JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Unveiling the Knowledge Structure and Emerging Research Trends of Google Earth Engine Applications in Remote Sensing: A Bibliometric Analysis of Scopus-Indexed Publications (2021–2026)

Supiyandi[1], Ramlah binti Mailok[2*]

[1]Fakultas Sains Komputasi dan Kecerdasan Digital, Information Technology, Universitas Pembangunan Panca Budi, Medan, Indonesia
[2]Faculty Of Computing and Meta-Technology, Information Systems and Management, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak - Malaysia
E-mail: [1]supiyandi@dosen.pancabudi.ac.id, [2*]mramlah@meta.upsi.edu.my
*E-mail Corresponding Author: mramlah@meta.upsi.edu.my
Received: August 8, 2026. | *Revision: August 11, 2026* | *Accepted: August 12, 2026*

Abstract

The convergence of the Random Forest (RF) ensemble classifier with the Google Earth Engine (GEE) planetary-scale computing platform has become one of the dominant operational paradigms in contemporary Earth observation. Despite the rapid accumulation of literature, no systematic bibliometric account of this specific intersection has been published. This study analyses 107 Scopus-indexed journal articles published between 2021 and 2026 that jointly apply RF and GEE to remote sensing problems. Using the bibliometrix R package and VOSviewer, the corpus was examined through descriptive indicators, performance analysis, and five science-mapping techniques: co-authorship, keyword co-occurrence, co-citation, bibliographic coupling, and thematic mapping. Results show a compound annual growth rate of 51.97% between 2021 and 2025, 1,685 total citations, an h-index of 24, and a highly concentrated publication landscape in which three journals Sustainability, International Journal of Digital Earth, and Ecological Informatics account for 59.8% of output. China dominates production with 48 affiliated documents, followed by the United States (20). Co-citation analysis identifies a compact intellectual core anchored by Gorelick et al. (2017), Breiman (2001), and Belgiu and Drăguţ (2016). Thematic analysis reveals a field that is methodologically mature but conceptually narrow: RF is applied across eight application domains, yet only 10.3% of documents address model transferability, 4.7% address uncertainty quantification, and 4.7% address reproducibility. Emerging fronts include explainable AI (SHAP), SAR–optical fusion, and deep-learning hybridisation. The study concludes by articulating seven research gaps and a corresponding agenda for the next phase of cloud-based Earth observation research.

Keywords: bibliometric analysis, Google Earth Engine, Random Forest, machine learning, remote sensing, science mapping, VOSviewer, bibliometrix

1. INTRODUCTION

Earth observation has undergone a structural transformation over the past decade. The opening of the Landsat archive in 2008, the launch of the Copernicus Sentinel constellation from 2014 onward, and the parallel emergence of cloud-based geospatial computing have collectively dissolved the two constraints that historically governed remote sensing practice: access to imagery and access to computational capacity [1], [2]. Where a continental-scale land cover analysis once demanded institutional infrastructure and months of preprocessing, the same task can now be specified in a browser-based script and executed against petabytes of analysis-ready data in minutes.

Google Earth Engine (GEE), introduced by Gorelick et al. in 2017, is the most consequential expression of this shift [1]. GEE combines a multi-petabyte catalogue of publicly available satellite imagery with a parallelised, server-side execution model and a JavaScript/Python application programming interface. Critically for the present study, GEE embeds a small library of supervised classifiers directly within its API, making machine learning available to users who possess domain expertise in Earth science but not necessarily in software engineering or high-performance computing. Tamiminia et al. [3] and Amani et al. [4] documented the platform's rapid uptake across environmental disciplines, and subsequent reviews have confirmed that GEE has moved from novelty to default infrastructure in large-area mapping. Within the classifier library that GEE exposes, Random Forest (RF) has attained a position of near-monopoly. Introduced by Breiman [5], RF is an ensemble of decision trees trained on bootstrap samples with randomised feature subsetting at each split. Its properties align almost exactly with the constraints of satellite image classification: it tolerates high-dimensional and collinear feature spaces, is robust to noisy training labels, requires minimal hyperparameter tuning, produces internal out-of-bag error estimates, and yields variable-importance rankings that are interpretable to domain scientists. Belgiu and Drăguţ [6] reviewed RF's remote sensing applications and established the empirical baseline against which subsequent studies position themselves; Rodriguez-Galiano et al. [7] provided the earlier accuracy benchmark that framed RF as a superior alternative to maximum-likelihood and single-tree classifiers.

The conjunction of these two elements an ensemble classifier that is computationally cheap and parameter-tolerant, and a cloud platform that removes data and compute barriers has produced a distinctive and highly productive research configuration. It is now routine for studies of wetland delineation, crop-type mapping, mangrove species discrimination, burned-area detection, above-ground biomass estimation, and urban heat island analysis to share a nearly identical

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

methodological skeleton: assemble a Sentinel-2 or Landsat time series in GEE, derive spectral indices and temporal composites, train an RF classifier on field or visually interpreted samples, and report overall accuracy and the Kappa coefficient.

This convergence is a substantial scientific achievement in that it has democratised large-area monitoring. It is also, potentially, a source of methodological inertia. Bibliometric analysis is well suited to interrogating this tension. As a quantitative method for mapping the structure and dynamics of a scientific field [8], [9], bibliometrics makes visible the properties that narrative reviews tend to obscure: the concentration of output across venues and countries, the degree of collaborative fragmentation, the size and composition of the shared intellectual core, and the trajectory of conceptual themes over time. Several bibliometric studies have addressed GEE broadly, and others have addressed machine learning in remote sensing broadly. To the authors' knowledge, however, no study has yet examined the specific RF–GEE intersection as a bounded research object. This omission matters, because it is precisely at this intersection that the field's methodological standardisation is most pronounced and its blind spots around transferability, uncertainty, and reproducibility-are most consequential. Accordingly, this study addresses the following research questions: RQ1-What are the volume, growth trajectory, and citation performance of RF–GEE remote sensing research between 2021 and 2026?. RQ2-How is scientific output distributed across sources, authors, institutions, and countries, and what does that distribution imply about the field's structure?. RQ3-What is the intellectual foundation of the field, as revealed by co-citation analysis, and how tightly is the current research front coupled to it?. RQ4-What conceptual themes structure the field, how have they evolved, and which topics constitute the emerging research front?. RQ5-What research gaps follow from the observed structure, and what agenda should guide subsequent work?

The contribution of this study is threefold. First, it provides the first quantitative structural account of RF–GEE research as a distinct field. Second, it operationalises seven science-mapping techniques on a single, carefully screened corpus, permitting cross-validation of findings across complementary analytical lenses. Third, it converts the observed structure into a specific, evidence-grounded research agenda rather than a generic call for further work. The remainder of the paper is organised as follows. Section 2 details the research design, data source, search strategy, screening protocol, preprocessing, analytical framework, and software. Section 3 presents and discusses results across fifteen analytical dimensions. Section 4 concludes.

2. RESEARCH METHODOLOGY

2.1 Materials and Methods

2.2.1 Research Design

This study adopts a quantitative, descriptive-exploratory bibliometric design. Bibliometrics applies statistical and network-analytic methods to bibliographic records in order to characterise the structure, dynamics, and intellectual organisation of a research domain [8]. Two complementary families of technique were employed, following the taxonomy proposed by Donthu et al. [9]. The first is performance analysis, which describes the productivity and impact of research constituents documents, authors, sources, institutions, and countries — using publication counts, citation counts, and derived indicators such as the h-index [10]. Performance analysis answers descriptive questions about who produces the research and how much influence it carries.

The second is science mapping, which reconstructs the relational structure of a field by treating bibliographic records as networks. Five mapping techniques were applied: co-authorship analysis (social structure), keyword co-occurrence analysis (conceptual structure), co-citation analysis (intellectual structure, retrospective), bibliographic coupling (intellectual structure, prospective), and thematic mapping with thematic evolution (conceptual dynamics). Science mapping answers relational questions about how the field is organised and where it is moving.

The two families are complementary rather than redundant. Performance analysis identifies the most prominent constituents; science mapping explains how those constituents relate. Reporting them together mitigates the risk, common in single-method bibliometric studies, of mistaking prominence for centrality. The overall design follows the PRISMA 2020 reporting framework for the identification, screening, and inclusion stages [11], adapted for bibliometric rather than effect-size synthesis. Reproducibility was prioritised throughout: the search string, screening decisions, preprocessing operations, and analytical parameters are reported in full so that the corpus can be reconstructed independently.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

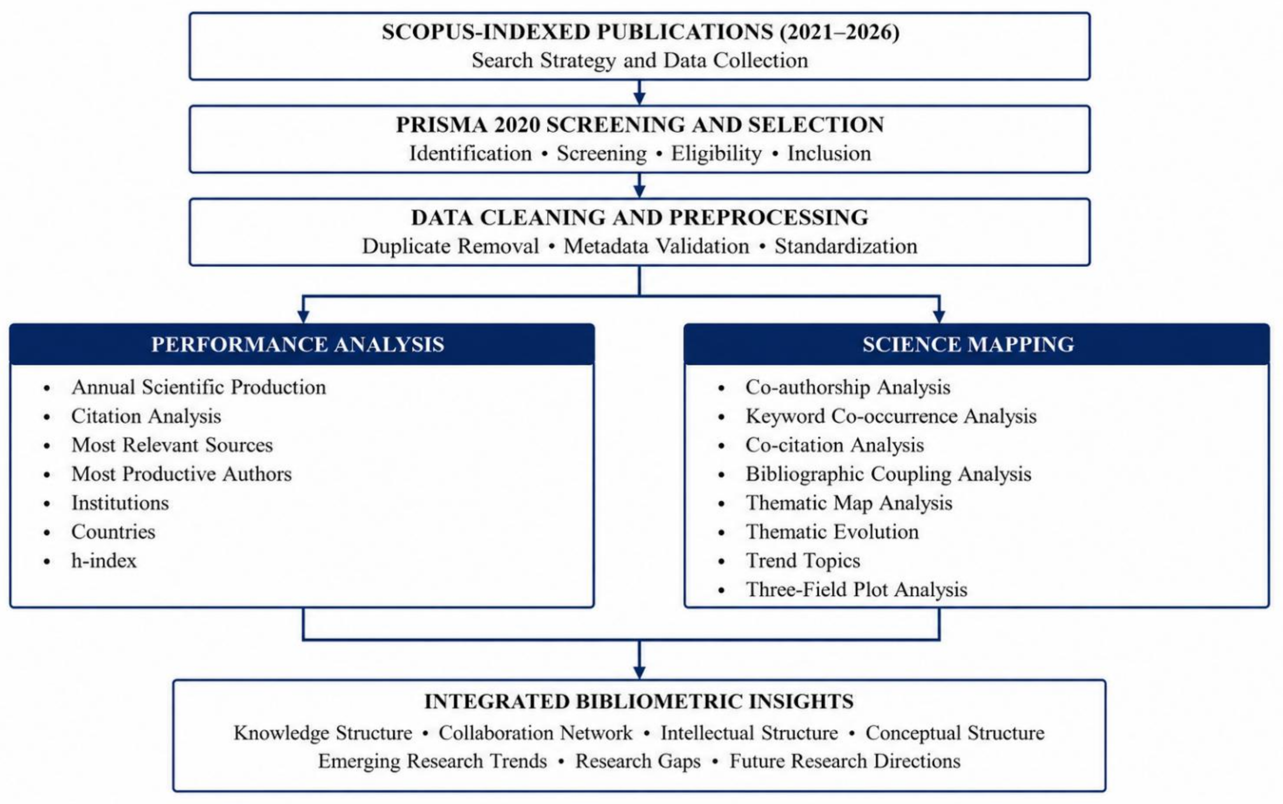


Figure 1. Conceptual framework of the bibliometric research design, showing the relationship between performance analysis and science-mapping components.

2.2.2 Data Source

Scopus (Elsevier) was selected as the sole bibliographic data source. Three considerations motivated this choice:

1. Coverage Scopus indexes a substantially larger set of peer-reviewed journals than Web of Science, with particularly strong representation of the applied remote sensing, environmental informatics, and geoinformation venues in which RF–GEE research predominantly appears Including Sustainability, Ecological Informatics, International Journal of Digital Earth, Big Earth Data, and Remote Sensing. A Web of Science–only corpus would have systematically under-represented output from Asian and Global South institutions, which Section 3.6 shows to be central to this field.
2. Metadata completeness and structure. Scopus exports supply complete, delimiter-separated cited-reference lists, author identifiers, affiliation strings, author keywords, and index keywords. Cited references are indispensable for co-citation and bibliographic coupling analysis; author and index keywords are indispensable for conceptual structure analysis. Google Scholar, though broader in coverage, does not provide structured, machine-readable metadata of this kind and includes non–peer-reviewed material without reliable filtering.
3. Single-database consistency. Merging Scopus and Web of Science records introduces duplicate-resolution artefacts and inconsistent reference-string formatting that propagate into co-citation networks as spurious node splitting. Because the analytical emphasis of this study is relational, internal consistency was prioritised over maximum recall.

The limitations of this choice exclusion of non-indexed regional journals, conference proceedings, book chapters, and grey literature are acknowledged in Section 4.

All records were retrieved on 6 August 2026. Because Scopus is updated continuously, this date is reported to allow exact temporal replication.

2.2.3 Search Strategy

The search string was constructed around two mandatory conceptual blocks joined by the Boolean AND operator, with terms within each block joined by OR:

1. Block A — Platform: Google Earth Engine and its common abbreviation.
2. Block B — Method: Random Forest and its common abbreviation, situated within a machine learning / remote sensing context.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

The query was executed against the TITLE-ABS-KEY field, which searches document titles, abstracts, and both author-supplied and indexer-assigned keywords. Field-restricted searching was preferred over full-text searching to maximise topical precision: a document that mentions RF or GEE only in a discussion paragraph is unlikely to be a genuine contribution to this literature. The following search terms were used in the first stage:

TITLE-ABS-KEY(
("Google Earth Engine")
AND
("Random Forest")
)

Using the advanced search function returned 2,621 documents, and then using the following function:

TITLE-ABS-KEY (("Google Earth Engine") AND ("Random Forest")) AND (LIMIT-TO (EXACTKEYWORD , "Google Earth Engine") OR LIMIT-TO (EXACTKEYWORD , "Remote Sensing") OR LIMIT-TO (EXACTKEYWORD , "Machine Learning") OR LIMIT-TO (EXACTKEYWORD , "Random Forest") OR LIMIT-TO (EXACTKEYWORD , "Sentinel-2")) AND (LIMIT-TO (SUBJAREA , "COMP")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (OA , "all")) AND PUBYEAR > 2020 AND PUBYEAR < 2027.

and 133 documents were identified, which were then filtered using the R-Code and RStudio programmes with the function:

scopus %>%
filter(DE!=' ', RP!=' ', ID!=' ') (1)

and can be seen in the R-Script in RStudio in the following image

```
scopus %>%
  filter(DE!='', RP!='', ID!='')
```

Where:
1. scopus = the database used
2. filter = filtering metadata using Biblioshiny
3. DE = metadata code with ‘Keywords’ description
4. RP = metadata code with ‘Corresponding Author’ description
5. ID = metadata code with ‘Keywords Plus’ description from.

and after filtering the data using an R-Script function in RStudio, a total of 107 documents were obtained.

The temporal window PUBYEAR > 2020 restricts the corpus to 2021–2026. This delimitation is deliberate rather than arbitrary. By 2021, GEE had passed its early-adoption phase and RF had been consolidated as the platform's default classifier; the resulting corpus therefore describes a mature, operational literature rather than a pioneering one. A six-year window is also short enough to permit meaningful two-period thematic evolution analysis while retaining sufficient documents in each period.

Table 1. Search string components, Boolean operators, and rationale for each query block

Query Block	Search String / Filter	Function	Expected Outcome	Rationale
Core Search Terms	TITLE-ABS-KEY ("Google Earth Engine") AND ("Random Forest")	Retrieves publications that simultaneously discuss Google Earth Engine and the Random Forest algorithm within the title, abstract, or author keywords.	A focused dataset containing studies that integrate Google Earth Engine with Random Forest, while excluding publications addressing only one of the two concepts.	Ensures the bibliometric analysis is centered on the intersection of cloud-based geospatial computing and machine learning applications.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Query Block	Search String / Filter	Function	Expected Outcome	Rationale
Keyword Restriction	LIMIT-TO (EXACTKEYWORD, "Google Earth Engine") OR LIMIT-TO (EXACTKEYWORD, "Remote Sensing") OR LIMIT-TO (EXACTKEYWORD, "Machine Learning") OR LIMIT-TO (EXACTKEYWORD, "Random Forest") OR LIMIT-TO (EXACTKEYWORD, "Sentinel-2")	Restricts the dataset to publications indexed with predefined relevant keywords.	Increased precision by eliminating unrelated records and retaining publications that explicitly belong to the targeted research domain.	Exact keywords improve thematic consistency and reduce retrieval noise in bibliometric datasets.
Subject Area Filter	LIMIT-TO (SUBJAREA, "COMP")	Restricts the search to the Computer Science subject area.	A dataset emphasizing computational techniques, cloud computing, artificial intelligence, and geospatial data processing.	The study focuses on computational methodologies underlying Google Earth Engine and Random Forest implementations.
Document Type	LIMIT-TO (DOCTYPE, "ar")	Includes only original research articles.	A homogeneous collection of peer-reviewed empirical studies suitable for bibliometric performance and science mapping analyses.	Research articles provide comprehensive metadata, references, and citation information compared to reviews or conference papers.
Language Filter	LIMIT-TO (LANGUAGE, "English")	Restricts the search to English-language publications.	Improved consistency and international comparability across retrieved documents.	English is the dominant language of scholarly communication and represents the majority of publications indexed in Scopus.
Source Type	LIMIT-TO (SRCTYPE, "j")	Includes only journal publications.	High-quality publications with standardized peer-review processes and reliable citation metadata.	Journal articles are considered the primary source for bibliometric evaluations because they generally exhibit higher scientific rigor than other publication types.
Publication Stage	LIMIT-TO (PUBSTAGE, "final")	Includes only finalized publications.	Stable bibliographic records with complete metadata and finalized citation information.	Excludes articles in press, which may undergo metadata or citation updates after publication.
Open Access Filter	LIMIT-TO (OA, "all")	Includes all categories of Open Access publications.	Comprehensive retrieval of openly accessible research regardless of Open Access model.	Prevents unnecessary exclusion of relevant publications while maintaining complete coverage of Open Access literature.
Publication Year	PUBYEAR > 2020 AND PUBYEAR < 2027	Restricts publications to the period 2021–2026.	A contemporary dataset reflecting recent scientific developments and emerging research trends.	The selected period captures the rapid expansion of Google Earth Engine and Random Forest applications following

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Query Block	Search String / Filter	Function	Expected Outcome	Rationale
				advances in cloud-based remote sensing platforms.
Boolean Operators	AND, OR	Combines search concepts and filtering criteria.	Balanced retrieval with high precision (AND) and comprehensive keyword coverage (OR).	The combination of Boolean operators ensures both specificity and completeness of the search strategy while minimizing irrelevant records.

2.2.4 Screening and Eligibility Criteria

The Initial data: 2,621 documents, the initial query returned 133 records. Screening proceeded in two stages and yielded a final corpus of 107 documents, an inclusion rate of 80.5%.

Stage 1. Automated filtering (database level). The following criteria were applied within Scopus:

Table 2. Criteria were applied within Scopus

Criterion	Inclusion	Exclusion
Document type	Journal article (ar)	Conference paper, review, book chapter, editorial, note, erratum
Language	English	All other languages
Publication stage	Final	Article-in-press
Access	Open access	—
Publication year	2021–2026	≤ 2020

All 133 retrieved records satisfied these criteria, since they were applied as query-level limiters. Screening and eligibility criteria. To ensure the relevance, consistency, and reproducibility of the bibliometric corpus, predefined inclusion and exclusion criteria were applied within the Scopus database. As summarized in Table 2, the search was restricted to journal articles, while conference papers, review articles, book chapters, editorials, notes, and errata were excluded. Only publications written in English were retained to ensure consistency in the interpretation of titles, abstracts, keywords, and other bibliographic metadata. The publication-stage criterion was limited to final publications, thereby excluding articles in press whose bibliographic information may still be subject to change.

The analysis was further restricted to Open Access publications, consistent with the Scopus query specification using LIMIT-TO (OA, "all"). The publication period was restricted to 2021–2026, corresponding to the predefined temporal scope of the study; publications published in 2020 or earlier were excluded. These criteria were applied systematically at the database-search stage to reduce retrieval noise and establish a homogeneous corpus before subsequent manual screening and data preprocessing. After the database-level filtering, the remaining records were subjected to manual eligibility assessment to identify publications that were suitable for the final bibliometric analysis. This sequential procedure helped ensure that the final corpus was both thematically relevant and methodologically consistent with the objectives of the study.The eligibility process was designed to balance retrieval precision and corpus completeness. Rather than relying solely on publication type or keyword restrictions, the screening procedure combined bibliographic filters with manual assessment of the retrieved records. This approach reduced the likelihood of retaining records that technically satisfied the database filters but were not sufficiently aligned with the research scope. The final eligible corpus was subsequently subjected to metadata validation, keyword harmonisation, and reference standardisation before bibliometric analysis. The complete selection pathway, including the number of records identified, excluded during screening, assessed for eligibility, and ultimately included in the analysis, is presented in Figure 2.

Stage 2. Manual eligibility screening (title, abstract, and keyword review). Each of the 133 records was assessed against three substantive criteria:

1. Joint application. The study must apply Random Forest within or in conjunction with the Google Earth Engine environment. Records that merely cited GEE as related work, or that used RF entirely outside a cloud-computing context, were excluded.
2. Remote sensing substance. The study must analyse satellite, airborne, or spaceborne Earth observation data. Purely simulation-based, purely tabular, or purely software-engineering contributions were excluded.
3. Analytical contribution. The study must report an original empirical analysis. Data-descriptor papers, software-impact notes, and dataset announcements were excluded, as their citation and reference behaviour differs systematically from that of research articles and would distort coupling and co-citation structure.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Twenty-six records (19.5%) were excluded at Stage 2. Screening was conducted on titles, abstracts, author keywords, and index keywords; where eligibility remained ambiguous, the full text was consulted. Exclusion decisions and their reasons were recorded in a screening log retained for audit.

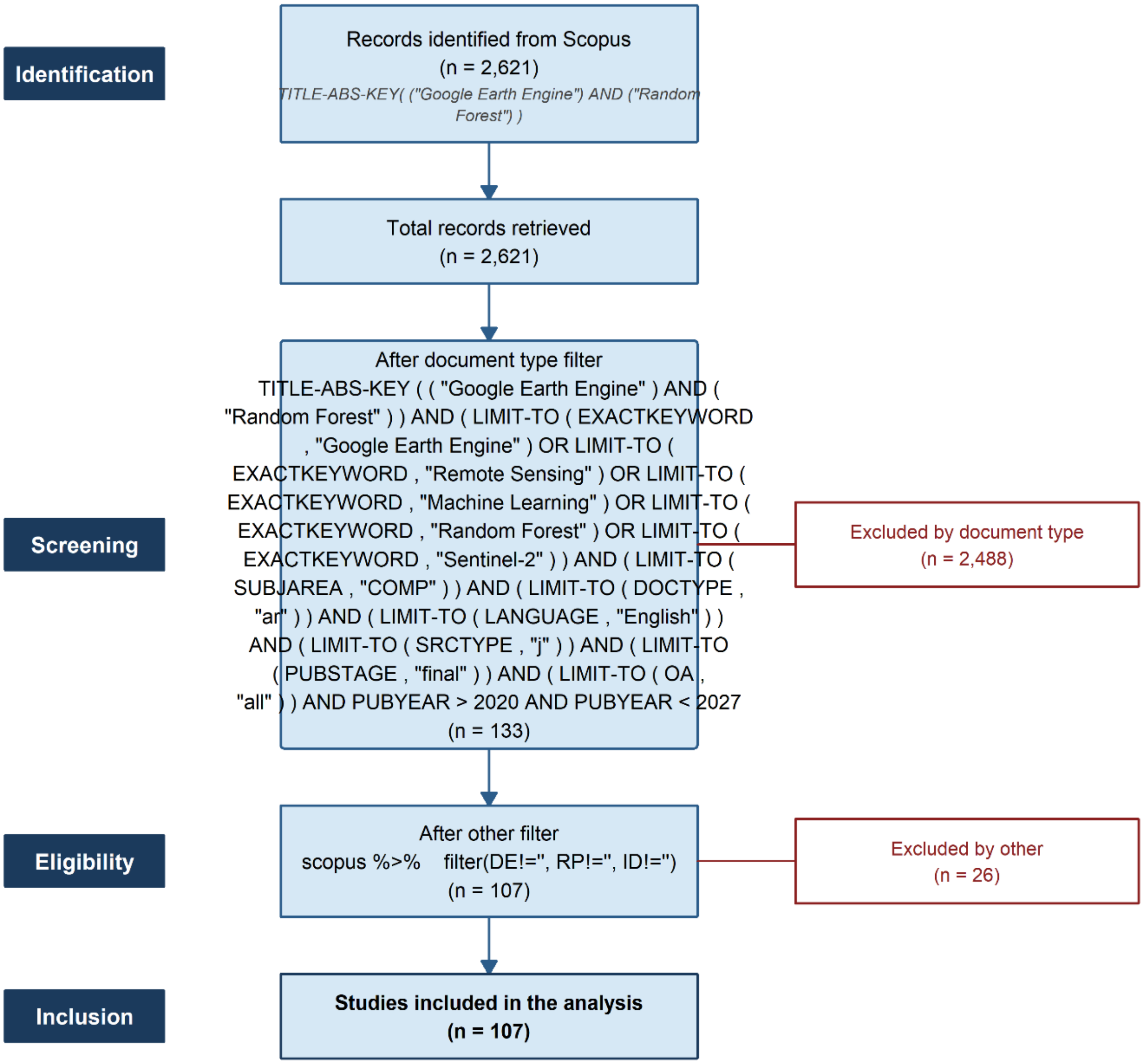


Figure 2. PRISMA 2020 flow diagram of the identification, screening, eligibility, and inclusion process (n = 133 identified; n = 26 excluded; n = 107 included).

Table 3. Inclusion and exclusion criteria with the number of records excluded under each.

Stage	Inclusion Criteria	Exclusion Criteria	Excluded (n)	Remaining (n)
Identification	Records retrieved from the Scopus database using the predefined search query.	None.	0	2,621
Screening	English-language, journal articles, Computer Science subject area, final publication stage, published between 2021 and 2026, indexed with the selected keywords.	Publications not meeting one or more search filters (e.g., document type, language, subject area, source type, publication stage, publication year, or keyword restrictions).	2,488	133

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Stage	Inclusion Criteria	Exclusion Criteria	Excluded (n)	Remaining (n)
Eligibility	Publications with complete bibliographic metadata suitable for bibliometric analysis.	Duplicate records, incomplete metadata, missing author keywords or references, and records unsuitable for Bibliometrix preprocessing.	26	107
Inclusion	Final Scopus-indexed publications included in the bibliometric analysis.	None.	0	107

2.2.5 Screening and Eligibility Criteria

Raw bibliographic exports contain systematic inconsistencies that, if left unaddressed, inflate node counts and fragment networks. Six preprocessing operations were performed on the 107-record corpus.

1. Field harmonisation. The Scopus CSV export was converted into a bibliometrix-compatible data frame in which each column is mapped to a standard bibliographic tag (AU authors, TI title, SO source, PY publication year, TC times cited, DE author keywords, ID index keywords, CR cited references, C1 affiliations, AB abstract, RP corresponding author).
2. Case normalisation. All text fields were converted to uppercase to prevent case-driven duplication (e.g. Google Earth Engine / GOOGLE EARTH ENGINE).
3. Author name disambiguation. Author strings were normalised to the SURNAME INITIAL convention. Scopus Author IDs were retained to support verification. It must be stated explicitly that homonym conflation remains a residual limitation for high-frequency East Asian surnames; the implications are discussed in Section 3.4.
4. Keyword normalisation. Author keywords were consolidated using a controlled mapping that merged orthographic and abbreviation variants of the same concept. The principal merges were:

Table 4. Standardization of Author Keywords by Merging Variant Forms into Normalized Terms

Variant forms	Normalised term
GOOGLE EARTH ENGINE (GEE), GEE	GOOGLE EARTH ENGINE
RF	RANDOM FOREST
MACHINE LEARNING ALGORITHMS	MACHINE LEARNING
SVM	SUPPORT VECTOR MACHINE
LAND USE/LAND COVER, LAND USE AND LAND COVER, LAND COVER CLASSIFICATION	LULC
REMOTE-SENSING	REMOTE SENSING

This procedure reduced the author-keyword vocabulary from 399 to 390 unique terms and, more importantly, consolidated the frequency of GOOGLE EARTH ENGINE from 51 to 66 occurrences a 29.4% correction that materially changes the co-occurrence network topology. The strongest co-occurrence link, GOOGLE EARTH ENGINE–RANDOM FOREST, likewise increased from 12 to 18 after consolidation. For index keywords, ten non-substantive indexer terms ARTICLE, HUMAN, HUMANS, PRIORITY JOURNAL, CONTROLLED STUDY, ENGINES, GOOGLE EARTHS, LEARNING SYSTEMS, LEARNING ALGORITHMS, and SATELLITES were removed as stop terms. These arise from Scopus's automated thesaurus assignment and carry no thematic information; retaining them reduces the index-keyword vocabulary from 1,005 to 996 terms and prevents spurious hub formation in the co-occurrence network.

5. Cited-reference normalisation. Scopus reference strings vary in journal-title abbreviation, volume-field presence, and punctuation, causing the same cited work to appear as multiple distinct nodes. A rule-based normalisation extracted the first author's surname and initial, the publication year, and the leading title fragment, then reconstructed a canonical key of the form SURNAME I. (YEAR) TITLE-FRAGMENT. The effect was substantial: Gorelick et al. (2017) appeared as three separate raw strings with 17, 15, and 11 occurrences and was consolidated into a single node with 49 occurrences; Breiman (2001) was consolidated from three strings (15, 12, 10) into one node with 47. Overall, 6,733 raw reference strings were reduced to 6,156 canonical references.
6. Affiliation and country extraction. Country names were parsed from the terminal segment of each semicolon-delimited affiliation string in the C1 field, with numeric prefixes and terminal punctuation stripped. Institution names were extracted using the bibliometrix university-normalisation routine. Where a document listed multiple affiliations from the same country, the country was counted once per document.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Following preprocessing, the corpus comprised 107 documents, 527 unique authors, 22 unique sources, 390 normalised author keywords, and 6,156 canonical cited references, with no missing values in the AU, TI, PY, SO, AB, C1, DE, or CR fields.

Table 3. Data Cleaning Operations, Procedures Applied, and Their Quantitative Effect on the Corpus

Data Cleaning Operation	Procedure Applied	Quantitative Effect	Final Status
Duplicate record removal	Duplicate bibliographic records were identified using DOI, title, and bibliographic metadata, and removed prior to analysis.	Duplicate records removed during preprocessing	Clean dataset retained
Metadata completeness assessment	Essential bibliographic fields (AU, TI, AB, PY, SO, C1, DE, CR, RP, DI, DT, LA, TC) were examined for missing values.	0 missing values (0%) across all essential metadata fields	Complete
Author keyword standardization	Synonyms, abbreviations, spelling variants, and lexical inconsistencies were merged into standardized keywords (e.g., *GEE → Google Earth Engine*; *RF → Random Forest*).	390 standardized author keywords	Standardized
Author and affiliation validation	Author names, corresponding authors, and institutional affiliations were checked for completeness and consistency.	0 missing values (0%)	Complete
Reference consistency check	Cited references were validated to ensure complete citation metadata for citation-based analyses.	6,156 canonical cited references retained	Complete
DOI verification	DOI fields were checked for completeness and formatting consistency.	0 missing values (0%)	Complete
Language verification	Publication language was verified to ensure all documents were written in English.	107 English-language publications (100%)	Complete
Publication metadata validation	Publication year, journal source, document type, and publication stage were validated against the predefined inclusion criteria.	107 records satisfied all criteria	Complete
Science category assessment	The Web of Science Categories (WC) field was unavailable because the dataset originated from Scopus rather than Web of Science.	107 missing values (100%)	Not applicable to Scopus

Table 3. Data cleaning operations, preprocessing procedures, and their quantitative effects on the final bibliometric corpus. Metadata quality assessment confirmed that all essential Scopus bibliographic fields required for Bibliometrix analyses were complete, while the Web of Science Categories (WC) field was absent because it is not included in Scopus metadata. This missing field did not affect the performance analysis or science mapping conducted in this study.

2.2.6 Bibliometric Analysis Framework

Seven analytical procedures were applied. Each is defined below together with its parameterisation.

1. Descriptive performance indicators. Annual scientific production, average citations per document, average citations per year, the h-index [10] and g-index of the corpus, the number of uncited documents, author-per-document ratio, and the proportion of internationally co-authored documents.
2. Productivity laws. Bradford's law of scattering [12] was applied to partition sources into three zones of approximately equal productivity. Lotka's law [13] was applied to the author productivity distribution, and Price's law was used to define the elite core as the square root of the total author population.
3. Co-authorship analysis. An undirected weighted graph was constructed in which nodes are authors and an edge is created for each co-authored document, with edge weight equal to the number of shared documents. Network order, size, density, component structure, largest-component share, and mean degree were computed.
4. Keyword co-occurrence analysis. An undirected weighted graph was constructed from normalised author keywords, with an edge created for each pair of keywords appearing in the same document. Analysis was performed at three minimum-occurrence thresholds (≥ 2, ≥ 3, ≥ 5) to test the robustness of the resulting structure.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

5. Co-citation analysis. Following Small [14], two references were linked when co-cited by a document in the corpus. This technique maps the intellectual foundation the retrospective knowledge base on which the field draws. Canonical reference keys from Section 2.5 (5) served as nodes.
6. Bibliographic coupling. Following Kessler [15], two corpus documents were linked in proportion to the number of references they share. Whereas co-citation maps the past, coupling maps the present research front: documents that share references are engaged with the same problem, irrespective of whether they cite each other. Total coupling strength per document was computed as the sum of shared-reference counts across all its links.
7. Thematic mapping and evolution. Following Cobo et al. [16] and the strategic-diagram logic of Callon et al. [17], keyword clusters were positioned on two axes. Centrality measures a cluster's external linkage strength and indexes its importance to the field; density measures internal linkage strength and indexes its degree of internal development. The resulting quadrants distinguish motor themes (high centrality, high density), basic and transversal themes (high centrality, low density), niche themes (low centrality, high density), and emerging or declining themes (low centrality, low density). Thematic evolution was assessed by partitioning the corpus into two periods 2021-2023 ($n = 37$) and 2024-2026 ($n = 70$) and comparing cluster composition across the cut point. Trend topics were derived from the median publication year of each keyword with frequency ≥ 3, with the first and third quartile years defining the term's active interval.
8. Three-field plot (Sankey diagram). A tripartite flow diagram linking countries (left), author keywords (centre), and publication sources (right) was constructed to visualise how geographic communities, conceptual foci, and publication venues align.

2.2.7 Software and Visualization Tools

Four software environments were used.

1. R (version 4.6.1) with 'biblioshiny' new in version 5.0 [18] served as the primary analytical engine for corpus conversion, descriptive indicators, productivity laws, thematic mapping, thematic evolution, trend topics, and the three-field plot. bibliometrix was selected because it implements the Callon centrality–density formulation and the Cobo thematic-evolution procedure natively, and because its biblioshiny interface supports interactive parameter exploration.
2. VOSviewer 1.6.20 [19] was used for network visualisation and clustering of the co-authorship, co-occurrence, co-citation, and bibliographic coupling networks. VOSviewer's association-strength normalisation and its unified mapping-and-clustering technique produce layouts in which spatial distance is directly interpretable as relatedness a property that force-directed layouts do not guarantee.
3. Python 3.11 with pandas, numpy, and collections was used for reference-string normalisation, keyword harmonisation, country parsing, and independent recomputation of all reported statistics. Cross-validating bibliometrix output against an independent Python implementation guards against silent errors in either pipeline.
4. Microsoft Excel was used for tabular formatting and for the preparation of bar and line figures.

Visualisation parameters are reported with each figure in Section 3. Where a minimum-occurrence threshold was applied, both the threshold and the resulting node count are stated, so that readers can reproduce the visualisation exactly.

Table 4. Software Tools, Versions, Analytical Functions Performed, and Outputs Generated

Software	Version	Primary Role	Analytical Functions	Outputs Generated
R (bibliometrix)	4.6.1 / 4.x	Performance analysis	Descriptive bibliometrics, productivity analysis, Bradford's Law, Lotka's Law, thematic mapping, thematic evolution, trend topics, three-field plot	Bibliometric indicators, thematic maps, thematic evolution, productivity analyses
Biblioshiny	4.x	Interactive bibliometric exploration	Parameter tuning, visualization, and exploratory analysis	Interactive dashboards and visual summaries
VOSviewer	1.6.20	Science mapping	Co-authorship, co-occurrence, co-citation, bibliographic coupling, clustering, overlay visualization	Network maps, density maps, cluster visualizations
Python	3.11	Data preprocessing and validation	Metadata cleaning, keyword harmonization, reference normalization, country parsing, independent verification	Standardized dataset and validated bibliometric statistics

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Software	Version	Primary Role	Analytical Functions	Outputs Generated
Microsoft Excel	Microsoft 365	Data presentation	Table formatting and descriptive chart preparation	Publication-ready tables and descriptive figures

Table 4. Software tools, software versions, analytical functions performed, and outputs generated throughout the bibliometric workflow. R with the bibliometrix package served as the primary platform for performance analysis and thematic analyses, VOSviewer was employed for science mapping and network visualization, Python supported metadata preprocessing and independent result validation, and Microsoft Excel was used for tabular formatting and presentation of descriptive figures.

2.2.8 Research Workflow

The study proceeded through five sequential phases.

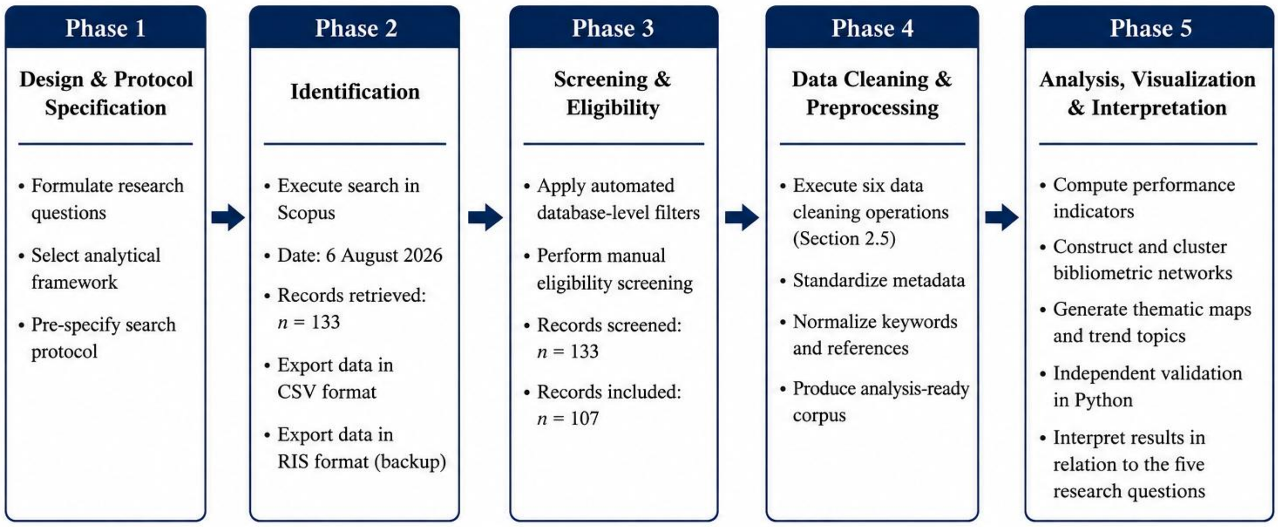


Figure 3. Research workflow of the bibliometric study.

Phase 1: Design and protocol specification. Research questions were formulated, the analytical framework was selected, and the search protocol was pre-specified.

Phase 2: Identification. The search string was executed in Scopus on 6 August 2026, returning 133 records. Records were exported in CSV format with all available fields enabled, and in RIS format as a redundant backup.

Phase 3: Screening and eligibility. Automated database-level filters were applied at query time; manual eligibility screening was then applied to all 133 records, yielding 107 included documents.

Phase 4: Preprocessing. The six cleaning operations described in Section 2.5 were executed, producing the analysis-ready corpus.

Phase 5: Analysis, visualisation, and interpretation. Performance indicators were computed, science-mapping networks were constructed and clustered, thematic maps were generated, and results were interpreted against the five research questions. All numerical results reported in Section 3 were recomputed independently in Python and reconciled against bibliometrix output before inclusion.

3. RESULTS AND DISCUSSION

3.1 Descriptive Bibliometric Overview

Table 5 summarises the principal characteristics of the corpus. The 107 documents were published between 2021 and 2026 across 22 distinct sources, were authored by 527 unique researchers, and had accumulated 1,685 citations at the time of extraction. Several features of the corpus warrant comment. The corpus is homogeneous by construction and by nature. All 107 documents are English-language, final-stage, open-access journal articles. Complete open-access coverage is unusual in bibliometric corpora and reflects both the query filter and the publishing economics of the field: the dominant

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

venues (Sustainability, Applied Sciences, Sensors, Journal of Imaging) are gold open-access MDPI titles, while the remainder (International Journal of Digital Earth, Ecological Informatics, Big Earth Data, IEEE Access) operate hybrid or gold models. The practical implication is that this literature is fully and freely accessible a condition that aligns with, and probably accelerates, the methodological convergence documented below.

Collaboration is the norm. With 5.37 authors per document and only 4 single-authored papers (3.7%), RF–GEE research is overwhelmingly team-based. This reflects the composite skill set the work requires: remote sensing domain knowledge, field data collection, and cloud-platform programming rarely coexist in one individual. Citation impact is substantial but unevenly distributed. Mean citations per document (15.75) and the corpus h-index (24) are strong for a six-year window. However, 20 documents (18.7%) remain uncited, and Section 3.2 shows this concentration to be an artefact of the citation-accrual lag affecting the 2025–2026 cohort rather than evidence of low-quality output.

The reference base is broad but the intellectual core is narrow. The corpus cites 6,156 distinct canonical references an average of 66.4 unique references per document yet only 40 references (0.65%) are co-cited five or more times, and only 8 are co-cited ten or more times. This disparity is the single most diagnostic finding of the descriptive analysis and is developed in Section 3.9.

Table 5. Main information about the corpus.

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2021:2026
Sources (Journals, Books, etc)	22
Documents	107
Annual Growth Rate %	25.93
Document Average Age	1.94
Average citations per doc	15.75
References	6732
DOCUMENT CONTENTS	
Keywords Plus (ID)	1005
Author's Keywords (DE)	399
AUTHORS	
Authors	527
Authors of single-authored docs	3
AUTHORS COLLABORATION	
Single-authored docs	4
Co-Authors per Doc	5.37
International co-authorships %	44.86
DOCUMENT TYPES	
article	107

Table 5. Main information about the corpus. Suggested rows: timespan (2021–2026); sources (22); documents (107); annual growth rate (51.97%); document average age; average citations per document (15.75); total citations (1,685); references, canonical (6,156); author keywords, normalised (390); index keywords (1,005 raw; 996 after stop-term removal); authors (527); authors of single-authored documents (4); single-authored documents (4); co-authors per document (5.37); international co-authorship (44.9%); document type (article, 100%); language (English, 100%); open access (100%); corpus h-index (24); corpus g-index (37); uncited documents (20; 18.7%).

3.2 Annual Scientific Production

Annual output rose from 6 documents in 2021 to 32 in 2025, then registered 19 documents for the partial year 2026 (records indexed through 6 August). The compound annual growth rate over the complete 2021–2025 interval is 51.97%, indicating a field in a phase of sustained exponential expansion.

The trajectory is not smooth. Output more than doubled between 2021 and 2022 (6 → 15), plateaued through 2023 and 2024 (16, 19), then rose 68.4% in 2025 (19 → 32). Two mechanisms plausibly explain the 2025 inflection. First, the Sentinel-2 archive reached a depth at which decade-scale dense time series became analytically routine, opening

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

a class of longitudinal studies that were previously infeasible. Second, the expansion of MDPI and Taylor & Francis open-access capacity in environmental informatics increased venue availability precisely as submission volume grew. The 19 documents already indexed for 2026 by early August are consistent with a full-year total in the range of 30–35, suggesting that output has reached a plateau rather than continuing to accelerate.

Citation dynamics move inversely to production, as expected. Mean citations per document fall monotonically from 58.83 (2021) through 22.33 (2022), 27.38 (2023), 20.42 (2024), 5.00 (2025), to 0.58 (2026). The 2023 uptick relative to 2022 is attributable to two exceptionally cited documents in that cohort. This pattern confirms that the 18.7% uncited share identified in Section 3.1 is a maturation artefact: 17 of the 20 uncited documents belong to the 2025–2026 cohorts. Interpretations of impact in this field must therefore be normalised by publication year; raw citation counts systematically privilege the 2021–2023 cohort.

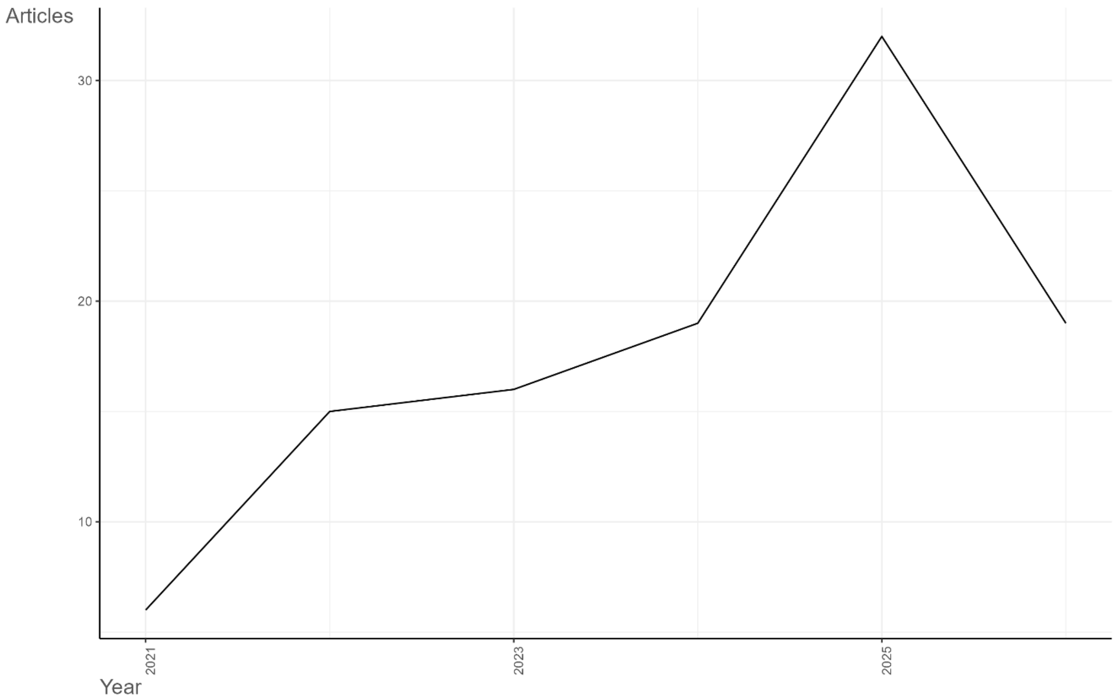


Figure 4. Annual scientific production and mean citations per document, 2021–2026.

Table 6. Annual scientific production

Year	Articles
2021	6
2022	15
2023	16
2024	19
2025	32
2026	19

3.3 Most Relevant Sources

The 107 documents are distributed across only 22 sources, and that distribution is severely skewed. Sustainability (Switzerland) published 27 documents (25.2%), the International Journal of Digital Earth 21 (19.6%), and Ecological Informatics 16 (15.0%). These three titles alone account for 64 documents 59.8% of the corpus. Adding Sensors (8) and Applied Sciences (6) raises the top-five share to 72.9%. Bradford's law partitions the sources into three zones of approximately equal output: Zone 1 comprises a single journal (Sustainability, 27 documents); Zone 2 comprises 2 journals (37 documents); Zone 3 comprises the remaining 19 journals (43 documents). A Zone 1 of one source is an extreme concentration even by Bradford standards and identifies Sustainability as the field's undisputed core venue. Source-level impact does not track source-level volume uniformly. Sustainability leads on both dimensions (h = 11; 512 citations), and the International Journal of Digital Earth follows closely (h = 10; 302 citations). IEEE Access, however, published only 4 documents yet accumulated 162 citations 40.5 per document, the highest per-document impact in the

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

corpus, driven by two heavily cited methodological comparisons. Conversely, Applied Sciences published 6 documents for 17 total citations, though its 2024 entry into the field means its cohort has had limited exposure time.

Two implications follow. Methodologically, the concentration means that a bibliometric corpus in this field is highly sensitive to venue indexing decisions; the exclusion of a single journal would materially alter the results. Substantively, the dominance of Sustainability a broad-scope, sustainability-framed multidisciplinary journal rather than a specialist remote sensing title indicates that RF–GEE research is increasingly positioned as applied environmental science addressed to a policy-adjacent audience, rather than as methodological remote sensing addressed to specialists. The comparatively modest presence of ISPRS Journal of Photogrammetry and Remote Sensing (5 documents), the field's leading methodological venue, reinforces this reading: genuinely novel algorithmic contributions are relatively rare within this corpus.

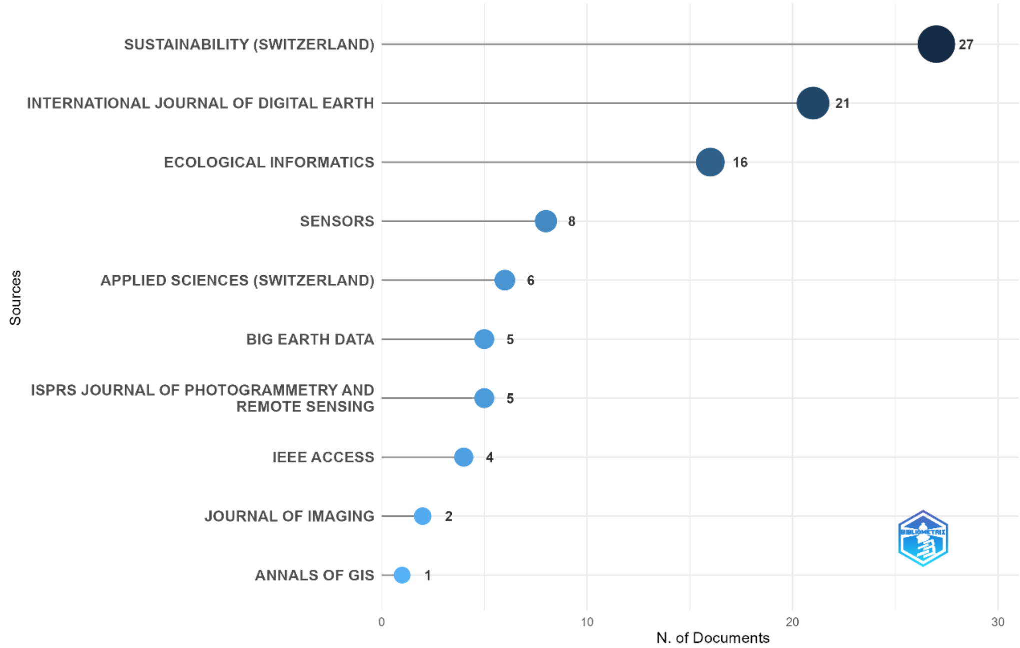


Figure 5. Most relevant sources by number of documents.

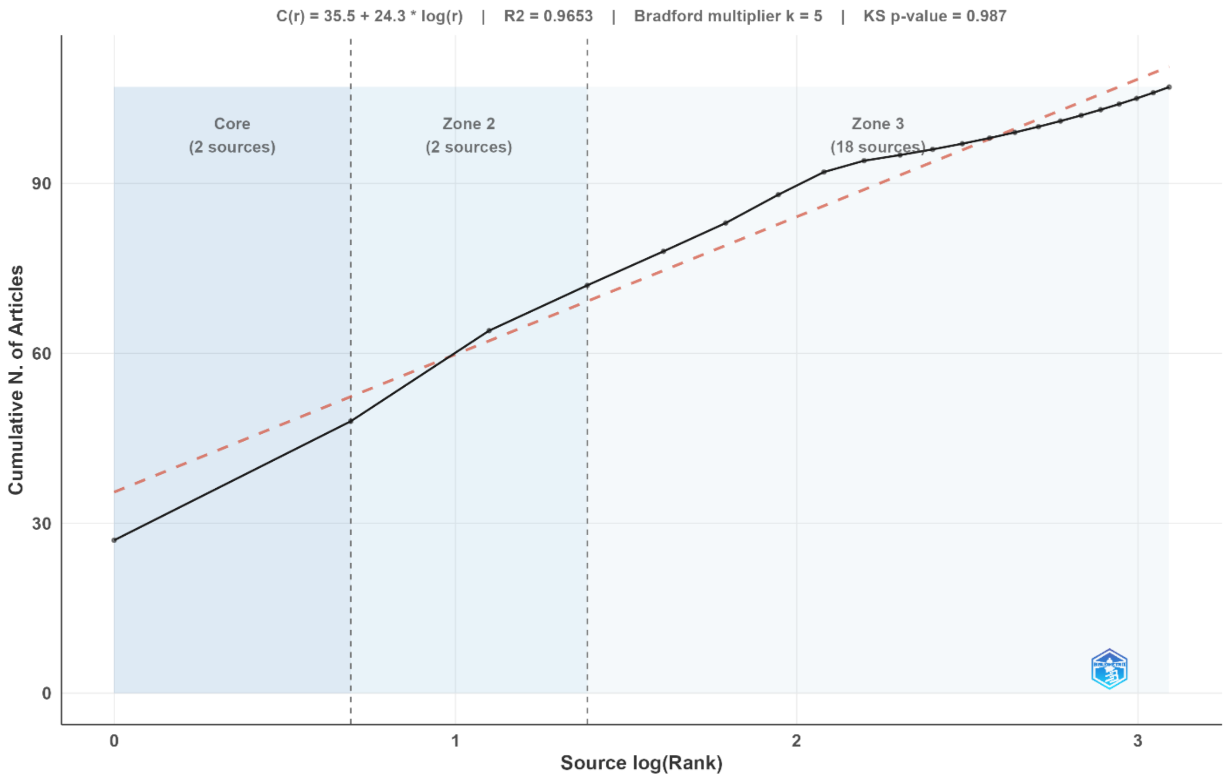


Figure 6. Bradford's law of scattering: core sources and zone partition.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Table 7. Annual scientific production and citation performance.

Year	MeanTCperArt	N	MeanTCperYear	CitableYears
2021	58.83	6	9.8	6
2022	22.33	15	4.47	5
2023	27.38	16	6.84	4
2024	20.42	19	6.81	3
2025	5	32	2.5	2
2026	0.58	19	0.58	1

3.4 Most Productive Authors

Author-level productivity is extremely dispersed. Of 527 unique authors, 492 (93.4%) contributed exactly one document; 25 contributed two, 8 contributed three, 1 contributed four, and 1 contributed five. The most productive author, Wang Z., has 5 documents; Wang Y. has 4; and a group of eight authors Li Z., Zhang Y., Wang W., Liu Y., Chen X., Mao D., Li S., and Liu L. each has 3.

This distribution conforms broadly to Lotka's inverse-square law, but the field is even more dispersed than the classical formulation predicts: Lotka anticipates roughly 60% single-publication authors, whereas this corpus exhibits 93.4%. Price's law defines the productive elite as $\sqrt{527} \approx 23$ authors, who should collectively produce approximately half of all output; in this corpus, the top 23 authors are associated with far less than half. The field therefore lacks an established productive core.

An important methodological caveat applies. The most productive author names Wang Z., Wang Y., Li Z., Zhang Y., Liu Y. are among the most common surname–initial combinations in Chinese-language authorship. Despite the retention of Scopus Author IDs for verification, homonym conflation cannot be fully excluded at this level of aggregation, and the reported counts should be regarded as upper bounds. This is a general limitation of surname–initial normalisation and one reason why author-level findings in this study are interpreted more cautiously than source-, institution-, or country-level findings.

Substantively, the dispersion indicates a field in which RF–GEE is a tool applied by many rather than a research programme pursued by few. Researchers enter the literature to solve a specific applied problem — mapping a particular wetland, estimating a particular crop's yield and typically publish once before moving on. This pattern is characteristic of a mature, low-barrier methodology and helps explain the methodological homogeneity documented in Sections 3.10 and 3.15: there is no sustained community of practice generating cumulative methodological refinement.

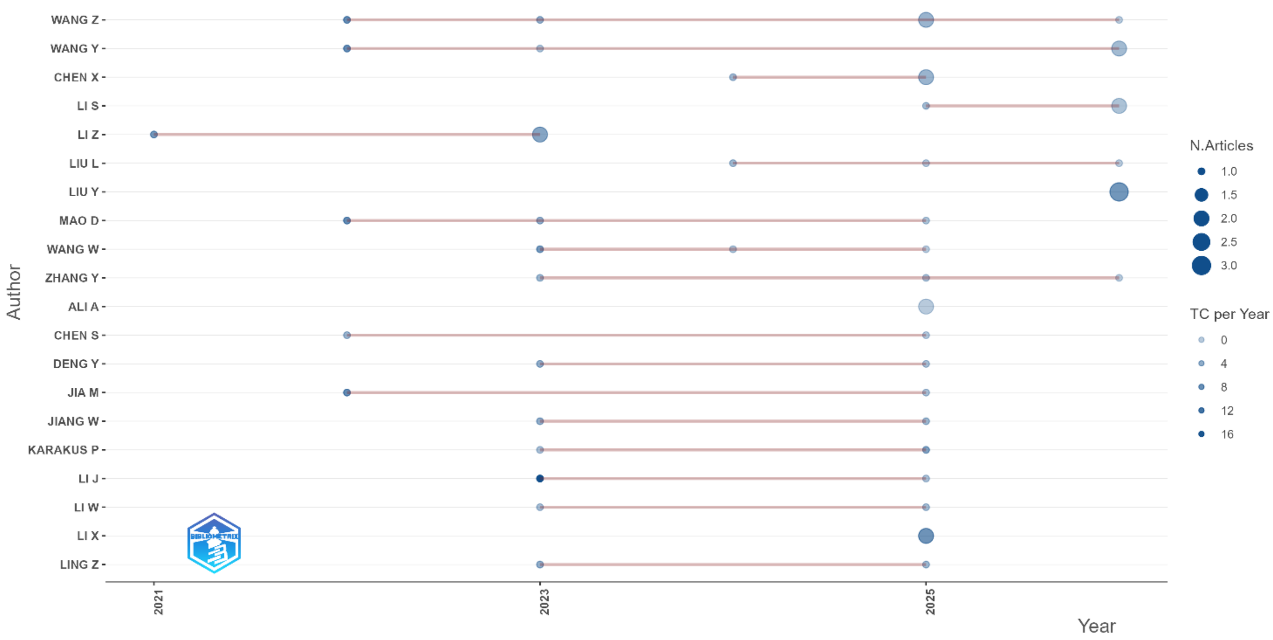


Figure 7. Most productive authors by number of documents with setting main configuration number of Auhtors are 20.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

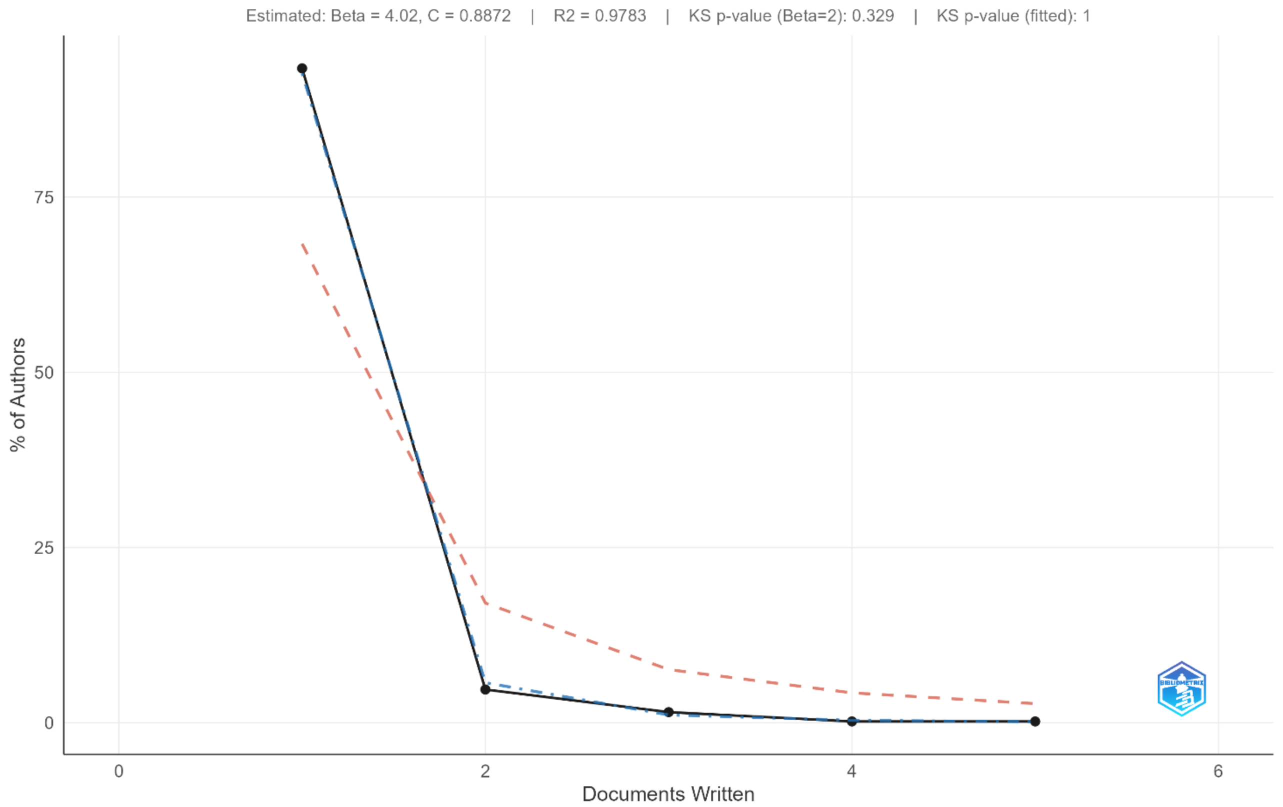


Figure 8. Author productivity distribution and Lotka's law fit.

Table 8. The 20 Most Productive Authors

Author	Year	Frequncy	Total Citation	Total Citattion per Year
LIU Y	2026	3	7	7
LI Z	2023	2	20	5
CHEN X	2025	2	6	3
WANG Z	2025	2	6	3
WANG Y	2026	2	2	2
LI S	2026	2	1	1
LI Z	2021	1	41	6.833333
MAO D	2022	1	42	8.4
WANG Y	2022	1	42	8.4
WANG Z	2022	1	42	8.4
WANG W	2023	1	27	6.75
MAO D	2023	1	15	3.75
WANG Z	2023	1	15	3.75
ZHANG Y	2023	1	10	2.5
WANG Y	2023	1	5	1.25
CHEN X	2024	1	8	2.666667
LIU L	2024	1	7	2.333333
WANG W	2024	1	3	1
LI S	2025	1	4	2
ZHANG Y	2025	1	3	1.5

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Author	Year	Frequncy	Total Citation	Total Citattion per Year
MAO D	2025	1	2	1
LIU L	2025	1	1	0.5
WANG W	2025	1	0	0
LIU L	2026	1	0	0
WANG Z	2026	1	0	0
ZHANG Y	2026	1	0	0

3.5 Institutional Contributions

Institutional output is likewise dispersed, with the leading institution accounting for fewer than 6% of documents. Beijing Normal University leads with 6 documents, followed by the Northeast Institute of Geography and Agroecology (Chinese Academy of Sciences) with 5. Four institutions have 4 documents each: North China University of Water Resources and Electric Power, Hohai University, China University of Geosciences, and the University of Chinese Academy of Sciences. Notably, RV College of Engineering (India) also records 4 documents, and a further group including the Forestry and Forest Products Research Institute (Japan), the Aerospace Information Research Institute (CAS), Shandong Agricultural University, Wuhan University, Shandong University, the University of Hong Kong, the University of Nebraska–Lincoln, Lund University, and Clemson University each contributes 3. Three patterns are evident:

1. Chinese institutional dominance is structural, not incidental. Seven of the top ten institutions are Chinese, and the presence of multiple Chinese Academy of Sciences units alongside major universities indicates coordinated, well-resourced national investment in Earth observation informatics rather than isolated research groups.
2. Hydrology and water-resource institutions are over-represented. North China University of Water Resources and Electric Power and Hohai University — both water-specialist institutions — rank in the top four. This aligns with the finding in Section 3.14 that water and wetland applications constitute a major thematic cluster.
3. Non-Chinese contributions are geographically scattered but institutionally coherent. The Japanese, Indian, Swedish, and United States institutions in the leading group are each associated with a specific applied niche (forest inventory, land use planning, ecosystem modelling, and agricultural monitoring respectively), suggesting parallel national research agendas rather than an integrated international programme.

The absence of any institution exceeding 6 documents reinforces the conclusion of Section 3.4: RF–GEE research has no institutional centre of gravity. It is a widely distributed methodological competence.

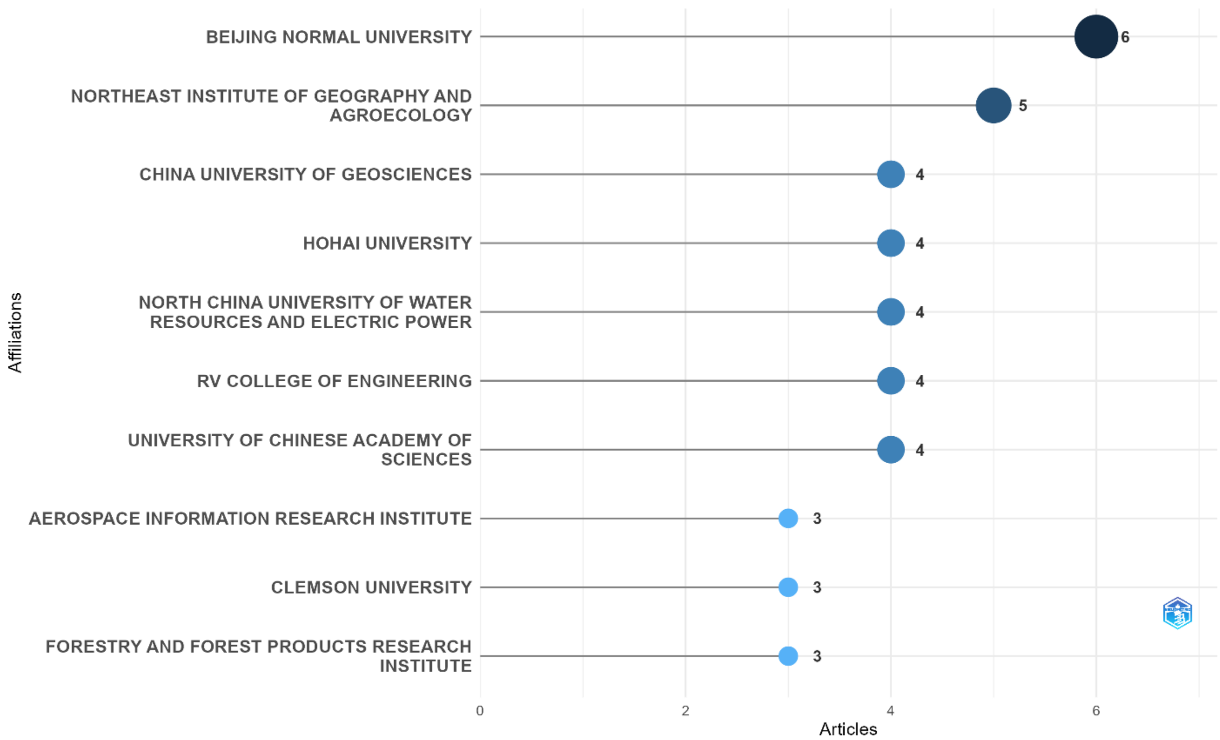


Figure 9. Most productive institutions by number of documents.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Table 9. The 20 Most Affiliations' Production over Time

Affiliation	Year	Articles
BEIJING NORMAL UNIVERSITY	2025	6
BEIJING NORMAL UNIVERSITY	2026	6
NORTHEAST INSTITUTE OF GEOGRAPHY AND AGROECOLOGY	2025	5
NORTHEAST INSTITUTE OF GEOGRAPHY AND AGROECOLOGY	2026	5
NORTH CHINA UNIVERSITY OF WATER RESOURCES AND ELECTRIC POWER	2026	4
UNIVERSITY OF CHINESE ACADEMY OF SCIENCES	2026	4
CHINA UNIVERSITY OF GEOSCIENCES	2025	4
CHINA UNIVERSITY OF GEOSCIENCES	2026	4
HOHAI UNIVERSITY	2025	4
HOHAI UNIVERSITY	2026	4
RV COLLEGE OF ENGINEERING	2025	4
RV COLLEGE OF ENGINEERING	2026	4
AEROSPACE INFORMATION RESEARCH INSTITUTE	2025	3
AEROSPACE INFORMATION RESEARCH INSTITUTE	2026	3
FORESTRY AND FOREST PRODUCTS RESEARCH INSTITUTE	2024	3
FORESTRY AND FOREST PRODUCTS RESEARCH INSTITUTE	2025	3
FORESTRY AND FOREST PRODUCTS RESEARCH INSTITUTE	2026	3
NORTHEAST INSTITUTE OF GEOGRAPHY AND AGROECOLOGY	2024	3
UNIVERSITY OF CHINESE ACADEMY OF SCIENCES	2025	3
BEIJING NORMAL UNIVERSITY	2024	3

3.6 Country Scientific Production

The geographical distribution of the research corpus reveals a pronounced but uneven concentration of scientific activity across countries. Based on the country-frequency analysis, China records the highest frequency with 153 occurrences, substantially exceeding the United States (34), Brazil (15), India (11), and Canada (10). Japan and Spain each record nine occurrences, whereas Italy, Kazakhstan, and Morocco each record eight. The distribution demonstrates that research on Google Earth Engine and Random Forest applications in remote sensing has developed across multiple geographical regions, although the intensity of research activity varies considerably among countries.

The dominant position of China is consistent with its substantial research capacity in remote sensing, Earth observation, geospatial information science, and computational environmental applications. The high frequency of Chinese affiliations suggests a strong institutional engagement with cloud-based geospatial analysis and machine-learning approaches for processing Earth observation data. This pattern may also reflect the scale of China's research system, the availability of extensive Earth observation applications, and the presence of diverse environmental and land-monitoring problems that require large-area and computationally intensive analysis. However, the country-frequency distribution should be interpreted as an indicator of the geographical presence of research contributions rather than as a direct measure of research quality, scientific impact, or methodological superiority. The United States represents the second most frequent country, with 34 occurrences, indicating a substantial contribution to the research domain. The presence of Brazil (15), India (11), and Canada (10) further demonstrates that research activity is not confined to a small group of technologically advanced countries. Japan and Spain, with nine occurrences each, together with Italy, Kazakhstan, and Morocco, with eight occurrences each, constitute an additional group of active contributors. This geographical distribution indicates that Google Earth Engine-based remote-sensing research has gained adoption across different environmental, geographical, and institutional contexts. The presence of countries from diverse geographical regions is particularly relevant to the development of cloud-based Earth observation research. Google Earth Engine provides access to extensive satellite-data archives and cloud-based computational capabilities, allowing researchers to conduct large-scale geospatial analyses without requiring equivalent levels of local data storage and computational infrastructure. Nevertheless, the observed country distribution should not be interpreted as direct evidence that Google Earth Engine itself has reduced research inequalities. Such a conclusion would require additional evidence concerning institutional resources, infrastructure accessibility, research funding, and longitudinal changes in research participation. The present bibliometric

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

evidence instead demonstrates a broad geographical diffusion of research activity involving Google Earth Engine and Random Forest.

The country-frequency distribution also highlights an important distinction between country occurrence and document-level scientific production. A country may appear multiple times in the affiliation metadata of a publication when several authors or institutions from the same country are represented. Therefore, a frequency of 153 for China should not be interpreted as 153 publications produced by China, particularly because the final corpus contains 107 documents. The frequency values should instead be understood as occurrences of country affiliations within the bibliographic records. Consequently, country-level scientific production and international collaboration should be examined using complementary indicators, such as the number of documents by country, single-country publications (SCP), multiple-country publications (MCP), and country collaboration networks. The international dimension of the corpus can therefore be assessed separately from country frequency. In the final dataset of 107 publications, 48 documents involve authors affiliated with more than one country, corresponding to 44.9% of the corpus, provided that this value is derived from the multiple-country publication (MCP) indicator. This proportion indicates a substantial level of international collaboration, although more than half of the publications remain associated with single-country research teams. The pattern suggests that Google Earth Engine and Random Forest research combines nationally focused investigations with a considerable degree of cross-border collaboration. International collaboration may be particularly relevant for studies involving transboundary environmental systems, comparative land-use analysis, large-scale Earth observation, and methodological transfer across geographical contexts.

Overall, the country analysis indicates a geographically diverse but highly asymmetric research landscape. China occupies a clearly dominant position in terms of country-frequency occurrence, followed by the United States and a broader group of countries with substantially lower frequencies. This concentration provides an important foundation for interpreting the institutional and collaboration structures of the field. The findings further suggest that the research landscape is simultaneously characterised by strong contributions from major research systems and increasing participation from countries across Asia, Europe, Africa, North America, and South America. The combination of country-level productivity, collaboration indicators, and network-based analyses is therefore necessary to obtain a more complete understanding of the global development of Google Earth Engine applications in remote sensing.

Taken together, the country-level evidence indicates that Google Earth Engine and Random Forest research has achieved broad international diffusion, while scientific production remains concentrated in a limited number of highly active national research systems.

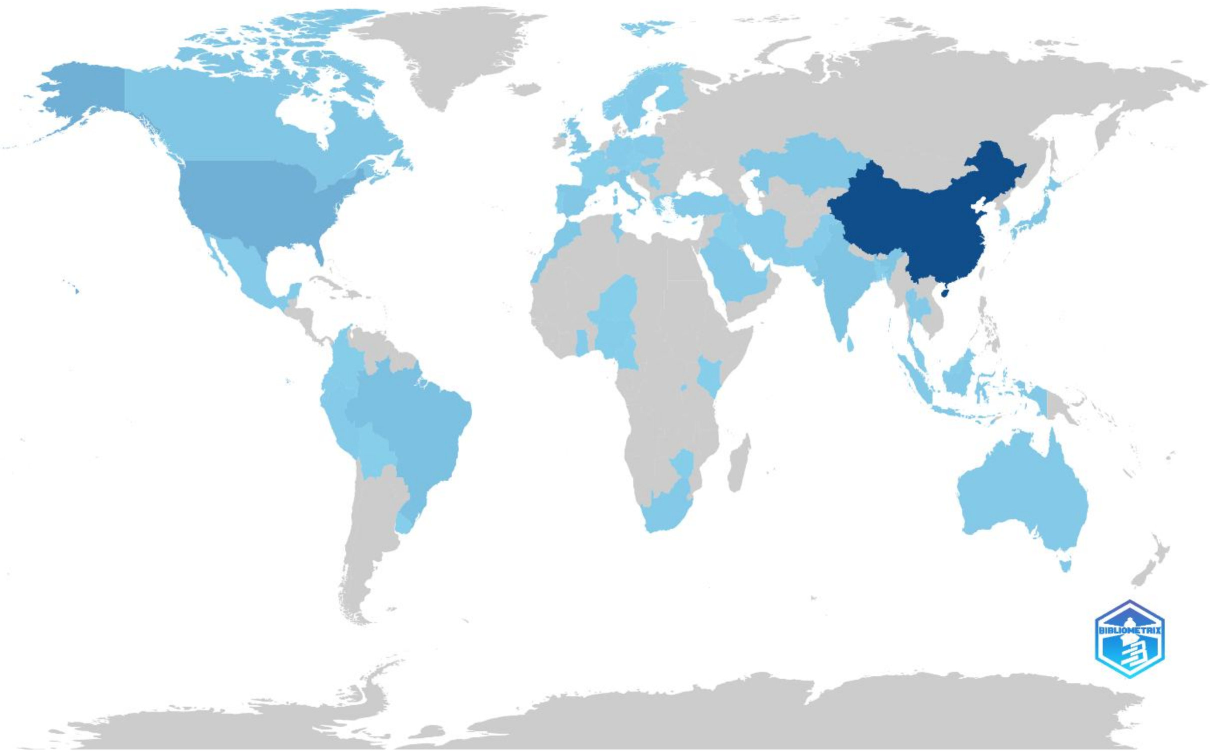


Figure 10. Country scientific production (corresponding-author and co-authorship affiliations), with China dominating

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

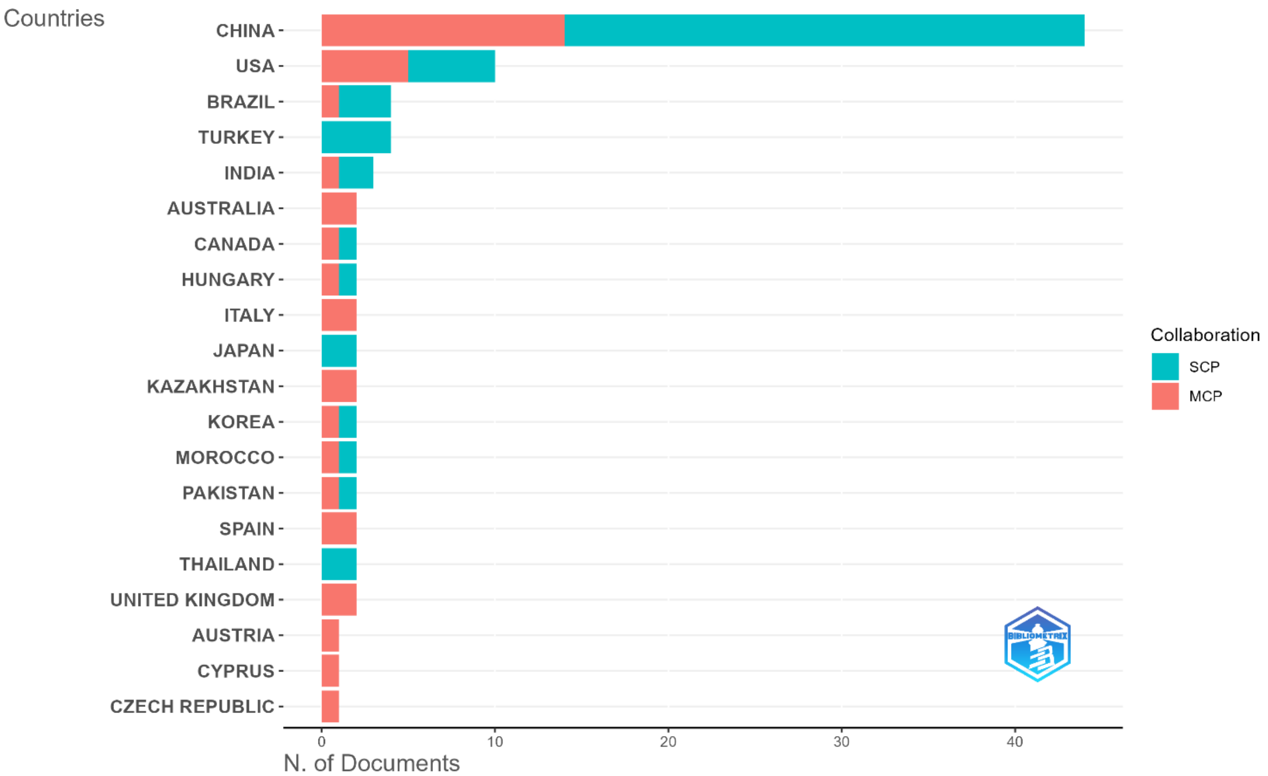


Figure 11. Corresponding-author country production: single-country versus multiple-country publications. The stacked horizontal bar chart of the top 20 countries showing single-country (SCP) and multiple-country (MCP) publications.

Table 10. Country scientific production and international collaboration ratio, the top 20 countries showing single-country (SCP) and multiple-country (MCP) publications

Country	Articles	Articles %	SCP	MCP	MCP %
CHINA	44	41.1215	30	14	31.81818
USA	10	9.345794	5	5	50
BRAZIL	4	3.738318	3	1	25
TURKEY	4	3.738318	4	0	0
INDIA	3	2.803738	2	1	33.33333
AUSTRALIA	2	1.869159	0	2	100
CANADA	2	1.869159	1	1	50
HUNGARY	2	1.869159	1	1	50
ITALY	2	1.869159	0	2	100
JAPAN	2	1.869159	2	0	0
KAZAKHSTAN	2	1.869159	0	2	100
KOREA	2	1.869159	1	1	50
MOROCCO	2	1.869159	1	1	50
PAKISTAN	2	1.869159	1	1	50
SPAIN	2	1.869159	0	2	100
THAILAND	2	1.869159	2	0	0
UNITED KINGDOM	2	1.869159	0	2	100
AUSTRIA	1	0.934579	0	1	100
CYPRUS	1	0.934579	0	1	100
CZECH REPUBLIC	1	0.934579	0	1	100

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

3.7 Co-authorship Analysis

The co-authorship network comprises 527 nodes and 1,555 edges, with a density of 0.0112 and a mean degree of 5.9. The network fragments into 68 disconnected components. The largest component contains 185 authors 35.1% of the author population while the remaining 64.9% are distributed across 67 smaller components and isolated teams.

This structure is diagnostic. A field with a well-integrated research community typically exhibits a giant component encompassing 60–80% of authors. At 35.1%, RF-GEE research is socially fragmented: it consists of many small, largely self-contained author teams that collaborate intensively within themselves but rarely across group boundaries. Only 35 authors (6.6%) appear in two or more documents, further confirming the absence of a persistent collaborative backbone. The high mean degree (5.9 collaborators per author) alongside low global density (0.0112) is characteristic of a network built from cliques: each document contributes a fully connected subgraph of its 5.37 average co-authors, but these cliques are only sparsely bridged. In effect, the network is a collection of research teams rather than a research community.

Two consequences follow. First, methodological innovation propagates slowly. Without dense inter-team bridges, an improvement developed by one group a better sampling design, an uncertainty-quantification procedure, a transferability testdiffuses through the literature only via citation, which is markedly slower than diffusion through collaboration. This helps explain the persistence of the standard RF-GEE workflow documented in Section 3.10. Second, the field is vulnerable to redundant effort: teams working on structurally similar problems in different regions are largely unaware of one another, a hypothesis directly supported by the bibliographic coupling results in Section 3.10, which show high reference overlap between documents whose authors have no collaborative connection.

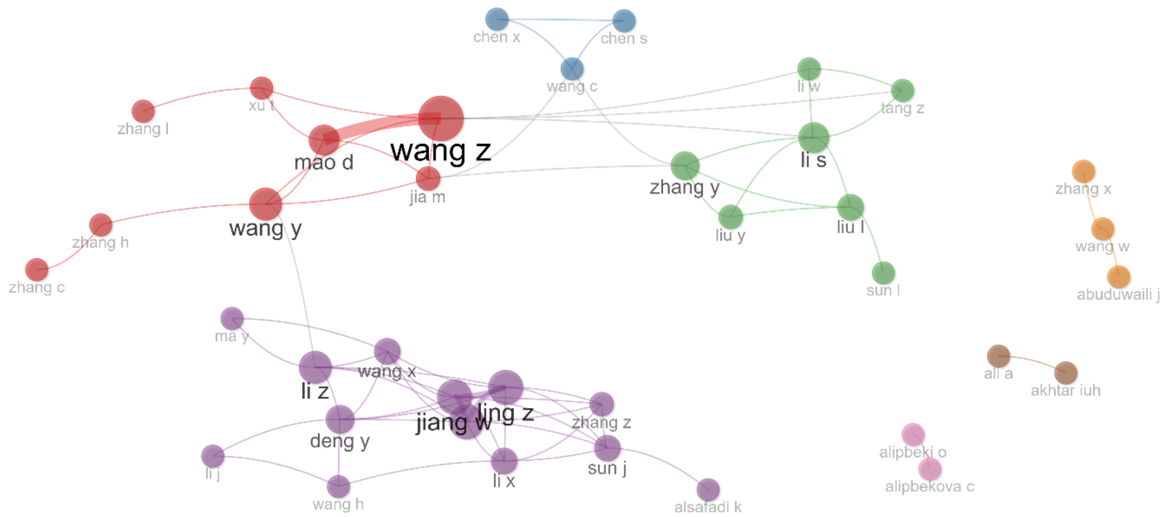


Figure 12. Co-authorship network of authors with two or more documents.

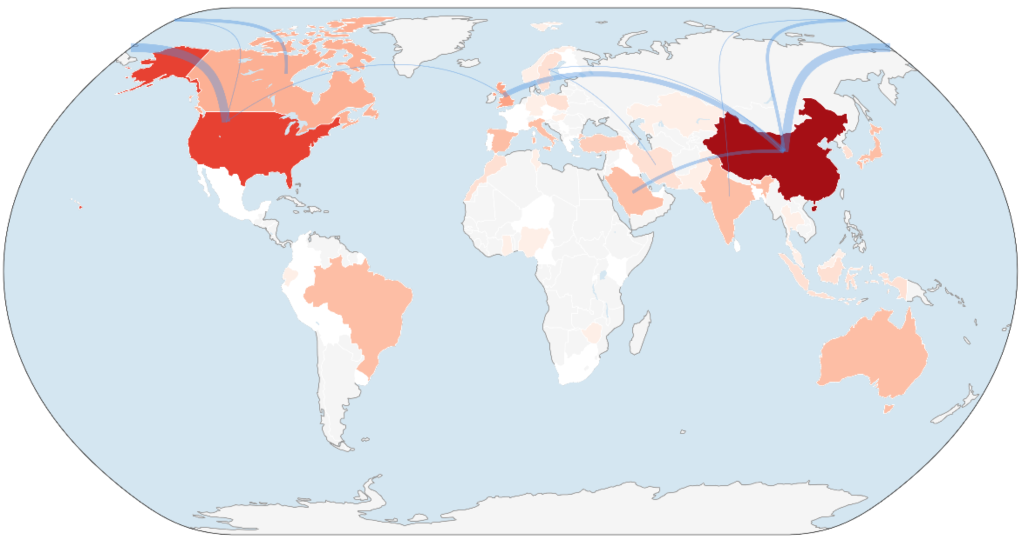


Figure 13. International co-authorship network by country.

.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Table 11. Co-authorship network metrics (nodes, edges, density, components, largest-component share, mean degree, authors with ≥ 2 documents, clustering coefficient)

Network Metric	Value	Interpretation
Number of nodes	38	Number of authors included in the co-authorship network.
Number of edges	69	Number of unique co-authorship links among the 38 authors.
Network density	0.098	Approximately 9.8% of all theoretically possible author-to-author connections are present.
Number of components	4	The network consists of four disconnected collaboration components.
Largest-component share	81.6%	31 of the 38 authors belong to the largest connected component.
Mean degree	3.63	Each author has, on average, approximately 3.63 direct co-authorship connections.
Authors with ≥2 documents	38	All authors represented in the network have at least two documents according to the exported author-weight vector.
Average clustering coefficient	0.451	Indicates a moderate tendency for an author's collaborators to collaborate with one another.

3.8 Co-occurrence Analysis

The normalised author-keyword vocabulary comprises 390 unique terms connected by 1,444 co-occurrence links. The frequency distribution is severely skewed: only 46 terms (11.8%) occur twice or more, 16 occur three times or more, and 12 occur five times or more. The great majority of keywords are used once, typically to name a specific study area, sensor product, or target class.

The high-frequency terms are: GOOGLE EARTH ENGINE (66), RANDOM FOREST (31), MACHINE LEARNING (29), REMOTE SENSING (21), SENTINEL-2 (15), LULC (10), then a group at 5 — SUPPORT VECTOR MACHINE, SENTINEL, CLASSIFICATION, TIME SERIES, LANDSAT, VEGETATION INDEX — followed by SENTINEL-1 and SAR (4 each) and SHAP and ENSEMBLE LEARNING (3 each).

The strongest co-occurrence links are GOOGLE EARTH ENGINE–RANDOM FOREST (18), GOOGLE EARTH ENGINE–MACHINE LEARNING (12), GOOGLE EARTH ENGINE–REMOTE SENSING (12), GOOGLE EARTH ENGINE–SENTINEL-2 (9), MACHINE LEARNING–RANDOM FOREST (6), MACHINE LEARNING–REMOTE SENSING (6), and GOOGLE EARTH ENGINE–LULC (6).

The network's most consequential property is its topology rather than its content. At every threshold tested (≥ 2, ≥ 3, ≥ 5), the network resolves into a single dense component centred on GOOGLE EARTH ENGINE, with no stable modular decomposition. Community-detection algorithms applied to the author-keyword network at threshold ≥ 3 and to the index-keyword network at threshold ≥ 8 both converge on a single cluster. This is not an artefact of algorithm choice; it is a substantive finding. Because virtually every document carries the same two or three anchor terms, the network has no structural basis for partition — the shared vocabulary is too dominant for thematic subcommunities to separate. The interpretation is that the field possesses a single, universally shared methodological vocabulary layered over a highly diverse but weakly interlinked set of application vocabularies. Application-specific terms (MANGROVE SPECIES MAPPING, CROP TYPE MAPPING, URBAN HEAT ISLAND, SOIL MOISTURE) appear once or twice each, connect to the methodological hub, and do not connect to each other. Conceptually, the field is a hub-and-spoke structure with an extremely strong hub and very weak rim.

Consistent with this, only three terms in the top tier are not methodological or sensor-related: LULC, CLASSIFICATION, and TIME SERIES — and all three are themselves generic. No substantive environmental concept (biodiversity, food security, carbon, adaptation) reaches the high-frequency tier. This is direct evidence that the literature is organised around a technique rather than around a problem.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

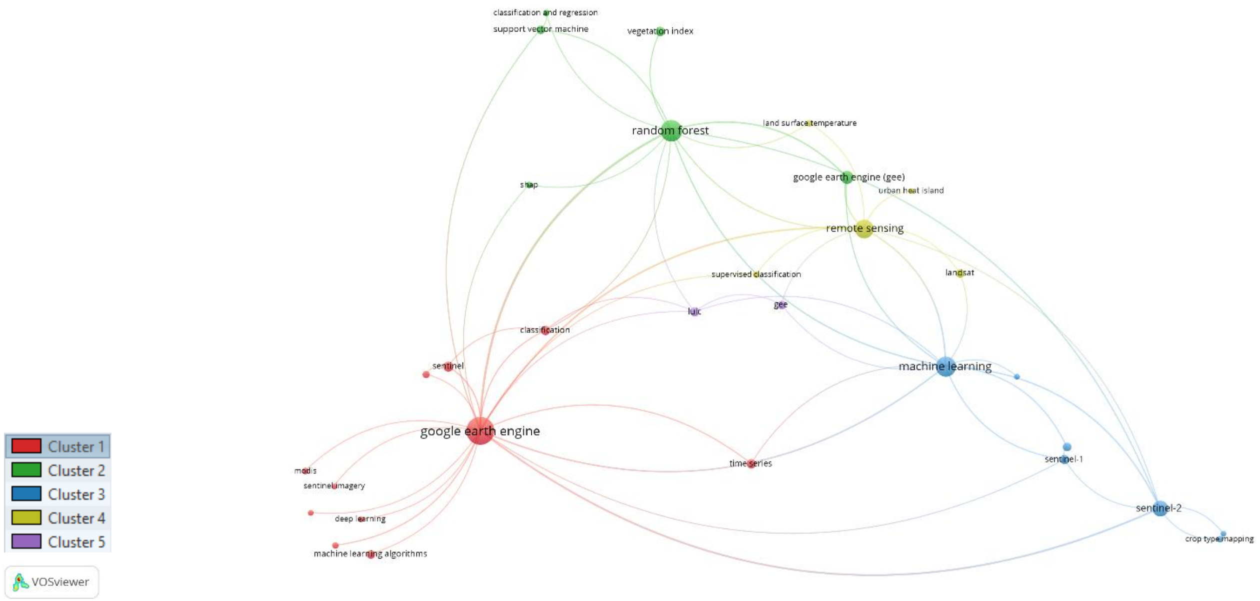


Figure 14. Author-keyword co-occurrence network.

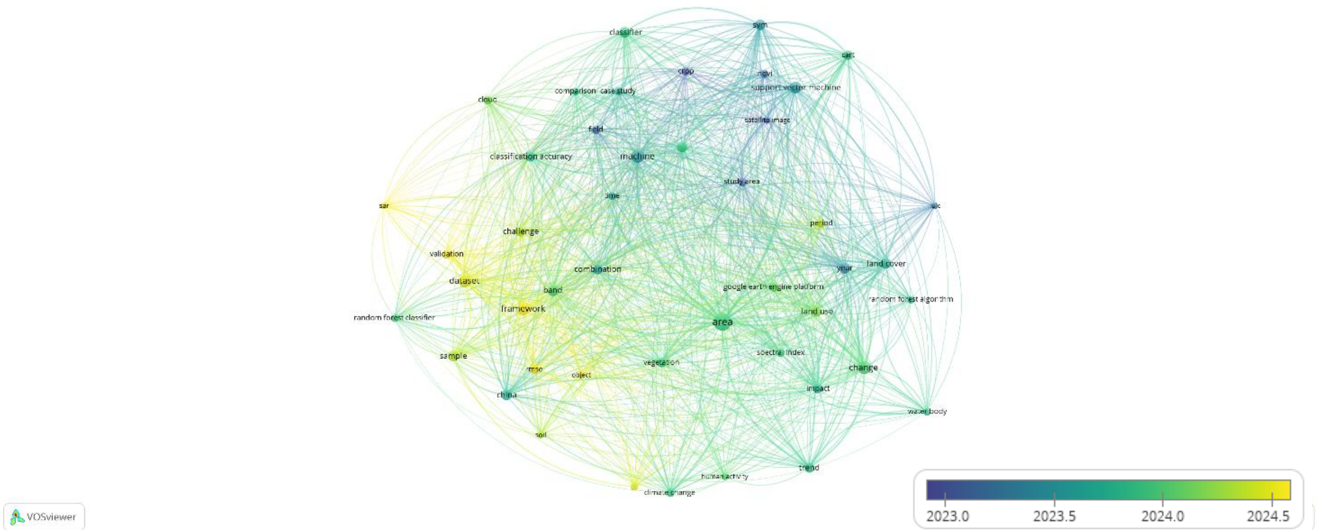


Figure 15. Keyword co-occurrence overlay by average publication year.

(a)

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

(b)

Figure 16. (a) Author-keyword word cloud, (b) Author-keyword TreeMap.

3.9 Co-citation Analysis

The 107 documents in the final corpus collectively cite 6,156 distinct canonical references, indicating a broad reference base underlying the research domain. Despite this breadth, citation activity is highly concentrated around a small group of foundational works. The three most frequently cited references are Gorelick et al. (2017), the Google Earth Engine platform paper, with 49 citations (45.8% of the corpus); Breiman (2001), the foundational Random Forest paper, with 47 citations (43.9%); and Belgiu and Drăguţ (2016), which reviews Random Forest applications in remote sensing, with 34 citations (31.8%).

The core is dominated by three works. Gorelick et al. (2017) was cited by 49 documents in the corpus (45.8%) [1]. Breiman (2001), the original Random Forest paper, is co-cited 47 times [5]. Belgiu and Drăguţ (2016), the review of RF in remote sensing, is co-cited 34 times [6]. Together these three constitute the field's canonical triad: platform, algorithm, and the review that legitimised their combination.

A second tier comprises Tamiminia et al. (2020), the GEE meta-analysis (17) [3]; McFeeters (1996), the NDWI index (13) [20]; Rodriguez-Galiano et al. (2012), the RF accuracy assessment (11) [7]; Tucker (1979), the foundational NDVI paper (11) [21]; and Amani et al. (2020), on GEE for remote sensing big data (10) [4].

A third tier of spectral-index and infrastructure references follows: Huete (1988) on SAVI (9) [22], Cortes and Vapnik (1995) on support vector networks (9) [23], Farr et al. (2007) on SRTM (8) [24], Breiman et al. (1984) on CART (8) [25], Xu (2006) on MNDWI (8) [26], Huete et al. (2002) on MODIS vegetation indices (8) [27], Hansen et al. (2013) on global forest cover change (8) [28], Drusch et al. (2012) on the Sentinel-2 mission (7) [29], Teluguntla et al. (2018) on Landsat-derived cropland extent (7) [30], Tassi and Vizzari (2020) on object-oriented LULC classification in GEE (7) [31], Congalton (1991) on accuracy assessment (7) [32], and Pedregosa et al. (2011) on scikit-learn (6) [33].

Three observations follow.

The foundation is infrastructural rather than theoretical. Nearly every core reference describes a platform, an algorithm, a sensor, a spectral index, or an accuracy protocol. There are no theoretical works on landscape ecology, land system science, or environmental change theory in the co-cited core. The field's shared knowledge base is a toolkit, not a body of theory.

The core is temporally bifurcated. It comprises either recent infrastructure papers (2016–2020) or classical index and algorithm papers (1979–2001), with a conspicuous gap between 2002 and 2015. The field has not built a cumulative intermediate literature of its own; it reaches directly from contemporary tools back to foundational primitives.

Concentration is extreme. With 45.8% of documents co-citing a single paper and 99.35% of references co-cited fewer than five times, the field exhibits a canonical core surrounded by an almost entirely non-shared periphery. Each study builds on the same three works plus its own largely unique application-specific literature. This structure explains the co-occurrence hub-and-spoke topology of Section 3.8 from the citation side: the shared foundation is methodological, and everything above it is idiosyncratic.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

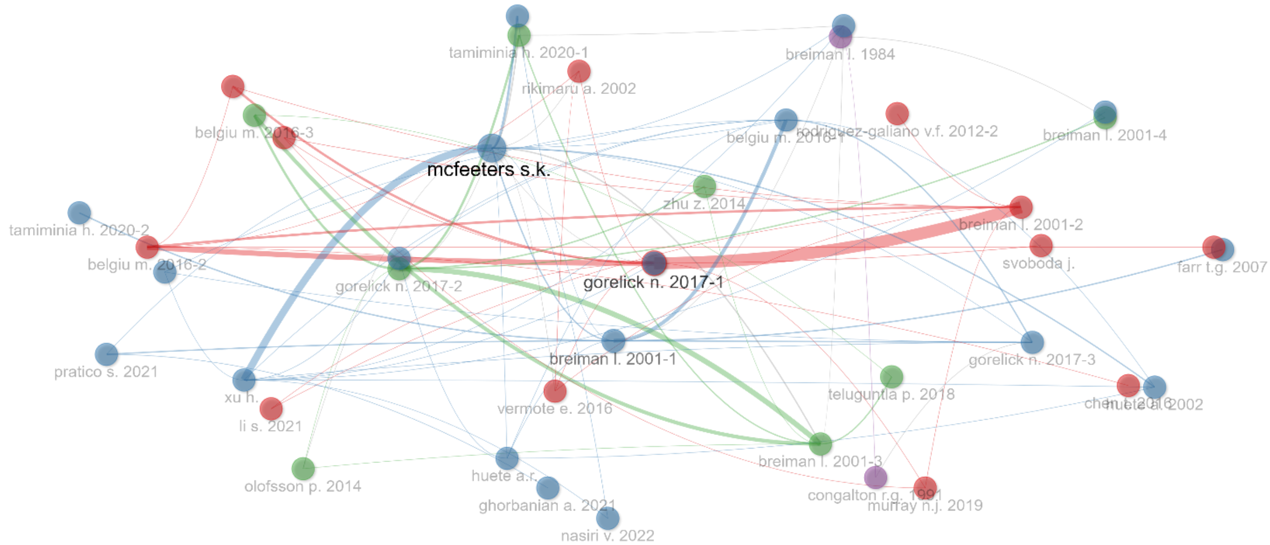


Figure 17. Co-citation network of cited references.

Rank	Reference	Year	Co-citation Count	Share of Corpus (%)	Functional Category	Primary Contribution
1	Gorelick et al.	2017	49	45.8	Platform	Google Earth Engine platform and cloud-based geospatial computation
2	Breiman	2001	47	43.9	Algorithm	Random Forest algorithm
3	Belgiu & Drăguţ	2016	34	31.8	Algorithm / Review	Random Forest applications in remote sensing
4	Tamiminia et al.	2020	17	15.9	Platform / Review	Review and meta-analysis of Google Earth Engine applications
5	McFeeters	1996	13	12.1	Index	Normalized Difference Water Index (NDWI)
6	Rodriguez-Galiano et al.	2012	11	10.3	Protocol	Random Forest classification and accuracy assessment
7	Tucker	1979	11	10.3	Index	Normalized Difference Vegetation Index (NDVI)
8	Amani et al.	2020	10	9.3	Platform / Application	Google Earth Engine for remote-sensing big-data analysis
9	Huete	1988	9	8.4	Index	Soil-Adjusted Vegetation Index (SAVI)
10	Cortes & Vapnik	1995	9	8.4	Algorithm	Support Vector Machine (SVM)
11	Farr et al.	2007	8	7.5	Sensor / Dataset	Shuttle Radar Topography Mission (SRTM)
12	Breiman et al.	1984	8	7.5	Algorithm	Classification and Regression Trees (CART)
13	Xu	2006	8	7.5	Index	Modified Normalized Difference Water Index (MNDWI)
14	Huete et al.	2002	8	7.5	Index / Dataset	MODIS vegetation-index products
15	Hansen et al.	2013	8	7.5	Dataset / Application	Global forest-cover change mapping
16	Drusch et al.	2012	7	6.5	Sensor	Sentinel-2 mission
17	Teluguntla et al.	2018	7	6.5	Dataset / Application	Landsat-derived cropland extent
18	Tassi & Vizzari	2020	7	6.5	Platform / Application	Object-oriented land-use/land-cover classification using GEE

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Rank	Reference	Year	Co-citation Count	Share of Corpus (%)	Functional Category	Primary Contribution
19	Congalton	1991	7	6.5	Protocol	Accuracy assessment of remote-sensing classifications
20	Pedregosa et al.	2011	6	5.6	Software / Infrastructure	Scikit-learn machine-learning framework

3.10 Bibliographic Coupling Analysis

Bibliographic coupling links documents that share references and thereby maps the contemporary research front. All 107 documents are coupled, forming 2,747 linked pairs 48.4% of all possible pairs. The maximum number of references shared by any pair is 18, and 59 pairs share five or more references.

Ranked by total coupling strength, the most central documents are Li S. et al. (2026) on UAV–satellite integration (222), Loukika et al. (2021) on LULC in the Munneru river basin (204), Safaei et al. (2025) on land-cover classification in Addo Elephant National Park (201), Zhang C. et al. (2023) on phenology-assisted paddy rice mapping (190), Fauzi et al. (2025) on scalable mangrove mapping (186), Deng et al. (2023) on Dongting Lake wetland mapping (181), Wang W. et al. (2023) on sand and dust storm source prediction (181), and Wimberly et al. (2022) on tropical forest degradation (177).

The substantive finding is the composition of this list. These eight documents address entirely different phenomena agricultural drones, river-basin land use, savanna ecology, rice phenology, mangroves, lake wetlands, dust storms, and tropical forests across four continents. They are nonetheless the most tightly coupled documents in the corpus. Coupling here therefore does not indicate a shared substantive research problem; it indicates a shared methodological apparatus.

This is arguably the single most important structural result of the study. Bibliographic coupling normally identifies emergent problem-focused research fronts. In this corpus it identifies a procedural front: studies converge because they cite the same platform paper, the same algorithm paper, the same review, the same spectral indices, and the same accuracy protocol not because they are addressing the same question. Coupling density (48.4% of all possible pairs linked) quantifies the degree of methodological standardisation.

Read alongside Section 3.7, the picture sharpens: the field is socially fragmented (35.1% largest component) yet methodologically unified (48.4% coupling density). Teams that never collaborate nonetheless produce near-identical workflows. Standardisation has been achieved through the platform itself rather than through community interaction — GEE's built-in classifier library and documentation function as a de facto methodological standard.

The benefits are real: comparability across studies, low entry barriers, and rapid deployment. The costs are equally real. When a workflow is standardised by infrastructure rather than by deliberate community consensus, its limitations are inherited uncritically. Section 3.15 develops this point.

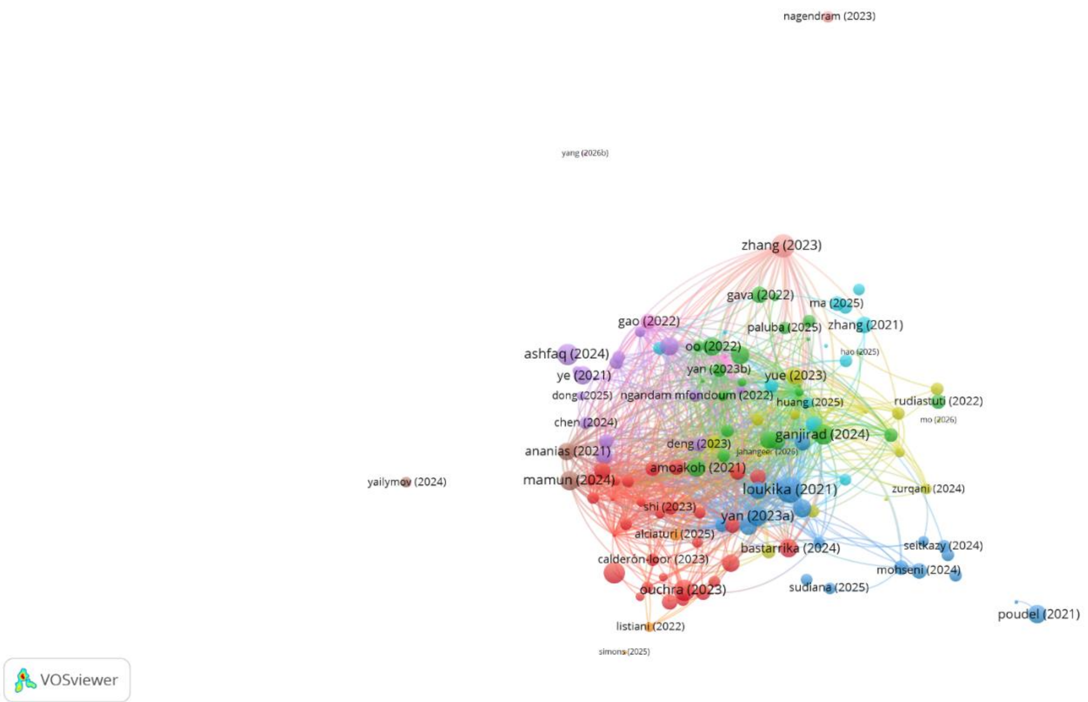


Figure 18. Bibliographic coupling network of corpus documents

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Table 14. Documents Ranked by Total Bibliographic Coupling Strength

Rank	Document	Year	Title	Total Link Strength (TLS)
1	Li S. et al.	2026	*From Plots to Region: Machine Learning-Based UAV-Satellite Integration for Mapping Fractional Coverage of Peanut Southern Blight*	222
2	Loukika K. N. et al.	2021	*Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India*	204
3	Safaei M. et al.	2025	*Land-Cover Classification in Addo Elephant National Park: Analyzing the Impact of Variables, Classifiers, and Object-Based Approach*	201
4	Zhang C. et al.	2023	*Phenology-Assisted Supervised Paddy Rice Mapping with the Landsat Imagery on Google Earth Engine: Experiments in Heilongjiang Province of China from 1990 to 2020*	190
5	Fauzi A. I. et al.	2025	*Exploring New Mangrove Horizons: A Scalable Remote Sensing Approach with Planet-NICFI and Sentinel-2 Images*	186
6	Deng Y. et al.	2023	*Automated and Refined Wetland Mapping of Dongting Lake Using Migrated Training Samples Based on Temporally Dense Sentinel-1/2 Imagery*	181
7	Wang W. et al.	2023	*Machine Learning-Based Prediction of Sand and Dust Storm Sources in Arid Central Asia*	181
8	Wimberly M. C. et al.	2022	*Historical Trends of Degradation, Loss, and Recovery in the Tropical Forest Reserves of Ghana*	177

Table 14. Documents ranked by total bibliographic coupling strength (TLS). Total link strength represents the cumulative strength of a document's bibliographic coupling relationships with other documents in the corpus, based on the number of cited references shared between publications.

Table 15. The 20 Most Globally Cited Documents in the Final Bibliometric Corpus

Paper	DOI	Total Citations	TC per Year	Normalized TC
LOUKIKA KN, 2021, SUSTAINABILITY	10.3390/su132413758	188	31.3	3.195467422
ZHANG C, 2023, COMPUT ELECTRON AGRIC	10.1016/j.compag.2023.108105	106	26.5	3.872146119
OUCHRA H, 2023, IEEE ACCESS	10.1109/ACCESS.2023.3293828	82	20.5	2.99543379
ASHFAQ M, 2024, IEEE ACCESS	10.1109/ACCESS.2024.3376735	71	23.7	3.476804124
YAN X, 2023, INT J DIGIT EARTH	10.1080/17538947.2023.2270459	68	17	2.484018265
GANJIRAD M, 2024, ECOL INFORMATICS	10.1016/j.ecoinf.2024.102498	59	19.7	2.889175258
MAMUN M, 2024, ECOL INFORMATICS	10.1016/j.ecoinf.2024.102608	50	16.7	2.448453608
ZHAO C, 2022, ISPRS J PHOTOGRAMM REMOTE SENS	10.1016/j.isprsjprs.2022.09.011	42	8.4	1.880597015
BASTARRIKA A, 2024, ISPRS J PHOTOGRAMM REMOTE SENS	10.1016/j.isprsjprs.2024.08.019	41	13.7	2.007731959
YE Z, 2021, SUSTAINABILITY	10.3390/su132414055	41	6.8	0.696883853
OO TK, 2022, SUSTAINABILITY	10.3390/su141710754	40	8	1.791044776

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Paper	DOI	Total Citations	TC per Year	Normalized TC
ZURQANI HA, 2025, ECOL INFORMATICS	10.1016/j.ecoinf.2025.103052	38	19	7.6
POUDEL U, 2021, SUSTAINABILITY	10.3390/su13147967	37	6.2	0.628895184
AMOAKOH AO, 2021, SENSORS	10.3390/s21103399	35	5.8	0.59490085
MASROOR M, 2022, SUSTAINABILITY	10.3390/su14020642	35	7	1.567164179
YUE L, 2023, INT J DIGIT EARTH	10.1080/17538947.2023.2166606	31	7.8	1.132420091
SAAD EL IMANNI H, 2022, J IMAGING	10.3390/jimaging8120316	30	6	1.343283582
PECH-MAY F, 2022, SENSORS	10.3390/s22134729	29	5.8	1.298507463
FERNANDO WAM, 2024, INF PROCESS AGRIC	10.1016/j.inpa.2023.02.009	29	9.7	1.420103093
ANANIAS PHM, 2021, INT J DIGIT EARTH	10.1080/17538947.2021.1907462	28	4.7	0.47592068

3.11 Thematic Map Analysis

Thematic mapping positions keyword clusters on the centrality–density plane [16], [17]. Applying the procedure to this corpus produced a result that is itself informative: at every tested configuration author keywords with minimum frequency 3 (16 terms), index keywords with minimum frequency 8 (42 terms), and intermediate thresholds the algorithm returned a single cluster. Rather than a limitation, this is a finding consistent with Sections 3.8 and 3.10. A strategic diagram requires modular structure; where the keyword network is a single dense component, no meaningful quadrant assignment is possible. Formally, the corpus contains one theme, positioned as a motor theme: maximally central (all inter-topic linkage passes through it) and highly dense (internal linkage density 105.0 for the author-keyword configuration, 54.7 for the index-keyword configuration), with a mean publication year of 2023.9–2024.0.

To recover interpretable thematic structure, a complementary content-based classification was applied. Each document was assigned to one or more application domains by regular-expression matching against the combined title, abstract, author-keyword, and index-keyword text. This procedure yields the following domain distribution (documents may belong to more than one domain):

Table 16. Application-Domain Distribution Identified Through Content-Based Classification

Application domain	Documents	% of corpus
Forest, vegetation and biomass	75	70.1
Agriculture and crop monitoring	50	46.7
LULC mapping and change detection	46	43.0
Climate, land surface temperature and atmosphere	35	32.7
Water and wetland	32	29.9
Soil and terrain	27	25.2
Methodology and explainable AI	24	22.4
Disaster and hazard	7	6.5

Table 16 presents the distribution of application domains identified through content-based classification of the 107 publications included in the final bibliometric corpus. The classification reveals that forest, vegetation, and biomass constitute the dominant application domain, appearing in 75 publications (70.1% of the corpus). This predominance indicates that Google Earth Engine (GEE) and Random Forest are particularly established for vegetation monitoring, forest assessment, biomass estimation, and related environmental applications. Agriculture and crop monitoring represents the second-largest domain, with 50 publications (46.7%), highlighting the increasing use of cloud-based Earth observation and machine-learning techniques for agricultural mapping and crop-related analysis. LULC mapping and

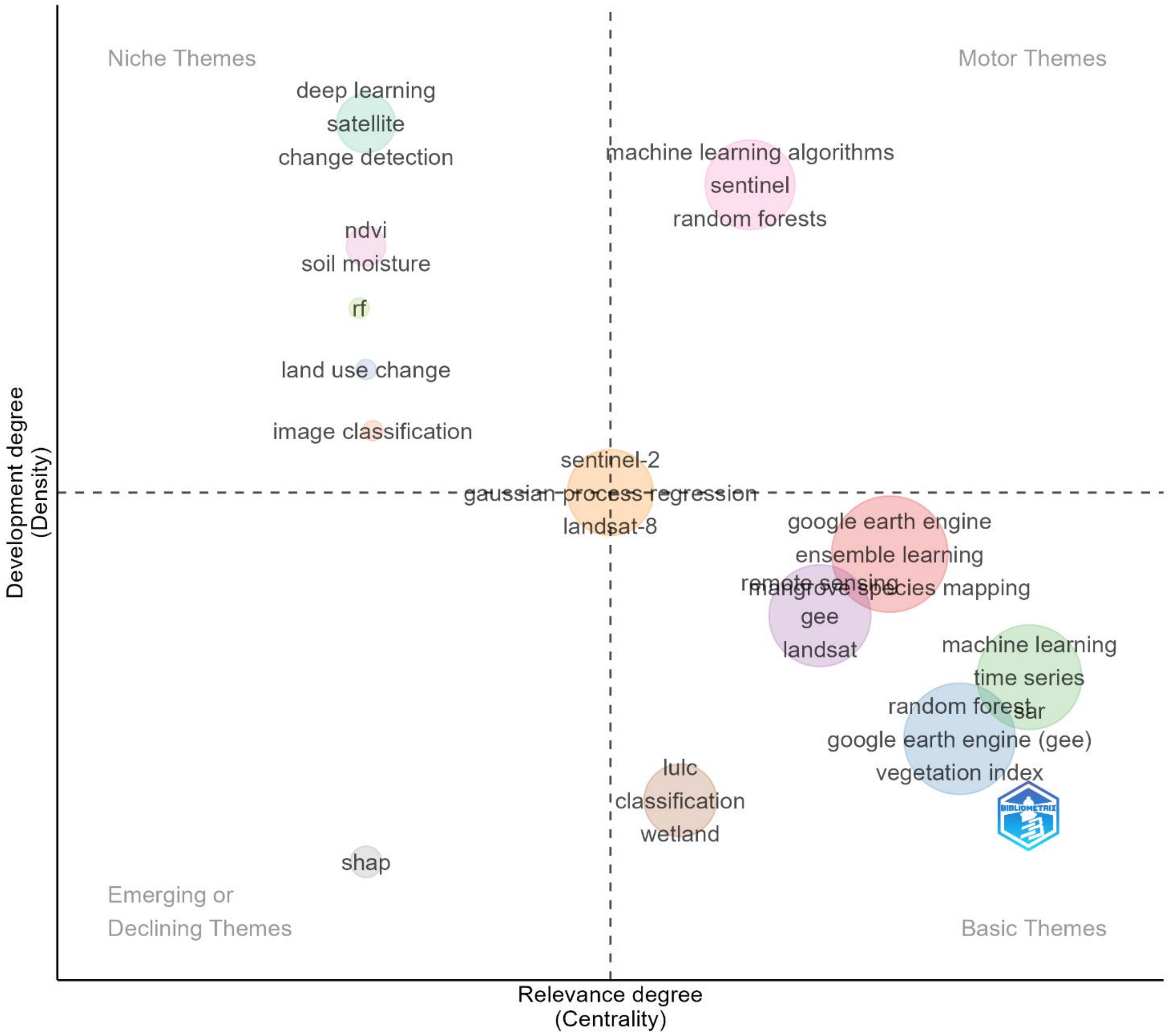
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

change detection follows closely, accounting for 46 publications (43.0%), confirming that land-cover classification and temporal change analysis remain central applications within the research domain.

The remaining domains demonstrate broader diversification of GEE and Random Forest applications. Climate, land surface temperature, and atmospheric applications are represented by 35 publications (32.7%), while water and wetland research accounts for 32 publications (29.9%). Soil and terrain applications occur in 27 publications (25.2%), indicating continued use of Earth observation data for environmental and geomorphological assessment. Meanwhile, methodology and explainable AI appears in 24 publications (22.4%), suggesting an emerging shift toward improving model transparency, interpretability, and methodological robustness. Disaster and hazard constitutes the smallest domain, with seven publications (6.5%), indicating a more specialised research direction.

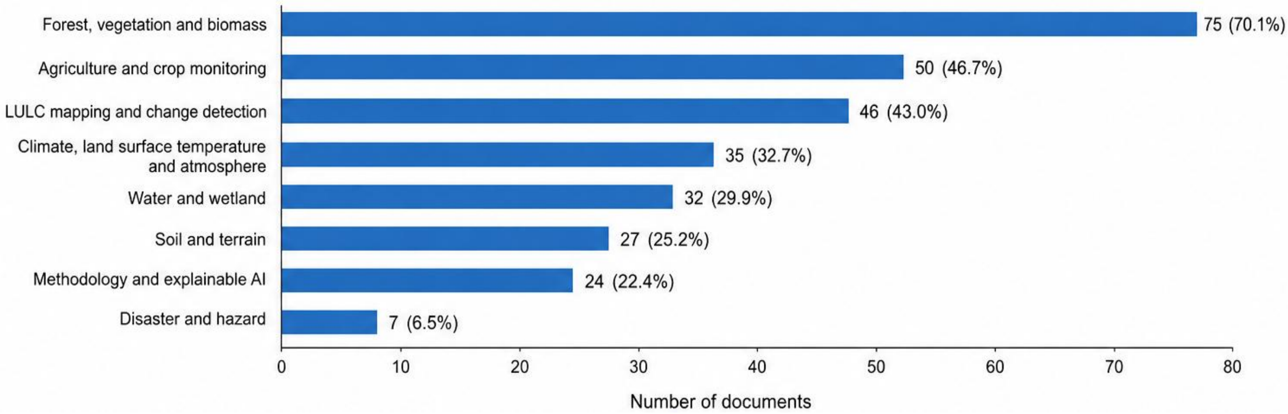
Importantly, these categories are not mutually exclusive, as individual publications may address multiple application domains. Therefore, the percentages do not sum to 100%. Overall, the distribution demonstrates that the field is dominated by environmental monitoring and land-surface applications while progressively expanding toward specialised environmental problems and methodological innovation. Interpreted through the strategic-diagram lens, these domains occupy distinguishable positions. Forest/vegetation, agriculture, and LULC function as basic and transversal themes: highly central, appearing throughout the corpus, but internally undifferentiated — they are the field's default application substrate. Water/wetland and climate/LST approach motor theme status: substantial in volume, internally coherent, and increasingly associated with dedicated institutional programmes (Section 3.5). Methodology and explainable AI constitutes an emerging theme: low volume (22.4%) but rising sharply, with SHAP carrying a median publication year of 2026. Disaster and hazard is a niche theme: small (6.5%) and internally specialised.

The sensor and algorithm distributions reinforce this reading. Sentinel-2 is referenced in 44 documents, Landsat in 42, Sentinel-1/SAR in 21, MODIS in 8, SRTM/DEM in 3, UAV in 2, and PlanetScope in 1. On the algorithmic side, Random Forest appears in all 107 documents (by construction), support vector machines in 33, CART in 17, gradient boosting and XGBoost in 14, deep learning architectures in 9, artificial neural networks in 6, and SHAP-based explainability in 5. Comparative benchmarking against RF is therefore common, but genuine algorithmic diversification is not.

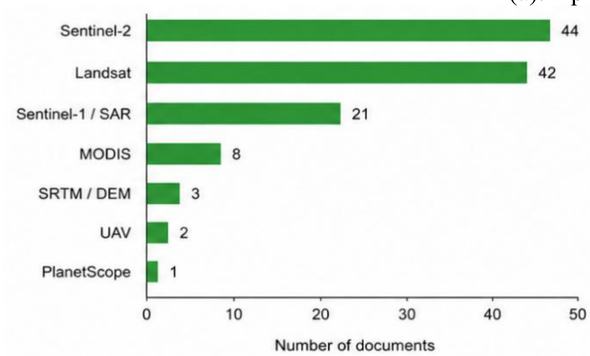


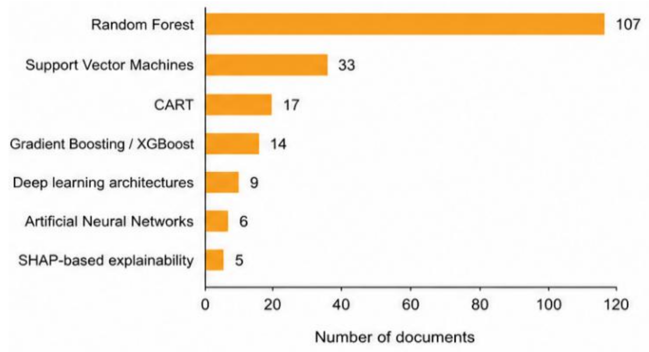
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Figure 19. Thematic map of the RF–GEE research field (Callon centrality × density)



(a). Aplication Domins





(b). Satellite Sensors / Data Sources

(c). Machine Learning Algorithms

3.12 Thematic Evolution

The corpus was partitioned at the end of 2023 into Period 1 (2021–2023, n = 37) and Period 2 (2024–2026, n = 70). Comparing keyword profiles across the cut point reveals both continuity and displacement.

Period 1 (2021–2023) was dominated by GOOGLE EARTH ENGINE (27), MACHINE LEARNING (12), RANDOM FOREST (10), REMOTE SENSING (7), SENTINEL-2 (6), SUPPORT VECTOR MACHINE (4), TIME SERIES (3), CLASSIFICATION (2), CLASSIFICATION AND REGRESSION TREES (2), and LULC (2).

Period 2 (2024–2026) is dominated by GOOGLE EARTH ENGINE (39), RANDOM FOREST (21), MACHINE LEARNING (17), REMOTE SENSING (14), SENTINEL-2 (9), LULC (8), SENTINEL (4), SHAP (3), SAR (3), and SENTINEL-1 (3).

Three transitions are evident:

1. Transition 1: from algorithm comparison to algorithm consolidation. In Period 1, SUPPORT VECTOR MACHINE and CLASSIFICATION AND REGRESSION TREES appear as named keywords, indicating an active comparative phase in which RF's superiority was still being established. Both recede in Period 2 while RANDOM FOREST rises from 10 to 21 occurrences. The comparative question has been settled or, more precisely, abandoned.
2. Transition 2: from optical-only to multi-sensor fusion. SENTINEL-1 and SAR are effectively absent from Period 1 and appear together in Period 2 with median publication years of 2025. Radar–optical fusion, which addresses the cloud-cover limitation that constrains optical monitoring in tropical and monsoonal regions, represents a genuine methodological advance rather than a relabelling.
3. Transition 3: from accuracy to interpretability. SHAP appears exclusively in Period 2, with a median publication year of 2026 the newest term in the corpus. ENSEMBLE LEARNING and DEEP LEARNING follow the same trajectory. This signals the beginning of a shift from how accurate is the map to why did the model produce this map, a transition already well advanced in other applied machine learning literatures.

Terms appearing for the first time in Period 2 include SHAP, WETLAND, LANDSAT-8, URBAN HEAT ISLAND, SUSTAINABILITY, GAUSSIAN PROCESS REGRESSION, QINGHAI–TIBET PLATEAU, IMAGE CLASSIFICATION, SOIL MOISTURE, WATER QUALITY, WETLANDS, MAPPING, PHOTOVOLTAIC POWER PLANTS, DEEP LEARNING, and CHANGE DETECTION. The appearance of PHOTOVOLTAIC POWER PLANTS and URBAN HEAT ISLAND marks an extension of the RF–GEE apparatus from natural-resource monitoring into energy infrastructure and urban climate application domains that were absent in Period 1.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Notably, the methodological core does not evolve. GOOGLE EARTH ENGINE and RANDOM FOREST increase in absolute and relative frequency across both periods. The field's evolution is therefore extensive (more applications, more sensors, more geographies) rather than intensive (deeper methodological refinement of the core workflow).

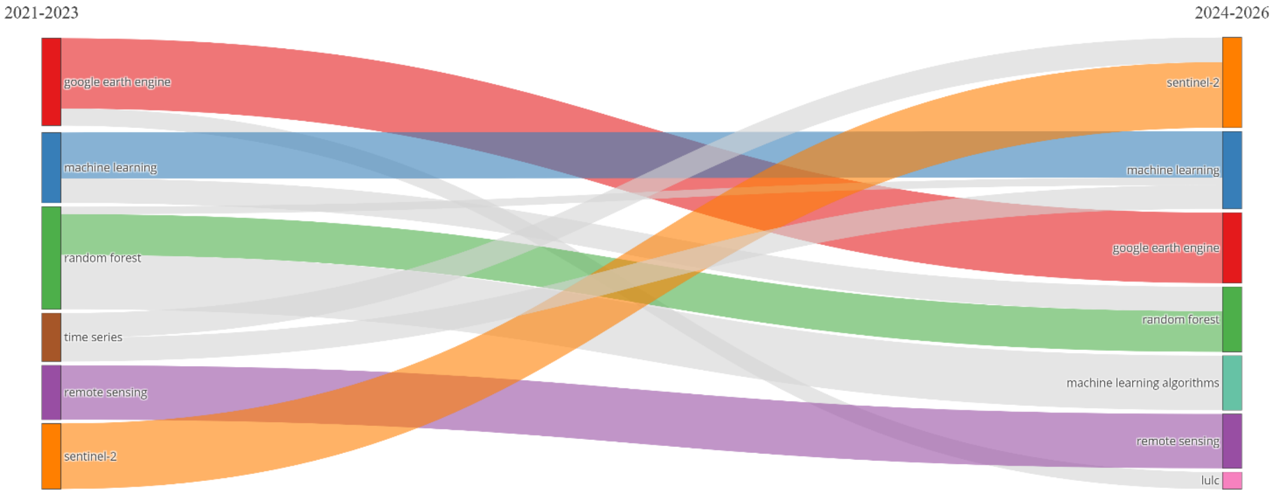


Figure 21. Thematic evolution between 2021-2023 and 2024-2026.

Figure 21 shows the thematic evolution of Google Earth Engine and Random Forest research between 2021–2023 and 2024–2026. The research structure demonstrates strong continuity, with Google Earth Engine, machine learning, Random Forest, and remote sensing remaining prominent across both periods. However, Sentinel-2 becomes more prominent in the later period, while LULC and machine learning algorithms emerge as more specific research orientations. Conversely, time series becomes less prominent. Overall, the pattern indicates that the field is evolving from a primarily platform- and methodology-oriented foundation toward more specific satellite-data applications and land-cover analysis, while retaining Random Forest and machine learning as core methodological foundations.

Table 17. Thematic transitions between periods.

From	To	Words	Weighted Inclusion Index	Inclusion Index	Occurrences	Stability Index	PageRank Index
google earth engine 2021-2023	google earth engine 2024-2026	google earth engine	0.913043478	0.5	21	0.25	0.895853557
sentinel-2 2021-2023	sentinel-2 2024-2026	sentinel-2	1	1	6	0.2	0.682876257
remote sensing 2021-2023	remote sensing 2024-2026	remote sensing	0.777777778	0.5	7	0.142857143	0.616294753
random forest 2021-2023	random forest 2024-2026	random forest	0.476190476	0.333333	10	0.142857143	0.567710532
machine learning 2021-2023	machine learning 2024-2026	machine learning	0.642857143	0.5	9	0.125	0.539501483
random forest 2021-2023	machine learning algorithms 2024-2026	machine learning algorithms	1	1	3	0.2	0.396649722
machine learning 2021-2023	random forest 2024-2026	google earth engine (gee)	0.357142857	0.5	5	0.25	0.266306459
time series 2021-2023	sentinel-2 2024-2026	landsat	0.4	0.5	2	0.166666667	0.23262321
time series 2021-2023	machine learning 2024-2026	time series	0.4	0.5	3	0.125	0.21002268
google earth engine 2021-2023	lulc 2024-2026	classification	0.25	0.5	2	0.333333333	0.185802453

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

random forest 2021-2023	machine learning 2024-2026	vegetation index	0.095238095	0.2	2	0.090909091	0.096150766

3.13 Trend Topics

Trend-topic analysis was conducted to identify the temporal development of frequently occurring terms within the corpus. Each term was positioned according to the median publication year of the documents in which it appeared, while the first and third quartiles were used to represent the temporal interval of its occurrence. Based on the applied frequency threshold, six terms were identified: machine learning algorithms, time series, Google Earth Engine, Random Forest, machine learning, and Sentinel.

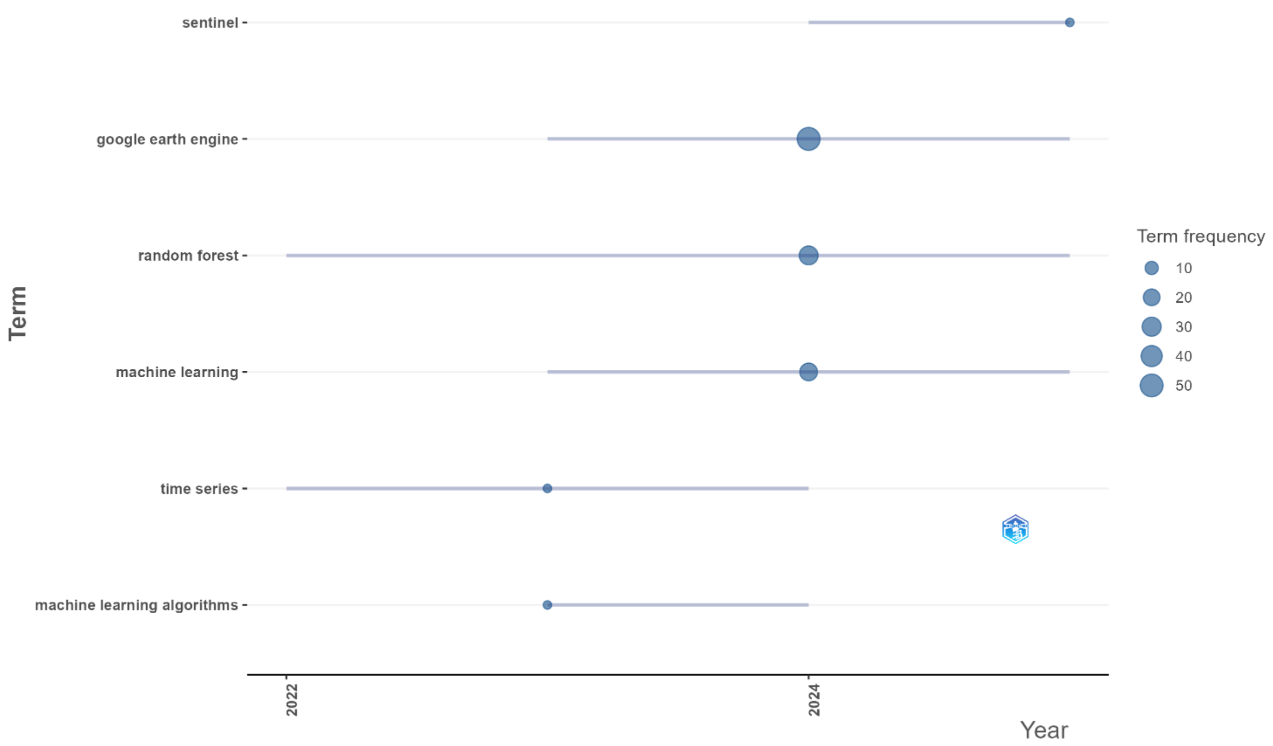


Figure 22. Trend topics based on median publication year and temporal publication interval.

The results reveal a clear distinction between earlier methodological orientations and the established conceptual core of the field. Machine learning algorithms has a frequency of five documents, with a median publication year of 2023 and an interquartile interval of 2023–2024. Similarly, time series appears in five documents, with a median year of 2023 and an interval of 2022–2024. Their relatively early temporal positioning suggests that these terms were more characteristic of the earlier development of the research domain and have become less prominent in the most recent publications.

The dominant methodological vocabulary is represented by Google Earth Engine, Random Forest, and machine learning. Google Earth Engine is the most frequent term, appearing in 51 documents, with a median publication year of 2024 and an interquartile interval of 2023–2025. Random Forest appears in 29 documents, also with a median year of 2024 and an interval of 2022–2025, while machine learning occurs in 24 documents, with a median year of 2024 and an interval of 2023–2025. The broad temporal intervals of these three terms indicate that they are not confined to a particular period but instead constitute persistent methodological components throughout the corpus.

The most recent term identified in the trend analysis is Sentinel, which occurs in five documents and has a median publication year of 2025, with an interquartile interval of 2024–2025. Its later temporal positioning indicates a more recent emphasis on Sentinel-based Earth observation within the research domain. Nevertheless, because the term occurs in only five documents, this pattern should be regarded as an emerging signal rather than evidence of an established long-term trend.

Taken together, the temporal distribution indicates that the research domain is characterised by strong methodological continuity with gradual diversification of data sources. Google Earth Engine, Random Forest, and machine learning remain the central methodological vocabulary, whereas time-series analysis and the broader term machine learning algorithms are positioned earlier in the observed temporal trajectory. The emergence of Sentinel at a median year of 2025 suggests increasing attention to Sentinel-based Earth observation in recent studies.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

The results also demonstrate that temporal prominence should not be confused with research importance. A term with an earlier median year is not necessarily declining scientifically, while a term with a later median year is not necessarily a sustained emerging trend. This distinction is particularly important for Sentinel, given its limited frequency. Furthermore, the 2026 publication window is incomplete because the dataset was retrieved on 6 August 2026. Therefore, the latest temporal patterns should be interpreted cautiously and treated as preliminary evidence of emerging research directions.

Overall, the trend-topic analysis suggests that the field has not experienced a fundamental methodological shift. Instead, recent research appears to build upon an established Google Earth Engine–machine-learning–Random Forest foundation while increasingly incorporating Sentinel-based Earth observation data. This pattern is consistent with the thematic-evolution results in Section 3.12 and supports the broader conclusion that the field is evolving primarily through incremental methodological and data-source diversification rather than replacement of its core analytical framework.

Table 18. Trend Topics by Frequency and Temporal Distribution

Term	Frequency	Year (Q1)	Median Year	Year (Q3)	Interpretation
Machine learning algorithms	5	2023	2023	2024	Earlier methodological orientation
Time series	5	2022	2023	2024	Earlier temporal-analysis orientation
Google Earth Engine	51	2023	2024	2025	Persistent core
Random Forest	29	2022	2024	2025	Persistent core methodology
Machine learning	24	2023	2024	2025	Persistent methodological core
Sentinel	5	2024	2025	2025	Recent/emerging data-source signal

Table 18. Trend topics ranked according to frequency and temporal distribution based on the first quartile, median, and third quartile publication years.

3.14 Three-Field Plot Analysis

The three-field plot links author keywords (DE), countries (AU_CO), and publication sources (SO), providing an integrated view of how conceptual topics are associated with national research production and journal outlets. The structure therefore complements the separate country, keyword co-occurrence, and source analyses by showing how these three dimensions interact within the same publication corpus.

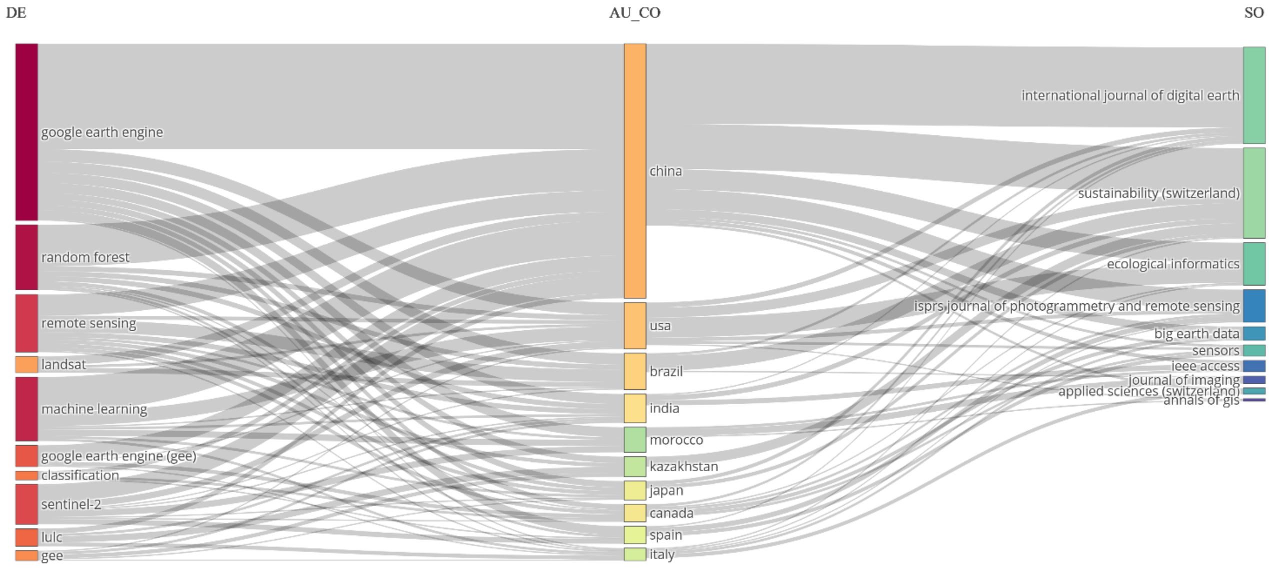


Figure 23. Three-field plot linking author keywords, countries, and publication sources.

The left field represents the principal author keywords, the middle field represents countries, and the right field represents publication sources. The dominant keyword–country relationship is Google Earth Engine → China, followed by Random Forest → China, Google Earth Engine → United States, Machine Learning → China, Machine Learning → United States, Remote Sensing → China, Google Earth Engine → United Kingdom, and Sentinel-2 → China. These flows indicate that China is strongly associated with the major methodological keywords defining the corpus, particularly Google Earth Engine, Random Forest, and machine learning.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

The three-field structure therefore reinforces the strong geographic concentration identified in the country-production analysis. China is not only the most productive country but is also strongly connected to the dominant conceptual and methodological vocabulary. The United States represents the second most visible national contributor and shows notable associations with Google Earth Engine and machine learning. Other countries, including Spain, India, Canada, and Saudi Arabia, appear through smaller flows, indicating broader international participation but with substantially lower volume.

The middle-to-right portion of the diagram illustrates the relationship between countries and publication sources. International Journal of Digital Earth, Sustainability, Ecological Informatics, ISPRS Journal of Photogrammetry and Remote Sensing, Big Earth Data, Sensors, IEEE Access, Journal of Imaging, Applied Sciences, and Annals of GIS constitute the principal outlets represented in the plot. The relatively broad distribution of country-to-source flows indicates that research produced by the major contributing countries is disseminated across several remote-sensing, environmental science, geoinformatics, and multidisciplinary journals.

An important structural characteristic is the presence of a convergent central pathway. The dominant author keywords are strongly connected to China and other major contributing countries, which in turn connect to several leading publication sources. This pattern suggests that the research field is not organised into clearly separated national–thematic–journal communities. Instead, researchers from different countries tend to operate around a relatively common methodological vocabulary centred on Google Earth Engine, Random Forest, machine learning, remote sensing, and satellite-based Earth observation.

The plot also indicates some degree of venue differentiation. Publications associated with Google Earth Engine and Random Forest extend across environmental and remote-sensing journals such as Sustainability, International Journal of Digital Earth, and Ecological Informatics, while keywords related to remote sensing and machine learning also connect with technically oriented outlets such as IEEE Access and ISPRS Journal of Photogrammetry and Remote Sensing. This suggests that similar computational foundations are being adapted to different disciplinary and publication contexts.

However, the three-field plot does not provide sufficient evidence to claim direct keyword-to-journal relationships independently of countries, because the visualisation explicitly represents the sequence keyword → country → source. Therefore, relationships between a keyword and a source should be interpreted as mediated through the country field rather than as direct flows. This distinction is important for avoiding overinterpretation of the diagram.

Overall, the three-field plot demonstrates that the research landscape is characterised by strong conceptual convergence, geographic concentration, and broad journal dissemination. The dominant keywords particularly Google Earth Engine, Random Forest, and machine learning—are strongly associated with the most productive countries, especially China and the United States, while the resulting publications are distributed across multiple international journals. This pattern supports the broader finding that the field is methodologically cohesive despite its geographic and application-level diversity.

Table 19. Principal Flows in the Three-Field Plot

Rank	Flow Type	From	To	Frequency
1	Author keyword → Country	Google Earth Engine	China	25
2	Author keyword → Country	Random Forest	China	11
3	Author keyword → Country	Google Earth Engine	United Kingdom	8
4	Author keyword → Country	Machine Learning	China	8
5	Author keyword → Country	Google Earth Engine	United States	7
6	Author keyword → Country	Remote Sensing	China	6
7	Author keyword → Country	Machine Learning	United States	6
8	Author keyword → Country	Sentinel-2	China	5
9	Author keyword → Country	Google Earth Engine (GEE)	China	5
10	Author keyword → Country	Google Earth Engine	Spain	4
11	Country → Publication Source	China	International Journal of Digital Earth	16
12	Country → Publication Source	China	Sustainability (Switzerland)	14
13	Country → Publication Source	United States	Ecological Informatics	7
14	Country → Publication Source	United States	Sustainability (Switzerland)	5
15	Country → Publication Source	China	Ecological Informatics	5
16	Country → Publication Source	China	ISPRS Journal of Photogrammetry and Remote Sensing	4

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

Rank	Flow Type	From	To	Frequency
17	Country → Publication Source	United States	International Journal of Digital Earth	3
18	Country → Publication Source	China	Big Earth Data	3
19	Country → Publication Source	Brazil	Sustainability (Switzerland)	3
20	Country → Publication Source	Saudi Arabia	Sustainability (Switzerland)	3

Table 19 highlights the principal relationships represented in the three-field plot. The strongest conceptual geographical association is between Google Earth Engine and China, occurring in 25 publications, followed by Random Forest-China (11), Google Earth Engine-United Kingdom (8), and Machine Learning-China (8). Google Earth Engine also shows substantial associations with the United States and Spain, while Sentinel-2 is particularly associated with China. These patterns confirm the strong concentration of the dominant methodological vocabulary within the most productive national research community.

At the country source level, China shows the strongest connection with the International Journal of Digital Earth (16 publications) and Sustainability (Switzerland) (14), while the United States is most strongly connected with Ecological Informatics (7). The distribution demonstrates that the dominant research countries disseminate their work across several international journals rather than being concentrated in a single publication outlet. Overall, the three-field structure reveals a convergent research landscape in which a common methodological vocabulary particularly Google Earth Engine, Random Forest, and machine learning is connected to major national contributors and subsequently disseminated through multiple remote-sensing, environmental, and geoinformatics journals.

3.15 Research Gaps and Future Research Directions

The preceding bibliometric analyses provide a coherent diagnosis of the research landscape surrounding Google Earth Engine (GEE) and Random Forest (RF) applications in remote sensing. The field has achieved substantial methodological and operational maturity, as reflected in the strong dominance of GEE, RF, machine learning, and remote-sensing terminology, the widespread use of freely accessible satellite observations, and the broad application of these workflows across environmental domains. At the same time, the analyses reveal a research landscape characterised by methodological standardisation, social fragmentation, and limited conceptual diversification. The co-occurrence and thematic analyses demonstrate a strong methodological core; co-citation analysis identifies a compact infrastructural and algorithmic foundation; bibliographic coupling indicates substantial similarity among otherwise weakly connected research teams; and thematic-evolution and trend-topic analyses suggest gradual diversification rather than a fundamental methodological transition.

Seven research gaps emerge directly from these findings. Table 20 summarises the corresponding bibliometric evidence, identified gaps, and proposed directions for future research.

Table 20. Research Gaps, Bibliometric Evidence, and Proposed Future Research

No.	Bibliometric Evidence	Identified Research Gap	Proposed Future Research Direction
1	Only 11 documents (10.3%) explicitly address transferability or generalisation.	Limited model transferability. Most studies validate RF models within a single study area, providing limited evidence of performance across regions, seasons, or sensor configurations.	Develop spatial and temporal cross-validation, independent out-of-region testing, cross-sensor validation, and explicit transferability reporting.
2	Only 5 documents (4.7%) explicitly address uncertainty.	Limited uncertainty quantification. Conventional accuracy measures dominate, while spatially explicit prediction uncertainty remains uncommon.	Integrate pixel-level probabilities, uncertainty propagation, conformal prediction, and probabilistic classification into GEE workflows.
3	Only 5 documents (4.7%) mention reproducibility, open-source practices, or code availability.	Underdeveloped reproducibility infrastructure. The portability of GEE workflows has not been fully exploited for reproducible research.	Promote script deposition, versioned training-data assets, documented workflows, and reusable open repositories.
4	RF occurs in 107 documents, compared with SVM (33), CART (17), gradient boosting/XGBoost (14), and deep learning (9).	Strong algorithmic concentration. The corpus is heavily centred on RF, limiting systematic evaluation of alternative approaches.	Benchmark RF against boosting, deep learning, hybrid models, and emerging geospatial foundation

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

No.	Bibliometric Evidence	Identified Research Gap	Proposed Future Research Direction
			models using comparable datasets and validation protocols.
5	No sampling-related keyword reaches the applied frequency threshold of ≥3.	Insufficient attention to training-sample design. Sample size, spatial dependence, class balance, and label quality receive limited explicit attention.	Investigate sampling strategies, sample-size requirements, spatial autocorrelation, class imbalance, active learning, and label noise.
6	The largest co-authorship component contains 35.1% of authors; 93.4% of authors contribute to one document, while 48.4% of document pairs are bibliographically coupled.	Social fragmentation and limited cumulative methodological development. Similar workflows are used by weakly connected research teams.	Establish benchmark datasets, shared validation sites, multi-institutional studies, and community-developed methodological standards.
7	The conceptual structure is dominated by GEE, RF, machine learning, and remote sensing, rather than explicit environmental or societal problem concepts.	Predominantly technique-driven research. Methodological implementation often receives greater emphasis than the decision problem being addressed.	Connect mapping outputs to land management, biodiversity, food security, climate adaptation, ecosystem services, environmental policy, and other decision contexts.

The first gap concerns model transferability. Only 11 documents (10.3%) explicitly address transferability or generalisation, indicating that most RF–GEE studies remain geographically and contextually bounded. A model that achieves high accuracy within one study area may not necessarily maintain its performance when applied to another region with different environmental conditions, class distributions, seasonal characteristics, or sensor configurations. This limitation is particularly significant because scalability is one of the principal advantages associated with cloud-based Earth observation. Future studies should therefore move beyond within-site validation by incorporating spatial and temporal cross-validation, independent out-of-region test sets, and cross-sensor validation. Explicit reporting of transferability should become an important criterion for studies claiming scalable classification frameworks.

The second gap concerns uncertainty quantification. Only five documents (4.7%) explicitly address uncertainty, demonstrating a strong emphasis on deterministic classification outputs and conventional accuracy indicators. While overall accuracy, F1-score, precision, recall, and related measures remain useful, they do not indicate where predictions are reliable or uncertain. This is particularly important when remote-sensing products are used for environmental monitoring and policy-oriented applications. Future research should therefore incorporate pixel-level prediction probabilities, class-specific confidence estimates, uncertainty propagation, and probabilistic approaches. Such methods would enable researchers to distinguish areas of high-confidence prediction from locations where additional observations or validation are required.

The third gap is reproducibility. Only five documents (4.7%) explicitly mention reproducibility, open-source practices, or code availability. This is notable because GEE provides an infrastructure in which computational workflows can be relatively easily documented and shared. The limited attention to reproducibility therefore represents an opportunity for methodological improvement. Future research should provide versioned GEE scripts, persistent training-data identifiers, detailed preprocessing descriptions, and reusable workflow repositories. Journals and research communities could further support this development by encouraging the deposition of executable scripts and associated metadata alongside publications.

The fourth gap is the strong algorithmic concentration around Random Forest. RF occurs in all 107 documents because it defines the methodological scope of the corpus, while alternative methods are substantially less frequent. SVM occurs in 33 documents, CART in 17, gradient boosting/XGBoost in 14, and deep-learning approaches in only nine. This distribution indicates that the field has developed a highly stable methodological core but has not yet achieved comparable diversification in algorithmic approaches. The appropriate response is not to assume that more complex algorithms are automatically superior. Instead, future studies should undertake controlled benchmarking of RF against alternative ensemble methods, deep-learning architectures, hybrid approaches, and emerging geospatial foundation models using matched datasets, consistent sampling designs, and equivalent validation protocols. Such studies could clarify the specific conditions under which RF provides advantages in accuracy, robustness, interpretability, and computational efficiency.

The fifth gap concerns training-sample design. No keyword associated with sampling design, training-set size, class balance, or label quality reaches the applied frequency threshold of three. This absence suggests that sampling methodology receives considerably less explicit attention than algorithm selection. Yet supervised classification performance is strongly influenced by the representativeness and spatial structure of training data. Future research should therefore investigate sample-size requirements, class imbalance, spatial autocorrelation, active-learning strategies, sampling design, and label noise. Particular attention should be given to spatially independent validation because random splitting of spatially correlated observations can result in overly optimistic estimates of model performance.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

The sixth gap concerns social fragmentation and limited cumulative methodological development. The co-authorship analysis shows that only 35.1% of authors belong to the largest connected component, while 93.4% of authors contribute to only one document. At the same time, the bibliographic coupling analysis shows that 48.4% of possible document pairs are connected through shared references. This combination indicates that researchers who rarely collaborate directly may nevertheless implement highly similar methodological workflows. Methodological convergence has therefore developed largely through shared platforms, algorithms, and references rather than through sustained collaboration. Future research could address this fragmentation through shared benchmark datasets, common validation sites, multi-institutional comparison studies, and deliberately developed methodological standards.

The seventh gap is the predominantly technique-driven character of the literature. The conceptual structure is strongly centred on GEE, RF, machine learning, remote sensing, and satellite-data terminology. In contrast, explicit environmental and societal problem concepts do not occupy a comparable position in the dominant conceptual structure. This suggests that a considerable proportion of the literature is organised around the implementation of a technique rather than around a clearly articulated decision problem. Future research should therefore strengthen the connection between remote-sensing products and concrete environmental or societal outcomes. Applications related to biodiversity conservation, food security, carbon management, climate adaptation, ecosystem services, land management, and environmental policy provide opportunities to shift evaluation from classification performance alone toward decision relevance and practical impact.

3.15.1 Emerging Research Directions

Beyond the seven gaps directly supported by the bibliometric evidence, the temporal and thematic analyses suggest several emerging research directions. These should be regarded as prospective opportunities rather than established trends in the corpus.

Explainable machine learning represents one such direction. Future research can move beyond conventional feature-importance measures toward spatially explicit explanations that identify where and why specific environmental variables influence model predictions. Such approaches could improve scientific interpretation and user confidence in machine-learning-derived environmental products.

Multi-sensor integration represents another important direction. The increasing prominence of Sentinel-related research provides a basis for integrating complementary optical and radar observations. In particular, combining Sentinel-1 SAR with Sentinel-2 optical data could improve monitoring under cloud-prone conditions and provide complementary information on surface structure, moisture, and vegetation characteristics.

A third direction is the development of hybrid machine-learning architectures. Rather than replacing RF with increasingly complex models, future studies could investigate architectures that combine RF's robustness and relatively low data requirements with advanced feature-extraction or ensemble techniques. Such approaches should be evaluated using transparent benchmarks that consider not only predictive accuracy but also transferability, computational cost, interpretability, and reproducibility.

3.15.2 Overall Research Agenda

Taken together, the identified gaps indicate that the next stage of RF–GEE research should move beyond routine application toward greater methodological robustness, transferability, reproducibility, and decision relevance. The field has already demonstrated substantial operational maturity through the widespread adoption of GEE, RF, and freely accessible Earth-observation data across diverse environmental applications. The principal challenge is therefore no longer simply to demonstrate that RF can be implemented in GEE, but to determine when, where, why, and under what conditions RF–GEE workflows remain reliable and transferable.

Future research should consequently prioritise cross-region validation, uncertainty quantification, reproducible workflows, systematic algorithm benchmarking, rigorous sampling design, collaborative benchmarking, and stronger problem-oriented environmental framing. Addressing these priorities would enable the field to progress from a highly standardised application-oriented literature toward a more cumulative, transparent, and scientifically mature research domain.

4. CONCLUSION

This study presented the first bibliometric analysis of the intersection between Random Forest and Google Earth Engine in remote sensing research, examining 107 Scopus-indexed open-access articles published between 2021 and 2026. Answering RQ1, the field is in sustained expansion, with a compound annual growth rate of 51.97% between 2021 and 2025, 1,685 total citations, a corpus h-index of 24, and a mean of 15.75 citations per document. Growth accelerated sharply in 2025 and appears to be reaching a plateau. Answering RQ2, output is highly concentrated by venue and by country but highly dispersed by author and institution. Three journals account for 59.8% of documents, and Bradford's Zone 1 contains a single title. China contributes 48 documents (44.9%), the United States 20. Yet 93.4% of the 527 authors published a single document, and no institution exceeds 6. The field has strong venue and national centres of gravity but no author or institutional core. Answering RQ3, the intellectual foundation is compact and infrastructural. Of 6,156 canonical references, only 40 are co-cited five or more times. Gorelick et al. (2017), Breiman (2001), and Belgiu

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

and Drăguţ (2016) form a canonical triad of platform, algorithm, and legitimating review. The core contains no theoretical works and exhibits a temporal gap between 2002 and 2015, indicating that the field reaches directly from contemporary tools to foundational primitives without an intermediate cumulative literature of its own. Answering RQ4, the field is conceptually unimodal. Keyword co-occurrence, index-keyword clustering, and thematic mapping all resolve to a single cluster centred on GOOGLE EARTH ENGINE; bibliographic coupling links 48.4% of all possible document pairs. Beneath this uniform methodological surface, content classification identifies eight application domains led by forest and vegetation (70.1%), agriculture (46.7%), and LULC (43.0%). Thematic evolution shows extensive rather than intensive development: more applications, sensors, and geographies, with an unchanging methodological core. The emerging front comprises SAR–optical fusion, ensemble learning, and explainable AI, with SHAP the newest term at median year 2026. Answering RQ5, seven gaps were identified: untested transferability (10.3% of documents), unquantified uncertainty (4.7%), absent reproducibility infrastructure (4.7%), algorithmic monoculture, undertheorised sampling design, social fragmentation impeding cumulative progress, and application without problem framing. The central contribution of this study is the identification of a specific structural configuration: a field that is socially fragmented yet methodologically unified. Teams that never collaborate only 35.1% of authors occupy the largest co-authorship component nonetheless produce near-identical workflows, because standardisation has been imposed by the platform rather than negotiated by the community. This configuration accounts simultaneously for the field's greatest achievement, the democratisation of large-area Earth observation, and its principal risk, the uncritical inheritance of a workflow whose limitations have not been collectively examined. Several limitations should be noted. The corpus is restricted to Scopus-indexed, English-language, open-access journal articles, excluding conference proceedings, non-indexed regional journals, and grey literature; the resulting picture is therefore of the formally indexed core rather than of the full literature. Author-level results are subject to residual homonym conflation, particularly for common East Asian surname–initial combinations. Coverage of 2026 is partial, extending only to 6 August, which compresses the most recent cohort. Citation-based indicators systematically disadvantage recent documents. Finally, keyword-based conceptual analysis depends on author keywording practice, which is uneven; the content-based domain classification introduced in Section 3.11 partially mitigates but does not eliminate this dependence. Future bibliometric work should extend the analysis to a merged Scopus–Web of Science corpus, incorporate full-text topic modelling to recover the thematic structure that keyword analysis cannot resolve in a hub-dominated network, and repeat the analysis after a further observation window to test whether the emerging explainability and sensor-fusion fronts consolidate. For the substantive field, the priority is clear: RF–GEE research has demonstrated that it can map almost anything, almost anywhere. The next phase must demonstrate that those maps are transferable, their uncertainties are known, their production is reproducible, and their purposes are explicit.

ACKNOWLEDGMENTS

The authors sincerely thank the researchers whose publications and scientific contributions form the knowledge base analysed in this study. We also acknowledge the developers and communities supporting Scopus, bibliometrix, Biblioshiny, VOSviewer, R, Python, and related analytical tools. AI-assisted tools were used for language refinement, structural organisation, and research-support tasks. All scientific analyses, interpretations, methodological decisions, and final statements were independently reviewed and validated by the authors, who remain fully responsible for the content and integrity of this manuscript.

REFERENCES

[1]. N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sensing of Environment, vol. 202, pp. 18–27, 2017. https://doi.org/10.1016/j.rse.2017.06.031

[2]. M. A. Wulder, J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock, "Opening the archive: How free data has enabled the science and monitoring promise of Landsat," Remote Sensing of Environment, vol. 122, pp. 2–10, 2012. https://doi.org/10.1016/j.rse.2012.01.010

[3]. H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, "Google Earth Engine for geo-big data applications: A meta-analysis and systematic review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 164, pp. 152–170, 2020. https://doi.org/10.1016/j.isprsjprs.2020.04.001

[4]. M. Amani et al., "Google Earth Engine cloud computing platform for remote sensing big data applications: A comprehensive review," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5326–5350, 2020. https://doi.org/10.1109/JSTARS.2020.3021052

[5]. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[6]. M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, pp. 24–31, 2016. https://doi.org/10.1016/j.isprsjprs.2016.01.011

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

[7]. V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 67, pp. 93–104, 2012. https://doi.org/10.1016/j.isprsjprs.2011.11.002
[8]. I. Zupic and T. Čater, "Bibliometric methods in management and organization," Organizational Research Methods, vol. 18, no. 3, pp. 429–472, 2015. https://doi.org/10.1177/1094428114562629
[9]. N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," Journal of Business Research, vol. 133, pp. 285–296, 2021. https://doi.org/10.1016/j.jbusres.2021.04.070
[10]. J. E. Hirsch, "An index to quantify an individual's scientific research output," Proceedings of the National Academy of Sciences, vol. 102, no. 46, pp. 16569–16572, 2005. https://doi.org/10.1073/pnas.0507655102
[11]. M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, art. n71, 2021. https://doi.org/10.1136/bmj.n71
[12]. S. C. Bradford, "Sources of information on specific subjects," Engineering, vol. 137, pp. 85–86, 1934.
[13]. A. J. Lotka, "The frequency distribution of scientific productivity," Journal of the Washington Academy of Sciences, vol. 16, no. 12, pp. 317–323, 1926. (DOI tidak tersedia / not indexed in Crossref)
[14]. H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," Journal of the American Society for Information Science, vol. 24, no. 4, pp. 265–269, 1973. https://doi.org/10.1002/asi.4630240406
[15]. M. M. Kessler, "Bibliographic coupling between scientific papers," American Documentation, vol. 14, no. 1, pp. 10–25, 1963. https://doi.org/10.1002/asi.5090140103
[16]. M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field," Journal of Informetrics, vol. 5, no. 1, pp. 146–166, 2011. https://doi.org/10.1016/j.joi.2010.10.002
[17]. M. Callon, J. P. Courtial, and F. Laville, "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry," Scientometrics, vol. 22, no. 1, pp. 155–205, 1991. https://doi.org/10.1007/BF02019280
[18]. M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," Journal of Informetrics, vol. 11, no. 4, pp. 959–975, 2017. https://doi.org/10.1016/j.joi.2017.08.007
[19]. N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," Scientometrics, vol. 84, no. 2, pp. 523–538, 2010. https://doi.org/10.1007/s11192-009-0146-3
[20]. S. K. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features," International Journal of Remote Sensing, vol. 17, no. 7, pp. 1425–1432, 1996. https://doi.org/10.1080/01431169608948714
[21]. C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," Remote Sensing of Environment, vol. 8, no. 2, pp. 127–150, 1979. https://doi.org/10.1016/0034-4257(79)90013-0
[22]. A. R. Huete, "A soil-adjusted vegetation index (SAVI)," Remote Sensing of Environment, vol. 25, no. 3, pp. 295–309, 1988. https://doi.org/10.1016/0034-4257(88)90106-X
[23]. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995. https://doi.org/10.1007/BF00994018
[24]. T. G. Farr et al., "The Shuttle Radar Topography Mission," Reviews of Geophysics, vol. 45, no. 2, art. RG2004, 2007. https://doi.org/10.1029/2005RG000183
[25]. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA, USA: Wadsworth, 1984. https://doi.org/10.1201/9781315139470 (DOI merujuk pada edisi cetak ulang Routledge/CRC 2017; edisi asli Wadsworth 1984 tidak memiliki DOI terdaftar)
[26]. H. Xu, "Modification of Normalised Difference Water Index (NDWI) to enhance open water features in remotely sensed imagery," International Journal of Remote Sensing, vol. 27, no. 14, pp. 3025–3033, 2006. https://doi.org/10.1080/01431160600589179
[27]. A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," Remote Sensing of Environment, vol. 83, no. 1–2, pp. 195–213, 2002. https://doi.org/10.1016/S0034-4257(02)00096-2
[28]. M. C. Hansen et al., "High-resolution global maps of 21st-century forest cover change," Science, vol. 342, no. 6160, pp. 850–853, 2013. https://doi.org/10.1126/science.1244693
[29]. M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," Remote Sensing of Environment, vol. 120, pp. 25–36, 2012. https://doi.org/10.1016/j.rse.2011.11.026
[30]. P. Teluguntla et al., "A 30-m Landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 144, pp. 325–340, 2018. https://doi.org/10.1016/j.isprsjprs.2018.07.017
[31]. A. Tassi and M. Vizzari, "Object-oriented LULC classification in Google Earth Engine combining SNIC, GLCM, and machine learning algorithms," Remote Sensing, vol. 12, no. 22, art. 3776, 2020. https://doi.org/10.3390/rs12223776

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Page: 1932-1971
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1511

[32]. R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," Remote Sensing of Environment, vol. 37, no. 1, pp. 35–46, 1991. https://doi.org/10.1016/0034-4257(91)90048-B
[33]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. (DOI tidak tersedia / not indexed in Crossref — JMLR tidak mendaftarkan DOI untuk artikel volume ini; tautan resmi: https://jmlr.org/papers/v12/pedregosa11a.html)
[34]. K. N. Loukika, V. R. Keesara, and V. Sridhar, "Analysis of land use and land cover using machine learning algorithms on Google Earth Engine for Munneru River Basin, India," Sustainability, vol. 13, no. 24, art. 13758, 2021. https://doi.org/10.3390/su132413758
[35]. C. Zhang et al., "Phenology-assisted supervised paddy rice mapping with the Landsat imagery on Google Earth Engine: Experiments in Heilongjiang Province of China from 1990 to 2020," Computers and Electronics in Agriculture, vol. 212, art. 108105, 2023. https://doi.org/10.1016/j.compag.2023.108105
[36]. H. Ouchra, A. Belangour, and A. Erraissi, "Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study," IEEE Access, vol. 11, pp. 71127–71142, 2023. https://doi.org/10.1109/ACCESS.2023.3293828
[37]. M. Ashfaq, I. Khan, and A. Alzahrani, "Accurate wheat yield prediction using machine learning and climate-NDVI data fusion," IEEE Access, vol. 12, pp. 40947–40961, 2024. https://doi.org/10.1109/ACCESS.2024.3376735
[38]. X. Yan, J. Li, and A. R. Smith, "Evaluation of machine learning methods and multi-source remote sensing data combinations to construct forest above-ground biomass models," International Journal of Digital Earth, vol. 16, no. 1, pp. 4471–4499, 2023. https://doi.org/10.1080/17538947.2023.2270459
[39]. M. Ganjirad and H. Bagheri, "Google Earth Engine-based mapping of land use and land cover for weather forecast models using Landsat 8 imagery," Ecological Informatics, vol. 80, art. 102498, 2024. https://doi.org/10.1016/j.ecoinf.2024.102498
[40]. M. Mamun, M. Hasan, and K.-G. An, "Advancing reservoirs water quality parameters estimation using Sentinel-2 and Landsat-8 satellite data with machine learning approaches," Ecological Informatics, vol. 81, art. 102608, 2024. https://doi.org/10.1016/j.ecoinf.2024.102608
[41]. C. Zhao, C.-Z. Qin, and Z. Wang, "Decision surface optimization in mapping exotic mangrove species (Sonneratia apetala) across latitudinal coastal areas of China," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 193, pp. 269–283, 2022. https://doi.org/10.1016/j.isprsjprs.2022.09.011
[42]. A. Bastarrika, A. Rodriguez-Montellano, and E. Roteta, "An automatic procedure for mapping burned areas globally using Sentinel-2 and VIIRS/MODIS active fires in Google Earth Engine," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 218, pp. 232–245, 2024. https://doi.org/10.1016/j.isprsjprs.2024.08.019
[43]. Z. Ye, Z. Sheng, and X. Liu, "Using machine learning algorithms based on GF-6 and Google Earth Engine to predict and map the spatial distribution of soil organic matter content," Sustainability, vol. 13, no. 24, art. 14055, 2021. https://doi.org/10.3390/su132414055
[44]. T. K. Oo, N. Arunrat, and S. Sereenonchai, "Comparing four machine learning algorithms for land cover classification in gold mining: A case study of Kyaukpahto gold mine, Northern Myanmar," Sustainability, vol. 14, no. 17, art. 10754, 2022. https://doi.org/10.3390/su141710754
[45]. H. A. Zurqani, "A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google Earth Engine platform," Ecological Informatics, vol. 86, art. 103052, 2025. https://doi.org/10.1016/j.ecoinf.2025.103052
[46]. A. O. Amoakoh, P. Aplin, and K. T. Awuah, "Testing the contribution of multi-source remote sensing features for random forest classification of the Greater Amanzule Tropical Peatland," Sensors, vol. 21, no. 10, art. 3399, 2021. https://doi.org/10.3390/s21103399