116 were assessed in full. Another 84 articles were filtered out at this point due to insufficient study focus (n = 38), a lack of methodology detail (n = 24), being a review article, a conference article, or being a duplicate (n = 20), or ultimately failing to meet the primary study inclusion criteria (n = 2). The remaining 32 peer-reviewed studies were included in the synthesis. Fig. 1 shows the entire PRISMA screening flow.

Data extraction

The following information was extracted systematically for each study included: lead author, publication year, target crop type, satellite platform and resolution, VI used, accuracy metrics reported, phenological stage explored, geographic area and scale of the farm system. This information was then used for the comparative analysis in Sections 3–7.

The selected studies span the years 2018–26, with no included study from 2019, reflecting either lower output or indexing gaps in that year. Publication frequency was non-monotonic: output peaked at five studies in 2022, dipped in 2023–24 and surged to nine studies in 2025 before declining to five in 2026 (Fig. 2). The overall upward trajectory from 2021 onward is broadly attributable to the growing availability of Sentinel-2 image archives, the increasing accessibility of cloud-based geospatial processing platforms such as GEE and the widening adoption of open-source machine learning frameworks within the agricultural remote sensing community.

Quality assessment

Scores for the methodological rigour of each included study were determined by answering five questions: was an independent validation dataset used, were ground-truth observations taken at locations that were spatially representative, were model hyperparameters reported transparently, were multiple accuracy metrics included and was the dataset cross-regionally and/or multi-seasonally applicable. The per-study scores are presented in Table 1, with studies scoring 4–5 considered high methodological quality, 3 moderate and below 3 low. This scoring guided the critical synthesis in Sections 5 and 6 and the identification of research gaps in Section 7. The rubric was developed specifically for this review in the absence of a standard risk-of-bias instrument validated for remote sensing yield prediction studies. The five criteria were selected to reflect the principal sources of performance inflation identified in the agricultural remote sensing literature: overfitting from spatial autocorrelation (Q1), spatial representativeness of ground truth (Q2), hyperparameter transparency (Q3), completeness of accuracy reporting (Q4) and generalisability (Q5). The threshold of ≥4 for high quality was set to require that a study satisfy at least four of five criteria, consistent with a rigorous but not perfectionistic standard for applied remote sensing research.

Out of the 32 studies included, 12 scored 4 or above (high quality), 9 scored 3–3.5 (moderate) and 11 scored less than 3 (low). Independent validation (Q1) was the most consistently met criterion, with 26 studies fully satisfying it. Only 8 studies reported cross-regional or multi-seasonal applicability (Q5), which is reflected in the discussion of the transferability gap in Section 7. Among the 26 primary empirical studies (excluding the six retained review articles), those scoring 4 or above showed a mean best R^2 of 0.87, while those scoring below 3 had a mean best R^2 of 0.61. This positive association between methodological rigour and predictive accuracy should be interpreted with caution, as higher-quality studies were disproportionately conducted in data-rich, mechanised farming environments where higher R^2 values are more achievable.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

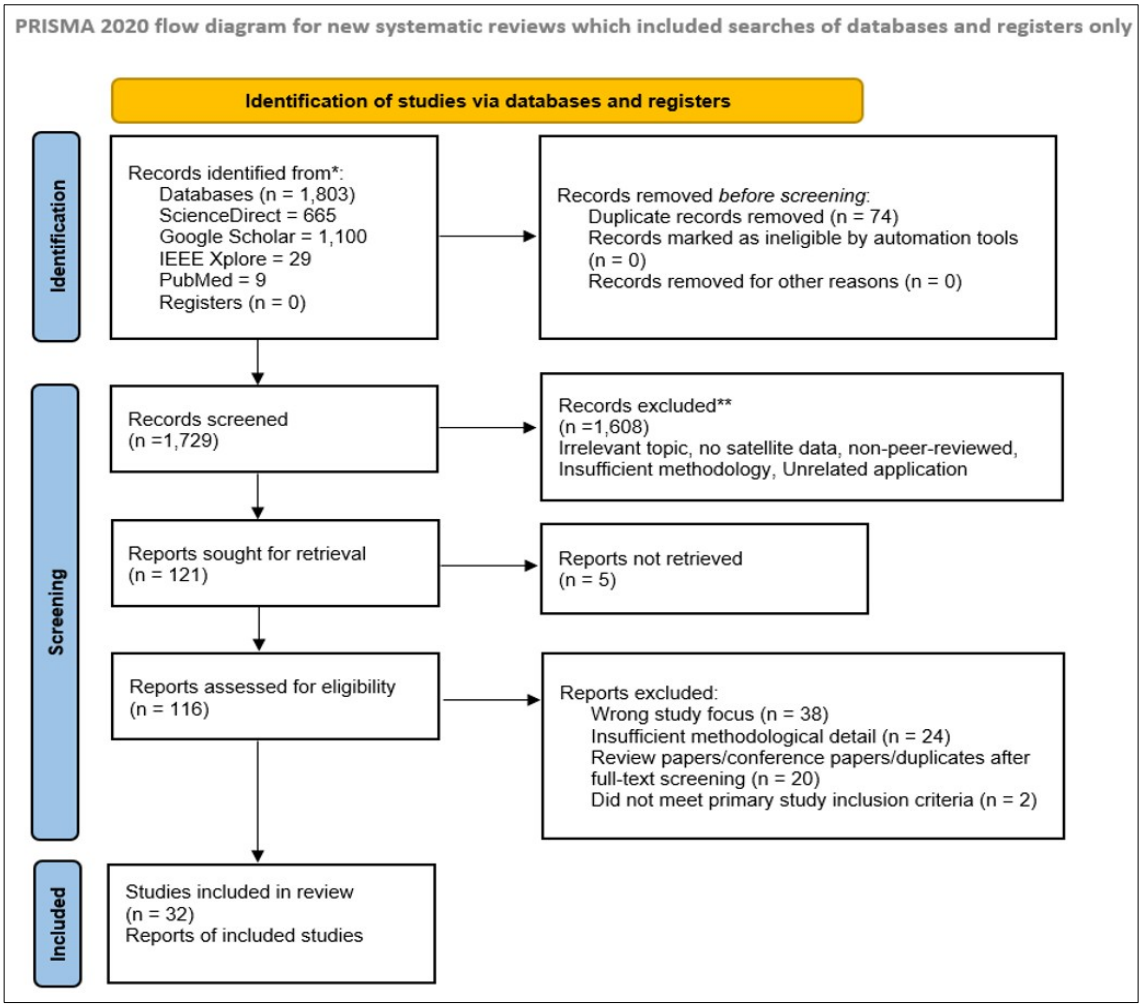


Fig. 1. PRISMA flow diagram illustrating the systematic search and study selection process across four electronic databases (2018–26).

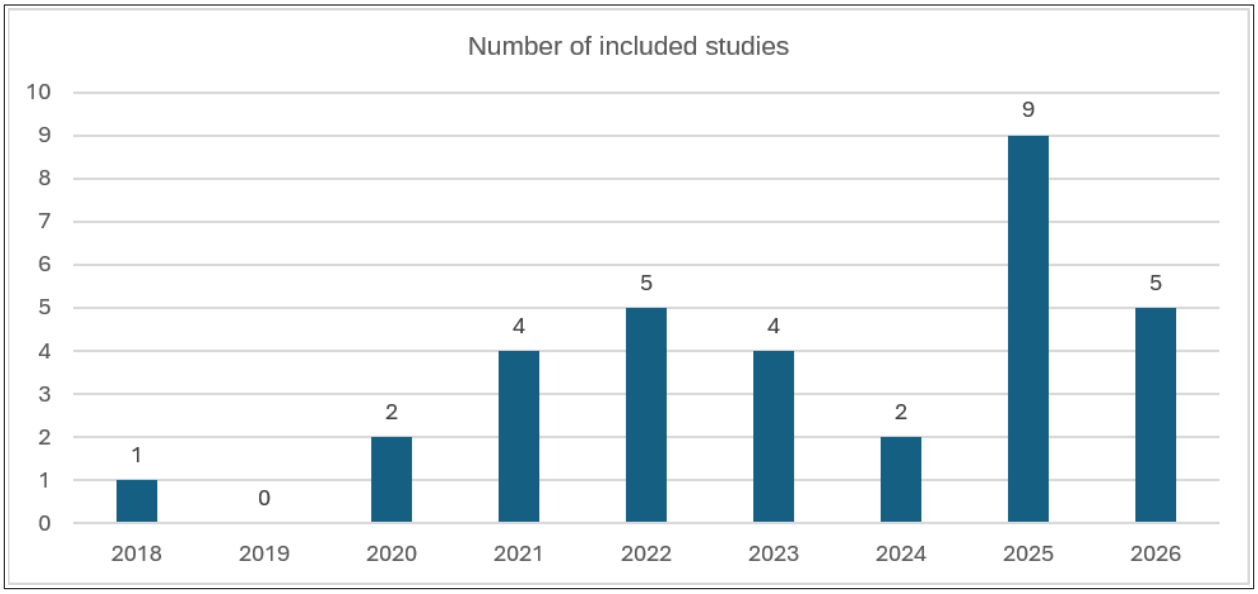


Fig. 2. Annual distribution of included studies by publication year (2018–26; n = 32).

No studies from 2019 met the inclusion criteria. Publication frequency was non-monotonic, with a pronounced peak in 2025 (n = 9); the general upward trend from 2021 onward reflects growing adoption of Sentinel-2, Google Earth Engine (GEE) and open-source machine learning (ML) tools in agricultural remote sensing research.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

3. Results

Overview: Vegetation indices used for food crop yield prediction

The VI are algebraically manipulated composite spectral reflectance measurements obtained from satellites that attempt to amplify the signal in response to vegetation condition and diminish the noise produced by the soil background, atmospheric influences and illumination geometry (7). Their usefulness for yield prediction is that they are correlated with several of the main biophysical parameters, such as LAI, chlorophyll content, the fraction of absorbed photosynthetically active radiation and canopy water content, which crucially affect photosynthetic productivity and ultimately the yield of harvestable grain (10). The overall integration of these indices in ML for food crop yield prediction is generalised from the 32 studies reviewed and depicted in Fig. 3.

In the 32 studies reviewed in this paper, nine major VI families were identified, such as NDVI, enhanced vegetation index (EVI), soil-adjusted vegetation index (SAVI) and its variants (MSAVI and MSAVI2), normalised difference red-edge index (NDRE), green normalised difference vegetation index (GNDVI), MERIS terrestrial chlorophyll index (MTCI), land surface water index (LSWI), kernel NDVI (kNDVI) and normalised difference water index (NDWI). These are explained in the following subsections that focus on their frequency of usage, biophysical basis and relative performance in various crop types. One common feature throughout the literature is that there is no single VI that is always best. The parameters for VI selection depend on the crop type, density of the crop at the appropriate growth stage, background reflectance of the soil and availability of specific spectral channels for the satellite platform used (5).

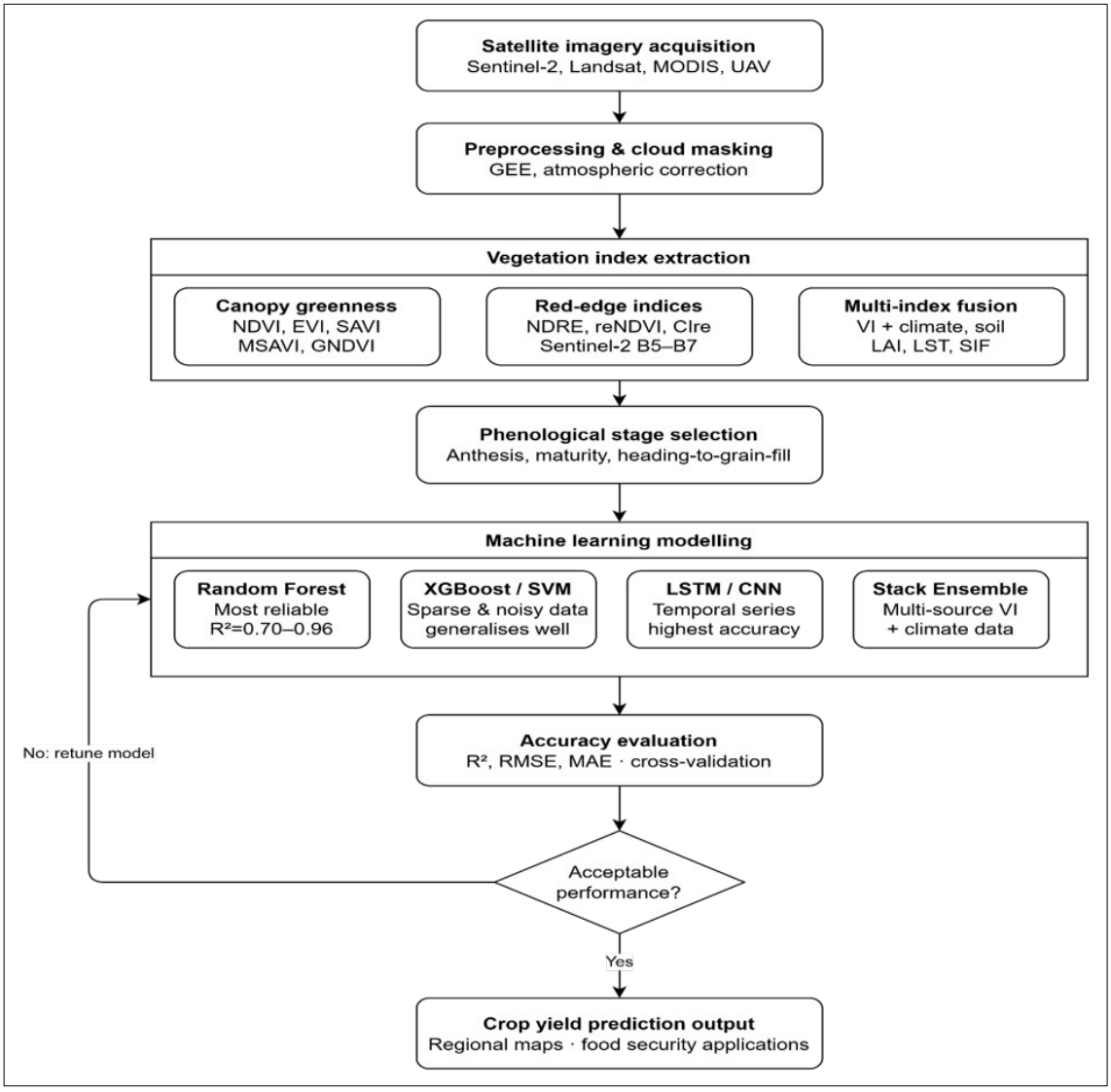


Fig. 3. Generalised vegetation index (VI) and machine learning (ML) pipeline for food crop yield prediction.

VI: vegetation index; NDVI: normalised difference vegetation index; EVI: enhanced vegetation index; SAVI: soil-adjusted vegetation index; NDRE: normalised difference red-edge index; ML: machine learning; LSTM: long short-term memory; CNN: convolutional neural network; PRISMA: preferred reporting items for systematic reviews and meta-analyses; RMSE: root mean square error; GEE: Google Earth Engine; MODIS: Moderate Resolution Imaging Spectroradiometer; CI: chlorophyll index; SIF: solar-induced chlorophyll fluorescence; LAI: leaf area index; LST: land surface temperature; MSI: moisture stress index; UAV: unmanned aerial vehicle.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

Normalised difference vegetation index: The dominant index and its documented limitations

The NDVI is the most used VI in all the studies reviewed, being used in the methodology of 28 out of 32 articles as the main predictor (4). It has been universally adopted due to its correlation with green biomass and chlorophyll content for all types of vegetation and is also computationally simple and available on all multispectral satellite platforms since Landsat-1.

The NDVI is reliable under early-growth or sparse-canopy conditions. It collected in the first half of May (1–15 May) proved to be the most significant predictor variable in the four ML models implemented for winter wheat in Lithuania, with RF resulting in the lowest mean absolute error of 0.951 t/ha (11). Wheat yield measurements also indicated a high correlation between NDVI derived from Sentinel-2 over four growing seasons in the Czech Republic, with a medium to strong correlation range of 0.51–0.89 (12). In the Bangladesh haor region, regional policy applications enabled simple NDVI-based regressions with R^2 values of 0.63 at maturity for boro rice under conditions of limited ground-truth availability (13).

The saturation limit of NDVI, however, is important for dense-canopy food crops at a peak biomass stage of growth. From a certain point, the NDVI does not undergo any further change if the biomass is further increased beyond a certain value (LAI of about 3), as the index is not sensitive to the increase in biomass, which contributes most strongly to final grain yield (5). This saturation was directly reported in rice studies in Guangdong Province (GD) (China), where the EVI and GRVI yielded useful yield-discriminating information at the growth stage where the NDVI was saturated (14). Based on the R^2 value, EVI showed better accuracy than NDVI in predicting organic rice yield in East Java, Indonesia (R^2 = 0.6773 for EVI and 0.616 for NDVI), with accuracy increasing at higher yield levels across two harvest periods (10). The implication is that NDVI is good for monitoring wheat in the early season and at regional scale, but in the case of tropical rice and in later growth stages of maize it is increasingly unsuitable as the predominant yield index and inertia in methodology may be responsible for this apparent fact still being perpetuated in the literature.

Enhanced vegetation index: Improved canopy sensitivity and its applications

To overcome two key limitations of the NDVI, the EVI was developed to add a soil adjustment factor and a blue-band atmospheric resistance term that renders it much more sensitive to changes in the canopy at high biomass (10). In general, the studies reviewed showed that for maize yield, EVI maintained the strongest and most consistent correlation with yield across the widest range of growth stages throughout the growing season, outperforming other single-VI predictors in temporal stability. The EVI exhibited the highest regional relationship to county maize yield in the US Corn Belt at early and peak growth stages and combinations of LAI, land surface temperature (LST) and EVI explained a mean R^2 of 0.67, higher than using LAI or LST alone (15). The positive modulation of the sensitivity of EVI to photosynthetic capacity (of vegetation) at critical crop growth stages demonstrates its functional advantage in high-biomass cropping systems in which NDVI loses its sensitivity.

In North Norfolk, UK, EVI showed the greatest value in April, May and June for winter barley and was the best predictor variable in the LSTM model compared to NDVI, SAVI and normalised difference moisture index (NDMI) in all seasonal windows tested (16). The data reveal that while EVI is positively related to barley yield, NDVI is negatively related. This is a highly counterintuitive result, which contradicts the practice of remote sensing systems for European cereals, where NDVI is used. This difference probably arises because EVI is more responsive to the small changes in canopy structure of the high- and low-yielding barley crops during the sensitive grain-filling stage.

Red-edge indices: The Sentinel-2 advantage

Many studies have shown that the red-edge-derived indices using the bands B5, B6 and B7, available on Sentinel-2, have delivered consistently better results in dense-canopy food crops compared to those derived from Landsat and MODIS. The red-edge spectral region has been identified as a better spectral area on which to base chlorophyll absorption as compared with the NDVI red band; thus, it is sensitive to canopy nitrogen status and photosynthetic efficiency at biomass levels where NDVI has reached saturation (5).

In Iran, the RF models that included SAVI, MSAVI and transformed vegetation index (TVI) together with the normalised difference red-edge index (NDRE) at the phenological stage of anthesis showed an R^2 of 0.90 and an RMSE of 0.29 t/ha and NDRE was among the best predictor variables of wheat yield (17). The MERIS terrestrial chlorophyll index (MTCI), which is computationally equivalent to the NDRE family and is derived from Sentinel-2 bands B5 and B7, ranked among the highest-importance predictor variables in RF models for wheat in Uzbekistan, where the full RF model achieved R^2 = 0.81 at the milk and dough stage (210–225 DAS) (18). In the maize crop in the USA, the NDRE calculated from UAV-derived red-edge bands (NIR842-RE717 ratio) was the most useful spectral band predictor when combined with other VI predictors, particularly at late vegetative and reproductive growth stages (19).

Comparing the indices mentioned above across the studies, it is important to note that the benefit of red-edge is not the same as the benefit of NDVI and that it is greatest when the crop is most dense, when the crop is in its reproductive growth stages, when most of the chlorophyll is redistributed from leaves to reproductive organs and in crops with high sensitivity to nitrogen, like rice and wheat and not in sparser dryland crops. Studies where NDRE was used selectively in the most suitable growth windows demonstrated significantly higher accuracy improvements over the application of a single NDRE as a substitute for NDVI throughout the entire growth period.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

Soil-adjusted indices in dryland systems

For dryland wheat and other crops in semi-arid regions such as Iran, Pakistan and Central Asia, where partial canopy coverage is common, NDVI may be substantially affected by soil background reflectance under sparse canopy conditions, which acts as a confounding factor in early-season yield modelling. The SAVI and its variants (MSAVI, MSAVI2) have been developed to reduce the confounding influence of the soil background reflectance on the vegetation signal (5). In several wheat studies from Iran, Pakistan and Central Asia, soil-adjusted indices showed higher predictive performance for wheat yield than NDVI alone, particularly during early- to mid-season observation periods when bare soil exposure is greatest.

In Hamedan province, Iran, GPR and RF models using SAVI, NDVI, EVI2, wide dynamic range vegetation index (WDRVI) and simple ratio (SR) as input features yielded the best result (R^2 = 0.73), with SAVI ranking among the highest-importance predictor variables (20). Including soil-adjusted indices (SAVI, MSAVI, MSAVI2, CTVI) besides NDVI in the input feature set enhanced R^2 from 0.54 at the tillering stage to 0.92 for the anthesis stage for the RF wheat model in the Qazvin plain of Iran, where during the tillering stage, the background soil bare surface is considerable (17). Effective use of VI for monitoring winter wheat in the Czech Republic (CZ) against Sentinel-2 data showed the following order of effectiveness: NDVI higher than GNDVI higher than EVI higher than NDRE higher than NRERI and the SAVI family indices worked well in fields with varying soil conditions.

Multi-index approaches and the case against single-index models

A consistent pattern across the reviewed literature is that multi-index approaches outperform single-index models when directly compared within the same study. Of the 32 studies reviewed, 14 explicitly compared multi-index with single-index inputs and reported accuracy metrics for both; in 13 of these 14 cases, the multi-index model returned a higher R^2 or lower RMSE. This pattern held across all crop types, all ML methods and all geographic regions represented in the review and was most pronounced in dense-canopy crops at reproductive growth stages.

The higher accuracy could be achieved by using the NDVI, GNDVI, MSAVI2, MTVI2, MTCI, reNDVI, EVI, SAVI and WDVI in RF models with environmental, topographic and climatic covariates, compared with any subset of the above indices in isolation in wheat studies, with the best phenological stage achieving an R^2 of 0.81 (18). In Northeast China, the CNN-LSTM-Attention model, including a new kNDVI fused with NDVI, EVI, NDWI and environmental factors, could provide better results than any single VI source or single-VI model, which implies that each VI family reflects a partially independent dimension of crop physiological status (21). In addition, for maize crops in Texas, the models using a combination of the spectral bands with VI features (band+VI) performed better than VI-only models at any crop growth stage and for each alternative ML algorithm (extra trees regressor, R^2 = 0.95 in the mid-vegetative growth stage, compared to 0.87 for VI-only) (19).

This consistent pattern can be explained mechanistically as individual VIs measure different and only part of the entire physiological spectrum of crops. The NDVI has been used to represent the general level of greenness and biomass, EVI has incorporated sensitivity to canopy structure, red-edge indices are sensitive to nitrogen and chlorophyll status at high LAI, soil-adjusted indices are designed to remove background interference and water-sensitive indices represent stress responses. None of the VIs aims at incorporating all these dimensions simultaneously and models that combine complementary indices are therefore more informationally complete and predictively robust.

Machine learning methods and predictive performance

Overview of machine learning methods across reviewed studies

The studies assessed in the paper utilised various ML and statistical techniques such as classical parametric regression, traditional ML techniques and advanced deep learning architectures. The most used ML method was RF, as reported by 25 of 32 representative studies summarising the ML methods, primary VIs and accuracy results in Table 2. In 12 studies, gradient boosting algorithms such as XGBoost and Light Gradient Boosting Machine (LightGBM) were used. In 8 studies, deep learning techniques such as CNN, LSTM and hybrid models were used, with deep learning adoption rising from 3 studies during 2018–22 to 5 studies during 2023–26, which is a near-doubling that reflects the growing availability of labelled training datasets and open-source deep learning infrastructure within the agricultural remote sensing community (3, 4).

Table 2. Summary of machine learning (ML) methods, primary vegetation indices and best accuracy metrics across representative reviewed studies

Reference	Year	Crop	Platform	Primary VIs	Best ML method	Best R^2	RMSE	Region
(17)	2026	Wheat	S2 + S1 SAR	SAVI, MSAVI, TVI	RF	0.90	0.29 t/ha	Iran
(22)	2025	Wheat	Landsat 8-9	NDVI+climate	RF	0.96	N/R	Pakistan
(18)	2023	Wheat	PS+S2	MTCI, reNDVI	RF	0.81	0.325 t/ha	Uzbekistan
(11)	2026	Wheat	S2	NDVI	RF	0.73*	MAE = 0.951	Lithuania

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

(20)	2022	Wheat	S2	SAVI, NDVI, EVI2	GPR, RF	0.73	0.28 t/ha	Iran
(23)	2025	Wheat	S2+APSIM	LAI, reflectance	LSTM	0.84	0.250 t/ha	Pakistan
(9)	2026	Rice	S2	NIR bands	XGBoost	0.739†	0.28 t/ha	Spain
(13)	2021	Rice	S2	NDVI, NDWI, MSI	ANN	0.84	N/R	Bangladesh
(24)	2023	Rice	MODIS+met	NDVI+auxiliary	Stack Ensemble	0.92*	0.328 t/ha	Nepal
(14)	2022	Rice	S2+GEE	NDVI, EVI, LSWI	RF	N/R	0.40 t/ha	China
(15)	2022	Maize	MODIS	EVI, LAI, LST	LSTM	0.67	N/R	US Corn Belt
(19)	2025	Maize	UAV	NDRE, Plant Senescence Reflectance Index PSRI	ETR	0.99‡	1.06 t/ha	Texas, USA
(25)	2025	Maize+Soy	S2	NDVI3RE, Plant Phenology Index (PPI)	RF	0.914*	N/R	Iowa and Illinois, USA
(21)	2024	Multi-crop	MODIS	kNDVI, SIF	CNN-LSTM-Att	High	N/R	NE China
(26)	2025	Sugarcane	PlanetScope	GNDVI	Heteroskedastic Gamma Regression (HGR)	0.89	N/R	Brazil
(27)	2021	Sorghum	S2+proximal	fAPAR, CI	Bayesian Ridge	0.35	N/R	Italy
(16)	2026	Barley	S2	EVI, SAVI, NDVI	LSTM	N/R	0.406 t/ha	UK

† Independent test set R^2 on 2024 validation data; training R^2 was 0.96 at 110 DAS.

* Cross-validated R^2; should be interpreted alongside RMSE and MAE for full performance context.

‡ R^2 value may reflect in-sample fit rather than an independent test set; interpret with caution.

VI: vegetation index; NDVI: normalised difference vegetation index; EVI: enhanced vegetation index; SAVI: soil-adjusted vegetation index; NDRE: normalised difference red-edge index; ML: machine learning; RF: Random Forest; LSTM: long short-term memory; CNN: convolutional neural network; PRISMA: preferred reporting items for systematic reviews and meta-analyses; RMSE: root mean square error; GEE: Google Earth Engine; MODIS: Moderate Resolution Imaging Spectroradiometer; CI: chlorophyll index; fAPAR: fraction of absorbed photosynthetically active radiation; SIF: solar-induced chlorophyll fluorescence; LAI: leaf area index; NDWI: normalised difference water index; LST: land surface temperature; MSI: moisture stress index.

Random Forest: The dominant and most reliable method

Random Forest is a supervised ensemble learning technique that builds a number of decision trees using bootstrapped samples of training data, combines predictions of these trees together, provides robustness against overfitting, is tolerant of multicollinearity between VI predictor variables and can capture nonlinear relationships between the VI and the yield without distributional assumptions (8). To predict agricultural yield, these attributes are especially advantageous since the VI inputs often measure the same phenomenon (vegetation greenness), the size of training data sets is often small and multiple biophysical and management variables control the relationship between the VI inputs and yield (4).

The R^2 values of all the studies reviewed lie between 0.70 and 0.96, in accordance with the crop type and phenological stage in which the VI is used and the methods of their selection, creating a stable and reliable range of performance. In the Pakistani wheat assessment, the highest RF accuracy was reported with the inclusion of NDVI with climate, soil and fertiliser covariate data, which gave an R^2 value of 0.9576 with regard to district-level wheat yield predictions, while Support Vector Machine (SVM) achieved an R^2 of 0.9264 on the same dataset (22). This is a case in point where the competitive but lower SVM performance on this dataset matches a well-established pattern that both methods do well with small-sized, tabular agricultural datasets, but both ensemble averaging with RF and SVM will be more stable over different partitioning of data. In a study covering Iowa and Illinois, the highest in-sample R-squared for the RF model across maize and soybean prediction was 0.914 (maize, 2022), but the spatially independent cross-validation R-squared peaked at only 0.409 across all years and crops (25), which the authors attributed to spatial autocorrelation in the training images rather than an inherent model failure. This key observation reinforces the need for spatially independent validation in agro-remote sensing studies and was secondarily reported in the literature reviewed.

Support vector machines: Performance in data-scarce contexts

Support vector machines (SVM) determine the best possible separating hyperplane in the transformed feature space using kernel functions and have proven to be consistently useful in the field of yield prediction in agriculture, especially when the size of the training datasets is small, which makes them suitable for the data-poor context in developing countries where the cost of collecting ground-truth information on agricultural yields is high (4). Wheat yield prediction in Multan district of

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

Pakistan using SVM with other changing climatic variables, NDVI data derived from Landsat images and soil data yielded an R^2 of 0.9264, which is quite good in the context of a district-level study with limited training observations, performing better than least absolute shrinkage and selection operator (LASSO) regression and close to RF. Support vector machines achieved an R^2 of 0.39 at tillering and 0.82 at anthesis for Iranian wheat in Hamedan, consistent with the performance of RF on the same dataset but with slightly higher RMSE values at all stages (20).

The main drawback of SVM in this context is that it operates as a black box, which means that, unlike RF, where variable importance can be directly linked with VI, in the case of SVM, decision boundaries are defined in high-dimensional kernel space, making them difficult to interpret in an agricultural context. This becomes an important limitation when using models for informing agronomic decisions for management or communicating to policymakers and therefore still favours RF over SVM in applied precision agriculture situations where the accuracy is the same (8).

XGBoost and gradient boosting methods

For datasets with irregular missing data patterns, complex feature interactions and high-dimensional input spaces, models built by XGBoost and other related gradient boosting approaches that build models sequentially, training on the residual errors generated from the previous models built by the ensemble, have become strong competitors to RF. Considering the reviewed studies, the evidence suggests that XGBoost has a clear strength over RF, that is, generalisation to truly unseen data from other years or locations, which is not necessarily superior in terms of accuracy on training distributions.

This was most evident in the Valencia rice study, where both models obtained very high training R^2 values (close to 0.96) at 110 days after sowing (DAS), but XGB was much more accurate on independently collected 2024 validation data (test R^2 = 0.739 vs 0.664 for RF), highlighting XGB's better ability to regularise and avoid overfitting to high-dimensional spectral input data (9). From a practical point of view, the takeaway is that when a model is trained on historical data and used for operational yield prediction systems on a year-to-year basis, XGB could be a better choice than RF despite the noticeable in-sample accuracy metrics. The ensemble approach performed better than any single algorithm, with a mean RMSE of 0.328 t/ha and MAE of 0.317 t/ha, achieved for rice yield prediction in Nepal using NDVI and auxiliary meteorological data, highlighting the importance of model diversity in capturing the influence of spatially heterogeneous metrics on rice yield prediction in a complex topography (24).

Deep learning: Convolutional neural network, long short-term memory and attention architectures

Some reviewed studies reported high predictive performance using deep learning; however, these methods require large, labelled training datasets, which, at their core, limit their usefulness when dealing with low-data-volume production agriculture signals. Convolutional neural network architectures, which themselves capture hierarchical spatial features directly from satellite imagery, are appropriate for within-field yield mapping where spatial texture and crop row pattern contain information for prediction. As a type of recurrent neural network (RNN) tailored for sequential data, LSTM networks are especially well placed to capture temporal dynamics throughout the growing season, as well as how the accumulated VI signals, as they pass through phenological stages, are translated into final yield (3).

The best deep learning result identified in the reviewed corpus was for an LSTM model predicting winter barley in North Norfolk (UK), which achieved an RMSE of 0.406 t/ha, substantially outperforming the CNN model applied to the same dataset, which recorded an RMSE of 8.579 t/ha across all tested seasonal windows (16). This difference was attributed to LSTM's capacity to integrate the full EVI and SAVI time series from November through June, rather than relying on single-date spectral inputs. The CNN RMSE value, which approaches the magnitude of typical UK winter barley yields (5–10 t/ha), reflects the fundamental unsuitability of a 1D CNN architecture for temporal sequence modelling of crop phenology, a result consistent with the broader finding that temporal sequence models outperform spatial feature extractors for deterministic grain yield prediction in temperate cereals. This difference was attributed to LSTM's capacity to integrate the full EVI and SAVI time series from November through June, rather than relying on single-date spectral inputs. This discovery serves as straightforward evidence that temporal sequence modelling and no single-date spectral optimisation, should be the proper modelling paradigm for deterministic grain yield prediction for temperate cereals. The CNN-LSTM-Attention model, which integrates the kernel NDVI in addition to solar-induced fluorescence (SIF) and environmental variables, outperformed both RF and XGBoost in a multi-crop prediction task covering maize, rice and soybean in Northeast China, which suggests that the attention mechanism can improve accuracy in a multi-crop prediction task when focusing on the most informationally important time steps of the VI time series compared with regular LSTM architecture models (21).

In the agricultural setting, transfer learning is a crucial emerging solution to the data challenge of deep learning. In this study, previous researchers showed how CNN models trained on US soybean MODIS data, after being transferred to Argentina through spatio-temporal alignment as shown in L2-SP and Batch Spectral Shrinking regularisation, resulted in a 19 % improvement in soybean yield prediction in terms of RMSE and a 39 % improvement in terms of R^2 against models trained on MODIS data for Argentina alone, thereby providing proof of how knowledge encoded in data-rich agriculture regions could be meaningfully transferred to more data-scarce regions (28). This finding directly tackles Gap 5 outlined in the systematic gap analysis and indicates a feasible path forward to enable the continuation of deep

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

learning capacity development in the least covered areas of smallholder farming.

Google earth engine as a processing enabler

The satellite data processing and VI extraction platform GEE was used as the primary processing platform in several reviewed studies, including studies on Chinese rice (14), Iranian wheat (17), multi-crop Northeast China (21) and Nepal rice (24). Cloud-based architecture removes the necessity of local storage of vast archives of satellite images, removes the need for automated cloud masking and VI time series extraction from the local environment and offers access to pre-processed analysis-ready data products. In low-resource institutional settings, especially in South Asia and Sub-Saharan Africa where the necessary computing infrastructure is not available, GEE has significantly reduced the technical threshold for research on yield prediction based on satellite data.

Crop-wise findings and regional patterns

Rice: The most extensively studied food crop

Rice is the best-documented food crop found in the available literature and is included in seven major studies in Bangladesh, Nepal, China, Spain and Indonesia (9, 10, 13, 14, 24, 29, 30). Of the seven rice studies reviewed, five used Sentinel-2 as the primary satellite platform and four applied RF as the primary ML method, making this the most common methodological combination for rice yield prediction in the reviewed literature. This reflects Sentinel-2's comparatively frequent revisit cycle relative to Landsat's 16-day repeat and its red-edge spectral bands that offer advantages over MODIS data at 250 m resolution for field-scale rice monitoring.

The main finding arising from the key VI experiments for rice is that, in general, EVI and the red-edge-based indices have outperformed NDVI only during the reproductive and grain-fill stages, when rice canopies typically develop into dense canopies, often developing a high LAI before maximum yield variation is expressed (5, 16). In the case of boro rice in Bangladesh, only an R^2 value of 0.84 was obtained at the maturity stage based on the multi-index (NDVI, NDWI, RGVI, MSI and LAI), ANN model, which is considerably superior to the single-index (NDVI) regression models (13). This was well illustrated in the Valencia rice study, whose RF and XGBoost models had steadily increased accuracy between 35 DAS and 110 DAS, at the latter of which, the XGBoost model achieved a training R^2 close to 0.96 at 110 DAS, with an independent test R^2 of 0.739 on separately collected 2024 validation data for japonica rice varieties (9). A spatial heterogeneity analysis showed that field edges and headlands were systematically less well predicted by all models, which can be attributed to the spatial limitations of combine yield monitoring, but also to mixed-pixel effects at the field edges in the 10 m imagery.

Geographic imbalances found by regional distribution analysis in rice yield prediction studies are important. High-quality mechanised farming systems studies are available for Spain, China and Nepal, as well as the haor study in Bangladesh, representing small-scale rice systems which form the backbone of rice production and food security risk in South and Southeast Asia (13). The Bangladesh study is limited to one wetland rice system (boro rice) and the results cannot be extrapolated to upland, rain-fed and transplanted rice systems that comprise most of the rice supply in India, Myanmar, Cambodia and Vietnam.

Wheat: Strong literature across diverse geographies

Among the various food crops reviewed, wheat is the next in the list with seven studies mainly performed in Iran, Pakistan, Lithuania, Uzbekistan and the Czech Republic (11, 12, 17, 18, 20, 22, 23). There is greater geographic variability in wheat research compared to rice, because wheat is grown in more climatic zones and plays a major role in food security in the Global North. This geographic variation also introduces considerable heterogeneity in VI performance quality, phenological timing and optimal modelling approach.

The main drawback of all wheat studies is the interference of phenological stage in imagery, irrespective of its influence on the prediction quality as compared to other VI or ML methods. The heading-to-grain-fill stage (physiologically around anthesis) always provided the best VI-yield relationship and model accuracy in all the areas investigated. The R^2 for RF in anthesis in Iran was 0.90, while it was 0.50 in the tillering stage (17). In Uzbekistan, the milk and dough stage (210–225 DAS) resulted in an RF R^2 of 0.81, whilst the emergence and early tillering stages yielded R^2 values below 0.50 (18). May NDVI was the main predictor among all growth stages in Lithuania (11). All the above show that systematic phenological targeting is a best practice for any satellite-derived wheat monitoring system, irrespective of the VIs or ML method used.

The second feature observed consistently is that while in homogeneous soil, high-fertility wheat systems, NDVI is as good as red-edge or soil-adjusted indices, in semi-arid or mixed soil wheat systems, the red-edge or soil-adjusted indices are superior. The difference between the Lithuanian and Iranian wheat studies is illustrative: in Lithuania, soils of sandy loam and silt loam texture support relatively uniform canopy development and an MAE of only 0.951 t/ha is obtained for an RF model using only NDVI as an input, whereas the Qazvin Plain of Iran is semi-arid and shows a significant variation in canopy development, which required soil-adjusted indices to get good early-season predictions (17).

Maize: Growing literature with emphasis on the United States

Maize yield prediction studies in the literature under review are clustered within the United States, with some work conducted in Northeast China, for a total of four studies (15, 19, 25, 31). The dominance of studies in the US Corn Belt is due to the high-quality county-level yield data available through USDA-NASS, as well as the presence of large and technologically advanced farms,

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

where data can be collected using UAVs, particularly through multispectral data acquisition.

The key result for maize is that temporal modelling of the growing season is much more effective than single-date or aggregated VIs. The mean R^2 of 0.67 achieved by the LSTM models using all monthly MODIS-EVI/LAI/LST time-series inputs in the US Corn Belt outperformed the results for RF (0.65), SVR (0.55) and LASSO (0.53) (15). The adoption of LAI and LST alongside EVI showed that not only canopy structure but also water stress information is important, with both affecting maize yield and being only partially represented by individual VIs. In RF models, combining features obtained from soil data (sand, silt, clay percentages, slope) with the spectral bands and VIs extracted from UAV images reduced the prediction error for precision maize prediction at the production scale in the southern US compared to all-VI RF models. The bands and indices that proved to be more informative were the triangular greenness index (TGI), HUE and green-normalised difference red-edge (GNDRE) during the tasselling stage (31). The UAV-based extra trees regressor applied achieved R^2 = 0.99 at the mid-vegetative stage (19). However, it is unclear whether this reflects an independent test set or an in-sample fit and this figure should be interpreted cautiously given the study's stated focus on robustness across varied growing environments.

Other food crops: Emerging literature and significant gaps

In addition to rice, wheat and maize, the reviewed literature includes studies on soybean (three; 6, 28, 32), sugarcane (one) (26), sorghum (two) (2, 27) and barley (one) (16). These crop types comprise a large proportion of the underrepresented global food crop portfolio and the emerging evidence from these studies points to important methodological and regional insights that warrant consideration.

Early researchers directly evaluated whether high-spatial-resolution imagery was better for yield prediction at the within-field scale in soybean (6). Results indicated that the accuracy of within-field soybean yield variability mapping by PlanetScope (PS) at 3 m was not consistently better than that of Sentinel-2 (S2) at 10 m, as the better spectral quality and radiometric consistency of both the SWIR and red-edge bands of S2 compensated for the coarser spatial resolution. This suggests that, especially when investing in commercial high-resolution imagery, it may not be worthwhile to use these data for soybean yield monitoring if Sentinel-2 data are optimised for the task and are freely available.

An R^2 of 0.89 was obtained for a heteroskedastic gamma regression model with mean GNDVI as the main variable during the growth phase, demonstrating that this method provided the best model for sugarcane yield in São Paulo State, Brazil (26). The comparison of a specialised statistical model with standard ML models highlights the heteroskedastic variance structure of the sugarcane yield dataset, which means that the variance of the yields increases with their means. This finding reinforces the overall message that the most prevalent methods in the agricultural remote sensing literature, namely RF and CNN, are not necessarily the best; characteristics of crop-specific data such as variance structure, temporal dynamics and phenological diversity may be better suited to other methods.

The results of the commercial farm study in Italy by Habyarimana and Baloch, which reported a cross-validated R^2 of 0.35 for biomass estimation from Sentinel-2-derived fraction of absorbed photosynthetically active radiation (fAPAR) and a proximal chlorophyll concentration index (27) and the smallholder study in Mali, which reported a calibrated R^2 of 0.25 or lower for satellite-based sorghum yield estimation (2), highlight the basic challenges of yield prediction in a rain-fed setting where smallholders are the producers. The low explanatory power of the satellite VI signal in these contexts is due to a combination of factors, including high yield variance within smallholder plots, mixed-pixel contamination across smallholder fields and variability in the harvest index, which confounds the signal from spectral biomass and objective measures of yield.

Regional distribution and the smallholder gap

The geographic distribution of the 32 studies reviewed highlights a strong geographic clustering of studies in North America, Western Europe and East Asia, accounting for approximately 65 % of the studies included (4, 5). These areas are typified by large-scale, mechanised, profit-oriented agriculture, accurate and high-quality yield-measurement devices, a well-developed database for agronomic research and institutional resources for research based on satellite information. These contexts are disproportionately represented in the literature, creating a systematic global evidence base that reflects farming systems in these regions rather than those experiencing the highest levels of food insecurity.

In Sub-Saharan Africa, there is only one sorghum study in Mali (2). In South Asia, although hundreds of millions of smallholder farmers face food insecurity, the studies are mainly from Pakistan and Bangladesh and even then, the studies predict yields at the district-level aggregate rather than monitoring individual smallholder plots. With more than 500 million people depending on rice cultivation in complex landscapes in the region, where rice diversity is widespread and smallholders dominate, Southeast Asia is largely underrepresented in the reviewed literature, apart from being mentioned indirectly in global review papers.

Satellite platforms and data considerations

Sentinel-2: The dominant platform

The most favoured satellite platform in the 32 studies reviewed was Sentinel-2, with more than 65 % of the studies using this platform. This is due to its combination of attributes that are not found together in any other freely available platform (5). The spatial resolution of 10 m in visible and NIR bands can be used for field-level analysis on farms larger than around 0.5 ha, including most commercial and semi-commercial farms. The revisit cycle

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

at equatorial latitudes is five days, which provides sufficiently high temporal sampling of important phenological intervals despite the omission of some scenes that are contaminated by clouds. The three dedicated red-edge bands (B5, B6, B7) support the calculation of NDRE, MTCI and other chlorophyll-sensitive indices, which have consistently been shown to perform better in dense-canopy food crops than the NDVI derived from other platforms.

Landsat and MODIS: Established but resolution-limited

The 30 m spatial resolution of Landsat is suitable for wheat and maize mapping but may not be high enough to analyse within-field yield variability in smallholder farming, where fields or parcels may be smaller than a single Landsat pixel (14). MODIS is only available at a 250–500 m resolution and is limited to yield prediction at the county level or above, since within-field or smallholder plot analyses cannot account for field-level heterogeneity (17, 26). The most important continuing value of both Landsat and MODIS platforms is their rich time series dating back to 1972 and 2000, respectively, allowing for long-term trend analysis and pre-Sentinel-2 climate-yield relationship modelling.

Multi-sensor fusion and its accuracy benefits

Sentinel-2 optical imagery coupled with radar data acquired from Sentinel-1 synthetic aperture radar is currently the most promising near-term approach to improving the accuracy of yield prediction, especially in cloudy tropical regions where the number of observations from optical sensors is limited during the growing season. The SAR C-band backscatter is less affected by atmospheric clouds and is sensitive to canopy characteristics and soil water conditions that cannot be measured by optical sensors (17). In the Iranian wheat study, the use of Sentinel-1 VV and VH polarisation data with 25 features derived from Sentinel-2 data and topographic data in an RF model led to a slight decrease in training RMSE from 0.14–0.13 t/ha, but the test RMSE increased slightly from 0.29–0.33 t/ha, indicating that the addition of SAR data provided some extra information but also introduced noise in the dry situations of the study area (17). Multi-source fusion approaches that combined optical VIs, precipitation, evapotranspiration and temperature with soil data performed better than optical-only models across a variety of crops under diverse conditions (20, 24, 26).

Cloud cover as an operational limitation

Cloud contamination of optical satellite imagery is the limiting operational constraint for VI-based yield monitoring in humid tropical and monsoon-dominated agricultural areas. Sentinel-2 Level-2A images were used for boro rice monitoring in Bangladesh under a range of cloud-cover conditions (from 2.4–26.9 %) and careful scene selection was applied, thereby decreasing the temporal sampling density (13). In Nepal, stacking ensembles were compared with models that relied only on meteorological parameters for forecasting because of gaps in satellite observations of the NDVI signal during monsoon cloud cover (24). These constraints affect the same tropical and subtropical zones where food security issues are strongest and where many satellite-based agricultural monitoring systems are disadvantaged.

4. Discussion

The systematic review summarised existing literature on 32 peer-reviewed studies investigating the integration of satellite-derived vegetation indices and machine learning techniques for food crop yield prediction. The results indicate that this is a rapidly growing area of research and that there are identifiable methodological trends, even though significant structural deficiencies limit the generalisability of current research.

Standardised benchmarking remains an unresolved challenge

The most salient constraint that restricts synthesis in the reviewed literature is the lack of standardised accuracy reporting. Of the 32 studies analysed, 29 reported R^2, 24 reported RMSE, 15 reported normalised RMSE (nRMSE) and four reported MAPE. Many reported only a subset of these metrics and used different spatial and temporal units for yield reporting (4, 5). Due to this heterogeneity, it is unclear whether it is possible to answer seemingly simple questions, such as which combination of VI and ML yields the best predictions for wheat, or whether deep learning is consistently better than RF for rice, without carefully considering the limitations of comparing different metrics and validation settings. A standardised framework specifying the minimum set of variables to report, such as R^2, RMSE, nRMSE, spatial resolution, phenological stage of imagery and spatial validation independence, would greatly facilitate the collective analysis of the research field in the future and make it possible to compare studies more reliably.

Geographic bias toward high-resource agricultural systems

High-accuracy studies have been largely concentrated in North America, Western Europe and East Asia, leaving the geographic scope of the study too limited to represent the agricultural regions and systems where yield prediction is most needed. This is due to an inherent problem in the data ecosystem of satellite-based models in the context of smallholder farming systems and is not directly attributable to the inability of the VI or the ML methods themselves. There is a great deal of heterogeneity in field management; multiple crop varieties differing in spectral signature are often inter-planted; plot boundaries are not as precisely delineated; and ground-truth yield measurements in smallholder systems are less accurate. This gap requires efforts to maintain or establish the infrastructure for collecting ground-truth data in South Asia, Southeast Asia and Sub-Saharan Africa alike, rather than necessarily refining the methods for more data-rich continents like North America or Europe.

Inconsistency in phenological stage selection

The crop phenological stage has been recognised as the key factor affecting the accuracy of VI yield models, yet the combinations of phenological stages selected for the extraction of VIs have not been consistent in the reviewed

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

studies, even for the same crop. While heading-to-grain-fill in wheat is often cited as the most critical period, the dates, growth staging codes and VI extraction methods have differed enough between studies that quantitative cross-study comparison of VI performance at the same growth stage is difficult to perform in practice. A feasible methodological step that could significantly increase the comparability of studies for the same crop is the development of crop-specific phenological VI extraction protocols standardised according to Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) growth stage codes and verified across different seasons and genotypes.

Model transferability and multi-source data integration

Nearly all the studies reviewed were conducted and evaluated in a single geographical area and were not designed to be applicable to other agro-climatic situations. Therefore, they did not evaluate whether the VI-yield relationships and ML model structures could be transferred. Previous researchers transfer learning study indicated a 19 % RMSE improvement for CNN models transferred from US soybean regions with high data availability to Argentina, while also highlighting the risk of negative transfer if the data are not regularised (28). Moreover, most research involving only VIs as input was found to underperform research that combined VIs with soil properties and meteorological and agronomic management data. Early results reported an R^2 improvement of 45 % for corn when soil properties were added to VI inputs (31). The optical data alone were not adequate for Pakistani wheat without climatic and fertiliser covariates (22) and found R^2 improvements of 0.05–0.08 when topographic and climatic inputs were added to optical-only models (18). The implementation of standardised data fusion pipelines within a GEE workflow to combine Sentinel-2 VI with gridded climate products, SoilGrids data and crop calendars would make integrated modelling with multi-source data accessible to a much larger research community.

Limitations

The review is subject to multiple limitations that need to be considered when appraising its results. First, there is a heavy bias towards studies conducted in North America, Western Europe and East Asia, which restricts the applicability of the conclusions to smallholder and tropical farming systems. Second, due to the lack of raw data, the synthesis relies solely on reported measurement accuracy metrics, the type, scale and validation design of which differ between studies and do not provide a formal basis for a meta-analysis. Third, the studies used a variety of VIs, phenological stages and spatial resolutions, which makes it hard to perform cross-study performance comparisons, even for the same crop type. Fourth, there is probably a publication bias, with an overemphasis on good model fits, because studies reporting poor model fits may be less likely to be published. Lastly, the quality rubric used in this instance was created specifically for this review and the scores are meant to be indicative rather than definitive.

Future directions

The synthesis raises a few high-priority research directions. First, multiple seasons of independent data from climatically similar regions with different geographical locations are urgently needed to validate VI-ML yield prediction models and determine the actual boundaries of their transferability. Second, some of the accuracy benefits shown by LSTM, CNN and attention-mechanism architectures are out of reach of research communities in food-insecure areas because of their need for large, labelled training sets. The three most promising avenues for making deep learning-based yield prediction accessible where it can most benefit from it are transfer learning, synthetic data augmentation from process-based crop models and participatory ground-truth data collection frameworks in smallholder regions (29). Third, future research should focus on establishing and standardising data-fusion pipelines that combine vegetation indices derived from Sentinel-2 data with data from SoilGrids, gridded meteorological data and national crop calendars in cloud-based processing platforms like Google Earth Engine. Finally, understanding of smallholder agriculture in Sub-Saharan Africa and South Asia also needs to be improved to ensure that satellite yield monitoring can be translated into more meaningful food security intelligence for the systems and farming communities that will most benefit from it.

5. Conclusion

Based on the findings of this systematic review (2018–26), 32 peer-reviewed studies have shown that, for major food crops, vegetation indices (VIs) derived from satellites and linked to a machine learning (ML) algorithm are a powerful and now widely adopted prediction tool, with important aspects of performance dependent on the crop, VIs used, phenological stage, geographic region and ML model.

In a variety of study settings, Random Forest (RF) is consistently effective and applicable across food crops and geographic settings, with R^2 ranging from 0.70–0.96. The use of XGBoost exhibits some strengths in terms of out-of-sample generalisation, while deep learning approaches (long short-term memory (LSTM) for temporal series and convolutional neural network (CNN)-LSTM attention hybrids for multi-source approaches) deliver the best absolute accuracy with sufficient training data. Overall, multi-index methods are superior to single-index methods for all crop types and the red-edge indices based on the specific spectral bands in Sentinel-2 data are superior to the normalised difference vegetation index (NDVI) for dense-canopy rice and wheat, especially during their reproductive growth stages. Sentinel-2 has emerged as the platform to rely on and the use of Google Earth Engine (GEE) has made computational resources and processing capabilities accessible to everyone.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

There is, however, a notable paucity of research, particularly regarding tropical smallholder crops and Sub-Saharan African agricultural systems and there is an overrepresentation of rice and wheat. The most easily realised enhancements to improve the collective value of this research field are the standardisation of accuracy reporting, phenological extraction protocols and cross-regional validation frameworks. The biggest methodological breakthroughs for advancing future research are the incorporation of VIs and data on soil, climate and management practices and the use of transfer learning techniques to expand the capabilities of deep learning methods across data-poor areas. While VI and ML systems based on satellite data have real transformative potential for monitoring food security globally, equitably realising that potential requires intentional investment in the most under-represented agricultural systems and research communities in the evidence base.

Acknowledgements

The authors sincerely acknowledge the Department of Remote Sensing and GIS, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India, for providing academic support, guidance and the necessary resources during the preparation of this review manuscript.

Authors' contributions

The idea for the review was conceived by SDSJ, who undertook the literature search and analysis and prepared the manuscript. KR guided the research and critically revised the manuscript. MD, PPC and BK were involved in designing the study, analysing the data and revising the manuscript. NRM helped with the literature search and manuscript editing. VS contributed to the preparation of the manuscript. All authors have read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest

Authors do not have any conflict of interests to declare.

Ethical issues

None

References

1. Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: a meta-review. Remote Sens Environ.2020;236:111402. http://doi.org/10.1016/j.rse.2019.111402
2. Boell DB, Di Tommaso S, You C, Djima IY, Burke M, Kilic T. Sight for sorghums: comparisons of satellite- and ground-based sorghum yield estimates in Mali. Remote Sens. 2020;12(1):100. http://doi.org/10.3390/rs12010100
3. Muruganantham P, Wibowo S, Grandhi S, Samrat NH, Islam N. A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sens. 2022;14(10):1990. http://doi.org/10.3390/rs14091990
4. Shawon SM, Ema FB, Mahi AK, Niha FL, Zubair HT. Crop yield prediction using machine learning: an extensive and systematic literature review. Smart Agric Technol. 2025;10:100718. http://doi.org/10.1016/j.atech.2024.100718
5. Narimani M, Pourreza A, Moghimi A, Farajpoor P. Sentinel-2 for crop yield estimation: a systematic review. Smart Agric Technol.2026;14:102405. http://doi.org/10.1016/j.atech.2026.102405
6. Amankulova K, Farmonov N, Akramova P, Tursunov I, Mucsi L. Comparison of PlanetScope, Sentinel-2 and Landsat 8 data in soybean yield estimation within-field variability with random forest regression. Heliyon. 2023;9(6). http://doi.org/10.1016/j.heliyon.2023.e17432
7. Ji Z, Pan Y, Zhu X, Wang J, Li Q. Prediction of crop yield using phenological information extracted from remote sensing vegetation index. Sensors. 2021;21(4):1406. http://doi.org/10.3390/s21041406
8. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. Sensors. 2018;18(8):2674. http://doi.org/10.3390/s18082674
9. Simeón R, Agenjos-Moreno A, Rubio C, Uris A, San Bautista A. Rice yield prediction model at pixel level using machine learning and multi-temporal Sentinel-2 data in Valencia, Spain. Agriculture. 2026;16(2):201. http://doi.org/10.3390/agriculture16020201
10. Pratiwi GR, Indarto I, Hakim FL, Sulistiono W, Roivansah MF, Helmanto H, et al. Rice yield estimation using vegetation indexes derived from Sentinel-2 imagery for sustainable agriculture. Discov Sustain. 2025;6:1048. http://doi.org/10.1007/s43621-025-01743-3
11. Okupska E, Juostas A, Gozdowski D, Wójcik-Gront E. Field-scale prediction of winter wheat yield using satellite-derived NDVI. Agronomy. 2026;16(6):670. http://doi.org/10.3390/agronomy16060670
12. Mezera J, Lukas V, Horniacek I, Smutny V, Elbl J. Comparison of proximal and remote sensing for the diagnosis of crop status in site-specific crop management. Sensors. 2022;22(1):19. http://doi.org/10.3390/s22010019
13. Islam MM, Matsuba S, Noguchi R, Ahamed T. Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. Remote Sens Appl Soc Environ. 2021;22:100494. http://doi.org/10.1016/j.rsase.2021.100494
14. Choudhary K, Shi W, Dong Y, Paringer R. Random forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. Adv Space Res. 2022;70(9):2443–57. http://doi.org/10.1016/j.asr.2022.06.073
15. Ji Z, Pan Y, Zhu X, Zhang D, Dai J. Prediction of corn yield in the USA Corn Belt using satellite data and machine learning: from an evapotranspiration perspective. Agriculture. 2022;12(8):1263. http://doi.org/10.3390/agriculture12081263
16. Li H-Y, Lawrence JA, Mason PJ, Ghail RC. A framework for accurate annual regional crop yield prediction. Remote Sens. 2026;18(8):1157. http://doi.org/10.3390/rs18081157

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.
https://doi.org/10.14719/pst.17573

17. Navidi MN, Fazli E, Kharazmi R, Seyedmohammadi J. Wheat yield prediction using integrated optical and radar remote sensing with machine learning across key phenological stages. Sci Rep. 2026;16:10470. http://doi.org/10.1038/s41598-026-41501-7

18. Farmonov N, Amankulova K, Szatmári J, Urinov J, Narmanov Z, Nosirov J, et al. Combining PlanetScope and Sentinel-2 images with environmental data for improved wheat yield estimation. Int J Digit Earth. 2023;16(1):847–67. http://doi.org/10.1080/17538947.2023.2186505

19. Sapkota BR, Baath GS, Flynn KC, Adhikari K, Hajda C, Smith DR. Machine learning algorithms for maize yield prediction with multispectral imagery: assessing robustness across varied growing environments. Sci Remote Sens. 2025;12:100267. http://doi.org/10.1016/j.srs.2025.100267

20. Ashourloo D, Manafifard M, Behifar M, Kohandel M. Wheat yield prediction based on Sentinel-2, regression and machine learning models in Hamedan, Iran. Scientia Iranica. 2022;29(6):3230–43. http://doi.org/10.24200/sci.2022.57809.5429

21. Lu J, Li J, Fu H, Tang X, Liu Z, Chen H, et al. Deep learning for multi-source data-driven crop yield prediction in Northeast China. Agriculture. 2024;14(6):794. http://doi.org/10.3390/agriculture14060794

22. Ashfaq M, Khan I, Afzal RF, Shah D, Ali S, Tahir M. Enhanced wheat yield prediction through integrated climate and satellite data using advanced AI techniques. Sci Rep. 2025;15:18093. http://doi.org/10.1038/s41598-025-02700-w

23. Ishaq RAF, Zhou G, Jing G, Shah SRA, Ali A, Imran M, et al. Geospatial robust wheat yield prediction using machine learning and integrated crop growth model and time-series satellite data. Remote Sens. 2025;17(7):1140. http://doi.org/10.3390/rs17071140

24. Islam MD, Di L, Qamer FM, Shrestha S, Guo L, Lin L, et al. Rapid rice yield estimation using integrated remote sensing and meteorological data and machine learning. Remote Sens. 2023;15(9):2374. http://doi.org/10.3390/rs15092374

25. Radočaj D, Plaščak I, Jurišić M. Phenology-based maize and soybean yield potential prediction using machine learning and Sentinel-2 imagery time-series. Appl Sci. 2025;15(13):7216. http://doi.org/10.3390/app15137216

26. Vasconcelos JCS, Arantes CS, Speranza EA, Antunes JFGC, Barbosa LAFB, Cançado GDMA. Predicting sugarcane yield through temporal analysis of satellite imagery during the growth phase. Agronomy. 2025;15(4):793. http://doi.org/10.3390/agronomy15040793

27. Habyarimana E, Baloch FS. Machine learning models based on remote and proximal sensing as potential methods for in-season biomass yields prediction in commercial sorghum fields. PLoS ONE. 2021;16(3). http://doi.org/10.1371/journal.pone.0249136

28. Huber F, Inderka A, Steinhage V. Leveraging remote sensing data for yield prediction with deep transfer learning. Sensors. 2024;24(3):770. http://doi.org/10.3390/s24030770

29. Morales A, Villalobos FJ. Using machine learning for crop yield prediction in the past or the future. Front Plant Sci. 2023;14:1128388. http://doi.org/10.3389/fpls.2023.1128388

30. Ajay PM, Ragunath KP, Pazhanivelan S, Muthumanickam D, Sivamurugan AP, Vanitha G. Rice yield predictions using remote sensing and machine learning algorithms: a review. Plant Sci Today. 2025; Early Access. http://doi.org/10.14719/pst.5976

31. Sarkar S, Osorio Leyton JM, Noa-Yarasca E, Adhikari K, Hajda CB, Smith DR. Integrating remote sensing and soil features for enhanced machine learning-based corn yield prediction in the Southern US. Sensors. 2025;25(2):543. http://doi.org/10.3390/s25020543

32. Yoosefzadeh-Najafabadi M, Tulpan D, Eskandari M. Using hybrid artificial intelligence and evolutionary optimization algorithms for estimating soybean yield and fresh biomass using hyperspectral vegetation indices. Remote Sens. 2021;13(13):2555. http://doi.org/10.3390/rs13132555

Additional information

Peer review: Publisher thanks the Sectional Editor and anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information: Available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy.

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting.

Copyright: © The Author(s). This is an open-access article distributed under the Creative Commons Attribution License CC BY 4.0.

Publisher information: Plant Science Today is published by Horizon e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.

© The Author(s) 2026. Creative Commons CC BY 4.0. Published by Horizon e-Publishing Group.

https://doi.org/10.14719/pst.17573