From neural networks to deep learning: Tracing the AI evolution in plant ecology

Akkacha Mehalhal[1,*], Mustapha Ainad Tabet [1]

[1]Laboratory of Ecology and Management of Natural Ecosystems, Faculty of Natural and Life Sciences and Earth and Universe Sciences, University of Tlemcen, P.O. Box 119, 13000, Tlemcen, Algeria.

[*]Corresponding author e-mail: mehaoka@gmail.com

Received: 28 September 2025 / Accepted: 19 March 2026

Abstract. The integration of artificial intelligence (AI) with plant ecology provides a powerful supplementary toolkit for analyzing complex environmental data, yet the structure of this rapidly growing field remains unmapped. This study employs a comprehensive bibliometric analysis of 1055 Scopus documents (1968-2025) to chart its conceptual, intellectual, and social landscape. Our analysis reveals a distinct three-phase evolution from foundational Artificial Neural Networks (ANNs) to a broad machine learning expansion, and a current surge in deep learning for applications like precision agriculture. The field's conceptual core is built on predictive modeling, particularly for species distribution, and advanced image analytics. We highlight that this evolution has been largely driven by the increasing availability of open-source programming libraries (e.g., in R and Python) and accessible software platforms, rather than just the theoretical advancement of AI itself. The global collaboration network is multi-polar, with the USA and China acting as prominent hubs. This research maps the trajectory of AI applications in ecology and emphasizes the emerging frontier of Explainable AI (XAI), which is essential for moving beyond "black-box" predictions toward interpretable ecological insights and fundamental scientific discovery.

Keywords: bibliometrics, science mapping, species distribution modeling, random forests, precision agriculture

Abbreviations:
AI: Artificial Intelligence
ANN: Artificial Neural Network
AUC: Area Under the Curve
CNN: Convolutional Neural Network
CSV: Comma-Separated Values
DL: Deep Learning
GAM: Generalized Additive Model
GLM: Generalized Linear Model

IPBES: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
ML: Machine Learning
RF: Random Forest
SDM: Species Distribution Modeling
SVM: Support Vector Machine
TSS: True Skill Statistic
XAI: Explainable Artificial Intelligence.

1. Introduction

The planet is at a critical juncture, facing unprecedented challenges of biodiversity loss and climate change, with plant ecosystems at the epicenter of these crises (IPBES, 2019). Plant ecology, the discipline dedicated to understanding the distribution, abundance, and interactions of plants and their environments, plays a vital role in informing strategies for a sustainable future. Ecological systems are characterized by intricate patterns across multiple scales (Levin, 1992). While ecology has a long history of being a quantitative, data-driven science exemplified by large-scale initiatives like the International Biological Program (IBP) long before the ubiquitous use of modern computing the modern deluge of multi-source data from remote sensing, genomic sequencing, and sensor networks has significantly pushed the limits of traditional analytical methods (Reichstein et al., 2019). To navigate this complexity, ecologists are increasingly adopting tools associated with the "Fourth Paradigm" of scientific discovery (Hey et al., 2009).

This shift has been significantly catalyzed not merely by the theoretical rise of artificial intelligence (AI) and machine learning (ML) (Jordan & Mitchell, 2015), but fundamentally by the increasing availability and accessibility of open-source programming libraries (e.g., in R and Python), user-friendly GIS software (e.g., QGIS), specific modeling packages (e.g., MaxEnt), and cloud-based computing platforms like Google Earth Engine. These accessible tools offer ecologists a powerful toolkit to uncover complex, non-linear patterns in vast datasets and build highly accurate predictive models (Thessen, 2016). From forecasting species distribution under future climate scenarios to automating plant identification from images, the application of these computational tools has unlocked new frontiers of research.

While the increasing application of these software-driven AI/ML tools in plant ecology has been highly productive, it has resulted in a vast, rapidly expanding, and multidisciplinary body of literature. (Jones et al., 2006). For researchers, especially those new to the field, navigating this extensive literature can be daunting. Key questions remain: What are the foundational concepts and dominant research themes? Who are the leading researchers and institutions shaping its trajectory? How have the core ideas and software tools evolved over time?

To address this knowledge gap, bibliometric analysis offers a rigorous and objective methodology for mapping the structure and dynamics of a scientific field (Zupic & Čater, 2015). By quantitatively analyzing the patterns within a large body of literature, it is possible to identify the intellectual and social structures that define a research domain (Donthu et al., 2021). Therefore, the objective of this study is to conduct a comprehensive bibliometric analysis of the global research on artificial intelligence in plant ecology. We aim to map its conceptual, intellectual, and social structures, identify its historical evolution and key contributors, and outline its emerging frontiers to provide a definitive guide for researchers navigating this vibrant and critically important field.

2. Material and Methods

To ensure transparency and reproducibility, the research methodology followed a structured, multi-step workflow. As illustrated in the roadmap (Figure 1), the process was divided into four key phases: (1) Data Identification, involving a comprehensive search strategy in the Scopus database; (2) Screening and Selection, where strict inclusion/exclusion criteria were applied to refine the corpus; (3) Bibliometric Analysis, utilizing the *bibliometrix* R-package for performance analysis; and (4) Science Mapping, employing VOSviewer and R to visualize conceptual and social networks.\

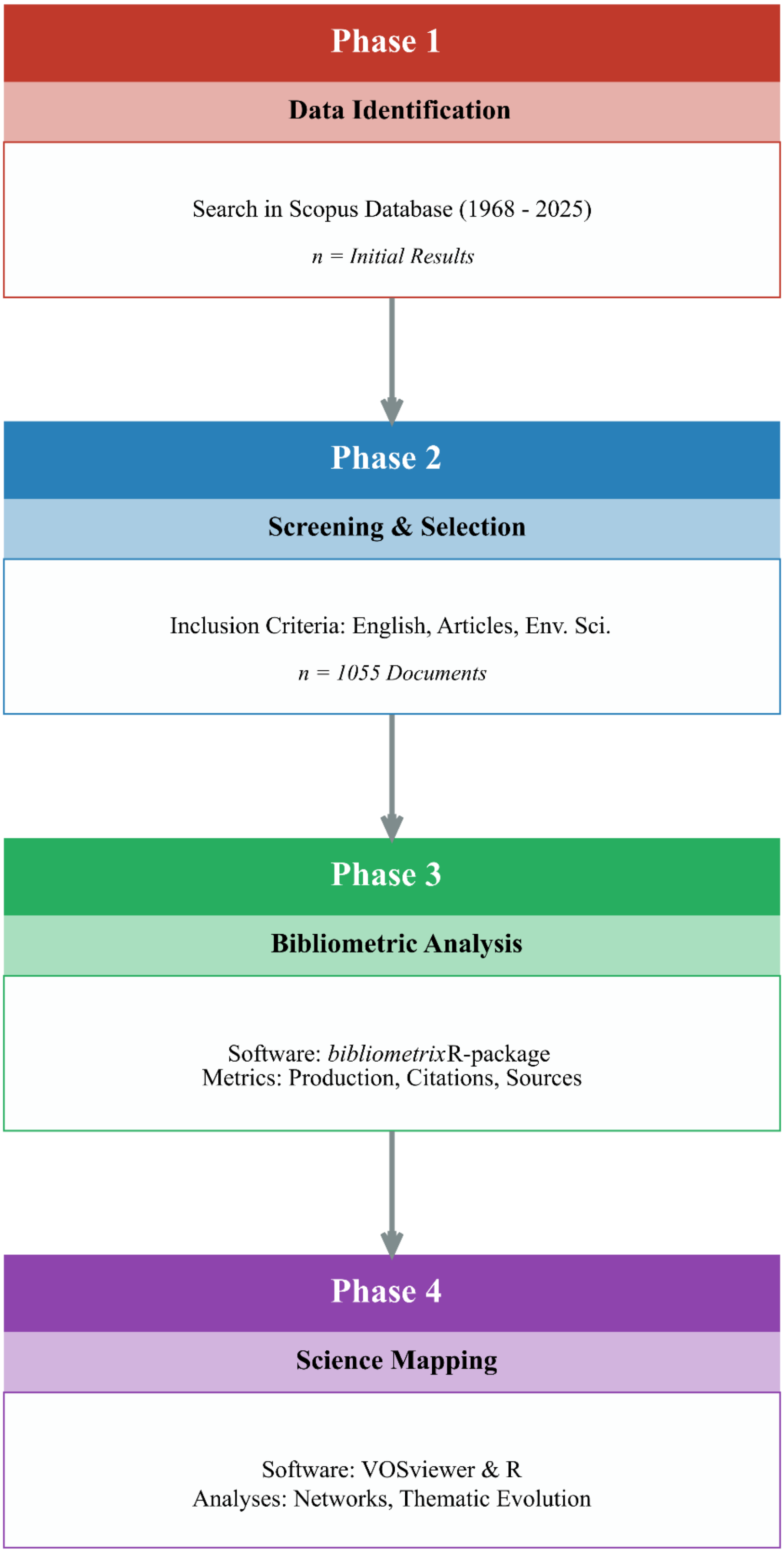

Figure 1. Roadmap of the bibliometric methodology illustrating the four-step workflow: Data Identification, Screening and Selection, Bibliometric Analysis, and Science Mapping.

2.1 Data Collection and Search Strategy

To construct a comprehensive bibliographic dataset, the Scopus database was selected as the sole data source due to its extensive coverage of peer-reviewed scientific literature. All data were retrieved on a single day, June 24, 2025, to ensure the consistency of the dataset and establish a fixed point for the analysis. The search query was meticulously designed to be exhaustive, incorporating a wide range of synonyms and common abbreviations to retrieve all relevant documents addressing the application of artificial intelligence (AI) and its subfields within the context of plant ecology. The final query, executed within the title, abstract, and keyword fields (TITLE-ABS-KEY), was as follows:

(("artificial intelligence" OR ai OR "machine learning" OR ml OR "deep learning" OR dl OR "neural networks" OR "NN" OR "AI-based models" OR "intelligent systems" OR "data-driven models" OR "computational intelligence") AND ("plant ecology" OR "vegetation ecology" OR "plant-environment interactions" OR "plant biodiversity" OR "plant distribution" OR "ecological modeling" OR "species distribution modeling" OR "vegetation analysis" OR "functional ecology" OR "remote sensing vegetation"))

The search included all publication years available in the database, resulting in a dataset spanning from the earliest document found in 1968 to those published up to our search date in mid-2025. The results were subsequently filtered using Scopus limiters to adhere to the following inclusion criteria: (a) documents published only in the English language; (b) document types limited to articles (ar), conference papers (cp), reviews (re), or books (bk); (c) documents in their "final" publication stage; and (d) documents from the subject areas of Environmental Science (ENVI), Agricultural and Biological Sciences (AGRI), or Earth and Planetary Sciences (EART). The initial exported dataset was further refined to remove duplicates and irrelevant entries, resulting in a final corpus of 1055 documents for analysis. All bibliographic data, including abstracts and cited references, were exported in Comma-Separated Values (CSV) format.

2.2 Data Analysis

All bibliometric analyses were performed using the bibliometrix R-package (Aria & Cuccurullo, 2017), an R-tool designed for comprehensive science mapping analysis. The analysis was structured into two primary stages: performance analysis and science mapping.

Science mapping was employed to visualize the intellectual structure and thematic evolution of the research field. This involved co-occurrence analysis of Author's

Keywords to identify key research themes and a thematic evolution analysis to track their development over time. To ensure the clarity of the thematic evolution diagram (Figure 3), synonymous keywords (e.g., "SDM" and "Species Distribution Modeling") were merged using a thesaurus file to avoid redundancy. The underlying network data for co-occurrence and collaboration analyses were generated using VOSviewer software (van Eck & Waltman, 2010). However, the final network figures presented in this paper were rendered using the R environment for statistical computing (R Core Team, 2023) and its associated graphical packages, specifically ggplot2 (Wickham, 2016) and igraph (Csardi & Nepusz, 2006), to achieve the highest visual quality for publication.

The specific metrics analyzed include: annual scientific production; the most productive and impactful sources, countries, institutions, and authors; and the intellectual structure as revealed by keyword co-occurrence, thematic evolution across the three identified time periods (1968-1997, 1998-2017, 2018-present), and collaboration networks.

3. Results

3.1 Publication Trends and Historical Evolution

The final dataset comprised 1055 documents published between 1968 and mid-2025. This body of work was authored by 4244 researchers, indicating a broad and active scientific community. The analysis revealed a highly collaborative research culture, with an average of nearly five co-authors per document, while single-authored papers constituted only 5.3% of the collection. The dataset demonstrated a significant scientific impact, averaging 79 citations per document. The annual growth rate of publications was approximately 8%, signifying a sustained and increasing scholarly interest in the topic.

The temporal distribution of publications reveals three distinct evolutionary phases in the history of this research field (Figure 2). The first phase, the foundational period (1968–1997), was characterized by sporadic and very limited publication activity, with typically fewer than five articles published per year. The second phase (1998–2017) marked a period of significant growth, although inconsistent, with annual publications generally ranging between 10 and 40 articles. The final and most recent phase (2018–present) shows a sharp, exponential increase in research output. Publications in this period consistently exceeded 40 articles per year, peaking at 117 articles in 2024. The apparent decline in 2025 is an artifact of the data collection timing, as the search was conducted mid-year.

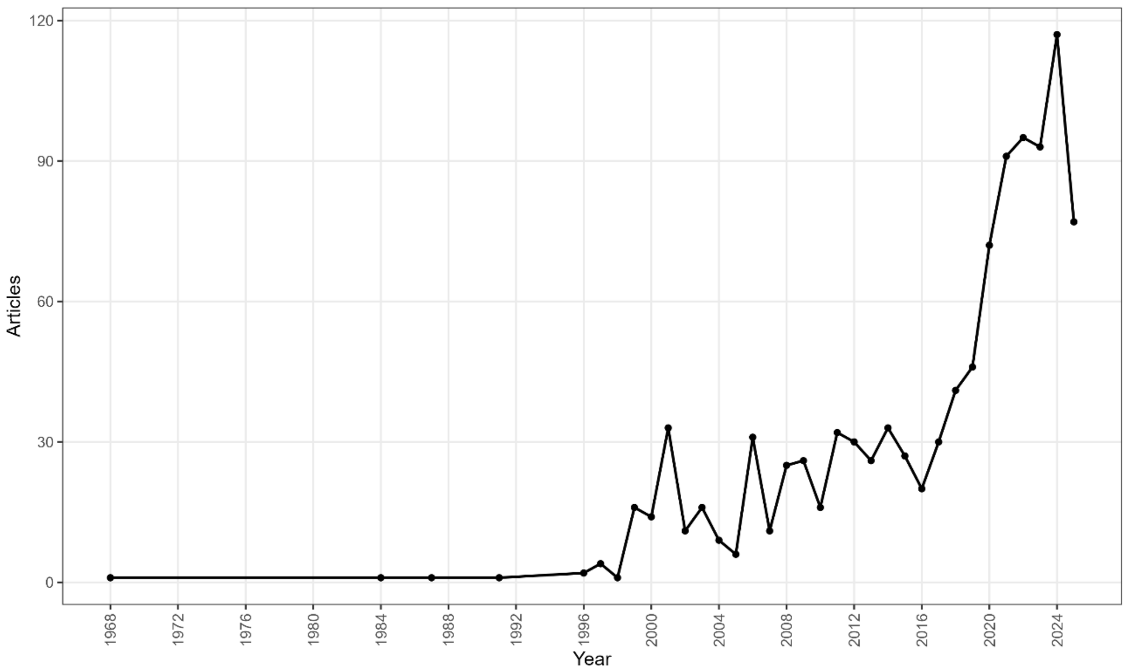

Figure 2. Annual scientific production of articles from 1968 to 2025. The data illustrates three distinct growth periods in the research field.

3.2 Thematic Evolution

To understand the dynamic shifts in research focus over the 57-year period, a thematic evolution analysis was conducted based on author keywords. The dataset was divided into three distinct chronological periods, as identified in the annual production analysis: the Foundational Period (1968-2010), the Machine Learning Expansion (2011-2018), and the Deep Learning Surge (2019-2025). Figure 3 illustrates the evolution of research themes across these periods, with the flow lines indicating the lineage and strength of connection between topics over time.

The first period (1968-2010) was dominated by the emergence of "Artificial Neural Network" and its application in "Ecological Modelling" and "Plant Ecology". These themes represented the core focus of the era, alongside "Species Distribution Modelling". The second period (2011-2018) shows a shift where "Machine Learning" emerged as a dominant theme, alongside the rise of specific algorithms like "Random Forest" and the continued relevance of "Neural Network". In the third period (2019-2025), "Deep Learning" appears as a major and rapidly developing theme. "Machine Learning" and "Random Forest" remain central, but they now connect to specialized applications such as "Precision Agriculture" and "Ecological Niche Modelling".

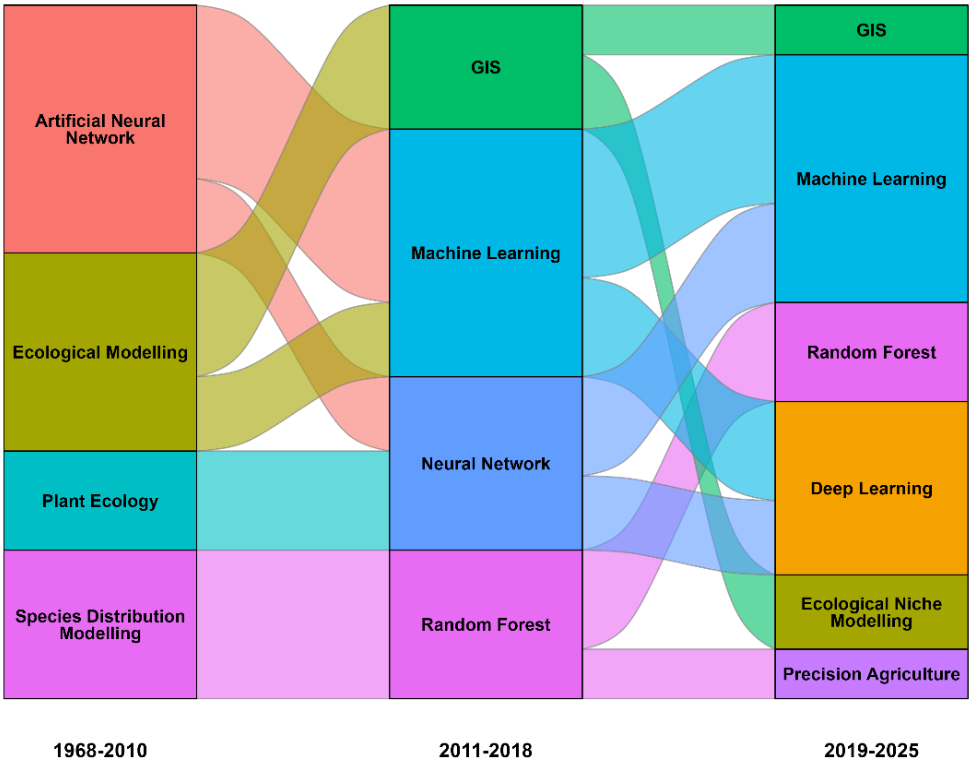

Figure 3. Thematic evolution of author keywords across three periods (1968-2010, 2011-2018, 2019-2025). The diagram illustrates the flow and evolution of research themes, highlighting the progression from foundational concepts to advanced machine learning and deep learning applications.

3.3 Performance Analysis and Leading Contributors

To further delineate the structure of the research landscape, a performance analysis was conducted to identify the most productive and influential sources, countries, and authors. The 1055 documents in the collection were published across 331 distinct sources, indicating a broad and interdisciplinary interest in the topic. The analysis of publication sources reveals that the most impactful articles are concentrated in journals specializing in ecological modeling, species distribution, quantitative ecology, and conservation biology. Table 1 lists the top 10 most frequently cited journals in the collection, highlighting the core publication venues that have shaped this research field.

Table 1. The top 10 most influential journals, ranked by their total number of publications or citations within the dataset.

Rank	Source	Number of Publications
1	Ecological Modelling	173
2	Ecological Informatics	89
3	Ecological Indicators	33
4	Science of The Total Environment	23
5	Remote Sensing	21
6	Computers and Electronics in Agriculture	16
7	Ecography	14
8	Environmental Modelling and Software	14
9	Forests	13
10	Methods in Ecology and Evolution	13

The geographical distribution of research, based on the corresponding author's affiliation, reveals that scientific production in this field is highly concentrated in a few key countries. The United States (169 articles) and China (134 articles) are the undisputed leaders in publication volume, establishing them as the dominant global hubs for this research domain. As illustrated in Figure 4, a second tier of highly active countries follows, which includes Germany (55 articles), Australia (43 articles), and France (42 articles).

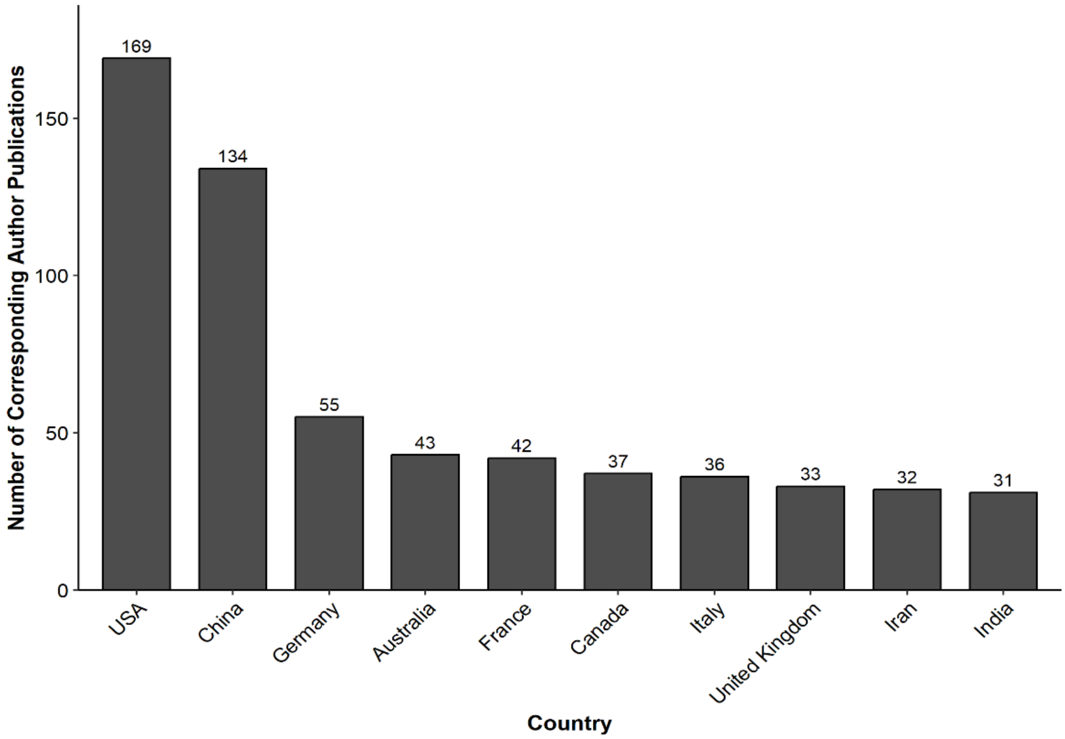

Figure 4. The top 10 most productive countries based on the number of corresponding author publications.

At the institutional level, the analysis pinpoints the specific academic centers driving research production. Ghent University (Belgium) emerges as the clear global leader with an exceptionally high output of 75 articles, significantly outpacing other institutions. A distinct group of highly productive universities follows, including the University of Oxford (UK) and Ocean University of China, each with 30 publications. Figure 5 illustrates the top 10 most productive institutions, revealing a global distribution of expertise with key centers in Europe, China, and the United States.

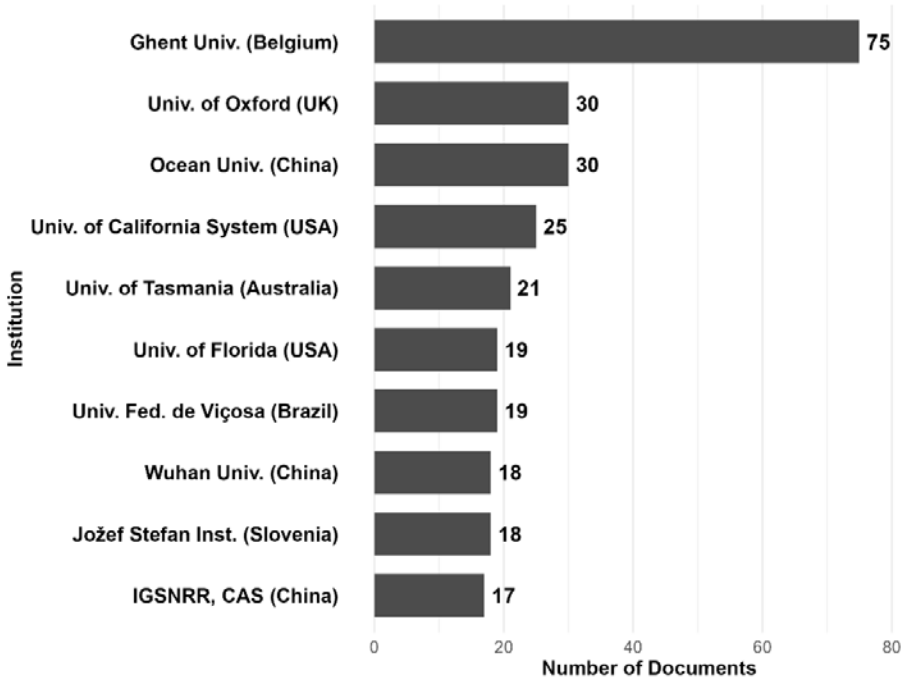

Figure 5. The top 10 most productive institutions, ranked by their total number of documents in the collection.

The analysis of the most productive authors identifies the key researchers driving the field's publication volume. Figure 6 presents the top 10 most prolific authors, ranked by their total number of publications within the collection. Sovan Lek emerges as the most productive author with 17 documents, followed by Sašo Džeroski (13 documents) and Peter L. M. Goethals (10 documents). This ranking provides a quantitative overview of the central contributors to this body of literature based on output volume.

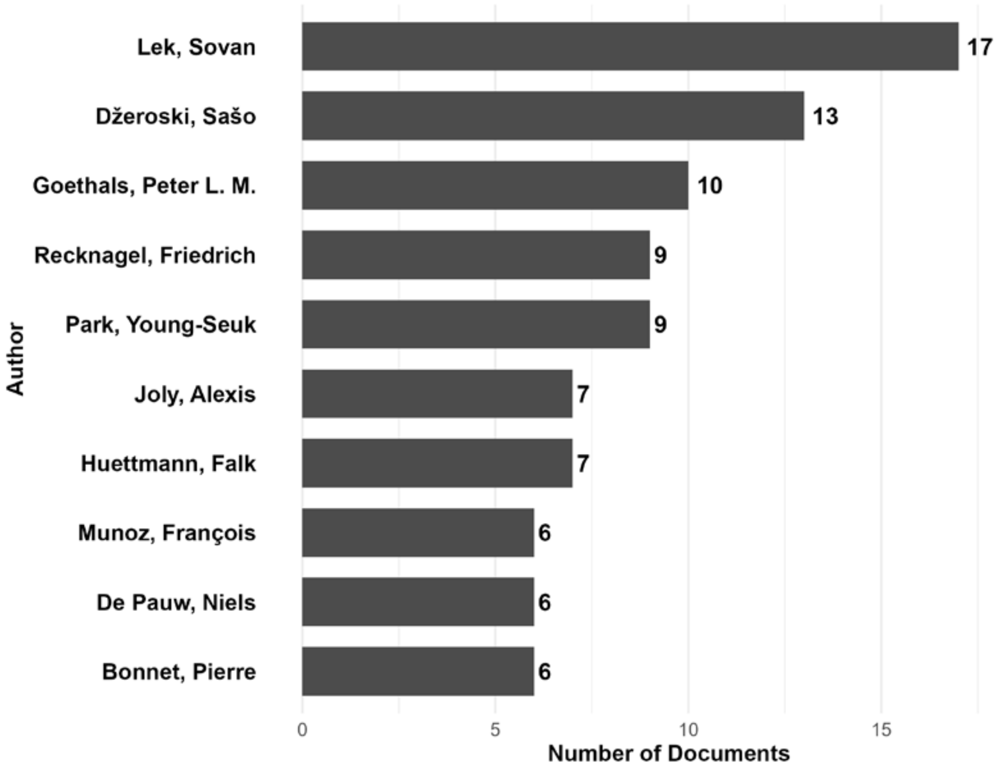

Figure 6. The top 10 most productive authors, ranked by their total number of documents in the collection.

3.4 The Conceptual Structure and Methodological Core

To map the intellectual core of the field, a keyword co-occurrence network was generated, revealing the primary research themes and their interconnections (Figure 7). The analysis identifies a highly structured landscape organized into three major thematic clusters, which represent the dominant schools of thought and application domains. The largest cluster (in orange), situated on the left, represents "Advanced Neural Networks and Image-Based Analytics". This theme is anchored by the term network, referring primarily to neural networks, and is strongly associated with cutting-edge techniques such as convolutional neural network and deep learning. The prominence of terms like image, detection, crop, and agriculture indicates a significant focus on applying these advanced AI methods to image analysis, particularly for applications in precision agriculture, such as crop monitoring and yield prediction. The second major cluster (in teal), occupying the middle and lower-right area, is focused on "Ecological Applications in Species Distribution and Climate Change Modeling". This application-driven theme is centered around core ecological challenges, with climate change, habitat suitability, and species distribution modeling (sdm) as its most prominent concepts. This cluster highlights the extensive use of predictive models, particularly maxent (Maximum Entropy), to forecast the potential distribution of species under various environmental scenarios, including the impact of invasive species. The third cluster (in purple), located in the upper-right, constitutes the"Statistical Modeling and Performance Evaluation Toolkit". This theme comprises the diverse set of statistical and machine learning algorithms that form the methodological backbone for many studies in the field. Key terms include traditional methods like generalized linear model (glm) and generalized additive model (gam), as well as powerful machine learning algorithms such as support vector machine (svm) and rf model (Random Forest). The strong presence of evaluation metrics like auc (Area Under the Curve) and tss (True Skill Statistic) underscores a rigorous focus on assessing model performance and accuracy. Notably, a small but critical group of terms (in pink) acts as a conceptual bridge, linking the major clusters. Terms like vegetation index, remote sensing data, and biodiversity conservation are central to the entire research domain, demonstrating how remote sensing data is a key resource used by advanced AI techniques (the orange cluster) to address large-scale ecological and conservation challenges (the teal cluster). Overall, the network illustrates the structural links between AI methodologies, data sources, and their specific ecological applications.

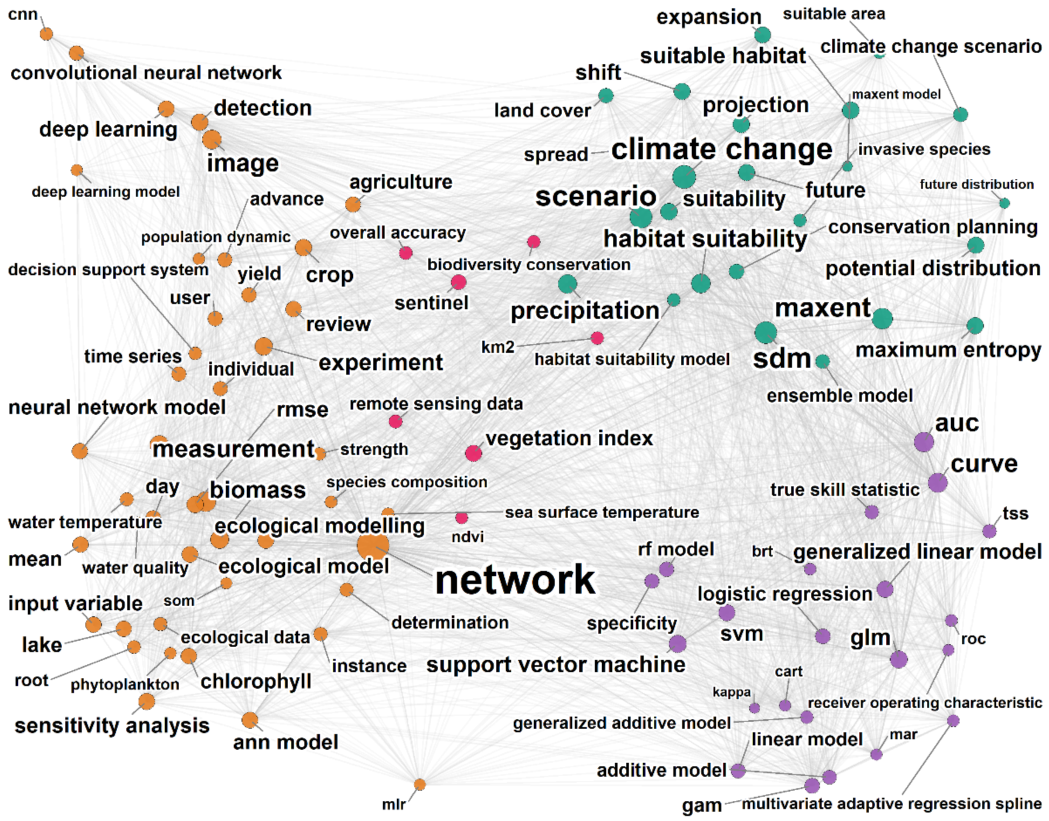

Figure 7. Conceptual structure of the research field based on keyword co-occurrence analysis. The network visualizes the main thematic clusters: Advanced Neural Networks and Image-Based Analytics (orange), Ecological Applications in SDM and Climate Change (teal), Statistical Modeling and Performance Evaluation (purple), and bridging concepts (pink).

3.5 Social Structure and Collaboration Patterns

To investigate the social structure of the field, collaboration networks were analyzed at both the author and country levels.

3.5.1 Author Collaboration Network

The author collaboration network reveals a clustered structure, characterized by several distinct, densely connected research groups with relatively few links between them (Figure 8). The analysis identifies one major, central component that is itself a super-cluster of several distinct groups. This includes a prominent group centered on Sovan Lek (teal, cluster 4), which is strongly connected to another highly active group featuring Peter L. M. Goethals (gold, cluster 2) and Niels De Pauw (light blue, cluster 5). The strong links between these authors form the collaborative backbone of the network. In addition to this central super-cluster, the network includes several smaller and more isolated groups. A notable example is the cluster led by Young-Seuk Park (red, cluster 1) at the periphery of the network. Overall, the network illustrates a research landscape organized around distinct collaborative circles.

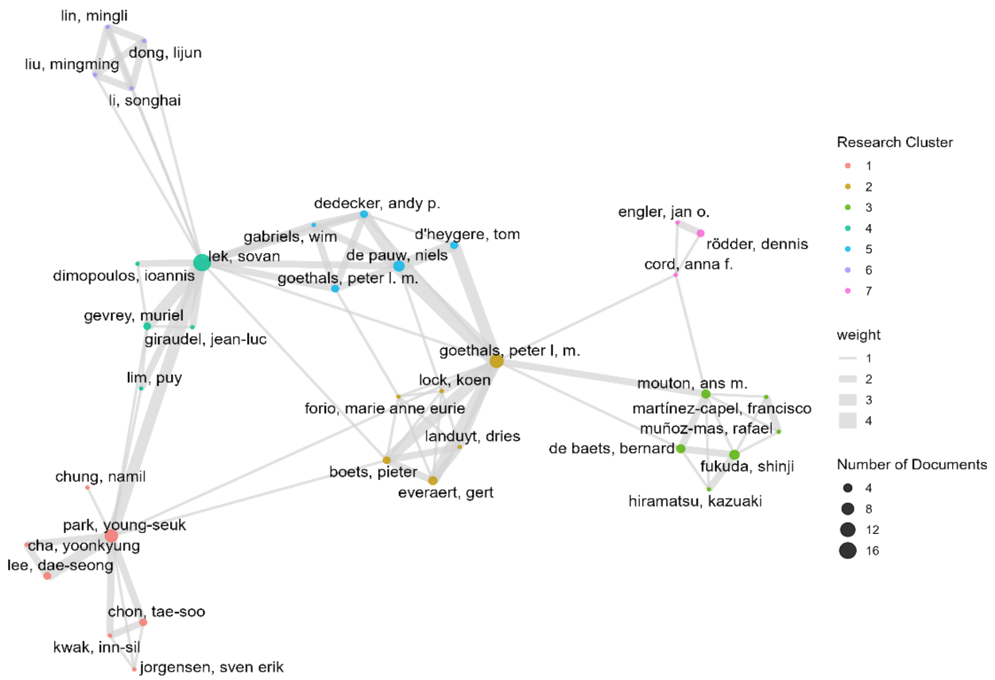

Figure 8. The author collaboration network based on the bibliometric dataset. The network displays a fragmented structure, characterized by several distinct research clusters with dense internal connections and sparse links between them.

3.5.2 Country Collaboration Network

The social structure at the international level reveals a highly connected yet multi-polar research landscape. The country collaboration network consists of the 35 most productive countries, each with a minimum of 9 documents. This core group of nations is interconnected by 308 distinct collaboration links, with a total link strength of 690. The network is dominated by the United States, which serves as the central hub, participating in the four strongest collaborative links: with China (total link strength: 33), the United Kingdom (18), Germany (17), and Canada (17). The analysis identified seven distinct research clusters, indicating a complex web of regional and intercontinental partnerships (Figure 9). The largest cluster is a global hub led by the United States (Cluster 1), which includes a diverse set of partners such as New Zealand and Iran. Two major European consortiums are evident: a Western European cluster centered around the United Kingdom and France (Cluster 4), and a North-Central European cluster led by Germany

(Cluster 3). Another significant group connects partners across the Atlantic and the Americas, including Canada and Spain (Cluster 2). A notable finding is that China forms an independent cluster of its own (Cluster 7). This multi-polar structure defines the global collaborative dynamics of the dataset.

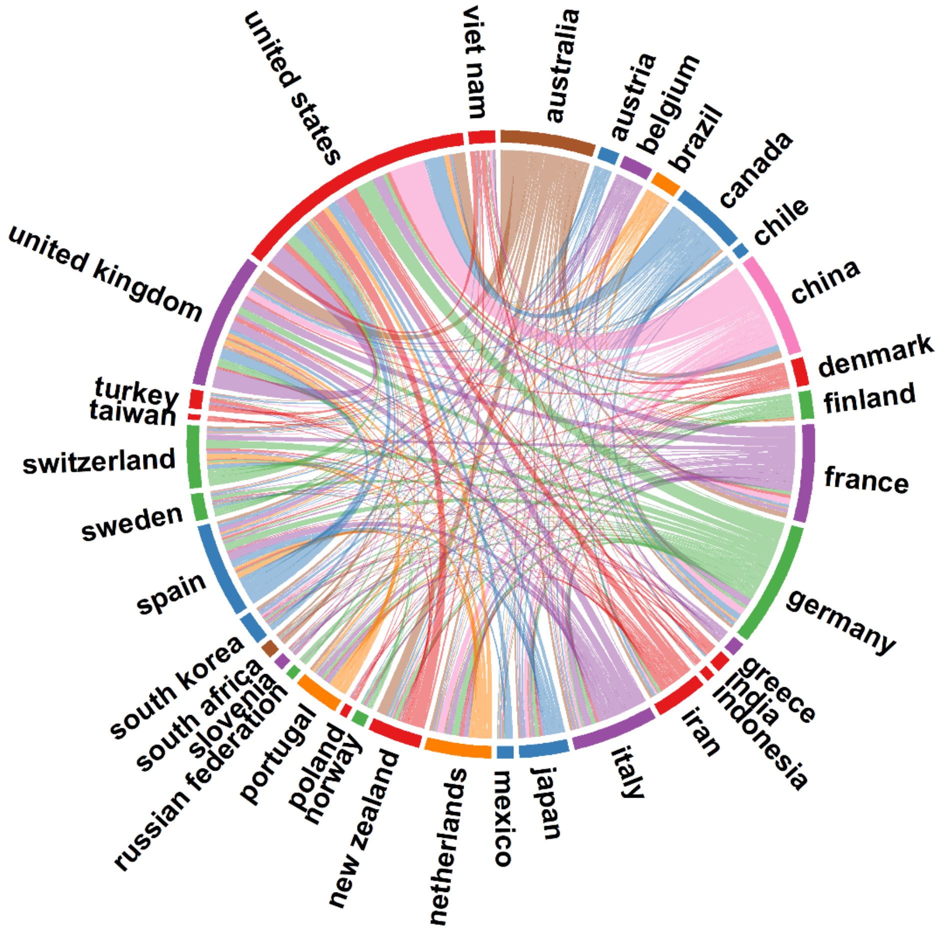

Figure 9. International research collaboration network among the 35 most productive countries. The size of each country's arc on the outer ring is proportional to its total number of documents. The internal chords represent co-authorship links, with their thickness indicating the strength of collaboration.

4. Discussion

4.1. The Software-Driven Evolution of the Intellectual Core

The bibliometric trajectory revealed in this study confirms that the fusion of AI and plant ecology has matured through three distinct phases, mirroring broader advancements in computational science (Michalski et al., 1983). As noted previously, ecology has a long history of being a data-driven science. The initial foundational period was characterized by nascent explorations where early computational models, particularly Artificial Neural Networks, were applied to predict ecological patterns and habitat distributions (Lek et al., 1996; Recknagel, 2001), laying the conceptual groundwork for the field (Guisan & Zimmermann, 2000). However, this era was marked by sporadic output as the ecological community began to grapple with these new, often complex paradigms (Fielding & Bell, 1997).

The subsequent Machine Learning Expansion heralded a transformative paradigm shift. This phase was defined by the widespread adoption of robust machine learning algorithms, which comparative studies demonstrated often outperformed traditional statistical models (Elith et al., 2006; Marmion et al., 2009). Crucially, the exponential surge in publications correlates strongly with the democratization of open-source programming libraries (e.g., in R and Python) and the implementation of algorithms like Random Forests in accessible software packages (Liaw & Wiener, 2002). Concurrently, powerful techniques like Maximum Entropy (MaxEnt) rose to prominence for species distribution modeling (Phillips et al., 2006). This methodological revolution, synthesized in influential works of the era (Heikkinen et al., 2006), was further encouraged by influential primers written specifically for the ecological community (Olden et al., 2008).

The current Deep Learning Surge represents the latest stage of this evolution. Driven by exponential growth in computational power and the availability of massive datasets, particularly from satellite remote sensing (Zhu et al., 2017), this period is defined by the adoption of sophisticated deep learning architectures (Christin et al., 2019). Although the need for vast, high-quality labeled training datasets remains a significant bottleneck (Lamba et al., 2019), these advanced models enable researchers to tackle previously intractable problems, such as individual tree crown detection from aerial imagery (Weinstein et al., 2019). This capacity for fine-grained, image-based analysis is fueling applications in precision agriculture and vegetation monitoring (Kattenborn et al., 2021), signifying the field's trajectory toward becoming a truly data-intensive science (Kamilaris & Prenafeta-Boldú, 2018).

4.2. Thematic Pillars and Data Integration

Beyond its temporal evolution, the intellectual core of this field is structured around three distinct yet interconnected thematic pillars. The first pillar represents the "Advanced AI and Image Analytics" frontier. This research is heavily focused on image-based tasks, driving innovations in automated crop monitoring and plant disease detection (Mohanty et al., 2016; Sladojevic et al., 2016), as well as high-throughput plant stress phenotyping (Singh et al., 2016). This trend reflects a broader scientific movement leveraging deep learning's power (LeCun et al., 2015).

The second pillar focuses on "Ecological Applications in Predictive Modeling," dominated by the challenge of species distribution modeling (SDM) (Elith & Leathwick, 2009) and conservation planning (Franklin, 2010). Researchers utilize predictive models to forecast how environmental variables shape habitats, often requiring specialized methods for presence-only data (Elith et al., 2011). These machine learning models frequently outperform traditional approaches (Norberg et al., 2019). Crucially, to feed these models, there is a growing emphasis on multi-source data integration, where field measurements (in situ) are fused with Geographic Information Systems (GIS) and Earth Observation (EO) data. Remote sensing acts as a critical link providing large-scale spatial data, a process greatly accelerated by the advent of open-access satellite data archives (Wulder et al., 2012).

4.3. The Methodological Toolkit: Algorithm Selection and Interpretability

The third pillar constitutes the "Statistical and Machine Learning Toolkit," emphasizing rigorous model performance evaluation (Olden et al., 2008). The prominence of algorithms like Random Forests reflects a strategic adoption of tools well-suited to ecological data (Guisan & Zimmermann, 2000). As an ensemble method, Random Forests are robust to overfitting and effectively capture non-linear patterns through aggregation (bagging) (Breiman, 2001; Cutler et al., 2007). They also provide interpretable outputs, such as variable importance metrics (Molnar, 2020; Prasad et al., 2006; Strobl et al., 2007), making it a "workhorse" algorithm (Valavi et al., 2021). However, newer gradient boosting methods like XGBoost are gaining traction. Unlike the aggregation approach of Random Forests, XGBoost relies on sequential learning. Recent studies demonstrate the efficacy of these advanced approaches in mapping complex vegetation patterns, such as management intensity types in grasslands, by synergistically using Sentinel-1 and Sentinel-2 imagery (Bartold et al., 2024).

While the field is increasingly sophisticated, it approaches a critical challenge: the trade-off between predictive power and interpretability (Olden & Jackson, 2002). The complexity that gives deep learning models their power often renders them "black boxes," making it difficult to understand the ecological reasoning behind predictions. In ecology, where understanding causal mechanisms is essential (Austin, 2007; Grace et al., 2010), this is a limitation. This paves the way for Explainable Artificial Intelligence (XAI)

(Arrieta et al., 2020). The demand is growing for inherently interpretable models (Rudin, 2019). Consequently, post-analysis methods to assess feature importance such as Boruta, Mean Decrease Accuracy (MDA), and XGBoost-SHAP are becoming central to model evaluation. For example, evaluating input variables using these interpretation methods is critical when explaining biophysical parameters like chlorophyll fluorescence in wetlands (Bartold & Kluczek, 2023). XAI is not merely a technical novelty; it is a critical tool for generating new scientific hypotheses (Langley, 1996; Roscher et al., 2020).

4.4. The Social and Global Collaborative Landscape

The development of this field is shaped by its distinct social structure. Our analysis of author collaboration networks reveals a "small-world" pattern characterized by densely connected clusters of researchers with sparse links between them, a well-documented feature of scientific collaboration (Newman, 2004; Watts & Strogatz, 1998) suggesting the presence of distinct "invisible colleges" (Crane, 1972; Zuccala, 2006). These clusters foster rapid, focused progress within specific niches (Bodin et al., 2006), related to the concepts of embeddedness and structural holes (Burt, 2004; Uzzi, 1997).

At the international level, this clustered pattern creates a multi-polar yet highly connected global landscape. The network is dominated by a few key hubs, with the United States acting as the primary nexus, consistent with its historical centrality (Adams, 2013). Concurrently, other major players like China have emerged as massive, self-contained research powerhouses (Royal Society, 2011; Zhou & Leydesdorff, 2006). This structure, with strong regional consortiums in Western Europe (Hoekman et al., 2009), suggests that a country's position is determined by its internal scientific capacity (King, 2004) and strategic international engagement (Narin et al., 1991).

5. Conclusion

In conclusion, this comprehensive bibliometric analysis demonstrates that the integration of artificial intelligence and plant ecology has matured into a vital supplementary toolkit for modern environmental research. Our study maps an evolutionary trajectory spanning over five decades, marked by three distinct phases: foundational ANN exploration, a transformative machine learning expansion driven by algorithms like Random Forests, and the current surge of deep learning applications. Importantly, this evolution has been catalyzed not merely by theoretical advancements, but fundamentally by the widespread accessibility of open-source programming libraries (e.g., in R and Python) and cloud-based platforms. The field's intellectual landscape is structured around advanced image analytics and predictive ecological modeling, heavily supported by multi-source data integration from remote sensing and GIS. The social structure remains a multi-polar "small-world" network, with the USA and China acting as prominent global hubs.

Moving forward, as the application of these data-intensive tools accelerates, the primary challenge for plant ecologists is balancing the immense predictive power of complex models with the fundamental scientific need for interpretability. Addressing this requires a deliberate shift from "black-box" predictions toward Explainable AI (XAI) methodologies. By utilizing interpretability techniques to understand feature importance and underlying mechanisms, researchers can transform machine learning into a transparent engine for true ecological insight and discovery.

Acknowledgements

The authors are grateful to the entire staff of the Laboratory of Ecology and Natural Ecosystems Management (N° 13) at the University of Tlemcen, Algeria, for their support. Special thanks are extended to the laboratory's director, Professor MERZOUK Abdessamad, for his valuable guidance and for fostering a supportive research environment.

Conflict of Interest

The authors declare no conflict of interest.

References

Adams, J. (2013). The fourth age of research. *Nature*, *497*(7451), 557–560. https://doi.org/10.1038/497557a

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Austin, M. P. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, *200*(1-2), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005

Bartold, M., & Kluczek, M. (2023). A Machine Learning Approach for Mapping Chlorophyll Fluorescence at Inland Wetlands. *Remote Sensing*, 15(9), 2392. https://doi.org/10.3390/rs15092392

Bartold, M., Wróblewski, K., Kluczek, M., & Dąbrowska-Zielińska, K. (2024). Mapping management intensity types in grasslands with synergistic use of Sentinel-1 and Sentinel-2 satellite images. *Scientific Reports*, 14, 32066. https://doi.org/10.1038/s41598-024-83699-4

Bodin, Ö., Crona, B., & Ernstson, H. (2006). Social networks in natural resource management: What is there to learn from a structural perspective?. *Ecology and Society*, *11*(2), r2. https://doi.org/10.5751/ES-01808-1102r02

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787

Christin, S., Hervet, E., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*(10), 1632–1644. https://doi.org/10.1111/2041-210X.13256

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. The University of Chicago Press.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. https://igraph.org

Cutler, D. R., Edwards, T. C. Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and*

Systematics, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*(1), 38–49. https://doi.org/10.1017/S0376892997000088

Franklin, J. (2010). *Mapping species distributions: spatial inference and prediction*. Cambridge University Press.

Grace, J. B., Anderson, T. M., Olff, H., & Scheiner, S. M. (2010). On the specification of structural equation models for ecological systems. *Ecological Monographs*, *80*(1), 67–87. https://doi.org/10.1890/09-0464.1

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2-3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Heikkinen, R. K., Luoto, M., Araújo, M. B., Virkkala, R., Thuiller, W., & Sykes, M. T. (2006). Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, *30*(6), 751–777. https://doi.org/10.1177/0309133306071957

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf

Hoekman, J., Frenken, K., & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *Annals of Regional Science*, *43*(3), 721–738. https://doi.org/10.1007/s00168-008-0252-9

IPBES (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES secretariat. https://doi.org/10.5281/zenodo.3831673

Jones, M. B., Schildhauer, M. P., Reichman, O. J., & Bowers, S. (2006). The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 519–544. https://doi.org/10.1146/annurev.ecolsys.37.091305.110031

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and challenges. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kattenborn, T., Leitloff, F., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010

King, D. A. (2004). The scientific impact of nations. *Nature*, *430*(6997), 311–316. https://doi.org/10.1038/430311a

Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation. *Current Biology*, *29*(19), R977–R982. https://doi.org/10.1016/j.cub.2019.08.016

Langley, P. (1996). *Elements of machine learning*. Morgan Kaufmann Publishers.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., & Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, *90*(1), 39–52. https://doi.org/10.1016/0304-3800(95)00142-5

Levin, S. A. (1992). The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture. *Ecology*, *73*(6), 1943–1967. https://doi.org/10.2307/1941447

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (1983). *Machine learning: An artificial intelligence approach*. Tioga Publishing Company. https://doi.org/10.1007/978-3-662-12405-5

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419

Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/

Narin, F., Stevens, K. S., & Whitlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, *21*(3), 313–323. https://doi.org/10.1007/BF02093973

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(Suppl 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., Husby, M., Kålås, J. A., Lehikoinen, A., Luoto, M., Mod, H. K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N. E., Gravel, D., & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, *89*(3), e01370. https://doi.org/10.1002/ecm.1370

Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1-2), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology*, *83*(2), 171–193. https://doi.org/10.1086/587826

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*, 181–199. https://doi.org/10.1007/s10021-005-0054-1

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling*, *146*(1-3), 303–310. https://doi.org/10.1016/S0304-3800(01)00316-7

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, *8*, 44963–44985. https://doi.org/10.1109/ACCESS.2020.2976199

Royal Society (2011). *Knowledge, networks and nations: Global scientific collaboration in the 21st century*. The Royal Society.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, *2016*, 3289801. https://doi.org/10.1155/2016/3289801

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Thessen, A. E. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, *1*, e8622. https://doi.org/10.3897/oneeco.1.e8621

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, *42*(1), 35–67. https://doi.org/10.2307/2393808

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, *44*(12), 1731–1742. https://doi.org/10.1111/ecog.05615

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., & White, E. (2019). Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, *11*(11), 1309. https://doi.org/10.3390/rs11111309

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

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

Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, *35*(1), 83–104. https://doi.org/10.1016/j.respol.2005.08.006

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307

Zuccala, A. (2006). Modeling the invisible college. *Journal of the American Society for Information Science and Technology*, *57*(2), 152–168. https://doi.org/10.1002/asi.20256

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629