

REVIEW

Open access

Data Science in Precision Agriculture: A Bibliometric and Thematic Review of Research Evolution, Knowledge Structures, and Emerging Frontiers

Mikko Lahtinen^{1*}, Elina Salo¹, Juhani Virtanen²

Abstract

Data science has become a central analytical foundation of precision agriculture, enabling researchers and practitioners to transform heterogeneous farm, crop, soil, climate, machinery, sensor, and remote-sensing data into decision-relevant knowledge. This original bibliometric and thematic review examines how precision agriculture data science has evolved as a research field, how its knowledge structures are organized, and which methodological and application frontiers are shaping its future direction. The review follows a transparent bibliometric logic by distinguishing database searching, eligibility screening, metadata cleaning, keyword normalization, bibliometric indicators, thematic mapping, and interpretive synthesis. Because verified database-export counts were not available in the present working environment, the manuscript does not report invented publication totals, citation rankings, cluster sizes, or network metrics. Instead, it develops a cautious bibliometric-thematic framework grounded in the approved reference set and in clearly stated methodological safeguards. The review focuses on research evolution from early sensor analytics and remote observation toward machine learning, deep learning, big data, IoT-enabled systems, data fusion, decision support, and sustainable smart farming. It also examines knowledge structures around journals, keywords, co-occurrence logic, thematic clusters, and emerging research fronts. The main contribution is an integrated synthesis of precision agriculture data science as a field shaped by technical innovation, methodological fragmentation, data-quality constraints, interoperability challenges, explainability needs, reproducibility demands, and farmer-centered implementation concerns. The review provides a future agenda for more robust, transparent, usable, and responsible data science in precision agriculture.

Keywords Precision agriculture, Machine learning, Smart farming, Data science, Bibliometric review, Thematic review

*Correspondence:

Mikko Lahtinen
mikko.lahtinen@helsinki.fi

¹ Department of Agricultural Data Science and Bibliometrics, Faculty of Agriculture, University of Helsinki, Helsinki, Finland

² Department of Knowledge Structures and Research Frontiers, Faculty of Agriculture, University of Eastern Finland, Kuopio, Finland

Introduction

Data science has become a central driver of precision agriculture because farm-management decisions increasingly depend on the ability to extract actionable knowledge from heterogeneous, high-volume, and spatially distributed data. Machine learning and deep learning have expanded the analytical repertoire of agricultural research by enabling classification, prediction, detection, estimation, and decision-support tasks across crop, soil, livestock, and production contexts [1, 2]. In this setting, precision agriculture is no longer defined only by variable-rate application or geospatial sensing; it is increasingly shaped by the full data pipeline linking acquisition, preprocessing, modelling, interpretation, and operational decision-making.

The emergence of data-intensive agriculture has also changed the scale and organization of research knowledge. Deep learning applications have strengthened image-based and high-dimensional analysis in agricultural systems, including computer-vision tasks associated with crop monitoring, disease detection, and production assessment [2]. These developments have intensified the need for bibliometric and thematic review approaches that can map not only the topics being studied, but also the relationships among methods, journals, keywords, application domains, and emerging fronts.

Smart farming further extends precision agriculture into connected data ecosystems in which sensors, platforms, analytics, and decision-support services interact across farm operations and agri-food value chains. Big data scholarship has shown that smart farming involves data capture, storage, sharing, interpretation, and governance questions that go beyond narrow algorithmic performance [3]. A bibliometric review of precision agriculture data science therefore needs to distinguish research visibility from research quality, technical novelty from operational usefulness, and keyword growth from practical implementation.

The field is also socio-technical. Digital agriculture scholarship emphasizes that technologies such as AI, robotics, sensor networks, decision-support systems, and data platforms are embedded in institutional, advisory, economic, ethical, and farmer-centered contexts [4]. The aim of this review is to map the evolution of data science in precision agriculture, identify visible knowledge structures and thematic clusters, interpret emerging research frontiers, and develop a future agenda for robust, transparent, reproducible, interoperable, and usable agricultural data science.

Data Science in Precision Agriculture

In precision agriculture, data science can be understood as the integrated use of data acquisition, data cleaning, data fusion, statistical modelling, machine learning, visualization, optimization, and decision-support methods to improve the interpretation of farm and environmental variability. Updated agricultural machine-learning literature indicates that supervised, unsupervised, and hybrid modelling approaches have been applied across crop monitoring, yield prediction, disease detection, irrigation support, soil analysis, and machinery-related decision contexts [5]. This positions data science as a connective layer between field observation and management action.

Crop yield prediction illustrates how precision agriculture data science links environmental data, agronomic variables, remote observations, and predictive modelling. Systematic evidence on yield prediction shows that machine-learning models are frequently used to integrate weather, soil, vegetation, management, and remote-sensing variables into predictive workflows [6]. However, the value of these workflows depends not only on model accuracy, but also on data quality, temporal transferability, spatial generalizability, uncertainty communication, and whether predictions can support decisions under real farming constraints.

Remote sensing and UAV analytics have expanded the observational foundations of precision agriculture. UAV-based precision agriculture applications demonstrate how imagery, vegetation indices, spatial sampling, and field-level monitoring can support crop assessment and management decisions [7]. Within a data science framework, such systems are important because they create high-resolution data streams that require preprocessing, feature extraction, model selection, validation, and translation into agronomic interpretation.

Data science in precision agriculture is therefore broader than the use of algorithms. It includes the design of data pipelines, the integration of multi-source data, the cleaning of inconsistent metadata, the selection of appropriate modelling approaches, the interpretation of outputs, and the communication of uncertainty to decision-makers. It also connects precision agriculture with smart farming, digital agriculture, AI-enabled decision support, and sustainability-oriented farm management, while raising persistent questions about reproducibility, interoperability, explainability, data ownership, and farmer-centered usability.

Bibliometric Method and Search Strategy

This review was designed as a bibliometric and thematic review rather than a generic narrative review. Bibliometric review design requires explicit decisions about database searching, metadata export, deduplication, screening, author-keyword cleaning, index-keyword normalization, indicator selection, and science-mapping interpretation. The bibliometric framework is relevant because it demonstrates how bibliographic records can be processed for descriptive bibliometrics, conceptual structure analysis, keyword analysis, and science mapping when reliable metadata exports are available [8].

The planned search strategy uses major bibliographic databases as primary sources and supplementary publisher or scholarly search interfaces only for checking metadata, DOI completeness, and topical relevance. Bibliometric indicators are interpreted cautiously because citation visibility and publication patterns can be influenced by indexing coverage, disciplinary citation practices, publication age, database selection, language, journal scope, and search-string construction [9]. Consequently, citation counts, journal visibility, or keyword frequency should not be treated as direct measures of research quality or practical importance.

The methodological workflow distinguishes the bibliometric dataset from the manuscript reference set. Software choice matters because tools differ in export compatibility, network construction, visualization logic, clustering options, and reproducibility support [10]. For this reason, the planned protocol specifies DOI-first deduplication, title-normalized duplicate checking, keyword cleaning rules, field-code adaptation by database, and transparent reporting of missing metadata. No record totals, screening counts, citation metrics, network centrality values, or cluster sizes are reported unless they are generated from an actual export.

The interpretive aim of the method is to move from bibliometric mapping to thematic synthesis. Bibliometric research should not stop at descriptive statistics; it should use mapping outputs to clarify intellectual structure, conceptual relationships, theoretical contribution, and future research direction [11]. In this manuscript, thematic interpretation is therefore tied to approved charted evidence, including machine learning, deep learning, big data, UAV analytics, remote sensing, IoT systems, crop data management, explainability, sustainability, socio-ethical concerns, and farmer-centered research priorities.

Table 1 summarises the bibliometric search, screening, data-cleaning, and thematic mapping protocol used to analyze data science research in precision agriculture.

Table 1. Bibliometric Search, Screening, Data-Cleaning, and Thematic Mapping Protocol for Data Science Research in Precision Agriculture: Databases, Search Terms, Eligibility Criteria, Deduplication, Metadata Fields, Bibliometric Indicators, Keyword Normalization, and Cluster Interpretation

Protocol component	Operational definition	Implementation rule	Data field required	Quality-control action	Risk of bias or limitation	Reporting requirement
Database selection	Identification of bibliographic sources for precision agriculture data science records	Use major bibliographic databases as primary sources; use publisher pages and scholarly search only for metadata checking	Database name; export date; record type	Record database used for each export	Database coverage may favor certain journals, languages, or disciplines	Report each database searched and avoid claims beyond database coverage

Search-string design	Boolean expression linking precision agriculture terms with data science terms	Combine precision agriculture, smart farming, digital agriculture, AI, ML, big data, analytics, remote sensing, IoT, and decision-support terms	Title; abstract; keywords	Adapt field codes to each database	Search terms may miss papers using different terminology	Report full search strings and field restrictions
Eligibility screening	Selection of records that match topic, document type, and metadata requirements	Include peer-reviewed journal articles focused on agricultural data science; exclude non-peer-reviewed and off-topic items	Title; abstract; keywords; journal; DOI	Title/abstract screening followed by metadata checking	Borderline records may be excluded if agricultural relevance is weak	State inclusion and exclusion criteria
Deduplication	Removal of duplicate records across databases	Apply DOI-first matching, then normalized title matching, then manual author-year-title inspection	DOI; title; authors; year	Preserve one complete metadata record per paper	Missing DOI or title variation can reduce duplicate detection	Report method and verified duplicate count only if generated
Metadata cleaning	Correction and standardization of exported bibliographic fields	Standardize author names, journal titles, years, DOIs, abstracts, and keywords	Authors; journal; year; DOI; abstract; keywords	Check missing or inconsistent fields	Incomplete records may reduce mapping reliability	Report excluded records only if screening log exists
Keyword normalization	Cleaning of author and index keywords before co-occurrence analysis	Lowercase terms; merge spelling variants; harmonize singular/plural forms; document synonym merges	Author keywords; index keywords	Maintain cleaning dictionary	Over-normalization may erase meaningful conceptual differences	Report major keyword-cleaning decisions
Bibliometric indicators	Measures used to interpret research evolution and knowledge structure	Use annual output, journal distribution, keyword co-occurrence, citation visibility, collaboration signals, and thematic density only when data support them	Year; source; keywords; references; citations; affiliations	Avoid unsupported rankings	Indicators reflect visibility and database coverage, not quality	Explain interpretive limits for each indicator
Thematic mapping	Interpretation of conceptual clusters from keywords and	Label clusters using top terms, representative	Keywords; abstract; title; selected full text	Cross-check labels against charted evidence	Algorithmic clusters may be sensitive to	Report clusters as interpretations,

	structured reading	records, and thematic logic			threshold and cleaning rules	not objective facts
Cluster interpretation	Narrative explanation of research themes and evolution	Link clusters to precision agriculture domains and data science methods	Keywords; abstracts; application fields	Avoid unsupported “hot topic” claims	Cluster names may overstate coherence	Use cautious wording such as “the dataset suggests”
Synthesis and agenda	Translation of mapped evidence into research gaps and future priorities	Integrate bibliometric patterns with thematic reading	Cluster labels; limitations; applications; methods	Distinguish evidence from author interpretation	Agenda may reflect reference-set scope	State limitations and avoid generalizing beyond evidence

Figure 1 presents the bibliometric review workflow used to identify, clean, analyze, and thematically interpret records on data science in precision agriculture.

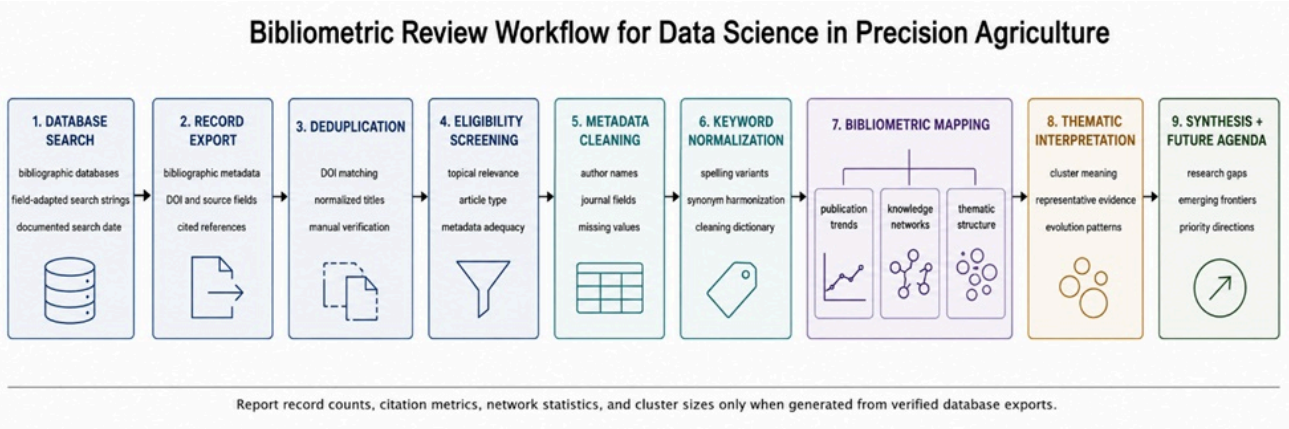


Figure 1. Bibliometric Review Workflow for Data Science in Precision Agriculture. The figure shows the workflow from database searching and record export to deduplication, eligibility screening, metadata cleaning, bibliometric analysis, thematic cluster interpretation, and synthesis of emerging frontiers. All record counts must be generated from the actual Part 1 search and screening process and must not be invented.

Alt-text description: A bibliometric review workflow diagram showing database searching, record export, duplicate removal, title and abstract screening, eligibility assessment, metadata cleaning, keyword normalization, bibliometric indicator analysis, thematic clustering, interpretation, and synthesis of research gaps and emerging frontiers. Count boxes must use only verified numbers from the bibliometric search process.

Q1-journal-publication-ready figure prompt: Create a clean, high-resolution bibliometric review workflow diagram titled “Bibliometric Review Workflow for Data Science in Precision Agriculture.” Use a professional Q1-journal style suitable for publication in precision agriculture, data science, or bibliometric review journals. The figure should show a left-to-right or top-to-bottom workflow with the following stages: database search, record export, duplicate removal, eligibility screening, metadata cleaning, keyword normalization, bibliometric indicator analysis, network or co-occurrence analysis, thematic cluster interpretation, bibliometric-thematic synthesis, and future agenda development. Include count boxes only if verified counts were generated in Part 1; otherwise use count-free process labels. Use clear boxes, arrows, consistent typography, and colorblind-safe colors. Do not use database logos, software logos, copied VOSviewer screenshots, copied Bibliometrix plots, maps, flags, photographs, or copyrighted imagery. Ensure the figure remains understandable in grayscale and suitable for 300 dpi journal export.

Publication Trends and Knowledge Structures

Publication-trend interpretation in this review is restricted to defensible bibliometric logic rather than invented numerical outputs. Prior digital-agriculture bibliometric studies show that science-mapping approaches can identify technology-centered streams, strategic themes, and research-evolution structures in digital agriculture when based on documented search strings, database exports, and keyword-network procedures [12, 13]. In the present manuscript, these prior studies support the methodological claim that publication evolution should be interpreted through transparent mapping rules rather than unsupported claims about growth rates, rankings, or dominance.

Knowledge-structure analysis is especially useful for identifying how specialized subfields form within the broader precision agriculture data science literature. Bibliometric analysis of deep learning in precision agriculture shows that deep-learning research can be treated as a distinct analytical subfield within AI-enabled agricultural research [14]. This supports the interpretation that precision agriculture data science contains method-specific knowledge structures, including machine learning, deep learning, remote sensing analytics, IoT data systems, and predictive decision-support applications.

Smart farming bibliometric mapping further indicates that digital agriculture research can be organized into thematic clusters rather than a single unified research stream. Theme mapping and bibliometric analysis of smart farming demonstrate how co-citation, co-word, and thematic mapping approaches can be used to interpret conceptual organization and research evolution across smart farming topics [15]. For precision agriculture data science, this means that knowledge structures should be read as interacting clusters of methods, data sources, applications, and socio-technical concerns.

Precision agriculture bibliometrics can also support research-agenda formation when bibliometric outputs are integrated with thematic interpretation. Recent bibliometric work on precision agriculture connects mapped research structures to future agenda development, including technology applications, challenges, and emerging opportunities [16]. This review follows that logic by treating publication trends and knowledge structures as inputs into a broader synthesis of methodological robustness, implementation readiness, explainability, interoperability, and sustainability.

Table 2 summarises the publication-trend and knowledge-structure indicators used to interpret the evolution of precision agriculture data science.

Table 2. Publication-Trend and Knowledge-Structure Indicators for Precision Agriculture Data Science: Annual Output, Journal Distribution, Author Collaboration, Citation Visibility, Keyword Co-Occurrence, Co-Citation Patterns, Country or Institution Signals, and Thematic Density

Bibliometric indicator	What it measures	Interpretive value	Required data source	Possible bias	Recommended interpretation	Relevance to research evolution
Annual output	Number of eligible records by publication year	Indicates temporal expansion or contraction of indexed literature	Verified bibliographic export with publication years	Database coverage and indexing delays	Interpret as publication visibility, not field maturity	Helps identify whether research activity appears to intensify or diversify
Journal distribution	Distribution of records across journals	Shows disciplinary venues and	Journal/source title field	Database and journal-indexing bias	Avoid ranking journals as “best” based only on frequency	Reveals whether the field is concentrated or

		publication outlets				dispersed across domains
Author collaboration	Co-authorship patterns and collaboration networks	Indicates social organization of research	Author and affiliation metadata	Name ambiguity and incomplete affiliation data	Treat as collaboration signal, not productivity quality	Shows whether data science research is interdisciplinary or fragmented
Citation visibility	Citation counts or cited-reference patterns where available	Suggests visibility and intellectual linkage	Citation metadata and reference lists	Age bias, language bias, field citation norms	Never treat citation count as direct quality evidence	Helps identify visible knowledge anchors and older foundational work
Keyword co-occurrence	Frequency with which terms appear together	Reveals conceptual proximity among methods and applications	Author keywords and index keywords	Keyword inconsistency and synonym variation	Interpret after documented keyword normalization	Supports identification of thematic clusters and methodological trajectories
Co-citation patterns	Sources cited together by later records	Indicates perceived intellectual relatedness	Cited-reference metadata	Requires complete reference lists; favors older works	Interpret as intellectual-structure signal	Helps identify foundational or bridging literature
Bibliographic coupling	Records sharing references	Indicates similarity among contemporary papers	Reference lists for each record	Sensitive to reference-list length and coverage	Use to detect current research fronts where data permit	Useful for emerging-frontier mapping
Country or institution signals	Geographic or organizational metadata in publications	Indicates distribution of indexed research activity	Affiliation metadata	Affiliation cleaning problems and database coverage	Avoid claims of national superiority	Shows possible institutional concentration or collaboration patterns
Thematic density	Internal development of a keyword cluster	Indicates whether a theme appears internally connected	Keyword network and clustering output	Threshold and cleaning sensitivity	Interpret relative to network method and dataset size	Helps distinguish consolidated themes from emerging or peripheral themes
Thematic centrality	Connection of a theme to other themes	Indicates whether a cluster bridges broader	Keyword network and thematic-map output	Algorithm sensitivity and term ambiguity	Treat as structural position, not importance	Helps identify cross-cutting areas such as

		knowledge structures				IoT, AI, and decision support
--	--	----------------------	--	--	--	-------------------------------

Thematic Clusters and Research Evolution

Remote sensing analytics represents a major thematic cluster in precision agriculture data science because it links field observation, spatial variability, crop condition assessment, and model-based interpretation. Agricultural remote sensing scholarship shows that satellite, airborne, and proximal sensing data can be used to estimate vegetation status, crop biophysical variables, and environmental conditions, while also requiring careful attention to scale, calibration, uncertainty, and agronomic interpretation [17]. In bibliometric and thematic terms, this cluster functions as a bridge between data acquisition and decision-oriented analytics.

A second cluster centers on integrated precision agriculture techniques, including sensing, positioning, geospatial analysis, IoT-enabled monitoring, and decision-support practices. Precision agriculture is best understood as a system of interacting technologies rather than as a single tool, because data-driven management depends on the coordination of sensors, field machinery, analytics, and operational decision rules [18]. This cluster explains why research evolution in the field has moved from isolated measurement toward integrated data-to-decision systems.

Big data analytics forms a third cluster that links heterogeneous agricultural data sources to predictive modelling, platform-based decision support, and multi-scale farm management. Agricultural big data research highlights the importance of data volume, variety, velocity, processing pipelines, storage, and analytical workflows in transforming agricultural observations into usable knowledge [19]. Within the mapped literature, this cluster supports the interpretation that precision agriculture data science increasingly depends on data integration and computational infrastructure.

Predictive agronomy is another visible research-evolution pathway, especially in relation to yield prediction, nutrient-status estimation, crop stress detection, and irrigation support. Machine-learning research on yield and nitrogen-status estimation shows that predictive analytics can connect remote sensing, soil information, crop status, and management variables, although model transferability and validation remain persistent concerns [20]. Thermal remote sensing further illustrates how sensor-derived signals can be translated into stress detection and water-management support, making it a specialized subcluster within remote sensing analytics and decision support [21].

Table 3 presents the thematic clusters and research-evolution patterns identified in the precision agriculture data science literature.

Table 3. Thematic Clusters and Research-Evolution Patterns in Precision Agriculture Data Science: Remote Sensing Analytics, Machine Learning Prediction, IoT and Big Data Systems, Decision Support, Data Fusion, Sustainability Applications, Explainable AI, and Farmer-Centered Digital Agriculture

Thematic cluster	Core research focus	Dominant data science methods	Typical agricultural application	Knowledge-structure interpretation	Evolution pattern	Methodological limitation	
Remote sensing analytics	Conversion of spatial and spectral observations into crop and environmental indicators	Image analytics, vegetation-index analysis, machine learning,	Crop monitoring, biomass estimation, stress detection,	Bridges observational data and agronomic interpretation	Moving from descriptive imagery toward predictive and decision-	Scale mismatch, atmospheric effects, ground-truth limitations, and	Mu un i

		temporal modelling	field variability mapping		oriented analytics	transferability constraints	
Machine learning prediction	Modelling relationships among crop, soil, climate, sensor, and management variables	Supervised learning, ensemble models, regression, classification, feature selection	Yield prediction, nutrient-status estimation, pest and disease detection	Forms a central analytical cluster in precision agriculture data science	Shifting from model comparison toward operational prediction and decision support	Risk of overfitting, limited external validation, and weak uncertainty communication	r
Deep learning analytics	Use of high-dimensional models for visual and complex agricultural data	CNNs, deep neural networks, representation learning, image classification	Disease detection, weed recognition, crop classification, image-based phenotyping	Specialized AI subfield connected to computer vision and remote sensing	Expanding from controlled image tasks toward field-scale monitoring	Data hunger, black-box behaviour, and limited generalizability across conditions	Ex le c
IoT and big data systems	Continuous farm-data capture, connectivity, storage, and analytics	Sensor analytics, stream processing, cloud analytics, edge analytics	Irrigation monitoring, machinery data, environmental sensing, livestock and crop monitoring	Infrastructure-oriented cluster supporting real-time decision systems	Moving from sensor deployment toward connected data ecosystems	Connectivity, interoperability, cybersecurity, and data-governance challenges	I c st in ar s
Data fusion	Integration of heterogeneous data streams across spatial, temporal, and operational scales	Multisource integration, feature fusion, sensor fusion, geospatial analytics	Combining UAV, satellite, soil, climate, and management data	Cross-cutting cluster linking data sources and modelling workflows	Increasingly important as farms generate multimodal data	Inconsistent metadata, incompatible formats, missing values, and resolution mismatch	r fus
Decision support	Translation of analytics into management recommendations	Predictive analytics, optimization, rule-based systems, dashboards, decision models	Irrigation scheduling, fertilization, pest control, yield planning	Application-facing cluster connecting models with farm decisions	Evolving from diagnostic analytics toward prescriptive and advisory systems	Limited usability, weak farmer participation, and unclear actionability	rec an c
Sustainability applications	Use of data science to support resource efficiency and	AI, optimization, remote sensing,	Input-use efficiency, climate-smart management,	Links technical analytics with	Moving from production optimization toward	Sustainability outcomes may be assumed	e i

	environmental objectives	climate analytics, scenario modelling	environmental monitoring	sustainability outcomes	environmental and resilience goals	rather than measured	
Explainable AI and farmer-centered digital agriculture	Transparency, usability, trust, and human interpretation of data-driven tools	Explainable AI, interpretable ML, participatory design, uncertainty communication	Farmer-facing decision support and advisory systems	Socio-technical cluster linking algorithmic outputs to human decisions	Emerging from concerns over black-box systems and adoption barriers	Explanations may be technically correct but practically unusable	Fa X

Emerging Frontiers in Precision Agriculture Data Science

Crop data management is an emerging frontier because precision agriculture increasingly depends on the ability to organize data across acquisition, storage, processing, interpretation, and action. Scholarship on crop data management argues that advanced smart farming requires data infrastructures capable of supporting decision-making across the full pipeline from data capture to management recommendation [22]. This makes data management a foundational frontier rather than a secondary technical issue.

Data-centric AI is also becoming important because agricultural AI performance depends heavily on the quality, representativeness, annotation, provenance, and reusability of datasets. Recent data-centric AI research in the agrifood domain shifts attention from model novelty alone toward dataset design, data curation, missing-data handling, augmentation, benchmarking, and evaluation practices [23]. This frontier is particularly relevant for precision agriculture because field data are often noisy, seasonal, site-specific, and difficult to standardize.

Explainable AI is another frontier for precision agriculture data science, especially when predictions influence farm-management decisions. General XAI research provides concepts for interpreting black-box systems, clarifying model behaviour, and communicating reasons behind algorithmic outputs [24]. In agricultural decision support, this frontier must be adapted carefully because useful explanations should be agronomically meaningful, uncertainty-aware, and understandable to farmers, advisors, and other decision-makers.

Sustainable precision agriculture connects data-driven analytics with resource efficiency, climate-smart decision support, environmental monitoring, and responsible technology adoption. Data-driven AI applications for sustainable precision agriculture show how analytics can support productivity and sustainability goals, but they also indicate the need for implementation-oriented evaluation [25]. Agricultural IoT research further shows that real-time farm data streams, sensor networks, connectivity, and edge-to-cloud infrastructures remain essential foundations for advanced smart farming, while raising persistent challenges around interoperability, scalability, and reliability [26].

Bibliometric and Thematic Synthesis

The bibliometric and thematic evidence suggests that precision agriculture data science is organized around interacting technical and socio-technical knowledge structures. Responsible innovation research in smart farming shows that digital agricultural systems create socio-ethical challenges related to stakeholder inclusion, power, accountability, responsibility, and the organization of innovation processes [27]. This means that thematic synthesis should not interpret AI, remote sensing, IoT, or big data as purely technical clusters; it should also examine how these clusters affect farmers, advisors, institutions, and agricultural decision systems.

The synthesis also indicates that data science research must confront translational hurdles between analytical possibility and practical farm value. Digital agriculture scholarship identifies barriers that can prevent technical advances from delivering their full promise, including gaps in evidence, infrastructure, skills, incentives, data access, and usable decision support [28]. For precision agriculture data science, these barriers are central because even technically strong models may fail to produce meaningful outcomes if they are not validated, interpretable, interoperable, and embedded in real decision contexts.

Sustainability-oriented interpretation requires particular caution. Future-oriented analyses of digital agriculture show that claims about digital transformation and sustainable food systems can reflect different visions of agricultural futures, including productivity-oriented, platform-centered, ecological, and socially inclusive pathways [29]. Therefore, the thematic movement toward sustainable smart farming should not be treated as automatic evidence that data science improves sustainability; instead, sustainability claims require explicit evaluation of environmental, agronomic, social, and economic outcomes.

The mapped field also remains uneven in who benefits from data-intensive agriculture. Research on big data in farming shows that perceptions of agricultural data systems include concerns about scale, access, value distribution, and whether digital tools serve large farms more effectively than diverse farming contexts [30]. This is important for bibliometric synthesis because visible research themes may overrepresent technologically intensive production systems while underrepresenting smallholder settings, advisory realities, farmer learning processes, and locally adapted decision practices.

Overall, the synthesis positions precision agriculture data science as an evolving field that has progressed from sensing and prediction toward connected, AI-enabled, and sustainability-oriented decision systems, while remaining fragmented by data-quality, interoperability, explainability, reproducibility, and adoption challenges. Socio-ethical and inclusion-oriented perspectives are therefore necessary to interpret technical clusters responsibly and to avoid overstating the practical reach of highly visible data-science themes [27, 30].

Table 4 synthesizes the main bibliometric patterns, thematic frontiers, research gaps, and future agenda for data science in precision agriculture.

Table 4. Bibliometric and Thematic Synthesis of Data Science in Precision Agriculture: Knowledge Structures, Emerging Frontiers, Methodological Gaps, Interoperability Challenges, Explainability Needs, Farmer-Centered Design, Reproducibility, and Future Research Agenda

Synthesis domain	Observed bibliometric or thematic pattern	Main evidence basis	Research gap	Practical implication	Future research priority	Caution for interpretation
Research evolution	Movement from sensor-centered analytics toward AI-enabled, integrated decision support	Charted evidence on ML, DL, remote sensing, UAVs, big data, and IoT	Limited longitudinal validation of operational decision outcomes	Tools may remain experimental if not tested across seasons and farms	Multi-year, multi-location validation studies	Do not infer maturity from topic visibility alone
Knowledge structures	Method-specific	Bibliometric and review	Fragmentation across	Knowledge may remain siloed	Cross-domain synthesis and	Keyword co-occurrence does

	clusters around ML, DL, remote sensing, big data, IoT, and decision support	evidence from mapped subfields	technical communities and application domains	across journals, tools, and disciplines	shared terminology	not prove conceptual integration
Remote sensing analytics	Strong connection between spatial observation and crop monitoring	Remote sensing and UAV review evidence	Scale mismatch, limited ground truth, and uncertainty handling	Remote-sensing outputs may be difficult to translate into management action	Fusion of satellite, UAV, proximal, and field data with uncertainty estimates	High-resolution imagery does not guarantee better decisions
Machine learning prediction	Predictive analytics for yield, nutrients, disease, and crop status	ML and yield-prediction review evidence	Overfitting, site specificity, and weak external validation	Models may not generalize beyond training conditions	Open benchmarks, transparent reporting, and external validation	Accuracy metrics alone do not demonstrate farm usefulness
IoT and big data systems	Increasing dependence on connected farm data infrastructures	Big data and IoT evidence	Interoperability, connectivity, metadata, and data-governance gaps	Data streams may remain underused without integration standards	Edge-to-cloud architectures and interoperable data models	Sensor abundance should not be equated with actionable intelligence
Data fusion	Growing need to integrate heterogeneous spatial, temporal, and operational data	Charted cross-cutting evidence	Missing metadata, incompatible formats, and uncertain provenance	Poor fusion can reduce model reliability	FAIR-aligned data pipelines and uncertainty-aware fusion methods	Fusion claims require transparent data lineage
Explainable AI	Rising need for interpretable decision support	XAI and farmer-centered evidence	Explanations often lack agronomic and user-context relevance	Farmers may distrust opaque recommendations	Farmer-centered explanation design and uncertainty communication	Technical interpretability is not the same as practical understandability
Reproducibility	Growing concern over repeatable data-science workflows	Data-centric AI and methodological evidence	Limited open code, datasets, preprocessing details, and evaluation protocols	Results may be difficult to compare or reuse	Reproducible workflows, open benchmarks, and reporting standards	Reproducibility cannot be assumed from publication alone

Sustainability	Increasing linkage between AI, smart farming, and sustainable agriculture	Sustainable PA and digital-agriculture futures evidence	Environmental benefits are often asserted rather than evaluated	Digital tools may shift burdens or create uneven benefits	Whole-system sustainability assessment and climate-smart evaluation	Sustainability claims require measured outcomes
Farmer-centered design	Recognition that decision support must fit real users and contexts	Socio-ethical, perception, and priority-setting evidence	Limited participatory design and usability testing	Tools may fail adoption despite technical sophistication	Co-design with farmers, advisors, and institutions	Adoption should not be reduced to individual farmer willingness

Figure 2 illustrates the integrated bibliometric and thematic synthesis linking research evolution, knowledge structures, thematic clusters, emerging frontiers, and future research priorities in precision agriculture data science.

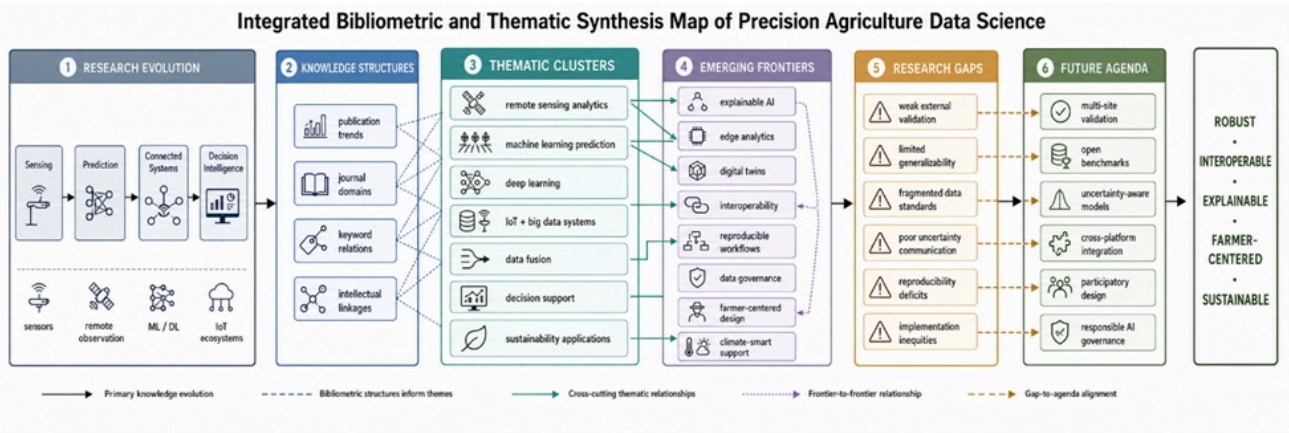


Figure 2. Integrated Bibliometric and Thematic Synthesis Map of Precision Agriculture Data Science. The figure links the evolution of data science research in precision agriculture to knowledge structures, thematic clusters, emerging frontiers, methodological gaps, and future research priorities. It summarizes how remote sensing analytics, machine learning, IoT, big data, data fusion, decision support, explainable AI, interoperability, sustainability, and farmer-centered design interact as research domains.

Alt-text description: A conceptual synthesis map showing precision agriculture data science at the center, connected to research evolution, knowledge structures, thematic clusters, and emerging frontiers. Surrounding nodes include remote sensing analytics, machine learning, deep learning, IoT and big data systems, data fusion, decision support, explainable AI, interoperability, reproducibility, sustainability, farmer-centered design, and future research priorities.

Research Gaps and Future Agenda

Future research should prioritize stakeholder-informed questions about data access, interoperability, evidence quality, usability, governance, environmental outcomes, and real-world implementation. Priority-setting work in digital agriculture shows that research agendas should include the concerns of farmers, advisors, researchers, industry, and policy actors rather than being defined only by technology developers or model-performance benchmarks [31]. For precision agriculture data science, this implies that future bibliometric and thematic work should examine not only what methods are published, but also whose problems those methods address.

A second priority is to connect data-science workflows to sustainable agricultural-system design. Digital agriculture can support sustainability only when analytics are linked to system-level goals such as input efficiency, resilience, environmental protection, productivity, and adaptive management [32]. Future research should therefore evaluate whether models improve farm decisions under real environmental and economic constraints, rather than relying only on retrospective accuracy or controlled-case demonstrations.

A third priority is to strengthen Agriculture 4.0 implementation readiness by addressing infrastructure, interoperability, skills, governance, and practical integration. Agriculture 4.0 scholarship identifies trends, challenges, and opportunities across sensors, automation, data platforms, AI, and smart farming systems, but also highlights the importance of aligning technological opportunity with adoption conditions and operational realities [33]. Precision agriculture data science should therefore advance from isolated proof-of-concept studies toward interoperable, validated, explainable, and maintainable decision-support ecosystems.

The future agenda should be organized around four connected directions. Methodologically, the field needs stronger data-quality reporting, external validation, open benchmarks, uncertainty quantification, and reproducible workflows. Practically, it needs farmer-centered interfaces, advisory integration, cross-platform data exchange, and decision-support tools that communicate uncertainty and recommended action clearly. Socio-technically, it needs participatory design, responsible AI governance, and attention to unequal access across farm types. Sustainability-oriented research should evaluate environmental and agronomic outcomes directly, while stakeholder-informed research-priority processes can help ensure that future precision agriculture data science addresses real decision needs [31].

Conclusion

This bibliometric and thematic review has mapped precision agriculture data science as an evolving field shaped by machine learning, deep learning, remote sensing analytics, UAV data, IoT systems, big data, data fusion, decision support, sustainability applications, and socio-technical concerns. The review distinguishes bibliometric dataset logic from manuscript reference use and avoids unsupported counts, rankings, citation metrics, cluster sizes, or claims of research dominance. Its main contribution is an integrated interpretation of research evolution, knowledge structures, thematic clusters, emerging frontiers, and future priorities. The synthesis shows that precision agriculture data science has moved beyond data collection toward connected decision systems, but its future relevance depends on stronger validation, data quality, interoperability, explainability, reproducibility, farmer-centered design, responsible data governance, and measured sustainability outcomes.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 17 Jul 2024

Revised: 09 Sep 2024

Accepted: 02 Nov 2024

Published online: 25 May 2025

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

- 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>.
- 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>.
- 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 Wagening J Life Sci*. 2019;90-91:100315.
<https://doi.org/10.1016/j.njas.2019.100315>.
- 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>.
- 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>.
- 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>.
- Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr*. 2017;11(4):959-75.
<https://doi.org/10.1016/j.joi.2017.08.007>.
- Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *J Bus Res*. 2021;133:285-96.
<https://doi.org/10.1016/j.jbusres.2021.04.070>.
- Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. *Prof Inf*. 2020;29(1):e290103.
<https://doi.org/10.3145/epi.2020.ene.03>.
- Mukherjee D, Lim WM, Kumar S, Donthu N. Guidelines for advancing theory and practice through bibliometric research. *J Bus Res*. 2022;148:101-15.
<https://doi.org/10.1016/j.jbusres.2022.04.042>.

Bertoglio R, Corbo C, Renga FM, Matteucci M. The digital agricultural revolution: A bibliometric analysis literature review. *IEEE Access*. 2021;9:134762-82.
<https://doi.org/10.1109/ACCESS.2021.3115258>.

Sott MK, Nascimento LdS, Foguesatto CR, Furstenau LB, Faccin K, Zawislak PA, et al. A bibliometric network analysis of recent publications on digital agriculture to depict strategic themes and evolution structure. *Sensors (Basel)*. 2021;21(23):7889.
<https://doi.org/10.3390/s21237889>.

Coulibaly S, Kamsu-Foguem B, Kamissoko D, Traore D. Deep learning for precision agriculture: A bibliometric analysis. *Intell Syst Appl*. 2022;16:200102.
<https://doi.org/10.1016/j.iswa.2022.200102>.

Kushartadi T, Mulyono AE, Al Hamdi AH, Rizki MA, Sadat Faidar MA, Harsanto WD, et al. Theme mapping and bibliometric analysis of two decades of smart farming. *Information*. 2023;14(7):396.
<https://doi.org/10.3390/info14070396>.

Rejeb A, Rejeb K, Abdollahi A, Hassoun A. Precision agriculture: A bibliometric analysis and research agenda. *Smart Agric Technol*. 2024;9:100684.
<https://doi.org/10.1016/j.atech.2024.100684>.

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>.

Shafi U, Mumtaz R, García-Nieto J, Hassan SA, Zaidi SAR, Iqbal N. Precision agriculture techniques and practices: From considerations to applications. *Sensors (Basel)*. 2019;19(17):3796.
<https://doi.org/10.3390/s19173796>.

Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX. A review on the practice of big data analysis in agriculture. *Comput Electron Agric*. 2017;143:23-37.
<https://doi.org/10.1016/j.compag.2017.09.037>.

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>.

Khanal S, Fulton J, Shearer S. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Comput Electron Agric*. 2017;139:22-32.
<https://doi.org/10.1016/j.compag.2017.05.001>.

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>.

Benfenati D, Amalfitano D, Russo C, Tommasino C, Rinaldi AM. Data centric artificial intelligence for agrifood domain: A systematic mapping study. *Comput Electron Agric*. 2025;239:110847.
<https://doi.org/10.1016/j.compag.2025.110847>.

Adadi A, Berrada M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*. 2018;6:52138-60.
<https://doi.org/10.1109/ACCESS.2018.2870052>.

Linaza MT, Posada J, Bund J, Eisert P, Quartulli M, Döllner J, et al. Data-driven artificial intelligence applications for sustainable precision agriculture. *Agronomy*. 2021;11(6):1227.
<https://doi.org/10.3390/agronomy11061227>.

Tzounis A, Katsoulas N, Bartzanas T, Kittas C. Internet of Things in agriculture, recent advances and future challenges. *Biosyst Eng*. 2017;164:31-48.
<https://doi.org/10.1016/j.biosystemseng.2017.09.007>.

Eastwood C, Klerkx L, Ayre M, Dela Rue B. Managing socio-ethical challenges in the development of smart farming: From a fragmented to a comprehensive approach for responsible research and innovation. *J Agric Environ Ethics*. 2019;32(5-6):741-68.
<https://doi.org/10.1007/s10806-017-9704-5>.

Shepherd M, Turner JA, Small B, Wheeler D. Priorities for science to overcome hurdles thwarting the full promise of the 'digital agriculture' revolution. *J Sci Food Agric*. 2020;100(14):5083-92.
<https://doi.org/10.1002/jsfa.9346>.

Lajoie-O'Malley A, Bronson K, van der Burg S, Klerkx L. The future(s) of digital agriculture and sustainable food systems: An analysis of high-level policy documents. *Ecosyst Serv*. 2020;45:101183.
<https://doi.org/10.1016/j.ecoser.2020.101183>.

Fleming A, Jakku E, Lim-Camacho L, Taylor B, Thorburn P. Is big data for big farming or for everyone? Perceptions in the Australian grains industry. *Agron Sustain Dev*. 2018;38(3):24.
<https://doi.org/10.1007/s13593-018-0501-y>.

Ingram J, Maye D, Bailie C, Barnes A, Bear C, Bell M, et al. What are the priority research questions for digital agriculture? *Land Use Policy*. 2022;114:105962.
<https://doi.org/10.1016/j.landusepol.2021.105962>.

Basso B, Antle J. Digital agriculture to design sustainable agricultural systems. *Nat Sustain*. 2020;3(4):254-6.
<https://doi.org/10.1038/s41893-020-0510-0>.

Araújo SO, Peres RS, Barata J, Lidon F, Ramalho JC. Characterising the Agriculture 4.0 landscape—emerging trends, challenges and opportunities. *Agronomy*. 2021;11(4):667.
<https://doi.org/10.3390/agronomy11040667>.