

REVIEW

Open access

Remote Sensing, Machine Learning, and Smart Farming: An Umbrella Review of Review-Level Evidence for Crop Monitoring, Disease Detection, and Yield Forecasting

Gabriel Costa^{1*}, Lucas Ribeiro¹, Ricardo Alves², Mariana Lopes³

Abstract

Remote sensing and machine learning have become prominent components of smart farming research, generating an expanding body of review literature across agricultural monitoring and prediction applications. This umbrella review synthesises review-level evidence concerning crop monitoring, disease detection, yield forecasting, and decision support while examining the methodological conditions that shape evidence strength and transferability. Peer-reviewed systematic reviews, scoping reviews, structured reviews, surveys, meta-reviews, and other eligible review-level articles were identified through searches of major academic databases and publisher platforms. Evidence was organised by application domain, sensing modality, analytical method, validation design, generalisation, uncertainty, reproducibility, and decision-support relevance. Crop-monitoring reviews describe extensive use of satellite, unmanned aerial vehicle, multispectral, hyperspectral, thermal, and multimodal data for crop mapping, growth assessment, phenotyping, stress monitoring, and related applications. Disease-detection reviews report rapid expansion of image-based machine learning and deep learning, but repeatedly identify data scarcity, class imbalance, curated-image dependence, and limited cross-context validation. Yield-forecasting reviews identify recurring integration of remote sensing, weather, soil, crop-model, and management information, although forecast horizons, spatial and temporal validation, uncertainty communication, and decision relevance remain inconsistent. Across domains, reported technical promise is moderated by heterogeneous datasets, benchmark inconsistency, review overlap, incomplete reproducibility, and limited evidence of transfer across crops, regions, seasons, and acquisition conditions. Review-level evidence supports a substantial research opportunity for remote sensing and machine learning in smart farming, but it does not independently establish field readiness, diagnostic reliability, operational forecasting performance, or deployment readiness. Stronger review reporting, external validation, transparent benchmarking, uncertainty assessment, reproducible workflows, and responsible decision-support evaluation are required.

Keywords Precision agriculture, Machine learning, Remote sensing, Smart farming, Umbrella review, Crop monitoring

*Correspondence:

Gabriel Costa
gabriel.costa@ufmg.br

¹ Department of Remote Sensing and Machine Learning for Agriculture, Faculty of Agricultural Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

² Department of Crop Monitoring and Disease Detection, Faculty of Agriculture, University of Coimbra, Coimbra, Portugal

³ Department of Yield Forecasting and Umbrella Review, Faculty of Agriculture, University of São Paulo, São Paulo, Brazil

Introduction

Review literature portrays smart farming as a data-intensive agricultural transformation in which machine learning and deep learning increasingly mediate heterogeneous sensor, image, and management data, while the evidentiary maturity of applications remains uneven [1–3]. Within this transformation, smart farming refers not merely to the use of isolated digital tools but to interconnected processes through which agricultural observations are acquired, transferred, analysed, interpreted, and potentially incorporated into management decisions. Machine-learning applications reviewed across agriculture include crop, soil, disease, weed, livestock, and yield-related tasks, whereas deep-learning surveys document the growing use of representation-learning methods for imagery and other high-dimensional data. These broad reviews establish the technical diversity of the field while also revealing substantial differences in datasets, analytical objectives, evaluation practices, and application contexts.

Remote sensing constitutes one of the most extensive evidence domains within this wider literature. A meta-review of agricultural remote sensing identified a broad range of sensing platforms, retrieved crop and environmental variables, processing approaches, and intended applications, demonstrating that the literature encompasses far more than a single technology or monitoring task [4]. Satellite, airborne, unmanned aerial vehicle, proximal, multispectral, hyperspectral, thermal, and structural sensing approaches have been reviewed in relation to crop identification, growth assessment, stress detection, phenotyping, resource monitoring, and yield-related applications. Nevertheless, the breadth of reviewed applications should not be interpreted as proof that the underlying systems produce reliable decisions across farms, crops, seasons, or regions. A meta-review maps recurring evidence patterns across earlier reviews; it does not independently verify the validity of every included primary study or establish operational reliability.

An umbrella-review perspective is therefore needed because the evidence base is increasingly mediated by multiple reviews with overlapping but non-identical scopes. Some reviews organise technologies according to sensor platform, whereas others focus on machine-learning families, crop traits, disease recognition, yield prediction, data infrastructures, or decision-support systems. Reviews also differ in whether they examine controlled-image classification, field-scale remote sensing, temporal forecasting, crop-model assimilation, socioeconomic conditions, or responsible deployment. Consequently, similar terms may refer to materially different research tasks, and apparently consistent conclusions may be based on overlapping primary studies, distinct validation designs, or unequal levels of methodological scrutiny. Treating all review conclusions as interchangeable would obscure these differences and could exaggerate the density or maturity of the evidence.

The objective of this umbrella review is to synthesise review-level evidence on remote sensing, machine learning, and smart farming across crop monitoring, disease detection, yield forecasting, and decision-support applications. Its contribution is threefold: first, it maps convergent and divergent findings across review domains; second, it evaluates recurring methodological limitations involving data quality, benchmarking, validation, generalisation, uncertainty, reproducibility, and review overlap; and third, it develops a bounded future research agenda connecting technical evaluation with decision-support translation and responsible deployment. The synthesis is deliberately interpretive rather than prescriptive. It assesses what the review literature supports, where evidence remains fragmented, and which claims require direct primary-study validation before field-readiness or deployment-readiness conclusions can be considered.

Smart Farming Evidence Across Review Literature

Review-level evidence presents smart farming and digital agriculture as sociotechnical developments rather than exclusively computational ones. Social-science reviews show that the meaning and consequences of digital agriculture depend on the roles of farmers, advisers, technology firms, data intermediaries, public institutions, and other actors, as well as on changes in labour, skills, knowledge, power, and governance [5]. This perspective qualifies purely technical accounts of remote sensing or predictive modelling. Even where sensing and analytical capabilities appear promising, their decision value may be influenced by data access, institutional arrangements, infrastructure, accountability, and the capacity of users to evaluate or act upon model outputs. Review-level synthesis must therefore distinguish evidence of technical capability from evidence concerning organisational integration, acceptance, equitable access, or practical use.

Crop-data-management reviews offer a complementary systems perspective by connecting sensing, data transmission, storage, analytics, interoperability, and management interfaces. Within this literature, data generated by field sensors, machinery, imagery, farm records, and external information services can be treated as components of an agricultural information architecture rather than independent inputs [6]. Such architectures may support monitoring and decision processes, but the review literature also identifies recurring challenges concerning heterogeneous formats, data ownership, connectivity, platform compatibility, temporal synchronisation, and the conversion of analytical outputs into interpretable actions. Consequently, the existence of multiple agricultural data streams does not itself establish effective data integration or useful decision support.

Updated machine-learning reviews indicate continued expansion in the number and variety of agricultural applications, including crop management, soil assessment, disease and pest recognition, weed detection, water management, livestock monitoring, and yield prediction [7]. The reviewed literature includes conventional statistical learning, random forests, support vector machines, neural networks, convolutional architectures, and other data-driven approaches. However, the apparent expansion of application categories is accompanied by substantial heterogeneity in sample construction, sensor conditions, preprocessing, feature selection, model tuning, comparison procedures, and validation. Review-level evidence can therefore identify recurring analytical approaches and application concentrations, but it rarely supports a context-independent ranking of algorithms or a general conclusion that one method is universally superior.

Precision-agriculture reviews similarly map machine-learning applications across soil, crop, disease, weed, irrigation, and yield domains while exposing uneven benchmarking and evaluation practices [8]. Differences in target variables, input resolutions, class definitions, spatial units, prediction horizons, and performance metrics limit direct comparison across reviewed studies. Review articles may also combine laboratory imagery, experimental plots, research farms, satellite products, and operational datasets within the same application category, despite their distinct validity conditions. The central review-level pattern is thus simultaneous growth and fragmentation: smart farming research is expanding across sensing and analytical domains, but methodological diversity, limited external validation, and inconsistent reporting constrain evidence integration. An umbrella synthesis is needed to preserve domain-specific distinctions while identifying limitations that recur across otherwise separate review literatures.

Umbrella Review Scope and Evidence Sources

The umbrella review was organised around six review-level questions addressing the overall evidence landscape, crop monitoring, disease and related detection tasks, yield forecasting and decision support, recurring methodological limitations, and future research priorities. An overview of reviews requires an explicit review-level question, a clearly bounded objective, and a defined unit of synthesis so that conclusions concern the findings of eligible reviews rather than an unstructured mixture of reviews and primary studies [9]. Accordingly, the unit of evidence was the peer-reviewed review article. Eligible sources included systematic reviews, scoping reviews, meta-reviews, structured critical reviews, surveys, integrative reviews, and bibliometric reviews only where they contributed substantive evidence mapping rather than publication counts alone.

Candidate sources were identified through targeted searches of ScienceDirect, SpringerLink and BMC journals, IEEE Xplore, Wiley Online Library, Taylor & Francis Online, the MDPI journal platform, and PubMed. Search concepts combined terms for umbrella reviews and review-level synthesis with smart farming, precision agriculture, digital agriculture, remote sensing, machine learning, deep learning, crop monitoring, plant disease detection, pest and weed detection, yield prediction, yield estimation, forecasting, crop models, data fusion, validation, generalisation, uncertainty, reproducibility, and decision support. Overview methods should be chosen transparently in relation to the purpose and structure of the evidence base [10]. The present synthesis therefore used structured qualitative comparison rather than a meta-analysis of primary-study effect sizes because the included reviews differed substantially in questions, application domains, study designs, outcomes, and reporting formats.

Eligibility assessment considered publication type, peer-review status, publication period, DOI availability, review design, agricultural relevance, and capacity to support at least one predefined umbrella-review domain. Primary model-development studies, field experiments, conference-only papers, preprints, books, theses, product materials, software documentation, vendor pages, standalone datasets, reports, standards, and other non-journal sources were excluded. Bibliometric articles without substantive evidence synthesis were not used as application evidence. Evidence extraction covered review aims, agricultural contexts, sensor modalities, analytical methods, application domains, reported evidence patterns, validation limitations, generalisation concerns, uncertainty, reproducibility, decision-support relevance, and stated research gaps. Because risk of bias and methodological limitations can affect conclusions at both review and primary-study levels, findings were not aggregated as though every included review provided evidence of equivalent strength [11].

Screening proceeded through bibliographic checking, title and abstract assessment, full-text eligibility review, confirmation of review type, evidence-domain mapping, and readiness for structured extraction. Particular attention was given to primary-study overlap because repeated inclusion of the same underlying studies across reviews can make an evidence domain appear denser without adding independent evidence [12]. Overlap awareness was incorporated qualitatively by comparing review scope, application focus, publication period, and frequently recurring evidence themes; no numerical overlap estimate or formal quality score is claimed. The resulting synthesis maps review-level patterns and limitations but does not independently validate primary datasets, sensor performance, model accuracy, diagnostic reliability, forecast reliability, farmer acceptance, or deployment outcomes. **Table 1** summarises the umbrella review scope, review questions, eligibility criteria, evidence sources, and review-level synthesis domains.

Table 1. Umbrella Review Scope and Evidence Sources for Remote Sensing, Machine Learning, and Smart Farming: Review Questions, Databases, Search Concepts, Eligibility Criteria, Review-Level Evidence Domains, Screening Logic, Overlap Awareness, and Synthesis Boundaries

Review component	Operational definition	Included review-level evidence	Excluded evidence	Purpose in the umbrella review	Boundary or limitation
Umbrella-review objective	Structured synthesis of findings, limitations, and gaps reported across eligible reviews	Reviews addressing remote sensing, machine learning, smart farming, precision agriculture, crop monitoring, disease detection, yield forecasting, or decision support	Unstructured commentary and unsupported technology claims	Establish a bounded review-level unit of analysis [9]	Does not independently reassess every primary-study observation
RQ1: Overall evidence landscape	Identification of the principal remote-sensing, machine-learning, and smart-farming review domains	Broad, integrative, structured, and systematic agricultural technology reviews	Generic computing reviews without agricultural relevance	Map the breadth and organisation of review-level evidence	Publication volume is not treated as evidence of maturity
RQ2: Crop-monitoring evidence	Synthesis of reviews concerning crop mapping, classification,	Reviews of satellite, UAV, airborne, proximal, multispectral, hyperspectral,	Primary sensor-validation studies considered in isolation	Compare application concentrations, modalities, and	Reported monitoring capability does not establish field readiness

	growth, canopy, stress, soil, water, nutrient, and phenotyping applications	thermal, and multimodal sensing		repeated limitations	
RQ3: Disease and related detection	Synthesis of reviews concerning plant disease, pest, and weed recognition	Image-processing, machine-learning, deep-learning, and remote-sensing reviews	Single-dataset classifier papers and non-reviewed diagnostic claims	Identify recurring data types, analytical methods, and validity concerns	Recognition research is not equated with validated diagnosis
RQ4: Yield and decision support	Synthesis of reviews concerning yield prediction, estimation, forecasting, crop models, and advisory translation	Reviews using weather, soil, management, imagery, crop models, and multimodal inputs	Isolated forecasting experiments or commercial product claims	Distinguish estimation, forecasting, and decision-support evidence	Technical prediction does not establish operational forecasting
RQ5: Methodological limitations	Cross-review assessment of data quality, benchmarking, validation, uncertainty, generalisation, bias, and reproducibility	Reviews explicitly reporting methodological constraints	Unsupported assumptions about methodological quality	Identify limitations recurring across application domains [11]	No formal quality score is assigned
RQ6: Future research agenda	Development of priorities grounded in repeated review-level gaps	Review recommendations concerning datasets, validation, reporting, explainability, uncertainty, and translation	Priorities unrelated to the extracted evidence	Connect identified gaps to bounded methodological advances	Priorities are research proposals, not implementation standards
Evidence sources	Academic databases and publisher platforms used to identify eligible journal reviews	ScienceDirect, SpringerLink and BMC journals, IEEE Xplore, Wiley Online Library, Taylor & Francis Online, MDPI journals, and PubMed	Google Scholar as a sole evidence source, vendor sites, and non-peer-reviewed webpages	Provide multidisciplinary coverage across agriculture, sensing, computing, and evidence synthesis	The search is not presented as a census of all possible reviews
Search concepts	Combined concepts for	Umbrella review, systematic review,	Searches limited to one algorithm,	Balance broad retrieval with	Terminology varies among

	review design, agricultural domain, sensing, analytics, applications, and methodological limitations	remote sensing, machine learning, deep learning, crop monitoring, disease, yield, validation, and related terms	crop, sensor, or commercial system	domain-specific evidence identification [10]	disciplines and journals
Eligible review types	Peer-reviewed review-level articles capable of supporting a planned synthesis claim	Systematic reviews, scoping reviews, meta-reviews, surveys, structured reviews, integrative reviews, and relevant meta-analyses	Primary studies except methodological or conceptual anchors	Maintain review-level synthesis as the central article design	Review labels used by authors may not imply uniform methods
Review-period criterion	Publication within the predefined eligibility period	Eligible peer-reviewed journal reviews published within the approved period	Sources outside the eligibility period	Focus the evidence base on contemporary smart-farming technologies	Earlier foundational work may be represented indirectly within reviews
DOI and bibliographic verification	Verification of title, authorship, journal, year, and DOI	DOI-bearing peer-reviewed journal articles	Records without verifiable journal publication or DOI	Support traceable reference management	DOI presence does not guarantee methodological quality
Screening logic	Sequential bibliographic, relevance, review-type, eligibility, and extraction-readiness assessment	Reviews meeting all inclusion requirements	Duplicates, non-reviews, irrelevant reviews, and sources lacking usable evidence	Ensure each included source supports a predefined claim	No invented screening precision or selection statistic is reported
Evidence extraction	Structured extraction of aims, methods, domains, findings, limitations, and research needs	Information clearly reported by the eligible reviews	Inferred model scores, datasets, or outcomes not reported by sources	Enable domain comparison and umbrella synthesis	"Not clearly reported" is preferred to unsupported inference
Review-overlap awareness	Qualitative attention to recurring primary studies and duplicated evidence themes	Reviews with partially overlapping technologies, applications, or periods	Assumption that every review represents an independent evidence set	Prevent inflation of apparent evidence density [12]	No numerical corrected-covered-area estimate is claimed

Synthesis approach	Structured qualitative comparison across evidence domains	Convergent findings, divergent findings, evidence limitations, and gaps	Meta-analysis of incomparable primary-study outcomes	Preserve methodological and contextual heterogeneity	Evidence strength is interpreted cautiously rather than scored
Decision-support boundary	Separation of technical research evidence from practical decision validity	Reviews addressing analytical outputs and potential management relevance	Operational prescriptions, vendor selection, and universal management recommendations	Clarify the translational meaning of review findings	No farmer approval, commercial implementation, or deployment readiness is inferred

Remote Sensing and Crop Monitoring Evidence

Review-level evidence maps satellite, airborne, and UAV remote sensing to crop-status assessment, stress surveillance, input-management support, and high-resolution field observation, but repeatedly conditions these opportunities on sensor calibration, preprocessing, scale, and validation design [13–15]. Across these reviews, remote sensing is used as an umbrella term for observations differing substantially in platform, spatial resolution, spectral resolution, temporal frequency, viewing geometry, atmospheric exposure, and processing requirements. Satellite systems provide repeated regional or landscape observations, whereas UAV platforms can provide finer spatial detail and flexible acquisition timing over smaller areas. Multispectral, hyperspectral, RGB, and thermal sensors are reviewed in relation to crop mapping, growth assessment, canopy condition, soil variability, water stress, disease or weed surveillance, and yield-related traits. These uses indicate a broad research base, but differences in acquisition and processing prevent a simple conclusion that one platform is uniformly preferable.

Field-based phenotyping illustrates both the analytical potential and the interpretive limits of high-resolution remote sensing. UAV reviews describe extraction of canopy cover, plant height, vegetation indices, temperature-related traits, biomass proxies, lodging characteristics, flowering information, and other phenotypic features [16]. Such measurements may increase temporal and spatial detail relative to manual observation, particularly where repeated acquisition is feasible. However, the reviewed evidence also identifies limitations involving sensor calibration, illumination, wind, geometric correction, plot delineation, occlusion, trait-reference measurements, varietal diversity, and scaling from research plots to heterogeneous fields. Reported associations between image-derived features and crop traits are therefore context-dependent. Phenotyping evidence supports continued evaluation of sensing and feature-extraction methods but does not establish that a model or trait-retrieval procedure will transfer across crop varieties, growth stages, environments, sensors, or seasons.

Reviews of remote-sensing contributions to crop models show a complementary relationship between spatial observation and process-based representation. Remote sensing can provide information on crop extent, vegetation condition, canopy development, leaf area, biomass-related variables, soil moisture, or other states that may initialise, calibrate, constrain, or update crop models [17]. Data assimilation and related integration methods can potentially reconcile observations with simulated crop processes, thereby supporting monitoring or yield estimation at spatial scales that field measurements alone may not cover. Nevertheless, the literature repeatedly identifies mismatches among observation timing, model state variables, spatial support, parameterisation, crop calendars, cloud contamination, and measurement uncertainty. Integration is therefore not inherently superior to either component used separately; its value depends on whether observation and model uncertainties are characterised and whether assimilation is evaluated under conditions distinct from those used for calibration.

Multimodal crop-monitoring reviews extend this logic by examining combinations of satellite, UAV, ground, spectral, thermal, structural, weather, and other data streams. Combining modalities may compensate for limitations of a single sensor, improve temporal coverage, or represent complementary crop properties, but the evidence remains methodologically heterogeneous [18]. Fusion may occur at raw-data, feature, decision, or model levels, with materially different requirements for registration, temporal alignment, missing-data treatment, scale harmonisation, and computational design. Review-level evidence suggests that multimodal monitoring is an important research direction, yet comparison is constrained by limited benchmark transparency and by inconsistent descriptions of preprocessing and fusion architectures. More inputs may also introduce additional sources of error, bias, or redundancy. Accordingly, multimodal complexity should not be treated as evidence of improved generalisation unless evaluated against transparent single-modality baselines and independent contexts.

Hyperspectral evidence illustrates the distinction between sensitive signal detection and validated field interpretation. Reviews report that narrow, contiguous spectral bands can support research on early plant stress and disease discrimination, including feature selection and classification of spectral responses that may precede clearly visible symptoms [19]. Yet spectral signatures can vary with crop species, cultivar, disease stage, water status, nutrient condition, canopy structure, illumination, background, atmosphere, sensor configuration, and acquisition geometry. Similar stressors may also produce overlapping spectral responses, making causal or diagnostic interpretation difficult without appropriate reference measurements. Across remote-sensing domains, repeated limitations include small or context-specific datasets, inconsistent ground truth, spatial and temporal dependence between training and testing data, incomplete uncertainty reporting, and limited external validation. The umbrella synthesis therefore treats remote sensing as a substantial crop-monitoring evidence base while maintaining a strict boundary between research capability and field readiness. **Table 2** summarises review-level evidence for remote sensing and crop monitoring in smart farming.

Table 2. Review-Level Evidence for Remote Sensing and Crop Monitoring in Smart Farming: Crop Mapping, Crop Classification, Growth Monitoring, Vegetation Indices, Canopy Monitoring, Phenotyping, Soil Monitoring, Water Stress, Drought, Nutrient Status, Satellite Imagery, UAV Imagery, Multispectral Imaging, Hyperspectral Imaging, Thermal Imaging, Data Fusion, and Field-Readiness Boundaries

Evidence domain	Core umbrella question	Review-level evidence pattern	Common data or sensor modality	Common analytical approach	Reported strength	Repeated limitation
Crop monitoring	What crop conditions are represented across remote-sensing reviews?	Reviews cover crop status, development, stress, resource condition, and management-related observations [13].	Satellite, UAV, airborne, proximal, multispectral, hyperspectral, RGB, thermal, and structural data	Vegetation indices, regression, classification, feature extraction, machine learning, and time-series analysis	Broad application coverage	Heterogeneous targets, scales, ground truth, and validation designs
Crop classification	How consistently are crop types distinguished?	Classification is widely reviewed using spectral, temporal, spatial, and structural	Multitemporal satellite imagery, UAV imagery, and spectral features	Supervised classification, random forests, support vector machines, neural networks,	Established research domain	Mixed pixels, class imbalance, crop-calendar variation, and label inconsistency

		differences among crop classes [13].		and deep learning		
Crop mapping	What evidence supports spatial delineation of crops?	Reviews describe field-to regional-scale crop mapping from repeated imagery and classification workflows [13].	Satellite time series, aerial imagery, and field boundaries	Pixel-based, object-based, temporal, and machine-learning classification	Strong evidence-mapping activity	Boundary errors, small fields, clouds, landscape fragmentation, and regional transfer
Crop growth monitoring	How is crop development monitored through time?	Repeated observations are used to track canopy development, phenology, biomass-related traits, and stress responses [13].	Satellite and UAV time series, vegetation indices, canopy structure, and temperature	Trend analysis, growth curves, regression, machine learning, and temporal modelling	Promising temporal evidence	Irregular acquisition, cloud gaps, growth-stage dependence, and limited multi-season testing
Vegetation indices	What role do spectral indices play across reviews?	Vegetation indices remain common proxies for greenness, canopy condition, chlorophyll-related status, and biomass [13].	Red, red-edge, near-infrared, and related multispectral bands	Index calculation, thresholding, regression, machine learning, and crop-model integration	Accessible and widely used	Saturation, soil background, atmospheric effects, cultivar differences, and non-specificity
Canopy monitoring	Which canopy properties are reviewed?	Reviews report canopy cover, height, temperature, architecture, density, and related structural or spectral traits [16].	RGB photogrammetry, multispectral imagery, thermal imagery, LiDAR, and structure-from-motion products	Segmentation, surface modelling, index analysis, regression, and machine learning	High spatial detail in research settings	Occlusion, geometric error, illumination, wind, and reference-measurement limitations
Phenotyping	Can remote sensing support field phenotyping?	UAV reviews identify diverse trait-extraction opportunities	High-resolution RGB, multispectral, thermal, and	Plot segmentation, feature extraction, three-	Strong methodological interest	Limited varietal diversity, plot-to-field scaling, and

		for breeding and crop research [16].	structural UAV data	dimensional reconstruction, and predictive modelling		inconsistent trait ground truth
Soil monitoring	What soil-related evidence is represented?	Reviews cover soil moisture, salinity, texture, organic matter, and surface variability, but support differs among properties [13].	Optical, thermal, microwave, radar, proximal, and derived terrain data	Spectral analysis, regression, spatial modelling, and machine learning	Moderate and property-dependent	Vegetation cover, depth mismatch, sparse samples, moisture effects, and calibration dependence
Water-stress monitoring	How is crop water stress assessed?	Thermal and spectral reviews describe canopy-temperature and reflectance-based indicators of water status [14].	Thermal infrared, multispectral, hyperspectral, weather, and canopy measurements	Stress indices, energy-balance approaches, regression, and machine learning	Promising under defined conditions	Weather sensitivity, reference-temperature selection, canopy cover, and timing
Drought monitoring	What is the evidence for drought-related assessment?	Remote-sensing reviews map vegetation and thermal responses associated with drought and water limitation [13].	Satellite vegetation, temperature, moisture, and meteorological products	Anomaly analysis, index combinations, temporal modelling, and spatial assessment	Broad regional observation potential	Confounding stressors, scale differences, baseline choice, and response lag
Nutrient-status monitoring	How well is nutrient condition represented?	Reviews identify spectral and canopy features associated with nitrogen and related nutrient status [14].	Multispectral, hyperspectral, chlorophyll-related, canopy, and field-reference data	Vegetation indices, spectral features, regression, and machine learning	Promising for research and targeted assessment	Variety, growth stage, water stress, disease, soil background, and reference-data variation
Satellite imagery	What contribution is	Satellite reviews	Optical, radar, thermal, and	Time-series analysis,	Regional coverage and	Cloud contamination,

	associated with satellite platforms?	emphasise repeated, wide-area observation for mapping, monitoring, and regional analysis [13].	multitemporal earth-observation data	classification, retrieval, spatial modelling, and data assimilation	historical continuity	revisit constraints, mixed pixels, and resolution mismatch
UAV imagery	What contribution is associated with UAV platforms?	UAV reviews emphasise flexible, high-resolution observation of fields, canopies, stress, weeds, and phenotypic traits [14].	RGB, multispectral, hyperspectral, thermal, and structural UAV imagery	Photogrammetry, segmentation, indices, classification, detection, and regression	Fine spatial detail and flexible timing	Calibration, flight consistency, weather, regulation, processing burden, and limited area coverage
Multispectral imaging	What evidence supports selected spectral-band sensing?	Multispectral data are widely reviewed for vegetation indices, classification, stress assessment, and crop-status estimation [13].	Visible, red-edge, near-infrared, and selected shortwave bands	Indices, band combinations, classification, regression, and machine learning	Mature and comparatively accessible modality	Spectral saturation, limited biochemical specificity, and cross-sensor differences
Hyperspectral imaging	What evidence supports dense spectral sensing?	Reviews describe sensitivity to subtle crop, stress, and disease-related spectral variation [19].	Contiguous narrow-band proximal, airborne, UAV, or laboratory spectra	Spectral preprocessing, feature selection, dimensionality reduction, classification, and regression	High information richness	High dimensionality, noise, cost, calibration, small datasets, and signature instability
Thermal imaging	What evidence supports canopy-temperature observation?	Thermal reviews describe water-stress, disease, physiological, and crop-condition applications [14].	Canopy and surface temperature from UAV, airborne, satellite, or proximal sensors	Thermal indices, temperature normalisation, energy-balance methods, and machine learning	Useful physiological sensitivity	Atmospheric conditions, time of day, emissivity, reference selection, and mixed background

Data fusion	Does combining modalities strengthen monitoring evidence?	Multimodal reviews suggest complementary information can reduce some single-sensor limitations [18].	Satellite, UAV, ground, spectral, thermal, structural, weather, and management data	Early fusion, feature fusion, decision fusion, multimodal learning, and assimilation	Promising cross-domain research direction	Registration, temporal alignment, scale mismatch, missingness, complexity, and benchmark scarcity
Validation and generalisation	Which methodological weaknesses recur?	Reviews repeatedly identify context-specific datasets, inconsistent reference measurements, dependent train-test samples, and limited external testing [18].	All remote-sensing modalities	Hold-out validation, cross-validation, spatial validation, temporal validation, and transfer testing	Mixed methodological maturity	Leakage, site dependence, sensor shift, seasonal variation, and incomplete uncertainty reporting
No field-readiness claim boundary	What may not be inferred from the review evidence?	Reviews map capabilities and limitations but do not jointly establish reliability, usability, maintenance, safety, or decision consequences.	All platforms and modalities	All analytical approaches	Evidence supports research synthesis	Sparse operational and prospective evaluation

Machine Learning for Disease Detection

Across reviews of image-based plant disease recognition, convolutional deep-learning pipelines dominate the reported literature, yet performance is strongly shaped by dataset construction, background conditions, class balance, and validation choices [20–22]. The evidence base is concentrated in RGB photographs of leaves or plant organs, frequently collected under controlled or semi-controlled conditions and organised as classification tasks. Convolutional neural networks and transfer-learning approaches are prominent because they can learn image features directly, reducing dependence on manually designed descriptors. However, reviews also describe substantial variation in disease classes, image acquisition, preprocessing, augmentation, train–test partitioning, and reported metrics. Consequently, apparently high within-dataset performance cannot be interpreted as equivalent evidence of reliable disease recognition under field conditions, where symptoms may be subtle, overlapping, partially occluded, or confounded by nutrient, water, weather, and pest stresses.

The disease-and-pest detection literature extends beyond image-level classification to object detection, lesion localisation, segmentation, severity estimation, and canopy-level surveillance. Deep-learning reviews identify growing methodological diversity, including region-based detectors, one-stage detectors, semantic segmentation, transfer learning, and combinations of learned and handcrafted features [23]. This expansion is important because practical monitoring may require locating affected tissue or plants rather than assigning one label to a pre-cropped image. Nevertheless, task expansion does not remove underlying validity concerns. Small or duplicated datasets, inconsistent symptom definitions, non-independent images from the same plant or acquisition event, and limited reporting of negative or mixed-stress examples can inflate apparent model performance. Cross-crop, cross-disease, cross-region, and cross-season transfer are discussed more often than they are demonstrated through independent testing.

Automated insect-pest detection and monitoring reviews describe heterogeneous systems based on trap images, close-range photographs, field imagery, and other sensing arrangements [24]. The reviewed approaches include conventional image processing, feature-based classification, convolutional networks, and object-detection methods intended to distinguish species, count individuals, or support monitoring. Their reported strengths lie in the possibility of repeated observation and reduction of manual image inspection. Yet field comparison is difficult because pest appearance varies by life stage, posture, damage, image scale, and environmental background, while rare species and low-prevalence classes complicate evaluation. Annotation quality, overlapping insects, debris, illumination, trap design, and inconsistent performance metrics further restrict comparability. Review-level evidence therefore supports continued automated-monitoring research but does not establish autonomous pest diagnosis or pest-management decisions.

Weed-detection reviews show similarly strong computer-vision activity across ground-based and aerial RGB or multispectral imagery [25]. Classification, object detection, and semantic segmentation are used to separate weeds from crops or background, with convolutional models increasingly replacing or supplementing handcrafted spectral, colour, texture, and shape features. Reported research strengths include fine-scale localisation and the potential to characterise spatial weed occurrence. Repeated limitations include crop–weed visual similarity, changing growth stages, canopy overlap, severe class imbalance, inconsistent weed taxonomies, and narrow field or regional coverage. Dataset and task definitions also differ according to whether the objective is species recognition, crop–weed separation, patch mapping, or plant-level localisation. Detection evidence alone cannot establish safe actuation, herbicide selection, or autonomous field intervention because downstream decision consequences, error tolerances, and failure handling are rarely evaluated by the reviewed literature.

Across disease, pest, and weed reviews, dataset properties remain a central determinant of reported performance. Disease-recognition outcomes are sensitive to sample size, class distribution, symptom similarity, image background, acquisition quality, and the relation between training and testing samples [26]. Augmentation and transfer learning may reduce some computational and sample constraints, but synthetic variation does not necessarily reproduce new cultivars, field backgrounds, weather conditions, sensor settings, disease stages, or mixed stresses. Stronger evidence would require provenance-rich datasets, per-class reporting, geographically and temporally separated validation, and explicit testing under realistic field variability. The umbrella synthesis therefore characterises these systems as recognition and monitoring research. It does not infer diagnostic readiness, cross-context generalisation, farmer approval, operational reliability, or deployment readiness from review-level model results.

Yield Forecasting and Decision-Support Evidence

Review-level evidence on yield prediction and nitrogen-status estimation shows sustained use of machine learning to relate crop outcomes to weather, soil, management, spectral, and crop-state variables [27]. Random forests, support vector machines, artificial neural networks, regression methods, and related models recur across the reviewed literature, but the analytical target is not always defined consistently. Some studies estimate yield near harvest from observed crop traits, whereas others attempt forecasts at earlier decision-relevant stages. This distinction matters because information available close to harvest may support accurate estimation without demonstrating useful forecasting at longer lead times.

Differences in crop type, spatial unit, season, input availability, forecast horizon, and evaluation metric also constrain cross-study comparison. Review-level evidence consequently supports a broad methodological opportunity but not a universal claim of reliable yield forecasting.

Systematic review evidence identifies weather, soil, management, and remote-sensing variables as recurring inputs to crop-yield prediction, with tree ensembles and neural approaches frequently represented [28]. Weather inputs can include rainfall, temperature, radiation, humidity, or derived agroclimatic indicators; soil inputs may include texture, moisture, fertility, or mapped attributes; and management inputs may include planting date, cultivar, irrigation, or fertilisation. Their presence in a model does not guarantee that they were measured at the relevant spatial depth, resolution, or time, nor that they would be available when a real forecast is issued. Inconsistent datasets, missing management records, static soil layers, differing yield definitions, and non-comparable metrics limit conclusions about algorithm superiority. Spatially or temporally dependent train–test partitions may also exaggerate performance by allowing closely related observations to appear on both sides of the evaluation.

Reviews focused on integrating agrarian factors reinforce the importance of contextual completeness. Yield is shaped by interacting weather, soil, genotype, crop condition, and management processes, making data fusion conceptually attractive but methodologically demanding [29]. Tabular machine learning, ensemble methods, deep architectures, and hybrid approaches can combine these sources, yet missingness, measurement error, proprietary farm records, inconsistent coding, and scale mismatch may reduce transportability. Temporal modelling is particularly vulnerable to leakage when predictors include information that would not have been known at the stated forecast date. Spatial modelling is similarly vulnerable when neighbouring fields, pixels, or administrative units are split randomly. Forecast evaluation should therefore preserve true temporal cut-offs and geographic independence while reporting data provenance and availability at prediction time.

Remote-sensing assimilation into crop models offers a distinct pathway for combining spatial observation with process knowledge. Reviews describe the use of remotely observed crop states to initialise, calibrate, constrain, or update process-based simulations, potentially linking canopy or vegetation information with weather, soil, cultivar, and management representations [30]. Such integration may improve consistency between observed and simulated crop development, but it also introduces uncertainty from sensor retrievals, crop-model structure, parameterisation, spatial support, observation timing, and assimilation design. A more complex hybrid model is not automatically a better forecast. Comparative evidence should distinguish observation uncertainty from process-model uncertainty, evaluate whether added data improve out-of-sample performance, and report sensitivity to assumptions and missing observations.

Deep-learning and remote-sensing yield reviews document expanding use of convolutional, recurrent, and hybrid architectures with satellite, UAV, weather, and temporal crop data [31]. Their research value lies in learning spatial or sequential representations that may be difficult to specify manually. However, reviewed comparisons remain limited by crop- and region-specific datasets, short time series, heterogeneous baselines, incomplete reporting, and sparse external validation across seasons and locations. Point-error metrics also provide little information about forecast calibration or decision risk unless accompanied by intervals, probabilistic scores, or scenario analysis. Decision-support claims therefore require evidence that forecasts are available at a meaningful lead time, interpretable within a farm workflow, and evaluated for uncertainty and downstream consequences. Table 3 maps review-level evidence for machine learning in disease detection, yield forecasting, and decision-support applications.

Table 3. Review-Level Evidence for Machine Learning in Disease Detection, Yield Forecasting, and Decision Support: Plant Disease Detection, Crop Disease Imaging, Pest Detection, Weed Detection, Yield Prediction, Yield Estimation, Remote Sensing Forecasting, Weather and Soil Integration, Temporal Modeling, Spatial Modeling, Uncertainty, Validation, and Deployment Boundaries

Evidence domain	Core evidence	Review-level evidence	Common input data	Common ML or DL	Reported strength	Repeated limitation	t
-----------------	---------------	-----------------------	-------------------	-----------------	-------------------	---------------------	---

	question	pattern		approach			
Plant disease detection	What does the review literature show about automated disease recognition?	Image-based classification and detection have expanded rapidly, with deep learning prominent [20].	RGB leaf and plant images; some field and spectral imagery	CNN classification, transfer learning, object detection	Large research concentration	Curated datasets, background bias, imbalance, and limited external testing	re re: c r
Crop disease imaging	Which image settings dominate the evidence?	Controlled or detached-leaf imagery remains more common than heterogeneous canopy and field imagery [21].	RGB photographs, close-range field images, multispectral and hyperspectral data	Preprocessing, segmentation, handcrafted features, CNNs	Strong controlled-image evidence	Illumination, occlusion, symptom-stage variation, and acquisition heterogeneity	C ima re v
Leaf image classification	How mature is the dominant disease task?	CNN evidence is concentrated in image classification using public or custom leaf datasets [22].	Pre-cropped leaf images and labelled disease classes	CNNs, transfer learning, augmentation	High reported within-dataset performance	Dataset redundancy, simple backgrounds, and non-independent splits	Cl pe ec far
Field disease detection	How much evidence addresses realistic environments?	Detection and segmentation are expanding, but field evidence is thinner and more variable [23].	In-situ RGB, canopy, UAV, and mixed-background imagery	Object detection, segmentation, domain adaptation	Emerging evidence	Complex backgrounds, mixed stresses, subtle symptoms, and weak prospective evaluation	r m field
Remote-sensing disease detection	What does remote sensing add?	UAV and spectral sensing can identify anomalous canopy patterns associated with disease or stress [19].	Hyperspectral, multispectral, thermal, and high-resolution canopy imagery	Spectral features, indices, SVMs, CNNs, anomaly detection	Promising experimental sensitivity	Ground-truth uncertainty, spatial resolution, timing, and etiological ambiguity	A sig alc dis
Hyperspectral disease	Can dense spectra	Reviews report early disease-	Hyperspectral cubes or	Dimensionality reduction,	High information	High dimensionality,	N

detection	support early discrimination?	and-stress discrimination using selected wavelengths and spectral patterns [19].	reflectance spectra	wavelength selection, SVMs, other classifiers	richness	instrument differences, confounding stressors, small datasets	si c th e
Multispectral disease detection	What is supported with fewer bands?	Multispectral features and indices appear in UAV and crop-stress applications [13].	Selected visible, red-edge, near-infrared, and thermal bands	Indices, random forests, SVMs, CNN variants	Moderate research support	Limited specificity and cross-sensor variation	gr c st
RGB disease imaging	Why is RGB dominant?	Accessibility and public image collections make RGB the principal modality in deep-learning reviews [26].	Smartphone, camera, and public leaf images	CNNs, object detection, segmentation	Very strong research concentration	Background leakage, image-quality variation, symptom similarity, domain shift	Av g rol r
Pest detection	What evidence exists for automated pest monitoring?	Reviews cover image-based recognition, counting, and trap monitoring across diverse species [24].	Trap images, close-range photographs, field imagery	Image processing, CNN classification, object detection	Moderate and growing	Species imbalance, clutter, life-stage variability, and annotation burden	au ma c
Weed detection	What evidence supports crop–weed discrimination?	Classification, object detection, and segmentation are active research areas in ground and aerial imagery [25].	RGB and multispectral ground or UAV images	CNNs, detectors, semantic segmentation	Strong computer-vision activity	Occlusion, crop–weed similarity, growth-stage change, class imbalance	I es a t c
Class imbalance and data scarcity	How do dataset composition and prevalence affect evidence?	Rare classes, small datasets, duplicated images, and narrow environmental coverage recur [26].	Public and custom disease, pest, and weed image collections	Augmentation, transfer learning, weighted loss, resampling	Well-recognised limitation	Aggregate accuracy can mask minority-class failure	I m pr ric da

Cross-crop and cross-region generalisation	Do models transfer among crops and locations?	Reviews provide limited consistent evidence of transfer to unseen crops, cultivars, regions, or acquisition systems [22].	Mixed crop and region-specific image collections	Transfer learning, domain adaptation, multi-task models	Weak review-level support	Morphology, symptom, climate, cultivar, prevalence, and sensor shifts	cla he in re
Yield forecasting	What evidence exists for pre-harvest forecasting?	Machine-learning reviews identify weather, crop-state, soil, management, and spectral predictors, but lead times vary [27].	Weather, crop observations, soil, management, satellite and UAV data	Random forests, SVMs, neural networks, boosting, regression	Promising and heterogeneous	Forecast horizons, target definitions, and external tests are inconsistent	o fo
Yield prediction	Which variables and methods recur?	Weather, soil, management, and vegetation variables are frequently combined in crop-specific prediction studies [28].	Agronomic tables, remote-sensing features, environmental covariates	Tree ensembles, neural networks, regression, ensembles	Broad research base	Dataset inconsistency, crop specificity, missing variables, metric heterogeneity	Fi co cr se a
Yield estimation and remote-sensing estimation	How does estimation differ from forecasting?	In-season or near-harvest crop traits and imagery are often used to estimate yield without long forecast lead times [31].	Vegetation indices, canopy traits, satellite or UAV sequences	Regression, CNNs, recurrent models, remote-sensing retrieval	Moderate to strong in selected contexts	Forecast timing may be conflated with estimation; site dependence persists	T ir co l
Weather, soil, and management integration	Does added context improve prediction?	Agrarian-factor reviews support multimodal integration, but availability and provenance are uneven [29].	Weather records, soil attributes, planting, cultivar, irrigation, fertilisation	Tabular ML, ensembles, multimodal and hybrid models	Conceptually strong	Missingness, proprietary records, spatial-depth mismatch, inconsistent coding	M au v tra

Temporal and spatial modelling	How are time and geography handled?	Sequential and spatial representations are expanding, especially with earth-observation time series [31].	Weather sequences, satellite stacks, field or regional grids	Recurrent networks, temporal CNNs, spatial CNNs, engineered time-series features	Emerging to moderate	Temporal leakage, neighbouring train–test samples, short histories	mu c g s
Machine learning, deep learning, crop models, and hybrids	Is any analytical family consistently superior?	Tree models, SVMs, neural networks, deep learning, crop models, and assimilative hybrids address different evidence structures [30].	Tabular, image, sequence, process-model, and multimodal inputs	RF, SVM, boosting, ANN, CNN, RNN, process simulation, assimilation	Diverse complementary methods	Incomparable tuning, baselines, uncertainty structures, and data contexts	No a n s
Decision support and uncertainty communication	Do predictions become useful decisions?	Reviews frequently state decision-support potential, while calibrated uncertainty and workflow evaluation remain limited [29].	Forecast outputs, farm records, environmental data, advisory interfaces	Ensembles, probabilistic models, intervals, rules, dashboards	Translation potential	Point estimates, unclear lead time, limited usability and consequence analysis	t re o p
Validation and deployment boundaries	What may not be inferred from reviewed model results?	Random splits, context-specific datasets, and limited prospective or external testing recur across domains [31].	All disease, pest, weed, and yield data	Hold-out tests, cross-validation, spatial, temporal, and external validation	Mixed methodological maturity	Leakage, domain shift, weak comparators, incomplete uncertainty and failure analysis	No o u ge au de r cla

Umbrella Synthesis

The integrated evidence base shows that technical research is embedded in a supply and service ecosystem that influences which technologies are developed, whose data are available, and how analytical outputs reach farms. Reviews of the digital agricultural supply side identify technology firms, equipment manufacturers, input providers, platforms, start-ups, advisers, research organisations, and public actors as important drivers of digital development [32]. Their roles

shape interoperability, business models, access, maintenance, data control, and the capacity to translate remote-sensing or machine-learning outputs into decisions. Consequently, decision-support value cannot be inferred from model performance alone. It also depends on whether data flows are sustainable, responsibilities are allocated, interfaces fit existing workflows, and users can understand, contest, or override analytical outputs.

Across application domains, a common pattern is that evidence strength declines as claims move from technical discrimination or prediction toward reliable action in open environments. Crop-monitoring reviews show extensive sensing and feature-extraction activity; disease reviews show concentrated image-recognition research; and yield reviews show diverse predictive and assimilative methods. Yet systemic agricultural-AI risk extends beyond average accuracy to include distributional effects, feedback loops, resource use, cybersecurity, market concentration, and harms arising from interactions among models, institutions, environments, and users [33]. The umbrella synthesis therefore interprets field readiness as a system property requiring evidence about data, models, infrastructure, human oversight, maintenance, and consequences rather than as a direct extension of benchmark performance.

Convergence is strongest around the expansion of remote sensing and machine learning, the centrality of data quality, and the need for validation beyond the development dataset. Fragmentation is greatest in benchmark definitions, sensor and crop contexts, disease ground truth, forecast horizons, uncertainty reporting, and links to practical decisions. Responsible digital-agriculture guidance further indicates that transparency, accountability, privacy, fairness, stakeholder participation, and governance should be considered throughout the analytical lifecycle [34]. These requirements connect methodological limitations to future research priorities: better cross-context testing is needed to assess transferability; transparent datasets and workflows are needed for comparison; uncertainty must be communicated at decision-relevant horizons; and user-centred evaluation is needed before advisory value is inferred. **Figure 1** synthesises review-level evidence across remote sensing, machine learning, crop monitoring, disease detection, yield forecasting, decision support, evidence limitations, and future research priorities.

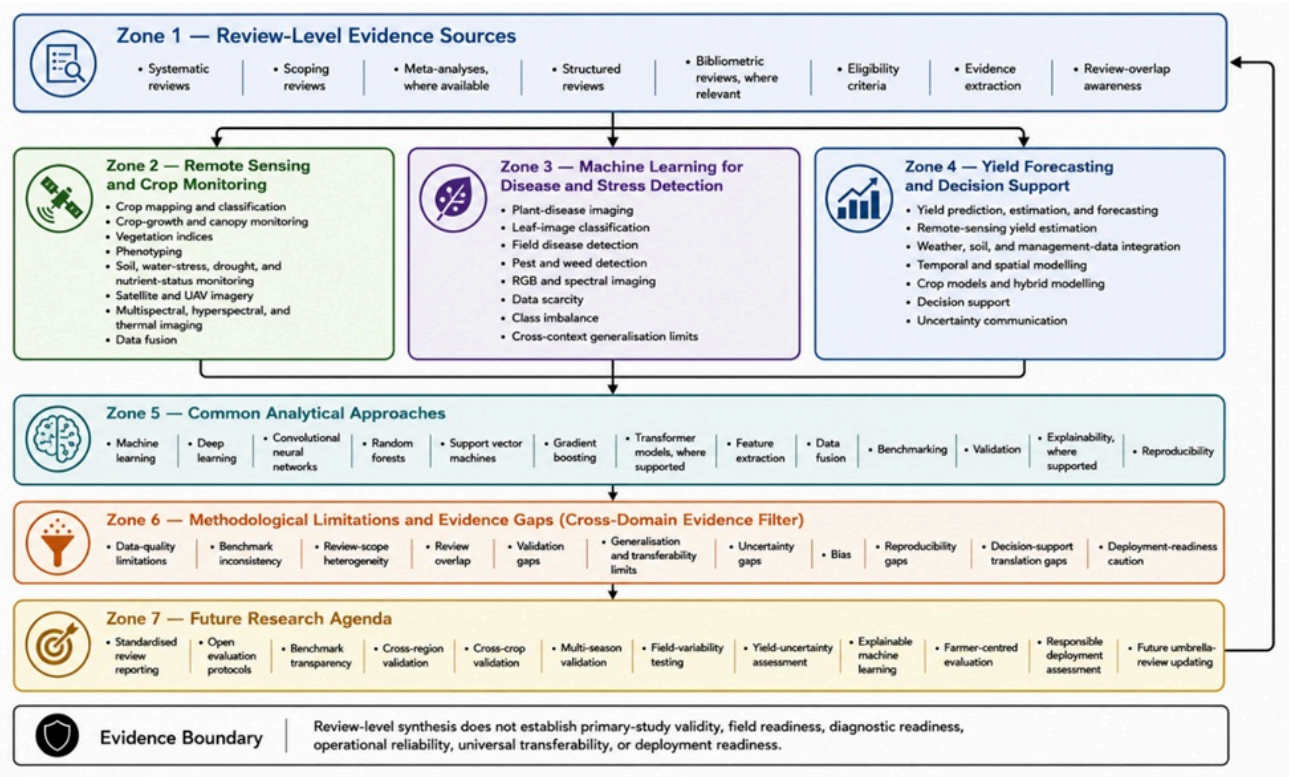


Figure 1. Umbrella Evidence Synthesis Map for Remote Sensing, Machine Learning, and Smart Farming

Alt-text description: A conceptual umbrella evidence synthesis map showing review-level evidence from remote sensing and machine learning reviews connected to crop monitoring, disease detection, yield forecasting, decision support, evidence strength, methodological limitations, gaps, and future research priorities.

Reproducibility is a cross-cutting condition for interpreting apparent evidence strength. Remote-sensing research can be difficult to reproduce when data versions, preprocessing decisions, code, computational environments, spatial partitions, parameter settings, or analytical workflows are incompletely reported [35]. Similar concerns apply to image-based disease detection and yield modelling, where undocumented augmentation, data cleaning, hyperparameter selection, or feature engineering can alter outcomes. Reproducibility does not prove validity, because a transparent workflow may still contain biased samples or inappropriate validation. It does, however, enable scrutiny, sensitivity analysis, independent reruns, and more credible comparison. Benchmarking should therefore include data provenance, contextual metadata, declared splits, and complete evaluation protocols rather than relying on leaderboard values alone.

Overall, the umbrella synthesis identifies a layered evidence structure. Breadth is greatest for remote-sensing applications and image-based machine learning; methodological concentration is strongest for crop monitoring and controlled-image disease classification; and evidence is more variable for transferable disease recognition, calibrated yield forecasting, and workflow-embedded decision support. Review overlap and scope heterogeneity mean that repeated positive conclusions cannot be counted as independent confirmations without examining their underlying studies [12]. The appropriate interpretation is that remote sensing and machine learning offer a substantial, rapidly developing research base whose decision-support value remains conditional on data provenance, independent validation, uncertainty, reproducibility, explainability, institutional context, and responsible deployment assessment. This synthesis supports comparative evidence mapping and research-priority development, not universal model selection or operational prescription.

Evidence Gaps and Future Research Agenda

The first priority is to improve the reliability of review-level inference itself. Future umbrella reviews should map overlap among included reviews whenever primary-study lists permit, because repeated inclusion can create apparent convergence without independent replication [12]. Review reporting should specify search coverage, review type, eligibility rules, primary-study designs, application definitions, appraisal methods, and limits to synthesis, consistent with the methodological need for explicitly planned overview questions and methods [9]. Benchmark transparency should extend beyond dataset names to versions, provenance, licensing, duplicate detection, class distributions, preprocessing, split logic, and known domain constraints. These advances would make evidence updates and comparisons more defensible, although they would not remove bias or validate the underlying primary studies by themselves.

A second priority is external validation designed around agricultural variation. Cross-region testing should separate training and evaluation geographically; cross-crop testing should hold out entire crop types or provide per-crop transport analyses; and multi-season testing should preserve chronology and include genuinely unseen seasons. Yield reviews particularly support the need for geographically and temporally independent evaluation because current comparisons are often crop-, region-, and dataset-specific [31]. Disease-recognition reviews likewise justify prospective field datasets containing complex backgrounds, cultivar diversity, mixed stressors, symptom stages, and expert-confirmed reference labels [22]. Yield forecasting should state the true information cut-off, forecast horizon, and data availability, and should complement point errors with calibrated intervals, probabilistic scores, or scenario sensitivity. These practices would clarify transportability and risk without implying global or universal generalisation.

A third priority is integration that remains testable, explainable, and reproducible. Weather, soil, management, crop-model, and remote-sensing data should be combined through modular designs that preserve source provenance, describe missing-data handling, and permit ablation of each modality [29]. Process-based and data-driven components should report their separate and propagated uncertainties rather than allowing hybrid complexity to obscure error sources [30]. Explainable machine learning should be evaluated for stability, faithfulness, agronomic coherence, and usefulness to

intended users; explanation alone must not substitute for predictive validity, fairness, or accountability [34]. Executable workflows, versioned data and code, declared environments, and independent reruns would strengthen reproducibility and comparison [35]. Open evaluation protocols should include spatial, temporal, imbalance-aware, and external tests, while avoiding the interpretation of leaderboard success as deployment certification.

A final priority is to study translation from analytical output to responsible decision support. Farmer-centred and workflow-embedded evaluations should examine timing, comprehensibility, burden, access, trust calibration, override behaviour, error consequences, and diversity among farm contexts, reflecting the social and institutional dimensions emphasised in digital-agriculture reviews [5]. Supply-side structures, service relationships, ownership arrangements, and platform dependencies should also be examined because they affect who can use a system and under what conditions [32]. Responsible deployment assessment should address privacy, fairness, accountability, security, environmental and market externalities, incident handling, and post-deployment monitoring where ethically appropriate [33]. Future umbrella-review updates should maintain versioned extraction records and reassess overlap, methodological quality, and emerging evidence rather than appending new reviews mechanically. **Table 4** presents evidence gaps and future research priorities emerging from the umbrella review.

Table 4. Evidence Gaps and Future Research Agenda for Remote Sensing, Machine Learning, and Smart Farming: Review Overlap, Standardized Reporting, Benchmark Transparency, Cross-Region Validation, Cross-Crop Validation, Multi-Season Validation, Field Variability, Yield Uncertainty, Explainability, Reproducibility, Open Protocols, Farmer-Centered Evaluation, Responsible Deployment, and Future Review Updating

Evidence gap or research priority	Problem addressed	Required methodological advance	Required data or infrastructure	Validation need	Review-level evidence basis	Expected contribution	Best practice
Review-overlap assessment	Repeated primary studies may create false convergence across reviews	Construct primary-study citation matrices and quantify overlap where records permit	Machine-readable bibliographies and persistent study identifiers	Compare synthesis results with and without highly overlapping reviews	Overlap-management guidance identifies this as a core overview problem [12]	More defensible interpretation of evidence density	Refine level of its validity
Standardized review reporting	Review scope, methods, evidence types, and limitations are inconsistently reported	Use domain-adapted review and overview reporting structures	Registered protocols, structured extraction forms, transparent amendments	Audit reporting completeness in future updates	Overview planning requires explicit questions and bounded objectives [9]	Improved comparability and updateability	Recheck removed the
Benchmark dataset transparency	Dataset versions, duplicates, classes, and splits may be unclear or repeatedly reused	Require provenance, versioning, licensing, class distributions, and split documentation	Dataset registries, data cards, persistent versions	Independent reruns using declared data versions and partitions	Reproducibility reviews emphasise transparent data and workflow records [35]	Lower leakage risk and more interpretable comparison	Trace back remain

Cross-region validation	Models may exploit local climate, cultivar, management, or acquisition conditions	Use geographically separated development and evaluation with transport analysis	Harmonised multi-region datasets and metadata partnerships	Independent regional external validation	Yield-review evidence identifies limited cross-region comparison [31]	Clearer evidence of geographic transferability	Selec does gen
Cross-crop validation	Crop morphology, symptoms, physiology, and yield relationships differ	Hold out crop species or use explicit multi-crop transfer protocols	Multi-crop image, spectral, weather, soil, and management datasets	Per-crop reporting on unseen crop tests	Disease-recognition evidence shows sensitivity to crop and dataset composition [26]	Better definition of reusable versus crop-specific models	Ac per mus cro
Multi-season validation	Seasonal weather and management shifts alter data distributions	Use temporally blocked, prospective, multi-year designs	Longitudinal field datasets and stable calibration records	Train on earlier seasons and test on unseen later seasons	Phenotyping reviews identify context and season dependencies [16]	Evidence about temporal robustness	Seve ma ra e
Disease detection under field variability	Controlled leaf images underrepresent clutter, mixed stresses, and symptom diversity	Collect prospective field and canopy datasets with realistic negatives	Expert-confirmed labels, multi-stressor sampling, acquisition metadata	Multi-site field comparison against appropriate reference diagnosis	CNN reviews identify controlled-dataset concentration and weak external validation [22]	More credible field-domain recognition evidence	R valid dia de a
Yield forecasting uncertainty assessment	Point metrics do not communicate calibration or decision risk	Produce probabilistic forecasts, intervals, calibration curves, and sensitivity analyses	Ensembles, weather scenarios, and decision-relevant thresholds	Out-of-time calibration using proper scoring rules	Deep-learning and remote-sensing yield reviews identify uncertainty and comparison gaps [31]	Forecasts whose uncertainty is interpretable at a stated lead time	Un estim guar d
Integration of weather, soil, and management data	Contextual variables are omitted or inconsistently encoded	Build modular multimodal models with provenance and missingness handling	Linked weather, soil, management, and earth-observation infrastructure	Ablation and external validation for each data source	Agrarian-factor reviews support contextual integration while noting data	Evidence about when added context improves prediction	Mo aut impr

					constraints [29]		
Explainable machine learning	Complex predictions may lack stable or meaningful reasons	Compare interpretable models with faithful, stability-tested explanation methods	Feature dictionaries, domain constraints, and explanation interfaces	Technical faithfulness plus expert and user evaluation	Responsible-AI guidance calls for transparency and stakeholder attention [34]	More inspectable decision-support research	Ex does va fai acc
Reproducible workflows	Incomplete code, preprocessing, environments, and splits obstruct replication	Publish executable pipelines, data lineage, parameter records, and environments	Versioned repositories, containers, workflow engines	Independent rerun and sensitivity analysis	Reproducibility guidance identifies transparent workflows as essential [35]	Greater confidence in reported comparisons	A re flaw rema
Open evaluation protocols	Metrics and split designs differ across studies and reviews	Create task-specific protocols for spatial, temporal, imbalance, and external testing	Public evaluation infrastructure or sealed test sets with metadata	Multi-team blind evaluation and detailed error analysis	Yield-prediction reviews show heterogeneous datasets and metrics [28]	More comparable evidence across methods	Lea n b de ce
Decision-support translation studies	Predictions are seldom evaluated for timeliness, actionability, and consequences	Conduct human-centred, workflow-embedded, mixed-method studies	Co-designed interfaces, use logs, and explicit outcome definitions	Prospective evaluation of use, override, errors, and downstream effects	Supply-side reviews identify organisational conditions shaping digital use [32]	Evidence connecting analytics with real decision processes	No c recor follo u va
Farmer-centered evaluation	Technical reviews underrepresent farmer priorities, access, and burden	Use participatory design and stratified evaluation across farm types	Trusted local partnerships, accessibility support, and compensated participation	Usability, trust calibration, equity, and opt-out assessment	Social-science reviews identify actor and institutional dimensions [5]	Better alignment with diverse users and contexts	Part not farm
Responsible deployment assessment	Accuracy-centred evaluation misses privacy, fairness, externalities, and accountability	Apply lifecycle risk assessment with explicit responsibility allocation	Governance records, incident processes, and stakeholder oversight	Pre-deployment scenarios and post-deployment monitoring where appropriate	Systemic-risk analysis shows that model accuracy captures only part of agricultural-AI risk [33]	Evidence about conditions and consequences of deployment	Ass not re safe

Future umbrella-review updating	Technologies, reviews, datasets, and methods will continue to change	Maintain a living candidate register and versioned overlap and extraction maps	Search alerts, persistent DOI records, and change logs	Independent screening checks and documented update decisions	Overview-method research supports method selection tailored to evidence structure [10]	Transparent updates without obscuring earlier evidence versions	Upd reas an ra a me
---------------------------------	--	--	--	--	--	---	---------------------

Conclusion

This umbrella review provides an integrated review-level synthesis of remote sensing, machine learning, and smart-farming evidence for crop monitoring, disease detection, yield forecasting, and decision support. The reviewed literature indicates substantial methodological development across satellite and UAV observation, multispectral and hyperspectral imaging, computer vision, conventional machine learning, deep learning, crop models, and multimodal forecasting. Its strongest contribution is evidence mapping: it identifies where research is concentrated, where findings converge, and where methodological limitations recur. Review-level evidence does not independently establish primary-study validity, reliable diagnosis, operational forecasting, universal model transfer, farmer approval, or deployment readiness. Claims of practical readiness should therefore remain conditional on review quality, primary-study overlap, transparent data provenance, independent spatial and temporal validation, cross-crop and cross-region testing, uncertainty assessment, reproducibility, explainability, workflow fit, and responsible governance. Progress toward decision-support value requires not only better models but also stronger evidence designs and explicit attention to the institutional and human settings in which analytical outputs may be used.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Wolfert S, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming - a review. *Agric Syst*. 2017;153:69-80.
<https://doi.org/10.1016/j.agsy.2017.01.023>.
- Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. *Sensors (Basel)*. 2018;18(8):2674.
<https://doi.org/10.3390/s18082674>.
- Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. *Comput Electron Agric*. 2018;147:70-90.
<https://doi.org/10.1016/j.compag.2018.02.016>.
- Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review. *Remote Sens Environ*. 2020;236:111402.
<https://doi.org/10.1016/j.rse.2019.111402>.
- Klerkx L, Jakku E, Labarthe P. A review of social science on digital agriculture, smart farming and Agriculture 4.0: New contributions and a future research agenda. *NJAS Wageningen J Life Sci*. 2019;90-91:100315.
<https://doi.org/10.1016/j.njas.2019.100315>.
- Saiz-Rubio V, Rovira-Más F. From smart farming towards Agriculture 5.0: A review on crop data management. *Agronomy*. 2020;10(2):207.
<https://doi.org/10.3390/agronomy10020207>.
- Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: A comprehensive updated review. *Sensors (Basel)*. 2021;21(11):3758.
<https://doi.org/10.3390/s21113758>.
- Sharma A, Jain A, Gupta P, Chowdary V. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*. 2021;9:4843-73.
<https://doi.org/10.1109/ACCESS.2020.3048415>.
- Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. *Syst Rev*. 2018;7(1):39.
<https://doi.org/10.1186/s13643-018-0695-8>.
- Pollock A, Campbell P, Brunton G, Hunt H, Estcourt L. Selecting and implementing overview methods: Implications from five exemplar overviews. *Syst Rev*. 2017;6(1):145.
<https://doi.org/10.1186/s13643-017-0534-3>.
- Ballard M, Montgomery P. Risk of bias in overviews of reviews: A scoping review of methodological guidance and four-item checklist. *Res Synth Methods*. 2017;8(1):92-108.
<https://doi.org/10.1002/jrsm.1229>.
- Lunny C, Pieper D, Thabet P, Kanji S. Managing overlap of primary study results across systematic reviews: Practical considerations for authors of overviews of reviews. *BMC Med Res Methodol*. 2021;21(1):140.
<https://doi.org/10.1186/s12874-021-01269-y>.

Sishodia RP, Ray RL, Singh SK. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* 2020;12(19):3136.
<https://doi.org/10.3390/rs12193136>.

Maes WH, Steppe K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* 2019;24(2):152-64.
<https://doi.org/10.1016/j.tplants.2018.11.007>.

Tsouros DC, Bibi S, Sarigiannidis PG. A review on UAV-based applications for precision agriculture. *Information.* 2019;10(11):349.
<https://doi.org/10.3390/info10110349>.

Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front Plant Sci.* 2017;8:1111.
<https://doi.org/10.3389/fpls.2017.01111>.

Kasampalis DA, Alexandridis TK, Deva C, Challinor A, Moshou D, Zalidis G. Contribution of remote sensing on crop models: A review. *J Imaging.* 2018;4(4):52.
<https://doi.org/10.3390/jimaging4040052>.

Karmakar P, Teng SW, Murshed M, Pang S, Li Y, Lin H. Crop monitoring by multimodal remote sensing: A review. *Remote Sens Appl Soc Environ.* 2024;33:101093.
<https://doi.org/10.1016/j.rsase.2023.101093>.

Lowe A, Harrison N, French AP. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods.* 2017;13(1):80.
<https://doi.org/10.1186/s13007-017-0233-z>.

Saleem MH, Potgieter J, Arif KM. Plant disease detection and classification by deep learning. *Plants (Basel).* 2019;8(11):468.
<https://doi.org/10.3390/plants8110468>.

Ngugi LC, Abelwahab M, Abo-Zahhad M. Recent advances in image processing techniques for automated leaf pest and disease recognition—a review. *Inf Process Agric.* 2021;8(1):27-51.
<https://doi.org/10.1016/j.inpa.2020.04.004>.

Abade A, Ferreira PA, Vidal FB. Plant diseases recognition on images using convolutional neural networks: A systematic review. *Comput Electron Agric.* 2021;185:106125.
<https://doi.org/10.1016/j.compag.2021.106125>.

Liu J, Wang X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods.* 2021;17(1):22.
<https://doi.org/10.1186/s13007-021-00722-9>.

Cardim Ferreira Lima M, Damascena de Almeida Leandro ME, Valero C, Pereira Coronel LC, Gonçalves Bazzo CO. Automatic detection and monitoring of insect pests—a review. *Agriculture.* 2020;10(5):161.
<https://doi.org/10.3390/agriculture10050161>.

Hasan ASMM, Sohel F, Diepeveen D, Laga H, Jones MGK. A survey of deep learning techniques for weed detection from images. *Comput Electron Agric.* 2021;184:106067.
<https://doi.org/10.1016/j.compag.2021.106067>.

Barbedo JGA. Factors influencing the use of deep learning for plant disease recognition. *Biosyst Eng.* 2018;172:84-91.
<https://doi.org/10.1016/j.biosystemseng.2018.05.013>.

Chlingaryan A, Sukkarieh S, Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput Electron Agric.* 2018;151:61-9.
<https://doi.org/10.1016/j.compag.2018.05.012>.

van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. *Comput Electron Agric.* 2020;177:105709.
<https://doi.org/10.1016/j.compag.2020.105709>.

Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput Electron Agric.* 2018;155:257-82.
<https://doi.org/10.1016/j.compag.2018.10.024>.

Jin X, Kumar L, Li Z, Feng H, Xu X, Yang G, et al. A review of data assimilation of remote sensing and crop models. *Eur J Agron.* 2018;92:141-52.
<https://doi.org/10.1016/j.eja.2017.11.002>.

Muruganantham P, Wibowo S, Grandhi S, Samrat NHL, Islam N. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sens.* 2022;14(9):1990.
<https://doi.org/10.3390/rs14091990>.

Birner R, Daum T, Pray C. Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges. *Appl Econ Perspect Policy.* 2021;43(4):1260-85.
<https://doi.org/10.1002/aepp.13145>.

Tzachor A, Devare M, King B, Avin S, Ó hÉigeartaigh S. Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nat Mach Intell.* 2022;4(2):104-9.
<https://doi.org/10.1038/s42256-022-00440-4>.

Dara R, Hazrati Fard SM, Kaur J. Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Front Artif Intell.* 2022;5:884192.
<https://doi.org/10.3389/frai.2022.884192>.

Kedron P, Frazier AE. How to improve the reproducibility, replicability, and extensibility of remote sensing research. *Remote Sens.* 2022;14(21):5471.
<https://doi.org/10.3390/rs14215471>.