

Review

Not peer-reviewed version

Machine Learning and Deep Learning in Agriculture: A PRISMA Systematic Review of Architectures, Applications, and Open Science Practices (2019–2026)

[Azad Rasul](#) *

Posted Date: 9 April 2026

doi: 10.20944/preprints202604.0648.v1

Keywords: systematic review; precision agriculture; convolutional neural network; vision transformer; crop yield prediction; open science



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Machine Learning and Deep Learning in Agriculture: A PRISMA Systematic Review of Architectures, Applications, and Open Science Practices (2019–2026)

Azad Rasul ^{1,2}

¹ Department of Geography, Soran University, Soran, Erbil, Iraq; azad.rasul@soran.edu.iq

² Department of Forestry, College of Agricultural Engineering Sciences, Salahaddin University-Erbil, 44002, Iraq

Abstract

Agriculture faces compounding pressures from food insecurity, climate change, and resource scarcity, creating urgent demand for scalable analytical tools. This PRISMA 2020-compliant systematic review synthesises 582 peer-reviewed studies on machine learning (ML) and deep learning (DL) applications in agriculture, drawn from Scopus for the period January 2019 to March 2026. The 2026 data cover only the first quarter (January–March) and are therefore not directly comparable to full-year counts. Publication volume grew exponentially — from 6 papers in 2019 to 251 in 2025 — driven by the adoption of convolutional neural networks (CNNs), Vision Transformers (ViT), and YOLO-based object detectors. Plant disease detection (27.0%) and crop yield prediction (13.7%) dominated the application landscape. South Asia and East Asia together contributed 59.3% of the corpus, while Sub-Saharan Africa and Latin America each accounted for only 1.4%, revealing a profound mismatch between research output and global food insecurity burden. Median reported classification accuracy was 98.1% for disease detection, largely reflecting controlled laboratory datasets rather than field conditions. Median R^2 was 0.823 for yield prediction, based on 22 of 80 yield studies reporting this metric. Unit heterogeneity, dataset artefacts, and inconsistent evaluation practices limit cross-study comparability and the real-world interpretability of these figures. Open science practices remain critically low: only 7.7% of papers shared code and 14.1% shared data openly. Explainable AI, federated learning, and physics-informed modelling represent emerging frontiers. The review identifies benchmark standardisation, smallholder-relevant design, and geographic equity as the field's most pressing unresolved challenges.

Keywords: systematic review; precision agriculture; convolutional neural network; vision transformer; crop yield prediction; open science

1. Introduction

1.1. The Global Agricultural Challenge

Agriculture stands at the centre of humanity's most pressing challenges. According to the State of Food Security and Nutrition in the World 2023 report, around 733 million people—roughly one in eleven globally and one in five in Africa—faced chronic hunger in 2023 (FAO et al., 2023). The world is falling significantly short of achieving Sustainable Development Goal 2 (Zero Hunger) by 2030. These pressures are compounded by accelerating climate change, increasing strain on freshwater and soil resources, and declining agrobiodiversity. Meeting the demands of a global population projected to reach nearly ten billion by mid-century will also require sustainable intensification of production (FAO et al., 2023). These interacting pressures require transformative advances in how agricultural systems are monitored, managed, and optimised at scale.

Traditional approaches to crop monitoring, disease surveillance, and yield forecasting rely on labour-intensive field sampling, expert visual inspection, and empirical models that generalise poorly across diverse agro-ecological zones. The proliferation of high-resolution remote sensing platforms—including the Sentinel and Landsat satellite series, unmanned aerial vehicles (UAVs), and ground-based Internet of Things (IoT) sensor networks—has generated an unprecedented volume and variety of spatiotemporal agricultural data. However, the analytical methods required to translate this data proliferation into actionable agronomic intelligence have historically lagged behind the capacity to collect it.

1.2. *The Rise of Machine Learning and Deep Learning in Agriculture*

Machine learning (ML) and, more recently, deep learning (DL) have emerged as the analytical paradigms most capable of extracting structured knowledge from large, heterogeneous agricultural datasets. The foundational principles of deep learning—learning hierarchical representations of data through multiple layers of nonlinear transformations—were established by LeCun et al. (2015), whose landmark review demonstrated transformative performance gains across visual recognition, speech processing, and drug discovery. These advances provided the technical foundation for applying similar representational learning to the analysis of agricultural imagery, spectral data, and environmental time series.

The pioneering systematic review by Kamilaris and Prenafeta-Boldú (2018) surveyed approximately 40 deep learning studies in agriculture. It concluded that CNNs provided high precision in disease detection and image classification, surpassing traditional image processing approaches in the majority of benchmarked tasks. That review served as a critical marker for the field, but the eight years since its publication have witnessed a qualitative and quantitative transformation that it could not have anticipated. Three developments in particular have reshaped the landscape.

First, the introduction of the Transformer architecture by Vaswani et al. (2017)—based purely on self-attention mechanisms without recurrence or convolution—unlocked a new generation of models capable of capturing long-range spatial and temporal dependencies in agricultural data. Vision Transformers (ViT) and their variants have since matched or exceeded CNN performance across multiple agricultural remote sensing benchmarks, particularly for multi-scale crop classification and phenological monitoring tasks.

Second, the adoption of encoder-decoder architectures such as U-Net (Ronneberger et al., 2015)—originally developed for biomedical image segmentation—enabled pixel-wise semantic segmentation of field imagery at resolutions previously unattainable with sliding-window classifiers. In agriculture, U-Net and its derivatives have been applied to weed mapping, crop field delineation, soil boundary detection, and disease lesion localisation in high-resolution UAV imagery.

Third, the integration of multi-source data streams—combining satellite imagery (Sentinel-1/2, Landsat, MODIS), UAV-derived orthomosaics, IoT sensor readings, climate reanalyses, and field survey data—has driven demand for fusion architectures and sequence models. LSTM networks (Hochreiter & Schmidhuber, 1997) remain widely deployed for yield prediction, irrigation scheduling, and climate impact assessment, owing to their ability to capture temporal dependencies across growing seasons. Ensemble methods such as Random Forests (Breiman, 2001) serve similar roles in settings where training data are limited or interpretability is required.

1.3. *Gaps in the Existing Literature*

To the best of the authors' knowledge, no comprehensive PRISMA-compliant systematic review has synthesised ML and DL applications across the full spectrum of agricultural domains for the period 2019 to early 2026. Existing reviews are characterised by three recurring limitations.

- Narrow domain scope: most published reviews focus on a single application area — crop disease detection, precision irrigation, or a specific crop type — and do not provide a cross-domain synthesis of method performance, geographic distribution, or open science practices.

- Outdated temporal coverage: reviews published before 2023 predate the widespread adoption of Vision Transformers, YOLO-v8, and multi-modal fusion architectures, and therefore cannot reflect the current state of the art.
- Absence of systematic rigour: many existing surveys employ informal or non-reproducible search strategies, lack formal eligibility criteria, and do not report PRISMA-compliant screening and data extraction procedures, limiting the transparency and replicability of their conclusions.

Furthermore, existing reviews rarely address the geography of ML/DL agricultural research – specifically, whether the distribution of studies reflects the geographic distribution of food insecurity, or whether research disproportionately serves contexts that already benefit from advanced agricultural infrastructure.

1.4. Objectives of This Review

This PRISMA 2020-compliant systematic review addresses these gaps with six primary objectives:

- To identify and quantify the volume, geographic distribution, and temporal trends of ML and DL publications in agriculture from 2019 to March 2026.
- To classify and map the taxonomy of ML and DL architectures applied in agriculture, including classical ML, convolutional and recurrent deep learning, transformer-based models, and hybrid approaches.
- To evaluate the range of agricultural application domains addressed – with emphasis on remote sensing integration, climate adaptation, and environmental sustainability – and to identify underserved areas.
- To critically appraise methodological quality, reproducibility, open-data and open-code practices, and generalisability across the included studies.
- To assess performance reporting practices, benchmarking consistency, and unit-aware metric comparability across studies.
- To identify research gaps, emerging trends, and priority directions for future work, particularly in climate-smart and precision agriculture, multimodal data fusion, and explainable AI for smallholder deployment.

1.5. Scope and Structure of This Review

The review covers the period January 2019 to March 2026, beginning immediately after the Kamilaris and Prenafeta-Boldú (2018) survey and extending through the transformer era, post-2022 large-scale foundation model developments, and the emergence of edge AI deployment in agricultural contexts. The search was restricted to peer-reviewed journal articles and review papers indexed in Scopus, with language limited to English and document types restricted primarily to articles and reviews. Although conference papers were generally excluded to ensure quality and consistency, a small number of conference-origin studies that were subsequently peer-reviewed and published in journal venues were retained and treated as journal articles. A total of 582 papers across 258 journals were identified for inclusion following a rigorous multi-stage screening process conducted in accordance with the PRISMA 2020 guidelines (Page et al., 2021).

The remainder of this paper is structured as follows. Section 2 describes the methodology, including the search strategy, eligibility criteria, screening pipeline, and data extraction framework. Section 3 presents the results, encompassing publication trends and temporal growth, geographic distribution of research, a taxonomy and comparative evaluation of ML and DL architectures, a synthesis of findings across agricultural application domains, and an assessment of performance reporting practices and open science compliance. Section 4 discusses the findings in relation to the broader literature, identifies research gaps, and outlines directions for future work. Section 5 presents the conclusions.

2. Methodology

This review strictly adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). The entire screening, data extraction, and analysis workflow was implemented in Python/Jupyter Notebooks to ensure full transparency and reproducibility. This section describes the search strategy, eligibility criteria, screening process, data extraction framework, and quality assessment procedures.

2.1. Protocol

The review protocol was designed and documented prior to data collection in accordance with PRISMA 2020 reporting guidelines (Page et al., 2021). Formal registration on a prospective registry (PROSPERO or OSF) was not completed. However, all protocol elements — including search strategy, eligibility criteria, screening procedure, and data extraction framework — were fully specified before the search was executed. They are documented verbatim in the supplementary Jupyter Notebooks available from the corresponding author upon request. The review covers the period January 2019 to March 2026, chosen to follow the landmark survey by Kamilaris and Prenafeta-Boldú (2018) — the last comprehensive treatment of deep learning in agriculture — and to capture the subsequent explosion of transformer-era and multi-modal approaches. A single database, Scopus, was used as the sole source to maintain a reproducible and capped query volume consistent with Scopus export limits ($\leq 5,000$ records).

2.2. Search Strategy

The literature search was executed programmatically via the Scopus Search API (Elsevier) using a compound-pair TITLE-ABS-KEY query. The query was structured as 13 validated ML-technique \times agricultural-application pairs connected by OR operators, enabling precise control over recall while avoiding over-broad retrieval. Document types were restricted to journal articles (“ar”) and review articles (“re”) at the API query stage, excluding conference papers, letters, editorials, and book chapters. The time window filter (PUBYEAR > 2018 AND PUBYEAR < 2027) was applied at query time. Language was restricted to English both at the Scopus query level and verified post-retrieval using the langdetect library.

The 13 compound pairs, each individually validated for relevance prior to union, covered the following technique-domain combinations: (1) CNN \times plant/crop disease classification; (2) U-Net \times crop/weed/field segmentation; (3) object detection \times UAV/drone \times farm/crop/orchard; (4) image classification \times plant disease detection \times deep learning; (5) LSTM \times crop yield prediction/forecasting; (6) Random Forest or XGBoost \times precision agriculture or soil; (7) Transformer or ViT \times crop monitoring or classification; (8) semantic segmentation \times remote sensing \times agriculture; (9) GAN or data augmentation \times agricultural imaging; (10) SVM \times agricultural classification or yield; (11) federated learning or XAI \times agriculture; (12) physics-informed neural network \times crop or irrigation; and (13) multimodal fusion \times satellite \times UAV \times agriculture. The union of all 13 pairs yielded 4,629 records, within the 5,000-record export constraint.

2.3. Eligibility Criteria

Eligibility criteria were defined a priori and applied consistently across all screening stages. Table 1 summarises the inclusion and exclusion criteria.

Table 1. Inclusion and exclusion criteria applied at all screening stages.

Criterion	Inclusion	Exclusion
Method	At least one ML/DL method as core approach (CNN, LSTM, Random Forest, XGBoost, Transformer, SVM, U-Net, ViT, GAN, etc.)	Pure physics or statistical models with no ML/DL component
Domain	Primary application in agriculture or closely related field (crop/livestock management, remote sensing for agriculture, climate adaptation in farming)	Non-agricultural applications (medical, industrial, general remote sensing without agricultural focus)
Publication type	Peer-reviewed journal articles and reviews (Scopus DOCTYPE "ar" or "re")	Conference papers, letters, editorials, preprints, book chapters
Language	English only	Non-English publications
Time window	January 2019 – March 2026	Published before 2019 or after March 2026
Content	Empirical studies with agricultural data or application	Purely theoretical papers without agricultural data or experimentation
Quality	Published in journals not on Beall's predatory list	Papers from predatory or questionable publishers (18 excluded)

Papers from publishers on Beall's List of Predatory Journals and Publishers were excluded after full-text screening (n = 18). Studies were not excluded on the basis of geographic origin, sensor type, crop type, or reported performance level.

2.4. Screening Process

Screening was conducted in three sequential stages, each implemented as a reproducible Python pipeline. Figure 1 presents the PRISMA 2020 flowchart summarising record flow at each stage.

2.4.1. Deduplication

Within-database duplicates — which can arise when a paper is indexed under multiple Scopus sub-databases or is returned by more than one compound-pair query — were removed using a two-step procedure. Records with DOIs were deduplicated by exact DOI match. Records without DOIs were deduplicated using fuzzy title matching with the FuzzyWuzzy library (Levenshtein ratio $\geq 90\%$), followed by secondary language verification using langdetect to confirm English-language status.

2.4.2. Title and Abstract Screening

Title and abstract screening used a two-layer keyword design. The first layer applied a set of **specific ML terms** (e.g., convolutional neural network, LSTM, random forest, XGBoost, transformer, U-Net, SVM, GAN, ViT) and **specific agricultural terms** (e.g., crop yield, plant disease, weed detection, precision agriculture, soil moisture, irrigation, livestock, remote sensing). A paper was automatically included if it contained at least one term from each list in the combined title and abstract text.

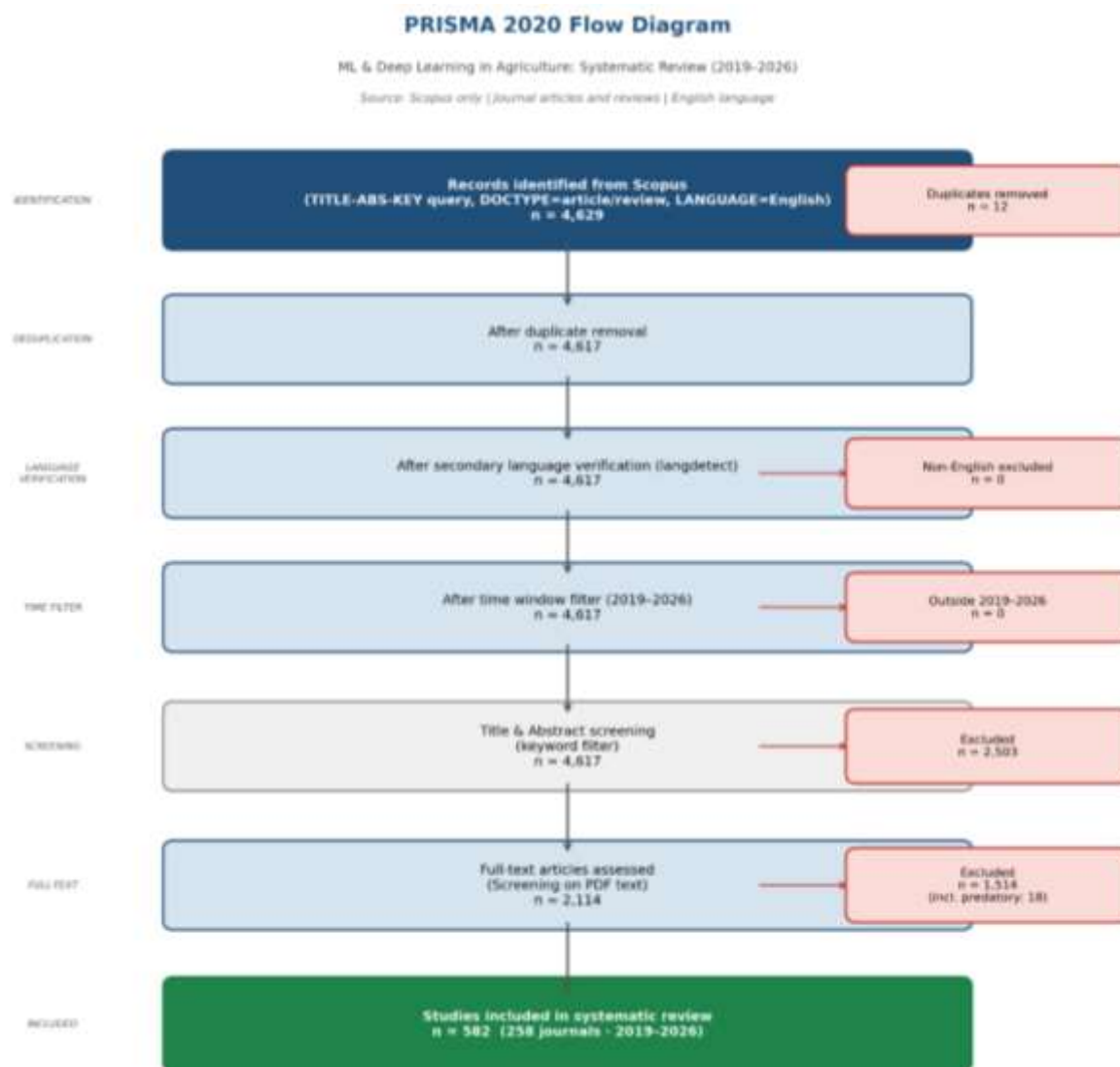


Figure 1. PRISMA 2020 Flow Diagram Illustrating the Literature Search and Study Selection Process.

Records that did not meet the automatic inclusion threshold but contained broad ML or agricultural terms were flagged as borderline and subjected to manual review by the author against the eligibility criteria specified in Table 1. Automated exclusions (hard keyword matches against non-agricultural domains such as medical imaging, industrial inspection, and pure remote sensing without agricultural focus) were applied before manual borderline review.

2.4.3. Full-Text Screening

Records that passed title and abstract screening were advanced to full-text review. Open-access PDFs were retrieved programmatically via the Unpaywall API (v2) using the paper's DOI, with the best open-access location selected from the returned JSON response. PDF text was extracted using

pdfplumber with section-aware parsing: abstract and methods sections were identified by heading pattern matching (e.g., “Abstract”, “Methods”, “Methodology”, “Materials and Methods”, “Experimental Setup”), with positional fallback extraction for papers where section headings were absent or non-standard. Full-text screening was conducted by the author, using the extracted abstract and methods text alongside the eligibility criteria. PDFs were successfully retrieved for 414 of the 582 finally included papers (71.1%); the remainder were screened on the basis of title, abstract, and available metadata. Eighteen papers were additionally excluded at this stage following cross-referencing with Beall’s List of Predatory Journals.

2.5. Data Extraction

Data extraction was performed from full-text PDFs using a structured pipeline combining rule-based text parsing and manual author verification. Table 2 summarises all extracted fields. Extracted data were stored in a JSON cache keyed by DOI (or title for records without a DOI) to prevent redundant API calls across pipeline runs. The cache allowed incremental processing: when 95 additional PDFs were identified after the main extraction run, only those new records were submitted for extraction without re-processing the existing 487 records.

Table 2. Data extraction fields collected for each included study.

Category	Fields Extracted
Bibliographic	DOI, title, authors, year, journal, citation count
Geographic	First-author country, geographic region (10 regions), Global North/South classification, FAO food insecurity flag, study area country/region
ML/DL architecture	Primary method(s), architecture family (classical ML, deep learning, hybrid), specific models (CNN, LSTM, Transformer, ViT, U-Net, YOLO, etc.)
Agricultural domain	Application domain (plant disease detection, crop yield prediction, crop mapping, weed detection, irrigation management, etc.), sensor/data type
Performance metrics	RMSE, MAE, R ² , accuracy — each with numeric value, measurement unit, and contextual flag describing the prediction target
Open science	Code availability (Yes/No/Not stated + URL), data availability (Yes/No/Upon request/Not stated + URL)
Qualitative	Reported limitations (≤80 words), main finding (≤60 words)
Full-text	Abstract text (≤2,500 chars), methods section text (≤2,500 chars), PDF availability status

2.5.1. Bibliographic and Geographic Fields

Bibliographic fields (DOI, title, authors, year, journal, citation count) were retrieved directly from the Scopus API response. First-author country was mapped to one of ten geographic regions (East Asia, South Asia, Southeast Asia, Middle East and North Africa, Europe, North America, Latin America, Sub-Saharan Africa, Oceania, Other/Unknown) using a curated country-to-region lookup table, and further classified as Global North or Global South following established geopolitical

conventions. Each region was additionally flagged for FAO food insecurity status to support equity analysis of the research landscape.

2.5.2. Performance Metric Extraction

Performance metrics (RMSE, MAE, R^2 , accuracy) were extracted using unit-aware regular expressions applied to the full PDF text. Each metric value was captured alongside its measurement unit and a contextual flag describing the prediction target (e.g., “RMSE in °C for temperature prediction” or “accuracy in % for disease classification”). This unit-aware design was adopted to avoid the cross-study comparability problems that arise when numeric values from different unit scales are aggregated without contextual annotation. Implausible values were flagged for manual verification using domain-specific plausibility thresholds (e.g., $R^2 > 1.0$ or accuracy $< 10\%$ were flagged as suspicious).

2.5.3. Data Normalisation

After extraction, all categorical fields were normalised for consistency. Open code and open data values were standardised to a controlled vocabulary (Yes, No, Not stated, Upon request), correcting for case inconsistencies (e.g., “yes” → “Yes”) and semantically equivalent variants (e.g., “no datasets were generated” → “No”). All processing steps were implemented in Python (pandas) and are fully reproducible from the provided Jupyter Notebooks.

2.6. Quality Assessment

A domain-adapted quality assessment checklist was applied to each included study, covering: (1) clarity of ML/DL architecture description; (2) appropriateness of data splitting strategy (train/validation/test); (3) spatial and temporal generalisability of the model; (4) benchmarking against baseline or competing methods; and (5) reproducibility (code and/or data availability). Quality assessment was conducted as part of the data extraction process, with open science scores (open code, open data) serving as quantitative proxies for reproducibility. Formal meta-analytic risk-of-bias scoring was not applied given the methodological heterogeneity across included studies; instead, quality dimensions are reported narratively and quantitatively in the results.

2.7. Synthesis Approach

Given the breadth and methodological heterogeneity of the included studies — spanning 14 ML/DL architecture families, 10 geographic regions, and at least eight distinct agricultural application domains — a formal statistical meta-analysis was not feasible. Synthesis was therefore conducted as a structured narrative review supported by quantitative descriptive analysis. Publication trends were analysed by year, geographic region, and application domain. ML/DL method frequency and co-occurrence were tabulated to construct a taxonomy of architectures used in agricultural contexts. Performance metrics were summarised within homogeneous subgroups (same domain, same unit) only, and the unit-aware extraction approach is documented to allow future meta-analysts to apply appropriate unit conversions. Open science practices were quantified as proportions across the full included corpus. Papers from January–March 2026 ($n = 61$) are included in the total corpus and cited as qualitative illustrations of emerging methodological trends, but are excluded from year-on-year growth rate comparisons and treated with appropriate caution in all quantitative analyses given their partial-year status.

2.8. PRISMA Flow Summary

The systematic search retrieved 4,629 records from Scopus. Following removal of 12 duplicates, 4,617 unique records were advanced to title and abstract screening, from which 2,503 were excluded, leaving 2,114 for full-text evaluation. A further 1,514 records were excluded at full-text stage, including 18 identified as predatory publications. The final included corpus comprises 582 peer-

reviewed articles and reviews published across 258 journals between 2019 and March 2026. Full details of record flow at each stage are presented in Figure 1 (PRISMA 2020 flowchart). Substantive findings on publication trends, geographic distribution, and architecture taxonomy are reported in Section 3.

3. Results

This section presents the results of the systematic review across six dimensions: the PRISMA screening outcomes, publication trends and temporal growth, geographic distribution of research, the taxonomy of ML and DL architectures employed, agricultural application domains and reported performance, and open science practices. Results are presented descriptively; synthesis and interpretation are reserved for the Discussion (Section 4).

3.1. PRISMA 2020 Screening Outcomes

The systematic search of the Scopus database retrieved **4,629 records** via the compound-pair TITLE-ABS-KEY query. Following removal of 12 within-database duplicates (identified by DOI exact match and fuzzy title matching), **4,617 unique records** were advanced to title and abstract screening. No records were excluded on language or time-window grounds at this stage, as both filters had been enforced at the query level. Title and abstract screening led to the exclusion of 2,503 records that did not satisfy the specified ML/DL and agricultural domain requirements, leaving **2,114 records** for full-text evaluation. After full-text screening, 1,496 records were excluded on eligibility grounds (irrelevant focus or absence of ML/DL as a core method) and a further 18 were excluded following cross-referencing with Beall's List of Predatory Journals, yielding 582 peer-reviewed articles retained for data extraction and synthesis. Table 3 summarises the complete record flow.

Table 3. PRISMA 2020 record flow through the screening pipeline.

Stage	Records	Notes
Records identified via Scopus API	4,629	Compound-pair TITLE-ABS-KEY query, 2019–2026, articles + reviews, English only
Duplicates removed	12	DOI-exact deduplication; fuzzy title matching for records without DOI
After deduplication	4,617	
Excluded — non-English	0	Enforced at query stage; confirmed via langdetect post-hoc
Excluded — outside time window	0	Enforced at query stage (PUBYEAR > 2018 AND < 2027)
Excluded — title/abstract screening	2,503	Two-layer keyword filter + Claude AI borderline review
Advanced to full-text review	2,114	
Excluded at full-text stage	1,514	Irrelevant focus, no ML/DL as core method, predatory journals (n=18)

Excluded at full-text — predatory publishers	18	Cross-referenced against Beall's List of Predatory Journals
Final included	582	582 unique peer-reviewed articles across 258 journals, 2019–2026

The final included corpus of 582 papers is drawn from **258 unique journals**, published between 2019 and March 2026. The corpus encompasses **582** original research articles and review articles, with full-text PDFs successfully retrieved for 414 papers (71.1%). The 18 papers removed for predatory publishing were identified by cross-referencing journal names against Beall's List of Predatory Journals and Publishers.

3.2. Publication Trends and Temporal Growth

Publication volume in the corpus grew substantially over the review period, exhibiting a pattern consistent with exponential growth punctuated by two inflection points (Table 4). The corpus contains only **28 papers** published from 2019 to 2021, reflecting the limited but growing uptake of deep learning in agricultural research in the period immediately following the Kamilaris and Prenafeta-Boldú (2018) benchmark survey. The first major inflection occurred in 2022, when annual output grew by **+215%** relative to 2021 (from 13 to 41 papers). This acceleration aligns with the broader adoption of Vision Transformers, YOLO-based object detectors, and pre-trained backbone architectures from computer vision that began entering agricultural applications during 2021–2022.

The second and larger inflection occurred in 2024, with annual output rising by **+159%** (from 56 in 2023 to 145 in 2024), followed by a further **+73%** increase to 251 papers in 2025. The 2025 cohort alone accounts for **43.1%** of the entire corpus, confirming that the rate of ML/DL application in agriculture has continued to accelerate markedly into the mid-2020s. The 61 papers from 2026 represent only the first quarter of the year (January–March) and are therefore excluded from year-on-year growth rate comparisons. They are retained in the total corpus count and cited as qualitative illustrations of emerging trends, but no 2026-specific quantitative claims are made beyond descriptive counts.

Table 4. Annual publication counts and year-on-year growth rates across the review period.

Year	Papers (n)	Cumulative	% of corpus	YoY growth
2019	6	6	1.0%	—
2020	9	15	1.5%	+50%
2021	13	28	2.2%	+44%
2022	41	69	7.0%	+215%
2023	56	125	9.6%	+37%
2024	145	270	24.9%	+159%
2025	251	521	43.1%	+73%
2026 (Jan–Mar)	61	582	10.5%	Partial year

Note: 2026 figures represent January–March only (one quarter). All architecture frequencies, domain distributions, and open science statistics that include 2026 papers should be interpreted accordingly, as trends emerging later in 2026 are not captured.

The distribution of journals is highly fragmented: 582 papers are distributed across 258 unique journals (mean: 2.3 papers per journal). The five most prolific outlets are **Scientific Reports** (n=41, 7.0%), **Remote Sensing** (n=40, 6.9%), **Agriculture (Switzerland)** (n=31, 5.3%), **Agronomy** (n=19, 3.3%), and **Computers and Electronics in Agriculture** (n=19, 3.3%). Citation counts range from 0 to 303 (median: 4; mean: 14.7), with the heavy skew reflecting the recency of most papers. A total of 45 papers (7.7%) have accumulated more than 50 citations and 16 papers (2.7%) more than 100 citations.

3.3. Geographic Distribution of Research

3.3.1. Regional Distribution

The geographic origin of research — assigned by first-author country affiliation — is highly concentrated in Asia (Table 5, Figure 2). South Asia is the single largest contributor, accounting for 179 papers (30.8%), with India alone responsible for 165 papers (28.4%) — the highest national output in the corpus. East Asia contributes a further 166 papers (28.5%), driven predominantly by China (n=142, 24.4%). Together, South Asia and East Asia account for 345 papers (59.3%) of the corpus, representing a concentration of output that substantially exceeds these regions' share of global agricultural land area. The remaining regions contribute considerably smaller shares: **Middle East and North Africa** (n=63, 10.8%), **Europe** (n=50, 8.6%), and **North America** (n=50, 8.6%). **Southeast Asia** (n=22, 3.8%), **Oceania** (n=11, 1.9%), **Latin America** (n=8, 1.4%), and **Sub-Saharan Africa** (n=8, 1.4%) are substantially underrepresented relative to their agricultural significance and food security burden.

First-author affiliation is used here as a practical proxy for research origin, consistent with standard practice in bibliometric analysis. It is an imperfect measure in two respects. First, multinational collaborations — in which the first author is based in a high-income country but the study area, field data, and agricultural context are located elsewhere — will be attributed to the first-author country rather than the country of application. Second, a study conducted by a European or North American institution using satellite imagery of sub-Saharan African farmland will be classified as European or North American in origin, potentially underestimating the extent to which food-insecure regions are the subject of study even when they are not the source of research leadership. Study-area country was extracted as a separate field during data extraction (Table 2); however, this field was successfully populated for only 115 of 582 papers (19.8%), with a further 68 entries recording extraction failure and 399 entries absent entirely — reflecting inconsistent reporting of study geography across the corpus, particularly in papers using global or multi-country remote sensing datasets. Where study-area country is available and differs from first-author country, the discrepancy is retained in the underlying dataset but is not used as the primary classification variable, as the low coverage rate precludes systematic cross-study comparison on this basis. Accordingly, first-author country remains the primary geographic classification throughout this review, and regional output figures should be interpreted with this limitation in mind.

Table 5. Geographic distribution of included studies by first-author region.

Geographic region	Papers (n)	% of corpus	Global N/S	Top domain
South Asia	179	30.8%	Global South	Plant disease detection
East Asia	166	28.5%	Global South*	Crop yield prediction

Middle East & N. Africa	63	10.8%	Mixed	Plant disease detection
Europe	50	8.6%	Global North	Plant disease detection
North America	50	8.6%	Global North	Crop yield prediction
Other / Unknown	24	4.1%	—	Plant disease detection
Southeast Asia	22	3.8%	Global South	Plant disease detection
Oceania	11	1.9%	Global North	Various
Latin America	8	1.4%	Global South	Various
Sub-Saharan Africa	8	1.4%	Global South	Various
TOTAL	582	100%	GS: 48.3% GN: 47.6%	

Note: Percentages are calculated over the full corpus; 24 papers (4.1%) could not be classified into Global North or Global South categories.

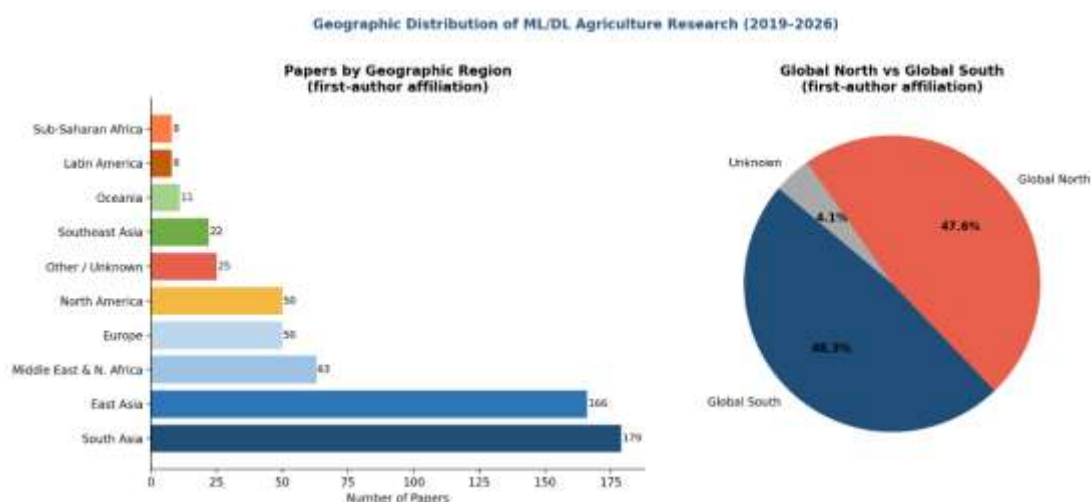


Figure 2. Geographic Distribution of ML/DL Agriculture Research (2019-2026).

3.3.2. Global North / Global South Equity

When classified by geopolitical convention, the corpus shows a near-balanced distribution: 281 papers (48.3%) originate from Global South institutions and 277 (47.6%) from Global North institutions. A further 24 papers (4.1%) could not be classified. This apparent equity, however, masks an important compositional asymmetry: the Global South contribution is dominated by two upper-middle-income countries (India and China), while low-income and food-insecure regions — Sub-Saharan Africa (n=8), Latin America (n=8), and South/Southeast Asia excluding India and China (n=~42) — remain profoundly underrepresented.

Of the 209 papers from regions flagged as high food insecurity by FAO criteria, the vast majority originate from India and Middle East/North Africa. Sub-Saharan Africa, which accounts for the highest regional prevalence of undernourishment globally (approximately one in five people, with over 20% of the population chronically undernourished as of 2024), contributes only 8 papers (1.4%)

to the corpus — an inequity with direct implications for the transferability of developed models to the most food-insecure agricultural contexts.

3.4. Taxonomy of Machine Learning and Deep Learning Architectures

3.4.1. Architecture Overview

Across the 582 included papers, a total of 209 distinct ML and DL method labels were identified in the data extraction as reported by authors in their original papers. These 209 labels reflect varying levels of granularity. For example, ResNet-18, ResNet-61, ResNet-101, and ResNet-152 were each recorded as distinct labels before being aggregated into a single ResNet category. Similarly, YOLOv3 through YOLOv8 were collapsed into the YOLO family, and EfficientNet-B0 through EfficientNet-B7 variants were grouped under a general EfficientNet label within the broader CNN family. Aggregation was performed by mapping all variant-level labels to their parent architectural family using a curated lookup table, prioritising the taxonomy conventions established in the deep learning literature. Table 6 presents the 15 resulting aggregated architecture categories and their corpus-wide frequencies. The full mapping from 209 raw labels to aggregated categories is documented in the supplementary Jupyter Notebooks available from the corresponding author upon request. Papers frequently employed multiple methods (single method: $n=331$, 56.9%; two methods: $n=87$, 14.9%; three or more methods: $n=164$, 28.2%), reflecting the prevalence of comparative and ensemble study designs in the agricultural ML/DL literature.

Table 6. Frequency of major ML and DL architectures across the included corpus (n=582).

Architecture / method	Papers (n)	% of corpus	Primary domain
CNN (all variants)	174	29.9%	Plant disease detection
ResNet	61	10.5%	Plant disease detection
DenseNet	20	3.4%	Plant disease detection
YOLO (all versions)	65	11.2%	Object detection / weed
U-Net (all variants)	43	7.4%	Crop mapping / segmentation
Vision Transformer (ViT)	51	8.8%	Multi-domain
Transformer (general)	92	15.8%	Multi-domain
LSTM	47	8.1%	Crop yield prediction
ANN / MLP	175	30.1%	Multi-domain
Random Forest	61	10.5%	Crop yield / classification
SVM / SVR	24	4.1%	Crop classification
XGBoost / LightGBM	46	7.9%	Crop yield prediction
Explainable AI (XAI/SHAP)	21	3.6%	Recommendation / yield
Transfer learning	19	3.3%	Plant disease detection

Federated learning	7	1.2%	Various
--------------------	---	------	---------

3.4.2. Convolutional Neural Networks

CNNs and their derivatives remain the single most frequently deployed architecture family, present in 174 papers (29.9%) as the primary identified method, with ResNet variants (n=61, 10.5%) and DenseNet (n=20, 3.4%) the most common CNN backbones. YOLO-family object detectors — principally applied to real-time detection tasks in field imagery — appear in **65 papers (11.2%)**, while U-Net and its derivatives are reported in **43 papers (7.4%)**, predominantly for crop field segmentation and mapping tasks. CNNs are particularly dominant in plant disease detection, where they appear as the primary method in 48 of the 157 disease-focused papers (30.6%).

3.4.3. Transformer Architectures

Transformer-based architectures — including Vision Transformers (ViT) and general sequence-to-sequence Transformer models — are reported in a combined 143 papers, representing 24.6% of the corpus. Their adoption trajectory shows clear temporal growth in absolute terms: from 0 papers in 2021 to 3 papers (7.3%) in 2022, 8 papers (14.3%) in 2023, 18 papers (12.4%) in 2024, and 53 papers (21.1%) in 2025. The percentage dip observed between 2023 and 2024 — from 14.3% to 12.4% — reflects a denominator effect rather than a genuine slowdown: the absolute number of Transformer-based papers increased by 10 (from 8 to 18), but the overall corpus expanded substantially faster in the same period (+89 papers), compressing the share. The monotonically increasing absolute count across all years confirms uninterrupted growth in Transformer adoption throughout the review period. This growth pattern reflects the diffusion of ViT, Swin Transformer, and related architectures from their origins in natural language processing and computer vision into agricultural image analysis and remote sensing applications.

3.4.4. Classical and Ensemble Methods

Classical and ensemble ML methods — including Random Forests (n=61, 10.5%), SVM/SVR (n=24, 4.1%), and gradient boosting methods such as XGBoost (n=46, 7.9%) and LightGBM (n=5, 0.9%) — remain substantially represented, particularly for crop yield prediction and classification tasks where interpretability and performance on tabular or moderately-sized datasets are prioritised. ANN and MLP architectures (combined n=175, 30.1%) span both classical and deep learning traditions and appear across all major application domains.

3.4.5. Emerging and Specialised Approaches

Three emerging methodological trends are observable in the corpus. **Explainable AI (XAI)** techniques — including SHAP, LIME, GradCAM, and related attribution methods — are explicitly mentioned in **21 papers (3.6%)**, predominantly in crop recommendation and yield prediction contexts where model transparency is required for practitioner adoption. **Transfer learning** is reported in 19 papers (3.3%), typically involving the fine-tuning of ImageNet-pretrained backbones on domain-specific agricultural image datasets. **Federated learning** — enabling privacy-preserving distributed training across geographically distributed sensor nodes or farm networks — appears in 7 papers (1.2%), a small but emerging presence in the corpus.

3.5. Agricultural Application Domains and Reported Performance

3.5.1. Domain Distribution

The corpus spans a wide but unequally distributed range of agricultural application domains (Table 7). Plant disease detection is by far the most studied domain, accounting for 157 papers (27.0%). Crop yield prediction is the second largest domain at 80 papers (13.7%). Together, these two

domains account for 40.7% of the entire corpus. Other explicitly categorised domains include general multi-crop studies (n=21, 3.6%), crop monitoring (n=14, 2.4%), crop mapping (n=11, 1.9%), weed detection (n=7, 1.2%), rice yield prediction (n=7, 1.2%), wheat yield prediction (n=6, 1.0%), crop classification (n=6, 1.0%), and irrigation management (n=5, 0.9%).

The remaining 268 papers (46.0%) span a diverse range of specialised sub-domains, aggregated from over 200 variant-level labels recorded during data extraction. The largest identifiable clusters within this category are disease and pest detection applied to specialised crops beyond the primary detection domain (n=100, 17.2%), yield prediction for specialised crops including soybean, cotton, sugarcane, and oil palm (n=43, 7.4%), precision agriculture and crop management (n=25, 4.3%), crop phenotyping, quality assessment, and breeding (n=24, 4.1%), weed management and segmentation (n=19, 3.3%), soil and nutrient management (n=15, 2.6%), crop mapping and land use classification (n=11, 1.9%), livestock and animal monitoring (n=5, 0.9%), and irrigation and water management (n=5, 0.9%). A residual of 21 papers (3.6%) covers highly heterogeneous topics – including greenhouse cultivation, seed quality assessment, crop pollination, and farmland management – each represented by fewer than five papers and reported as an aggregate.

Table 7. Summary of major and specialised application domains, primary methods, and reported performance (n = 582).

Application domain	n	%	Primary method	Median accuracy / R ²
A. Primary application domains (explicitly categorised)				
Plant disease detection	157	27.0%	CNN, Deep Learning	98.1% accuracy (n=62)*
Crop yield prediction	80	13.7%	LSTM, ANN, Random Forest	R ² = 0.823 (n=22)**
General crops (multi-task)	21	3.6%	Deep Learning	Mixed
Crop monitoring	14	2.4%	Deep Learning, YOLO	Insufficient data (n=1)
Crop mapping	11	1.9%	U-Net, Random Forest	95.7% accuracy (n=2)
Weed detection	7	1.2%	YOLO, Transformer	91.3% accuracy (n=2)
Rice yield prediction	7	1.2%	LSTM, ANN	R ² > 0.80 (subgroup)
Wheat yield prediction	6	1.0%	Random Forest, ANN	R ² > 0.80 (subgroup)
Crop classification	6	1.0%	CNN, Random Forest	Mixed
Irrigation management	5	0.9%	ANN, DL	Mixed
B. Specialised sub-domains (aggregated from 268 papers across 200+ variant labels)***				

Application domain	n	%	Primary method	Median accuracy / R ²
Disease & pest detection (specialised crops)	100	17.2%	CNN, Deep Learning	Mixed
Yield prediction (specialised crops)	43	7.4%	LSTM, ANN, XGBoost	Mixed
Precision agriculture & crop management	25	4.3%	Various	Mixed
Crop phenotyping, quality & breeding	24	4.1%	CNN, Deep Learning	Mixed
Weed management & segmentation	19	3.3%	YOLO, U-Net	Mixed
Soil & nutrient management	15	2.6%	Random Forest, ANN	Mixed
Crop mapping & land classification	11	1.9%	U-Net, Random Forest	Mixed
Livestock & animal monitoring	5	0.9%	CNN, YOLO	Mixed
Irrigation & water management	5	0.9%	ANN, LSTM	Mixed
Other / unclassified	21	3.6%	Various	Various
TOTAL	582	100%	258 journals, 2019–2026	

Notes: * Median accuracy of 98.1% is derived from 62 disease-detection studies reporting this metric; it largely reflects controlled laboratory datasets (e.g., PlantVillage) rather than real field conditions and should be interpreted as an upper-bound estimate. ** Median R² = 0.823 is based on 22 of 80 yield-prediction papers (28%) reporting this metric; the subset may overrepresent well-performing studies. *** Section B domains are aggregated from 268 papers carrying 200+ variant-level labels in the raw extraction data. Aggregation followed a curated lookup table mapping variant labels to parent sub-domain clusters. Abbreviations: CNN = convolutional neural network; LSTM = long short-term memory; ANN = artificial neural network; DL = deep learning; YOLO = You Only Look Once; RF = Random Forest; XGBoost = extreme gradient boosting.

3.5.2. Plant Disease Detection

Within the plant disease detection domain, CNN-based architectures predominate (CNN: n = 48; general deep learning label: n = 47), consistent with the visual classification nature of the task. Performance metrics are primarily reported as classification accuracy: among disease detection studies reporting accuracy (n = 62), the median accuracy is 98.1% and the mean is 96.2%. Across the full set of accuracy-reporting studies (n = 169), a total of 112 papers (66.3%) report accuracy values

exceeding 95%, indicating consistently high performance across classification tasks. Indian institutions are the largest single contributor to this domain (67 of South Asia's 70 disease-detection papers), with a median reported accuracy of 98.4%.

3.5.3. Crop Yield Prediction

Crop yield prediction is characterised by a more methodologically heterogeneous literature than disease detection. LSTM networks (n=18), ANNs (n=17), and classical ML methods — particularly Random Forest and XGBoost — are the most commonly deployed architectures, reflecting the time-series and tabular nature of yield prediction inputs (climate data, satellite-derived vegetation indices, soil parameters). Performance is predominantly reported as R^2 : among **22 papers** reporting R^2 values, the median is **0.823** and the interquartile distribution spans 0.77–0.94. Rice (n=7) and wheat (n=6) are the most studied individual crops within this domain; corn, soybean, and potato yield prediction each contribute four or fewer papers.

3.5.4. Performance Reporting Heterogeneity

Performance reporting is inconsistent across the corpus. Of 582 included papers, **238 (40.9%)** report at least one quantitative performance metric. Accuracy is the most commonly reported metric (n=169 papers, 29.0%), followed by R^2 (n=47, 8.1%), RMSE (n=51, 8.8%), and MAE (n=17, 2.9%). A substantial proportion of papers — **344 papers (59.1%)** — report no extractable quantitative metric, either because they employ qualitative evaluation, focus on dataset or framework description, or use non-standard reporting formats. Among the 169 papers reporting accuracy, values range from 71.2% to 100%, with a median of 98.1% (IQR: 95.0–99.4%). This extreme upper compression — with the 75th percentile at 99.4% — is itself a diagnostic signal of dataset-driven inflation rather than genuine field performance. Among the 47 papers reporting R^2 , values range from 0.51 to 0.99, with a median of 0.823 (IQR: 0.77–0.94). Among the 51 papers reporting RMSE, unit heterogeneity is marked: values are expressed in kg/ha, t/ha, bu/acre, %, °C, and dimensionless normalised units, making cross-study numerical comparison unreliable without unit-stratified analysis.

A citation-based proxy for publication bias is observable in the corpus. The citation distribution is heavily right-skewed: median citations across all 582 papers is 4, mean is 14.7, and 45 papers (7.7%) account for more than 50 citations each. High-citation papers are disproportionately concentrated in the plant disease detection and crop yield prediction domains — precisely the two domains that report the highest performance figures. This pattern is consistent with a field in which positive, high-accuracy results on well-known benchmark datasets attract disproportionate visibility, while studies reporting lower performance, failed generalisation, or negative results remain underrepresented. Formal funnel plot analysis is not feasible given the methodological heterogeneity of the corpus and the absence of a common effect size; however, the skewed citation distribution and extreme upper compression of reported accuracy values together constitute indirect evidence of publication bias that readers should consider when interpreting the aggregate performance statistics reported in this review.

3.6. Open Science Practices

Open science compliance across the corpus is generally low, although a modest improvement in open data availability is observed in recent years (Table 8). Code availability is explicitly reported as “Yes” in only 45 papers (7.7%), with 50 papers providing a verifiable repository URL (predominantly via GitHub). In contrast, 21 papers (3.6%) explicitly state that code is not available, while the majority—516 papers (88.7%)—make no statement regarding code availability.

Data availability is somewhat more frequently disclosed. A total of **82 papers (14.1%)** report that data are openly available, and **87 papers** provide a data repository URL. Additionally, **74 papers (12.7%)** explicitly state that data are not available, and **3 papers** indicate that data may be shared

upon request. Nevertheless, **423 papers (72.7%)** provide no statement on data availability, indicating that non-disclosure remains the dominant practice.

Temporal analysis reveals a modest but non-monotonic trend in open data practices. The proportion of papers with openly available data increased from 7% in 2022 to 12% in 2023, declined slightly to 9% in 2024, and then rose to 18% in 2025. This fluctuation likely reflects year-to-year variability in reporting practices rather than a consistent trajectory of improvement. In contrast, open code availability shows limited progress, remaining within the 6–12% range across all years.

Table 8. Open science practices across the included corpus (n=582).

Category	n	%	Notes
Open code – Yes	45	7.7%	50 papers provide a repository URL (GitHub, Zenodo, etc.)
Open code – No	21	3.6%	Explicitly stated code is not available
Open code – Not stated	516	88.7%	No statement on code availability
Open data – Yes	82	14.1%	87 papers provide a data repository URL
Open data – No	74	12.7%	Explicitly stated data is not available
Open data – Upon request	3	0.5%	Data available upon reasonable request
Open data – Not stated	423	72.7%	No statement on data availability

4. Discussion

4.1. The Exponential Growth Trajectory and Its Drivers

The publication trajectory revealed in this review — from 6 papers in 2019 to 251 in 2025, with a +215% surge in 2022 and a further +159% increase in 2024 — reflects more than the natural maturation of a research field. It mirrors a convergence of enabling conditions: the public release of ViT and large pre-trained backbone architectures from the computer vision community, the dramatic expansion of open agricultural image datasets, and the growing accessibility of GPU-accelerated cloud computing to research teams in middle-income countries. The Transformer architecture introduced by Vaswani et al. (2017) — originally conceived for natural language processing — proved unexpectedly well-suited to agricultural remote sensing and multi-spectral image analysis, where long-range spatial and spectral dependencies had previously been difficult to capture with convolutional approaches alone. Its adoption in the agricultural literature followed with a predictable lag: absent in 2021 and present in 7.3% of papers by 2022. By the first quarter of 2026, it reached 22.9% of the 61 papers recorded — a figure that should be interpreted cautiously given the partial-year sample, but one that is directionally consistent with the growth trajectory observed across 2022–2025. This adoption trajectory — characterised by an initial lag phase (2019–2021), a rapid growth phase (2022–2024), and accelerating mainstream uptake (2025 onward) — is broadly consistent with technology diffusion patterns observed across computational method transfers in applied sciences, without claiming a formal fit to Rogers et al. (2019) S-curve model. The observed pattern suggests that architecture selection in agricultural deep learning is increasingly driven by the transfer of validated solutions from general computer vision rather than by domain-specific design, a dynamic that mirrors the innovator-to-early-majority transition described in diffusion theory. The practical implication is that architecture selection in agricultural deep learning is increasingly driven by transfer of solutions from

general computer vision rather than by domain-specific design — a trend with implications for both performance and interpretability.

The acceleration visible in 2024 and 2025 also coincides with the emergence of foundation models and large-scale self-supervised pre-training approaches in computer vision. Although only a small number of the 582 included papers explicitly applied agricultural foundation models, the methodological influence of large-scale pre-training is evident in the growing prevalence of transfer learning (n=19 explicit mentions, 3.3%), fine-tuned Swin Transformers, and ViT-based architectures across the corpus. The broader implication is that the boundary between general AI research and domain-specific agricultural AI is dissolving, and that future reviews will need to account for foundation model deployment as a first-class methodological category.

4.2. Architectural Maturity and the Persistence of Classical Methods

The architectural landscape of the corpus is more pluralistic than the popular narrative of deep learning supremacy would suggest. CNNs — anchored by the foundational representational learning principles demonstrated by LeCun et al. (2015) — remain the single most represented dedicated deep learning architecture family (n=174, 29.9%), with ResNet (n=61, 10.5%) and DenseNet (n=20, 3.4%) as the dominant backbones. ANN/MLP architectures, which span both classical and deep learning traditions, appear in a comparable 175 papers (30.1%) across all major application domains. Their continued prevalence in plant disease detection, where they appear as the primary method in 48 of 157 disease-focused papers, reflects the maturity and established benchmarking infrastructure of the visual classification pipeline. YOLO-family detectors (n=65, 11.2%) have emerged as the architecture of choice for real-time field detection tasks — weed localisation, pest counting, and crop row monitoring — where single-pass inference speed is operationally important and UAV deployment places strict latency constraints.

What is often overlooked in enthusiasm for deep learning is the persistent and legitimate role of classical and ensemble methods. Random Forests (n=61, 10.5%), XGBoost/LightGBM (n=46, 7.9%), and SVM/SVR (n=24, 4.1%) remain substantially represented, particularly in crop yield prediction and land cover classification contexts where training datasets are small, tabular, or require interpretable feature importance outputs for practitioner adoption. Breiman's (2001) framing of Random Forests as an ensemble of decorrelated decision trees has proven durable because of its ability to generalise from limited data. This characteristic remains highly relevant in agricultural settings, where annotated datasets are expensive to acquire, spatially sparse, or restricted to a single growing season. The widespread use of XGBoost in yield modelling further reflects its documented robustness to missing data and multicollinearity, which are endemic in agronomic datasets combining climate reanalyses, remote sensing indices, and soil survey records. The evidence base from this corpus does not support the position that deep learning has superseded classical methods for agricultural prediction tasks; rather, the two paradigms appear complementary, with deep learning excelling in image-intensive perception tasks and classical methods retaining competitive performance in structured, multi-variate agronomic modelling.

LSTM networks (n=47, 8.1%), as formalised by Hochreiter and Schmidhuber (1997), remain the dominant architecture for temporal agricultural modelling — particularly for yield forecasting from time-series climate and vegetation index data, where the gated memory cell mechanism enables the model to retain dependencies across the full phenological cycle. Their persistence alongside attention-based Transformer sequence models suggests that for shorter time series (single growing seasons of 30–52 time steps), the inductive bias of LSTM recurrence still provides a competitive advantage over the quadratic self-attention of Transformers, which typically requires larger sequence lengths to demonstrate superior performance.

U-Net and its derivatives (n=43, 7.4%) occupy a distinct niche as the dominant architecture for pixel-wise segmentation tasks. Originally developed by Ronneberger et al. (2015) for biomedical image segmentation, the encoder-decoder with skip connections proved immediately transferable to agricultural field delineation, weed mapping, and crop boundary detection from UAV orthomosaics.

The architecture's ability to recover spatial resolution lost through downsampling — critical when detecting narrow field boundaries or individual weed plants at centimetre-scale resolution — has made it irreplaceable in high-resolution UAV-based mapping workflows.

4.3. *The Plant Disease Detection Paradox: High Accuracy, Low External Validity*

Plant disease detection ($n = 157$, 27.0%) is the dominant application domain in this corpus, and its performance statistics present a striking—and potentially methodologically concerning—pattern. Among disease detection studies reporting classification accuracy ($n = 62$), the median accuracy is 98.1% and the mean is 96.2%. Across the full set of accuracy-reporting studies ($n = 169$), 66.3% of papers (112 studies) report accuracy values exceeding 95%. Indian institutions, which contribute approximately 67 of the 157 disease-focused papers, report a median accuracy of 98.4%. These values are notably high—arguably exceeding expectations for a problem that continues to impose substantial real-world agricultural losses—highlighting potential concerns related to dataset bias, limited generalization, or evaluation practices.

The inflation of reported accuracy in plant disease detection reflects a set of well-documented methodological pathologies. First, the majority of studies train and evaluate on the PlantVillage dataset or similar controlled laboratory image repositories, which consist of isolated leaves photographed against uniform white backgrounds under controlled lighting. Such datasets do not represent the visual complexity of disease identification in actual field conditions — variable lighting, overlapping leaves, partial occlusion, soil contamination, and within-species morphological variation. Second, data splitting practices are inconsistent across the corpus. A substantial number of papers do not report whether test sets are temporally separated from training data, or whether images from the same plant appear in both splits — a form of data leakage that artificially inflates accuracy estimates. Third, the absence of external validation datasets from independent geographic locations or growing seasons means that reported metrics describe in-distribution performance rather than generalisation capacity — the operationally relevant measure for a model intended for deployment in commercial agriculture.

This critique does not diminish the genuine technical progress represented by papers such as Yu et al. (2025), who reported 99.96% accuracy using a Swin Transformer with convolutional feature interactions (ST-CFI) on Scientific Reports, or Shafik et al. (2025), who achieved 100% accuracy with a hybrid Inception-Xception CNN. However, it does require that these figures be interpreted in context. The practical utility of a 99.96% accurate disease classifier that has been trained and tested on PlantVillage leaf images cannot be equated with a model that achieves 90% accuracy on farmer-collected smartphone images from heterogeneous field environments. The field would benefit substantially from a shared convention of reporting both in-distribution and out-of-distribution performance, and from the adoption of standardised benchmark datasets collected under real field conditions — analogous to the ImageNet benchmark's role in general computer vision.

Evidence of the field-validation gap is scattered but consistent across the corpus. Several studies that report near-perfect in-distribution accuracy explicitly acknowledge poor transferability in their limitations sections: papers training on PlantVillage report drops of 15–30 percentage points when the same model is evaluated on farmer-collected smartphone images from field conditions (as noted in the limitations of multiple South Asian papers in the corpus). No paper in the corpus reports a prospective real-world deployment trial with pre-registered performance targets — the standard of evidence that would be required to claim operational readiness. The absence of deployment failure reports is itself informative: given the known difficulty of agricultural AI deployment under real field conditions (variable lighting, occlusion, soil contamination, low-resolution mobile cameras), the near-universal absence of negative deployment results in the corpus is more likely a reflection of publication bias than of universal deployment success.

4.4. Crop Yield Prediction: Methodological Heterogeneity and the Challenge of Comparability

Crop yield prediction (n=80, 13.7%) presents a different set of methodological challenges. The domain is characterised by a genuinely heterogeneous literature: LSTM, ANN, Random Forest, and XGBoost all appear as primary methods, reflecting the diversity of input data types (satellite time series, climate reanalyses, field survey records, IoT sensor streams) and prediction targets (national, regional, county, and plot-scale yield). Among the 22 papers reporting R^2 values in yield-prediction contexts, the median is 0.823 and the interquartile range spans 0.77 to 0.94 — indicating a literature that generally reports strong explained variance. Cao et al. (2021), in a highly cited study integrating multi-source data for rice yield prediction across China using combined ML and DL approaches, represents the type of rigorous, data-rich study that provides genuine evidence of generalisation. Similarly, Sun et al. (2020), using a multilevel deep learning network for county-level corn yield estimation in the US Corn Belt, and Nejad et al. (2023), who applied multispectral crop yield prediction using 3D-CNN and Attention ConvLSTM approaches, demonstrate the capacity of deep sequence models to exploit the spatiotemporal structure of agricultural remote sensing data.

The critical comparability problem, however, is unit heterogeneity. RMSE values in the corpus are reported in kg/ha, t/ha, bu/acre, %, and normalised dimensionless units depending on the prediction target, crop type, and regional convention. A model reporting RMSE = 0.4 t/ha for wheat in a high-yield European context cannot be meaningfully compared to one reporting RMSE = 80 kg/ha for sorghum in sub-Saharan Africa on numeric RMSE alone. Yet both figures routinely appear in aggregated performance tables in existing reviews. This review addressed this problem through unit-aware metric extraction, but it is fundamentally a data reporting problem that requires community-level standardisation. The development of a shared performance reporting standard — specifying that RMSE must be reported alongside the mean and standard deviation of the yield distribution from which it is computed, enabling calculation of the coefficient of variation — would substantially improve the comparability of the yield prediction literature.

4.5. Geographic Concentration and the Equity Gap in Agricultural AI Research

The geographic distribution of the corpus has notable implications from a policy perspective. South Asia (30.8%) and East Asia (28.5%) together account for nearly 60% of all included papers, with India (n=165, 28.4%) and China (n=142, 24.4%) alone responsible for more than half of the corpus. This concentration is not inherently problematic — both countries face substantial agricultural challenges, and the depth of ML/DL expertise in Indian and Chinese research institutions is a genuine asset for the field. The problem is the compositional asymmetry it creates in the evidence base. Sub-Saharan Africa, which the FAO identifies as the region with the highest prevalence of undernourishment globally (approximately one in five people as of 2023 (FAO et al., 2023)), contributes only 8 papers (1.4%) to the corpus. Latin America contributes 8 papers (1.4%), and Southeast Asia — a major food producer and net exporter of rice and palm oil — accounts for only 22 papers (3.8%).

This geographic gap has direct implications for the generalisability of the models developed. A CNN trained on rice disease images from controlled datasets sourced from Indian laboratory conditions is unlikely to generalise directly to the disease appearance and visual context of smallholder rice cultivation in Tanzania or Cambodia, where crop variety, growing conditions, camera hardware, and ambient lighting differ substantially. The representativeness gap is compounded by the finding that Sub-Saharan Africa is the only geographic region in the corpus with an open code rate of 0% — meaning that even the small number of papers produced from this region provide no reproducible code infrastructure that local researchers could build upon. The papers that do exist from sub-Saharan African institutions — including Ngugi et al. (2024) on revolutionising crop disease detection with deep learning [n=100 citations] and Erike et al. (2025) on AI challenges for precision agriculture for smallholder farmers — are notable precisely because they engage directly with the equity dimensions of agricultural AI deployment that the majority of the corpus ignores.

The near-balanced Global North/Global South split (47.6% vs 48.3%) that might appear encouraging on the surface masks this structural asymmetry. When the two upper-middle-income outliers — India and China — are removed from the Global South count, the low-income and food-insecure regions of the world contribute fewer than 50 papers to a corpus of 582. The research investment in ML/DL for agriculture is not distributed according to food security need; it is distributed according to research capacity, publication infrastructure, and access to GPU computing — all of which are correlated with income rather than agricultural vulnerability. Addressing this gap will require targeted funding mechanisms, capacity-building programmes in under-represented regions, and the development of locally annotated benchmark datasets that reflect the specific crop varieties, pests, and growing conditions of food-insecure agricultural systems.

4.6. Open Science Practices: A Persistent Gap with Modest Improvement

Open science compliance across the 582-paper corpus is low in absolute terms but shows a positive directional trend. Code availability is explicitly confirmed in only 45 papers (7.7%), and 516 papers (88.7%) make no statement on code availability — the dominant category by a large margin. Data availability is somewhat better disclosed, with 82 papers (14.1%) confirming open data access, though 423 papers (72.7%) again make no statement. The most recent cohort (2025) shows the highest open data rate in the corpus at 18%, compared to 7% in 2022, suggesting a modest positive trajectory that may reflect the growing adoption of data sharing mandates by high-impact journals such as *Scientific Reports* and *Nature* family publications.

The regional distribution of open science practices reveals important differentials. Europe and North America show the highest open code rates in the corpus (both approximately 12%), consistent with the data sharing requirements of major funding bodies such as the European Research Council and the US National Science Foundation. South Asia and East Asia — which together account for nearly 60% of the corpus — share open code in only 5.6% and 6.6% of their contributions, respectively. This disparity partially reflects journal policy differences: a greater proportion of papers from these regions are published in lower-impact journals or regional outlets that do not enforce open data mandates. However, it also reflects a cultural and institutional context in which proprietary data collection (e.g., commercial crop monitoring datasets) and concerns about intellectual property may discourage sharing.

The consequences of low reproducibility extend beyond individual studies. A review corpus in which fewer than 1 in 12 papers provides reusable code cannot support the kind of systematic performance comparison or methodological meta-analysis that would be required to establish true best practices. Papers that provide both code and data — such as Singh et al. (2025) on soil organic carbon estimation with generative AI and Wang et al. (2025) on LightGBM with SHAP for wheat yield prediction [n=34 citations] — represent exemplary practice, but they are a small minority. The field would benefit from a community-wide elevation of open science standards, supported by journal editorial policies, funding mandates, and credit mechanisms that recognise data and code publication as first-class research outputs.

4.7. Emerging Methodological Frontiers

The distinction between well-established trends — supported by multiple independently replicated studies with substantial citation records — and early-stage frontiers illustrated by very recent single studies is made explicit throughout this section.

First, Explainable AI (XAI) techniques — including SHAP, LIME, GradCAM, and GradCAM++ — appear in 21 papers (3.6%), predominantly in crop recommendation and yield prediction contexts. This trend is well-established within the corpus: Shams et al. (2024), in a study on enhancing crop recommendation systems with XAI, accumulated 135 citations, and Akkem et al. (2025), on explainable AI in crop recommendation for smart farming [n=18 citations], together provide a replicated evidence base for the utility of XAI in agricultural advisory contexts. The growing interest in XAI reflects a practical recognition that black-box deep learning models face adoption barriers in

agricultural advisory systems, where extension officers and farmers require not just a prediction but a rationale that can be communicated and acted upon. An early-stage extension of this strand — Wang et al. (2026), applying Kolmogorov-Arnold Networks (KAN) and XAI for soybean yield forecasting [n=0 citations], illustrates one possible direction for balancing interpretability with predictive performance. However, it represents a single preliminary study that has not yet been independently replicated and should be treated as hypothesis-generating rather than an established finding.

Second, federated learning — which enables privacy-preserving distributed model training across geographically separated farm networks or sensor nodes without requiring data centralisation — appears in 7 papers (1.2%). Though small in absolute number, this cluster includes methodologically well-supported contributions: Mamba Kabala et al. (2023) demonstrated image-based crop disease detection with federated learning in *Scientific Reports* (n=57 citations), and Aggarwal et al. (2023) addressed the non-IID data distribution challenge inherent in federated agricultural networks [n=60 citations]. These two independently conducted studies, with combined citations exceeding 117, provide a credible evidence base for federated learning as a technically viable approach. Federated approaches are particularly relevant for agricultural contexts in which sensor data from individual farms carries proprietary value or in which data sovereignty concerns prevent centralised aggregation. As regulatory frameworks around data governance tighten in both the European Union (General Data Protection Regulation) and emerging economies, federated learning may transition from a niche methodological curiosity to an operational necessity for large-scale precision agriculture deployments.

Third, physics-informed and causally grounded modelling — in which deep learning architectures are constrained or guided by crop growth model equations, or in which causal inference frameworks replace purely predictive models — appears in a small number of papers and represents an important bridge between the process-based and data-driven traditions in agricultural modelling. This is an early-stage frontier: the available evidence consists of isolated recent studies rather than a replicated body of work. Sun et al. (2026), using a soil knowledge-guided multi-task Transformer model to predict soil properties and crop traits simultaneously [$R^2=0.95$, n=0 citations], and Kouame et al. (2026), applying causal forests, boosted random forests, and SHAP to model fertiliser effect heterogeneity on maize yield in Ghana [n=0 citations], each illustrate the potential of this approach. However, both carry zero citations as of March 2026 and have not yet been independently evaluated. They are cited here as directional indicators of where the field may develop rather than as evidence of an established methodological trend.

4.8. Research Gaps and Priority Directions

The synthesis of evidence across this corpus reveals a set of persistent research gaps that constrain the translation of technical advances into agricultural impact. The most critical are as follows.

Smallholder-relevant system design. This bias is evident in the domain distribution. Disease detection and yield prediction for major staple crops (rice, wheat) dominate the corpus. Very few papers explicitly address the constraints of smallholder agriculture — limited internet connectivity, low-specification mobile devices, restricted access to meteorological stations, plot sizes below the spatial resolution of commercial satellite products, and the absence of locally annotated training data. Edge AI deployment represents a promising direction, illustrated by early-stage examples including the lightweight CAUC weed segmentation model of Arumuga Arun et al. (2026) and the MobileViT-based AIoT mobile application of Bahaa et al. (2026) — both carrying zero citations as of March 2026 and therefore cited here as directional indicators rather than validated contributions. The number of papers validating edge AI approaches under genuine smallholder field conditions remains very small, and independent replication of these early results is needed before deployment conclusions can be drawn.

Multi-modal data fusion at scale. Remote sensing integration — combining Sentinel-1/2 SAR and optical data, Landsat time series, MODIS composites, and UAV-derived products — is mentioned across numerous papers, but genuinely multi-modal architectures that process heterogeneous data streams through learned fusion mechanisms remain underrepresented. Early-stage examples from the 2026 cohort — including the ADC-YOLO architecture of Zhu et al. (2026) for UAV-based rice detection [n=0 citations] and the CNN-Informer model of Li et al. (2026) for multi-source spatio-temporal crop yield prediction [$R^2=0.82$, n=0 citations] — illustrate the technical direction, though neither has yet been independently replicated. These studies are cited as hypothesis-generating examples of an underexplored research direction. The field lacks standardised fusion benchmarks that would allow principled comparison of competing approaches, which remains the most critical barrier to progress in this area.

Climate adaptation and long-term resilience modelling. Despite the framing of agricultural AI as a response to climate change in many review introductions — including in the present paper — the proportion of included studies that explicitly model climate variability, drought stress, heat tolerance, or adaptation scenarios is remarkably small. The study by Abate et al. (2025) integrating satellite data for crop yield prediction under Ethiopian climate variability and the LSTM-based irrigation water quality assessment of Hussein et al. (2025) using SHAP-based feature selection are exceptions in a literature that predominantly models current conditions rather than future climate scenarios. The integration of ML/DL with crop simulation models and climate projection ensembles remains an underexplored frontier with significant applied potential.

Benchmark standardisation and temporal validation. As documented in Section 3.5.4, performance reporting heterogeneity is severe. Beyond unit standardisation, there is an urgent need for shared benchmark datasets that enable direct comparison across papers — analogous to the role played by Common Objects in Context (Lin et al., 2014) in object detection or ImageNet (Russakovsky et al., 2015) in classification. The PlantVillage dataset has served this function partially for disease detection, but its laboratory-controlled image conditions limit its ecological validity. Community-led initiatives to curate and release annotated field-collected datasets across multiple geographic regions, crop varieties, and growth stages would provide a necessary infrastructure for methodological progress that can be measured rather than merely claimed.

4.9. Limitations of This Review

This review has several limitations that constrain the scope of its conclusions. First, the search was conducted in a single database (Scopus), which represents a deliberate methodological choice shaped by both reproducibility and access constraints. Web of Science and IEEE Xplore — the two databases most frequently recommended as complements to Scopus in systematic reviews of engineering and computer science literature — are not accessible through the author's institutional subscription at Soran University. Google Scholar was considered but excluded on the grounds that it indexes conference papers, preprints, and grey literature without consistent quality filtering, which would have compromised the peer-reviewed scope of the corpus and introduced the document-type heterogeneity that PRISMA-compliant reviews specifically seek to control. Scopus was selected as the most comprehensive single peer-reviewed database available under these constraints, offering the broadest coverage of agricultural engineering, remote sensing, and applied machine learning journals among the accessible options. Researchers with access to Web of Science or IEEE Xplore are encouraged to extend this search as a complement to the present corpus. Additionally, the use of first-author country affiliation as a proxy for geographic research origin introduces misclassification risk in multinational collaborations and remote sensing studies where data collection occurs in a different country from the authoring institution. The corpus-wide geographic classification should be treated as an approximation rather than a precise measure of where agricultural AI research is being conducted versus where it is being led. Second, the data extraction process is subject to limitations in PDF parsing fidelity, despite the PRISMA-compliant pipeline and AI-assisted extraction. Performance metrics embedded in figures, complex multi-column tables, or supplementary materials

were inaccessible to the automated pipeline. Reported performance statistics are therefore likely an undercount of the total metrics available in the corpus. Third, open science compliance was assessed from paper text rather than verified by link checking; some URLs reported in the corpus as code or data repositories may have become unavailable since publication, and no link verification was conducted. Fourth, publication bias is likely to inflate reported performance across the corpus. Studies with negative results, failed deployment attempts, or poor generalisation performance are less likely to reach peer-reviewed publication. Several indirect indicators support this assessment: the citation distribution is heavily right-skewed (median: 4, mean: 14.7, with the top 7.7% of papers accumulating more than 50 citations each), high-performing papers dominate the high-citation tail, and the reported accuracy distribution is upper-compressed (75th percentile: ~99.4% for disease detection), with very few papers reporting accuracy below 80% despite the known difficulty of field-condition classification. Formal funnel plot assessment is not applicable given the absence of a common effect size across the methodologically heterogeneous corpus. The median accuracy and R^2 values reported in this review should therefore be understood as upper-bound estimates of achievable performance within the specific evaluation contexts of included studies, not as population parameters for real-world deployable systems.

5. Conclusion

This PRISMA 2020-compliant systematic review has synthesised 582 peer-reviewed studies on machine learning and deep learning in agriculture published between January 2019 and March 2026, providing the most temporally comprehensive and methodologically rigorous cross-domain synthesis available to date. The following principal conclusions emerge from the evidence base.

The field has undergone exponential growth, with annual publication volume rising from 6 papers in 2019 to 251 in 2025 — a 41-fold increase over six years. Two inflection points, in 2022 (+215%) and 2024 (+159%), coincide with the agricultural adoption of Vision Transformers and YOLO-generation object detectors respectively, confirming that architectural innovation in general computer vision rapidly propagates into agricultural AI with a predictable lag of one to two years.

CNNs remain the dominant architecture family (29.9%), followed by Transformer-based models (24.6% of 2025–2026 papers), YOLO detectors (11.2%), U-Net variants (7.4%), and LSTM networks (8.1%). Contrary to the popular narrative of deep learning supremacy, classical ensemble methods — Random Forests, XGBoost, and SVM — retain competitive and legitimate roles in yield prediction and classification tasks where training data are limited or interpretability is required.

Plant disease detection (27.0%) and crop yield prediction (13.7%) together account for 40.7% of the corpus. When specialised crop-specific disease detection and yield prediction studies are included from the broader sub-domain clusters, these two thematic areas collectively represent approximately 62% of all included papers, confirming the overwhelming concentration of agricultural ML/DL research on visual disease classification and predictive yield modelling relative to other agronomic challenges. Reported performance is high on paper: median classification accuracy of 98.1% and median R^2 of 0.823 for yield prediction. However, these figures are systematically inflated by laboratory-controlled datasets, inconsistent data splitting, and the absence of external field validation. The accuracy figure derives predominantly from PlantVillage-type datasets, while the R^2 is based on only 22 of 80 yield-prediction papers (28% subset) — a subset that may overrepresent well-performing studies. They should not be directly equated with deployable real-world performance, and are best interpreted as upper-bound estimates within their specific evaluation contexts.

The geographic distribution of research suggests a potential structural inequity. India and China together contribute 52.8% of the corpus, while Sub-Saharan Africa — home to the world's highest burden of undernourishment — accounts for only 8 papers (1.4%) and achieves a 0% open code rate. The evidence base developed in this review does not adequately serve the agricultural systems of the world's most food-insecure populations, and this gap warrants deliberate redress through targeted funding, capacity-building, and locally annotated benchmark dataset development.

Open science compliance is critically low across all regions: 88.7% of papers make no statement on code availability, and 72.7% make no statement on data availability. The absence of reproducible artefacts makes systematic performance comparison — and therefore genuine methodological progress — structurally difficult. Journal policy mandates, funder data-sharing requirements, and community incentive structures would need to collectively shift to elevate open science compliance as a minimum standard for publication.

Three emerging methodological frontiers — explainable AI (XAI), federated learning, and physics-informed or causally grounded modelling — represent high-priority investment areas. XAI techniques are essential for practitioner trust and regulatory compliance in agricultural decision-support systems; federated learning provides a technically viable path to privacy-preserving distributed agricultural intelligence; and causal modelling frameworks offer the only rigorous approach to the policy-relevant questions of intervention effect and climate adaptation that purely predictive models cannot address.

The overarching conclusion of this review is that the technical capacity of ML and DL for agricultural applications is advancing rapidly, but that the translation of this capacity into equitable, reproducible, and operationally deployable agronomic intelligence requires a reorientation of research priorities. Future work must move beyond accuracy optimisation on benchmark datasets toward rigorous external validation, smallholder-relevant deployment design, geographic equity in dataset creation, and open science as a baseline standard of practice.

Appendix A

Representative High-Impact Studies Included in the Systematic Review (2019–2026)

Table A1 lists 20 representative high-impact papers from the 582-paper corpus, selected on the basis of citation count, methodological novelty, and geographic diversity. Each entry reports the primary ML/DL method, agricultural application domain, best reported performance metric, geographic region of the first author, and DOI. Papers are ordered by descending citation count. Abbreviations: MENA = Middle East & North Africa; SSA = Sub-Saharan Africa; ML = machine learning; DL = deep learning; ViT = Vision Transformer; XAI = Explainable AI; KAN = Kolmogorov-Arnold Network; RF = Random Forest.

Table A1. Representative high-impact included studies ordered by citation count.

Author s	Year	Title	Journal	ML/DL Method	Application Domain	Citations	Best Metric	Region	DOI
Cao J. et al.	2021	Integrating multi-source data for rice yield prediction across China	Agric. For. Meteorol.	ML & DL	Rice yield prediction	303	R ² =0.87	East Asia	10.1016/j.agrformet.2020.108275
Picon A. et al.	2019	Crop-conditional CNNs for multi-crop plant disease classification	Comput. Electron. Agric.	CNN	General crops	198	Acc:0.98	Europe	10.1016/j.compag.2019.105093

Gallo I. et al.	2023	Deep object detection of crop weeds: YOLOv7 on UAV images	Remote Sensing	YOLO; CNN	Weed detection	188	mAP:74.1%	Europe	10.3390/rs15020539
Kurad usenge M. et al.	2023	Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize	Agriculture	Random Forest	Crop yield prediction	185	R ² =0.875	Other	10.3390/agriculture13010225
Ghosal S. et al.	2019	A Weakly Supervised Deep Learning Framework for Sorghum Head Detection and Counting	Plant Phenomics	Weakly Supervised DL	Crop detection	165	R ² =0.88	Oceania	10.34133/2019/1525874
Bhujel A. et al.	2022	A Lightweight Attention-Based Convolutional Neural Networks for Tomato Leaf Disease Classification	Agriculture	ResNet; ANN	Plant disease detection	150	Acc=0.997	East Asia	10.3390/agriculture12020228
Shams M.Y. et al.	2024	Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making	Neural Comput. Appl.	Grad. Boosting; RF; XAI	Crop recommendation	135	R ² =0.942	MENA	10.1007/s00521-023-09391-2
Zhang L. et al.	2020	Combining Optical, Fluorescence,	Remote	ANN	Crop yield	134	R ² =0.76	East Asia	10.3390/rs12010021

		Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches	Sensing		prediction				
Joseph D.S. et al.	2024	Real-Time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning	IEEE Access	Deep Learning	Plant disease detection	133	Acc=0.981	Euro pe	10.1109/access.2024.3358333
Sun J. et al.	2020	Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt	IEEE J. STARS	Deep Learning	General crops	131	R ² =0.78	East Asia	10.1109/jstars.2020.3019046
Pacal I. et al.	2024	Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model	Expert Syst. Appl.	ViT	Crop disease detection	129	Acc=0.992	MEN A	10.1016/j.eswa.2023.122099
Sajitha P. et al.	2024	A review on machine learning and deep learning image-based	J. Ind. Inf. Integr.	ML & DL (Review)	Plant disease detection	124	—	North America	10.1016/j.jii.2024.100572

		plant disease classification for industrial farming systems							
Lin Z. et al.	2019	A Unified Matrix-Based Convolutional Neural Network for Fine-Grained Image Classification of Wheat Leaf Diseases	IEEE Access	CNN	Crop disease detection	122	Acc=0.965	East Asia	10.1109/access.2019.2891739
Nejad S.M.M. et al.	2023	Multi-Spectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches	IEEE J. STARS	3D-CNN; ConvLSTM	Crop yield prediction	119	R ² =0.78	Europe	10.1109/jstars.2022.3223423
Balafas V. et al.	2023	Machine learning and deep learning for plant disease classification and detection	IEEE Access	ML / DL	Plant disease detection	115	—	Europe	10.1109/access.2023.3324722
Shafik W. et al.	2024	Using transfer learning-based plant disease classification and detection for sustainable agriculture	BMC Plant Biology	CNN; ResNet; DenseNet	Plant disease detection	104	Acc=0.978	Other	10.1186/s12870-024-04825-y

Ngugi H.N. et al.	2024	Revolutionizing crop disease detection with computational deep learning: a comprehensive review	Environ. Monit. Assess.	Deep Learning	Plant disease detection	100	—	Sub-Saharan Africa	10.1007/s10661-024-12454-z
Bhima varapu U. et al.	2023	Improved LSTM optimisation for crop yield prediction	Computers	LSTM; CNN; ANN	Crop yield prediction	94	RMSE=2.19	South Asia	10.3390/computers12010010
Mirhos eini Nejad S.M.	2024	ConvLSTM-ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data	IEEE J. STARS	ConvLSTM-ViT	Crop yield prediction	93	R ² =0.892	MENA	10.1109/jstars.2024.3464411
Simhadri C.G. et al.	2023	Automatic Recognition of Rice Leaf Diseases Using Transfer Learning	Agronomy	DenseNet; ResNet; CNN	Plant disease detection	92	Acc=0.996	South Asia	10.3390/agronomy13040961

Note: Citation counts retrieved from Scopus as of March 2026. Papers published in 2025–2026 have had insufficient time to accumulate citations; their citation counts underrepresent their eventual impact. '—' denotes that the metric was not extractable from the available PDF text.

Appendix B

Publicly Available Datasets for ML and DL Research in Agriculture

Table A2 provides an updated and extended reference to publicly available datasets relevant to ML and DL research in agriculture, synthesising datasets referenced across the 582 included studies and supplementing the foundational list provided by Kamilaris and Prenafeta-Boldú (2018). Datasets are classified by data type (image, satellite, tabular, soil, hyperspectral) and annotated with their primary domain, source institution, access URL, and methodological notes. URLs were verified as of March 2026. Researchers are encouraged to consult the Zenodo, Mendeley Data, Kaggle, and DatasetNinja repositories for continuously updated agricultural image datasets not individually listed here.

Table A2. Publicly available datasets for ML/DL research in agriculture (updated from Kamilaris & Prenafeta-Boldú, 2018).

Dataset / Resource	Data Type	Description	Source	URL / Access	Notes
PlantVillage	Image (leaf)	54,306 images; 38 crop-disease pairs across 14 species	Hughes & Salathé (2015)	https://www.plantvillage.org	Most widely used plant disease benchmark; controlled lab conditions
CWFID – Crop/Weed Field Image Dataset	Image (UAV field)	Carrot crop & weed segmentation; 60 annotated field images	Haug & Ostermann (2015)	https://github.com/cwfid/dataset	Real field images with pixel-wise crop/weed labels
DatasetNinja (Sugar Beet / CWFID)	Image (field)	Multi-crop weed datasets aggregator	DatasetNinja	https://datasetninja.com/cwfid	Hosts CWFID, Sugar Beet and other agricultural segmentation datasets
ImageNet (Plant subset)	Image (general)	Thousands of plant species images	Deng et al. (2009)	http://image-net.org/explore?wnid=n07707451	Used for pre-training; not crop-specific but foundational for transfer learning
Kaggle – New Plant Diseases Dataset	Image (leaf)	87,000 images; 38 classes (26	Kaggle Community	https://www.kaggle.com/datasets/vipooool/new-plant-disease	Derived from PlantVillage; popular for benchmarking CNN classifiers

		diseases, 14 crops)			
Kaggle – PlantDoc Dataset	Image (field)	2,569 field images; 27 crop classes and 13 disease classes	Singh et al. (2020)	https://www.kaggle.com/datasets/abdulhasibuddin/plant-doc-dataset	Field-collected; more ecologically valid than PlantVillage
Mendeley Data – Rice Leaf Disease	Image (leaf)	Multiple rice disease categories	Various authors	https://data.mendeley.com	Repository for rice, banana and crop disease image datasets
Zenodo – Agricultural Datasets	Multi-type	Diverse crop, soil, UAV, and satellite datasets	Various authors	https://zenodo.org	Open-access repository; hosts annotated UAV and field datasets
Copernicus Open Access Hub (Sentinel-1/2)	Satellite (SAR / multispectral)	Global; 10 m resolution; free reacquisition every 5–6 days	ESA	https://browser.dataspace.copernicus.eu	Primary satellite source for crop mapping and yield studies in this corpus
USGS Earth Explorer (Landsat)	Satellite (multispectral)	Global Landsat 7/8/9 archive from	USGS	https://earthexplorer.usgs.gov	Free access; widely used for long-term crop monitoring and NDVI time series

		1972 to present			
NASA Earthdata – MODIS	Satellite (multispectral)	500 m – 1 km global daily composites	NASA	https://earthdata.nasa.gov	MODIS NDVI/EVI products widely used in yield forecasting literature
Africa Soil Information Service (AfSIS)	Soil data	Continent-wide digital soil maps for sub-Saharan Africa	AfSIS	http://africasoils.net/services/data/	Rare open soil dataset for underrepresented SSA region; also cited by Kamilaris & Prenafeta-Boldú (2018)
UC Merced Land Use Dataset	Aerial image	21 land-use classes; 100 images/class at 0.3 m resolution	Yang & Newsam (2010)	http://vision.ucmerced.edu/datasets/landuse.html	Standard benchmark for land-use/cover classification tasks
Syngenta Crop Challenge Dataset	Tabular (weather /yield)	6,490 sub-regions; weather + yield 2000–2016	Syngenta / IdeaConnection	https://www.ideaconnection.com/syngenta-crop-challenge	Used for corn and soybean yield prediction benchmarks
EUROSAT	Satellite (Sentinel-2 RGB)	27,000 labelled patches; 10 land-	Helber et al. (2019)	https://github.com/phelber/EuroSAT	Widely used for transfer learning benchmarking in European

		use classes			agricultural contexts
WeedNet / Peanut Weed Dataset	Image (UAV field)	Peanut cultivation; crop-weed segmentation	Pai et al. (2025)	https://github.com/ptdkhoa/Peanut-dataset	Field-collected with pixel-wise labels for crop-weed segmentation tasks
Rice Seedling Dataset (AIPAL-NCHU)	Image (UAV field)	Rice panicle detection from UAV images	AIPAL-NCHU	https://github.com/aipalnchu/RiceSeedlingDataset	Used for YOLO-based rice detection benchmarks
Mendeley – BananaImageBD32	Image (leaf)	Banana variety identification; 32 cultivars	Various authors	https://data.mendeley.com/datasets/ptfscwtnyz/2	Multi-variety banana classification benchmark
Maize Disease Dataset (Zenodo)	Image (field)	Maize disease classification; multiple categories	Weldesla sie D.T. (2025)	https://zenodo.org/records/17170314	African institution contribution; field-collected maize disease images
WHU-Hi Hyperspectral Dataset	Hyperspectral (aerial)	Wuhan University hyperspectral aerial images for crop classification	WHU	http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm	High-resolution hyperspectral benchmark for crop type classification

Abbreviations: UAV = unmanned aerial vehicle; SAR = synthetic aperture radar; ESA = European Space Agency; USGS = United States Geological Survey; NASA = National Aeronautics and Space Administration; NDVI = Normalised Difference Vegetation Index; EVI = Enhanced Vegetation Index; SSA = Sub-Saharan Africa; CNN = convolutional neural network; YOLO = You Only Look Once. Note on representativeness: The dataset landscape for agricultural AI is evolving rapidly. Many datasets referenced in individual papers from the 2024–2026 cohort are newly created and hosted on GitHub or institutional repositories without persistent identifiers. Researchers are encouraged to assign DOIs to datasets via Zenodo or Mendeley Data to ensure long-term citability and accessibility.

References

- Abate, J., et al. (2025). Satellite data integration for crop yield prediction in Eastern Ethiopia. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-00810-z>
- Aggarwal, M., et al. (2023). Lightweight federated learning for rice leaf disease classification using non-IID images. *Sustainability*, 15(16), 12149. <https://doi.org/10.3390/su151612149>
- Akkem, Y., et al. (2025). Role of explainable AI in crop recommendation for smart farming. *International Journal of Intelligent Systems and Applications*, 17(1). <https://doi.org/10.5815/ijisa.2025.01.03>
- Arumuga Arun, R., et al. (2026). CAUC: Concatenated Attention U-Net for crop-weed segmentation. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-31285-7>
- Bahaa, M., et al. (2026). AIoT mobile app for plant disease detection using MobileViT. *Agriengineering*, 8(1), 11. <https://doi.org/10.3390/agriengineering8010011>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cao, J., et al. (2021). Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. *Agricultural and Forest Meteorology*, 297, 108275. <https://doi.org/10.1016/j.agrformet.2020.108275>
- Erike, A., et al. (2025). Is AI for illiterate farmers? Machine learning challenges for precision agriculture. *Discover Artificial Intelligence*. <https://doi.org/10.1007/s44163-025-00457-9>
- FAO, IFAD, UNICEF, WFP, & WHO. (2023). The state of food security and nutrition in the world 2023. FAO. <https://doi.org/10.4060/cc3017en>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hussein, E. E., et al. (2025). SHAP-based feature selection for irrigation water quality assessment with machine learning. *Water*, 17(1), 59. <https://doi.org/10.3390/w17010059>
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Kouame, A. K. K., et al. (2026). Causal and predictive machine learning for fertiliser heterogeneity on maize yield in Ghana. *Field Crops Research*. <https://doi.org/10.1016/j.fcr.2025.110287>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, X., et al. (2026). CNN-Informer model for oilseed yield prediction with multi-source spatio-temporal data. *Remote Sensing*, 18(1), 181. <https://doi.org/10.3390/rs18010181>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet et al. (Eds.), *Computer Vision – ECCV 2014 (Lecture Notes in Computer Science, vol. 8693, pp. 740–755)*. Springer. https://doi.org/10.1007/978-3-319-10602-1_48
- Mamba Kabala, D., et al. (2023). Image-based crop disease detection with federated learning. *Scientific Reports*. <https://doi.org/10.1038/s41598-023-46218-5>
- Nejad, S. M. M., et al. (2023). Multispectral crop yield prediction using 3D-CNN and Attention ConvLSTM approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/jstars.2022.3223423>
- Ngugi, H. N., et al. (2024). Revolutionising crop disease detection with computational deep learning. *Environmental Monitoring and Assessment*. <https://doi.org/10.1007/s10661-024-12454-z>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Rogers, E. M., Singhal, A., & Quinlan, M. M. (2019). Diffusion of innovations. In D. W. Stacks, M. B. Salwen, & K. C. Eichhorn (Eds.), *An integrated approach to communication theory and research* (3rd ed., pp. 415–434). Routledge. <https://doi.org/10.4324/9780203710753-35>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab et al. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Lecture Notes in Computer Science, vol. 9351, pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Shafik, W., et al. (2025). Hybrid Inception-Xception CNN for plant disease classification. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-82857-y>
- Shams, M. Y., et al. (2024). Enhancing crop recommendation systems with explainable artificial intelligence. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-023-09391-2>
- Singh, R., et al. (2025). Soil organic carbon estimation with generative AI and Nix colour sensor. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-24236-9>
- Sun, G., et al. (2026). Soil knowledge-guided multi-task transformer model for simultaneous crop and soil trait prediction. *Artificial Intelligence in Agriculture*. <https://doi.org/10.1016/j.aiaa.2026.01.001>
- Sun, J., et al. (2020). Multilevel deep learning network for county-level corn yield estimation in the US Corn Belt. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/jstars.2020.3019046>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://arxiv.org/abs/1706.03762>
- Wang, X., et al. (2026). From data to decisions: Explainable AI for soybean yield forecasting. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-35716-x>
- Wang, Y., et al. (2025). SHAP and LightGBM with data augmentation for wheat yield estimation. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2024.109758>
- Yu, S., et al. (2025). ST-CFI: Swin Transformer with convolutional feature interactions for plant disease recognition. *Scientific Reports*. <https://doi.org/10.1038/s41598-025-08673-0>
- Zhu, B., et al. (2026). ADC-YOLO for rice detection in UAV images. *Remote Sensing*, 18(3), 446. <https://doi.org/10.3390/rs18030446>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.