Journal of Cheminformatics

RESEARCH Open Access

The published role of artificial intelligence in drug discovery and development: a bibliometric and social network analysis from 1990 to 2023

Murat Koçak[1*] and Zafer Akçalı[1]

Abstract

Today, drug discovery and development is one of the fields where Artificial Intelligence (AI) is used extensively. Therefore, this study aims to systematically analyze the scientific literature on the application of AI in drug discovery and development to understand the evolution, trends, and key contributors within this rapidly growing field. By leveraging various bibliometric indicators and visualization techniques, we seek to explore the growth patterns, influential authors and institutions, collaboration networks, and emerging research trends within this domain. Bibliometric and network analysis methods (co-occurrence, co-authorship, and collaboration, etc.) were used to achieve this goal. Bibliometric visualization tools such as Bibliometrix R package software, VOSviewer, and Litmaps were used for comprehensive data analysis. Scientific publications on AI in drug discovery and development were retrieved from the Web of Science Core Collection (WoS CC) database covering 1990–2023. In addition to visualization programs, the InCites database was also used for analysis and visualization. A total of 4059 scientific publications written by 13,932 authors and published in 1071 journals were included in the analysis. The results reveal that the most prolific authors are Ekins (n = 67), Schneider (n = 52), Hou Tj (n = 43), and Cao Ds (n = 34), while the most active institutions are the "Chinese Academy of Science" and "University of California." The leading scientific journals are "Journal of Chemical Information and Modelling," "Briefings in Bioinformatics," and "Journal of Cheminformatics." The most frequently used author keywords include "protein folding," "QSAR," "gene expression data," "coronavirus," and "genome rearrangement." The average number of citations per scientific publication is 28.62, indicating a high impact of research in this field. A significant increase in publications was observed after 2014, with a peak in 2022, followed by a slight decline. International collaboration accounts for 28.06% of the publications, with the USA and China leading in both productivity and influence. The study also identifies key funding organizations, such as the National Natural Science Foundation of China (NSFC) and the United States Department of Health & Human Services, which have significantly supported advancements in this field. In conclusion, this study highlights the transformative role of AI in drug discovery and development, showcasing its potential to accelerate innovation and improve efficiency. The findings provide valuable insights into the current state of research, emerging trends, and future directions, offering a roadmap for researchers, industry professionals, and policymakers to further explore and leverage AI technologies in this domain.

*Correspondence:
Murat Koçak
muratkocak25@gmail.com

Scientific contribution This study provides a comprehensive bibliometric analysis of 4,059 scientific publications (1990–2023) to map the evolution, trends, and key contributors in AI-driven drug discovery, identifying prolific authors (e.g., Ekins, Schneider), leading institutions (e.g., Chinese Academy of Sciences, University of California), and high-impact journals (*Journal of Chemical Information and Modelling*). It reveals critical collaboration patterns (28.06% international co-authorships), dominant funding sources (e.g., NSFC, NIH), and emerging research hotspots (e.g., *protein folding, QSAR, coronavirus*), while highlighting the transformative role of deep learning post-2014. By synthesizing these insights, the study offers a strategic roadmap for researchers and policymakers to optimize AI applications in drug development, addressing both current challenges and future opportunities in the field.

Keywords Drug discovery, Drug development, Artificial intelligence (AI), Bibliometric analysis, Network analysis, Scientific publications

Introduction

The discovery and development of new drugs is a complex, time-consuming, and resource-intensive process, often taking years and costing billions of dollars [1]. In this context, integrating AI techniques into drug discovery and development has emerged as a promising approach to address these challenges and accelerate innovation [2, 3].

The term "artificial intelligence" was first coined by John McCarthy [4] in 1956 during the Dartmouth Conference, a seminal event widely considered the birthplace of AI as a field of academic study [4]. Specifically, McCarthy used the term when organizing the conference, which brought together leading researchers to explore the possibility of creating machines that could simulate human intelligence. The first major AI application specifically targeting drug discovery was MYCIN, developed at Stanford between 1972 and 1976 [5]. While primarily designed as a medical diagnosis system for bacterial infections, MYCIN demonstrated how rule-based expert systems could assist in medical decision-making, laying crucial groundwork for future AI applications in pharmaceutical research.

By the 1980s, researchers began more systematically exploring AI techniques like machine learning and expert systems for identifying potential drug candidates, molecular modeling, and predicting drug interactions. This period marked the beginning of computational approaches that would eventually evolve into today's sophisticated AI-driven drug discovery platforms [6, 7].

AI encompasses a wide range of computational methods and techniques, including machine learning, deep learning, natural language processing, and computer vision [6, 8]. These techniques have demonstrated remarkable potential in various drug discovery and development pipeline stages, such as virtual screening, lead optimization, predictive modeling, and clinical trial design [3, 9]. By leveraging the ability of AI to process and analyze vast amounts of data, identify patterns, and make predictions, researchers can gain valuable insights, streamline decision-making processes, and ultimately enhance the efficiency and effectiveness of drug discovery efforts [7, 10].

Computer-aided drug design (CADD) and AI in drug discovery are both computational approaches to accelerate the drug development process, but they differ in their core methodologies. CADD encompasses a broader range of techniques, often based on established scientific principles and equations, to design and optimize drug molecules. AI, particularly machine learning, utilizes algorithms that learn complex patterns from large datasets to make predictions and decisions, offering the potential to uncover new insights and handle more complex scenarios than traditional CADD methods. While CADD emphasizes molecular design, AI is expanding into areas like target identification and clinical trial prediction. Both approaches, however, share the common goals of predicting molecular properties, optimizing drug-target interactions, and ultimately, developing more effective and safer medicines. They are often used synergistically, leveraging the strengths of each approach to maximize the efficiency and success of drug discovery [2, 11].

As the application of AI in drug discovery and development continues to grow, it is crucial to assess the field's current state, identify emerging trends, and understand the impact of AI on various aspects of the pharmaceutical industry. Bibliometric analysis, which involves the quantitative study of published literature, provides a powerful tool to achieve these objectives [12, 13]. By analyzing the bibliometric data related to AI in drug discovery and development, researchers can gain insights into the evolution of this field, the most influential publications, the key players driving innovation, and the potential future directions [14].

The primary aim of this study is to systematically analyze the scientific literature on the application of artificial intelligence (AI) in drug discovery and development to understand the evolution, trends, and key contributors within this rapidly growing field. By leveraging various

bibliometric indicators and visualization techniques, we seek to explore the growth patterns, influential authors and institutions, collaboration networks, and emerging research trends within this field. The findings from this analysis will contribute to a better understanding of the role of AI in drug discovery and development and provide valuable insights for researchers, industry professionals, and policymakers involved in this domain. AI is based on creating machines that think like humans and is seen as the pinnacle of science [15]. AI (AI), which emerged in different fields in the twenty-first century, has become a trend in many fields, such as medicine, science, and business [16] and has also manifested itself in medicine. AI-based algorithms are particularly well suited for problems where the physical laws that determine the molecular properties to be predicted are not fully known or are too complex to establish experimental relationships [9, 11]. AI-based approaches are utilized in many fields. One of these areas is drug development.

AI has revolutionized computer science, medicine, economics, engineering, and many other disciplines. Especially in health sciences, the potential of AI is used in a wide range of fields, from the diagnosis of diseases to the development of treatment methods [17]. The drug discovery and development process benefits significantly from AI technologies, enabling the rapid discovery of new drug candidates, improving drug efficacy and safety profile, and optimizing development processes [9, 18].

Since 1990, there has been a dramatic increase in the number of studies on AI and drug discovery. This increase has been driven by advances in the capacity of deep learning, machine learning, and other advanced AI techniques to process biomedical data [19, 11]. However, the literature in this field needs to be analyzed comprehensively and systematically. Bibliometric analyses of 4059 articles on AI and drug discovery in the WoS CC database form the basis of this study.

While the literature on the applications of artificial intelligence in drug discovery and development processes is gradually expanding, bibliometric studies that comprehensively examine research trends in this field have also been carried out. Tran et al. [10] analyzed publications between 1990 and 2020, revealing the historical development of the field and the main research clusters. Yang et al. [20] examined the current trends and collaboration networks in AI-assisted drug development research, focusing on the period 2014–2019. Zhang et al. [11] analyzed publications between 2000 and 2021 and evaluated methodological developments and application areas in the field. Our current study updates and extends these previous bibliometric analyses and provides a more up-to-date analysis covering the period up to 2023. It also examines the role of AI in drug discovery and development processes using more detailed subcategories and new analytical approaches. While bibliometric analysis is seen as inevitable for many fields in this concept [21, 22], academic studies on the impact of AI applications in medicine should be examined through bibliometric analysis. The increasing popularity and widespread use of AI causes the subject to become a need in terms of health.

Overall, these papers highlight the importance of AI in improving Drug Discovery and emphasize the need for further research in this area.

This study aims to address critical questions, including:

- Which scientific journals are most influential in this field?
- Who are the leading authors contributing to AI-driven drug discovery?
- What are the most cited articles?
- Which keywords dominate the research landscape?
- Which institutions and countries lead in AI research for drug discovery?
- What are the primary funding sources supporting these advancements?

The core research question is: How has the field of AI in drug discovery and development evolved, and what are the key trends, influential contributors, and areas of focus within this research domain?

Method

The related subject is analyzed statistically and mathematically in bibliometrics, and a framework is drawn for the course. In addition, bibliometric analysis can also reveal the effectiveness of studies conducted through statistical data [23]. Since it is necessary to use one of the appropriate research methods for evaluation and prediction, bibliometric analysis is one of the analyses that serves this purpose. Bibliometrics is used as a type of analysis that systematically analyses datasets [13, 24]. This type of analysis requires a systematic literature review and meaningful structuring of a large data set [13]. Bibliometric analysis is a method used to analyze studies in a particular field, and its use has increased in recent years.

The data for this study were retrieved from the Web of Science Core Collection (WoS CC) online database on 21.11.2023. The choice of WoS CC as the primary data source for this bibliometric analysis is justified by several key factors that distinguish it from other scientific databases such as Scopus, PubMed, and Google Scholar.

Firstly, WoS CC is renowned for its comprehensive coverage of high-quality, peer-reviewed journals across a wide range of disciplines, including science, technology,

medicine, and social sciences. This extensive coverage ensures that the dataset used in this study is both robust and representative of the global research output in the field of AI in drug discovery and development. Unlike Google Scholar, which includes a broader range of sources including preprints and less rigorously reviewed content, WoS CC maintains a stringent selection process for journal inclusion, which enhances the reliability and credibility of the data [25].

Secondly, WoS CC provides advanced bibliometric tools and metrics that are specifically designed for large-scale bibliometric analyses. These tools allow for detailed tracking of citation networks, collaboration patterns, and research trends over time. The ability to generate precise citation metrics and visualize complex bibliometric networks is a significant advantage of WoS CC over other databases, which may not offer the same level of analytical depth [12].

However, it is important to acknowledge the limitations of relying solely on WoS CC. One of the primary challenges is the potential exclusion of relevant studies that may be indexed in other databases such as Scopus or PubMed. Scopus, for instance, is known for its broader coverage of international journals, particularly in non-English speaking countries, which could provide additional insights into global research trends. Similarly, PubMed is a critical resource for biomedical research, and its exclusion might result in the omission of important studies in the field of drug discovery.

Moreover, integrating data from multiple databases presents its own set of challenges. Differences in indexing criteria, citation tracking methods, and data formats can complicate the harmonization of datasets, leading to potential inconsistencies in the analysis. For example, Google Scholar includes a wider range of document types, such as conference papers and theses, which can introduce noise into the dataset and make it difficult to compare findings across databases. Additionally, the lack of standardized metadata across different platforms can hinder the accurate identification of authors, institutions, and funding sources, which are crucial for bibliometric analyses.

Despite these challenges, the decision to use WoS CC as the primary data source for this study is based on its established reputation for quality and its advanced analytical capabilities. While the inclusion of additional databases could provide a more comprehensive view of the research landscape, the complexities involved in integrating and harmonizing data from multiple sources would likely outweigh the benefits in this context. Therefore, WoS CC was chosen as the most suitable tool for conducting a detailed and reliable bibliometric analysis of AI in drug discovery and development.

The search strategy of this study was meticulously designed to ensure the retrieval of relevant publications. The Advanced Search feature of WoS CC was utilized to construct a comprehensive query that captures the breadth of research on AI in drug discovery and development. The search query was structured as follows:

WoS CC Advanced Search Query: (TS = ("deep-Learn*" OR "machine learn*" OR "deep learn*" OR "AI" OR "artificial neural network*" OR "deep neural network*")) AND (TS = ("drug discovery")) AND (PY = 1990–2023). "TS" refers to the Topic field, which searches titles, abstracts, author keywords, and Keywords Plus. The asterisk (*) was used as a wildcard to capture variations of the terms (e.g., "deep learning" and "deep learnings"). The time frame was restricted to publications between 1990 and 2023 to capture the evolution of AI in drug discovery over the past three decades.

The time horizon for the research was set as 1990–2023 and this decision was based on the following criteria:

- Technological evolution: The time frame from 1990 to 2023 allows for the analysis of the evolution of AI technologies in drug discovery, from early rule-based systems to contemporary deep learning models. This period captures the transition from traditional computational approaches to more sophisticated AI-driven methods, providing a comprehensive view of the field's development.
- Recent trends: Including the year 2023 ensures that the most recent advancements and trends in AI-driven drug discovery are captured. This is particularly important given the rapid pace of innovation in AI and its increasing impact on pharmaceutical research in recent years.
- Data availability: The WoS CC database provides comprehensive coverage of scientific literature from 1990 onwards, ensuring that the dataset used in this study is both extensive and reliable. By limiting the search to this time frame, we ensure that the analysis is based on a consistent and well-documented dataset.

To ensure the accuracy and reproducibility of the search results, the following steps were taken:

1. Indexing and filtering: The search results were indexed using WoS CC's built-in indexing tools, which allow for the exclusion of irrelevant document types such as editorials, letters, and meeting abstracts. Only peer-reviewed articles, conference papers, and reviews were included in the final dataset.

2. Inclusion and exclusion criteria: After the screening process, a total of 4059 articles were included in the final dataset. During the screening process, 110 articles were excluded based on the inclusion and exclusion criteria

   – Inclusion criteria: Publications must focus on the application of AI technologies in drug discovery and development. Only peer-reviewed articles, conference papers, and reviews published between 1990 and 2023 were included. Studies that explicitly mention methodologies related to AI, such as machine learning, deep learning, or neural networks, were prioritized.
   – Exclusion criteria: Articles not related to drug discovery or development were excluded. Publications that do not utilize AI methods or technologies were also omitted. Duplicates and non-English publications were filtered out during the data collection process.

3. Data extraction: Metadata from the selected documents, including active authors, journal sources, countries, institutions, and funding sources, were directly accessed from the WoS CC database. This metadata was exported in plain text and tab delimited formats for further analysis.
4. Data processing and cleaning: The raw data exported from WoS CC [26] and InCites [27] underwent a rigorous cleaning and preprocessing phase to ensure the accuracy and consistency of the dataset. The following steps were taken during this phase:

   – Duplicate removal: Duplicate entries were identified and removed using automated scripts and manual verification to ensure that each publication was represented only once in the dataset.
   – Keyword normalization: Author keywords and Keywords Plus were normalized to account for synonyms and variations. For instance, "machine learning" and "ML" were treated as equivalent terms.

5. Data analysis tools: The analysis was conducted using several specialized tools:

   – Bibliometrix package[28]: This tool was used to construct and visualize bibliometric networks, perform co-authorship analysis, create keyword co-occurrence networks, and generate density visualizations and cluster analyses.
   – VOSviewer[29]: This software was used to convert raw data into structured bibliometric data frames, generate descriptive analyses of bibliographic attributes, and create citation networks and collaboration patterns.
   – Litmaps[30]: This platform was used to create interactive visualizations of citation networks, identify key literature and research fronts, and map knowledge flows between publications.

The methodological procedures adopted in this study were designed to ensure transparency and reproducibility. By providing a detailed description of the search strategy, inclusion and exclusion criteria, and data analysis tools, this study aims to enable other researchers to replicate the method and conduct similar bibliometric analyses in the future.

Results

The findings of the bibliometric analysis conducted within the scope of the study are presented in this section. The results are organized into several subsections, each focusing on a specific aspect of the research, including general information about the dataset, authorship patterns, collaboration networks, keyword analysis, and institutional contributions. These findings provide a comprehensive overview of the evolution, trends, and key contributors in the field of AI in drug discovery and development.

General information

The dataset used in this study comprises 4059 documents published between 1990 and 2023, retrieved from the Web of Science Core Collection (WoS CC) database. These documents were written by 13,932 authors and published in 1071 journals. The average annual growth rate of publications in this field is 6.5%, indicating a steady increase in research activity over the past three decades. The average number of citations per document is 28.62, reflecting the high impact of research in this domain.

Table 1 provides an overview of the dataset, including key bibliometric indicators such as the number of sources, documents, authors, and international collaborations. The dataset includes 5869 Keywords Plus (ID) and 6592 author keywords (DE), which were used to identify research themes and trends. The analysis reveals that 28.06% of the documents involve international co-authorships, highlighting the global nature of research in this field.

The graph shows the number of WoS CC documents published and the number of times those documents were cited, both related to AI, from 1992 to 2023.

Figure 1 reveals a steady increase in both metrics from the early 2000s until around 2014. This suggests a gradual growth of interest and research in AI during that period.

Table 1 Overview of the dataset (Bibliometrix & R software, 2024)

Description	Results
Main Information about the Data	
Time interval	1990:2023
Sources (Journals, Books, etc.)	1071
Documents	4059
Annual Growth Rate %	6.5
Average citations per document	28.62
Document Content	
Keywords Plus (ID)	5869
Author's Keywords (DE)	6592
Authors	
Authors	13,932
Authors of single-author documents	215
Co-authors per Document	5.29
International co-authorships %	28.06

However, a significant shift occurred in 2014, with a sharp spike in published documents and citations. This dramatic increase indicates a surge in the popularity and impact of AI research.

The trend of increasing citations is particularly notable after 2018, with a steep upward trajectory. This signifies AI research's growing influence and significance within the scientific community. The rapid acceleration of citations suggests that AI research increasingly produces impactful and influential work.

Overall, the graph provides a visual representation of the rapid rise of AI as a field of study and its increasing impact on various scientific disciplines.

The trend of the journal impact factor quartile of articles is shown in Fig. 2. From 2016 to 2021, the number of documents in Q1 articles showed a rapid growth trend; after 2017, the number of documents in Q1 increased steadily. For the first time, a decrease in the number of publications was observed after 2022. This means that

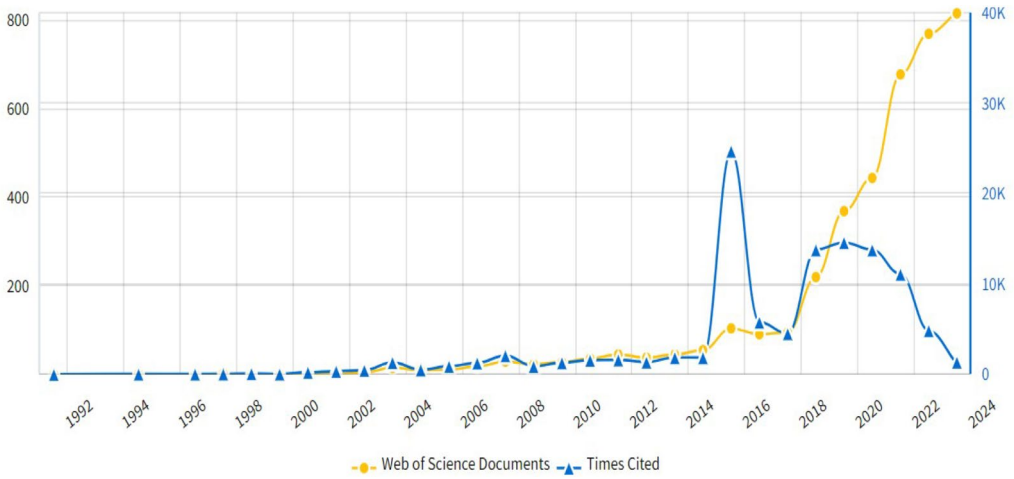

Fig. 1 Distribution of WOS Articles and Citations on AI and Drug Discovery between 1990 and 2023 (n = 4059) [16]

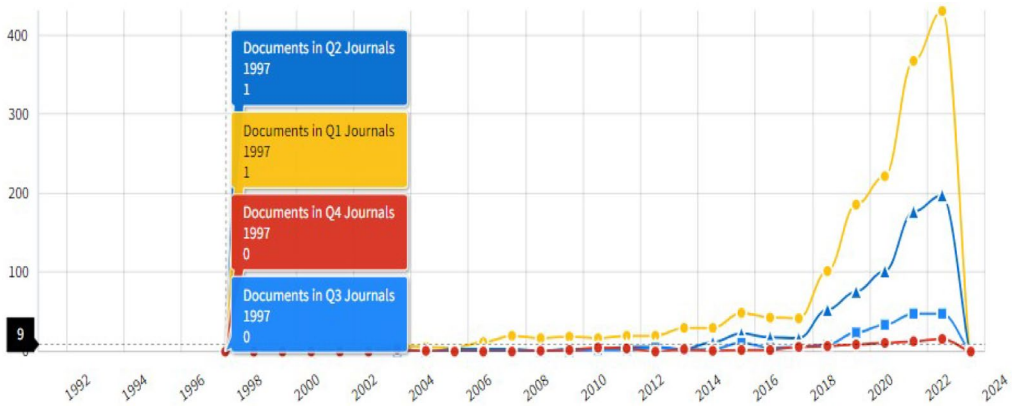

Fig. 2 Distribution of Documents in Q1, Q2, Q3 and Q4 Journals in the Field of Aland Drug Discovery between 1990 and 2023 [16]

scientific journals have started to reduce their acceptance rates in this field. The lowest rate in this field is for papers in the Q4 category. This indicator shows the popularity of this field.

WoS CC was used to identify and categorise citation topics rigorously. All findings from the search query were included in the Review without further filtering. Citation topics were narrowly focused and categorized according to the classifications recently published by the WoS CC, covering more than 2500 detailed citation topics. This categorization operates hierarchically below the WoS CC subject categories and citation topics at a broader level. It provides a precise and unbiased assessment of the technologies used in the search query.

Figure 3 presents the most frequently mentioned micro-citation topics in articles where AI and drug discovery are used together. These topics were identified using the WoS Citation Index, which categorizes citation topics hierarchically. The most represented topics include "protein folding," "QSAR" (Quantitative Structure-Activity Relationship), "gene expression data," "coronavirus," and "genome rearrangement." These topics reflect the key areas of research within the field of AI-driven drug discovery

Authorship patterns

The analysis of authorship patterns in this study leverages Lotka's Law [31], a foundational principle in bibliometrics that describes the distribution of scientific productivity among authors in a given field. Lotka's Law posits that a small proportion of authors contribute the majority of publications, following an inverse square relationship. Specifically, the law predicts that approximately 60% of authors will publish one paper, 15% will publish two papers, 7% will publish three papers, and so on, with productivity decreasing exponentially as the number of publications per author increases.

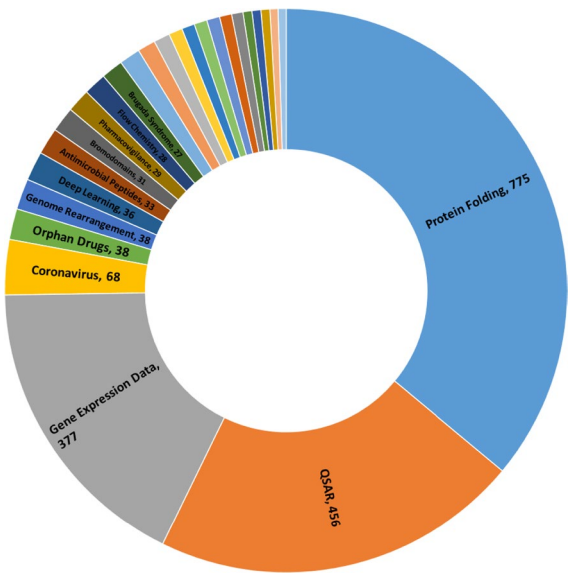

Fig. 3 Most frequently mentioned micro-citation topics in Articles where AI and Drug Discovery are used together. (WoS CC, 2023)

Table 2 and Fig. 4 illustrate how the productivity of authors in AI-driven drug discovery aligns with Lotka's Law. The results show that 75% of authors contributed a single publication, 14% contributed two publications, and 4% contributed three publications. While these values slightly deviate from the theoretical predictions of Lotka's Law (e.g., 60% for one publication), the overall trend confirms the law's applicability to this field. Such deviations are common in empirical studies and may reflect the collaborative nature of modern research, where teams often produce multiple works collectively.

By applying Lotka's Law, this study not only validates its relevance to AI in drug discovery but also provides a framework for identifying influential researchers and understanding productivity dynamics in rapidly evolving scientific domains.

Figure 5a highlights the most prolific authors in the field, with Ekins (n = 67), Schneider (n = 52), Hou Tj (n = 43), and Cao Ds (n = 34) being the leading contributors. These authors account for 29.8% of the total publications in the dataset, underscoring their significant influence on the field.

Figure 5b shows the most cited authors, with Schneider G, Zhavoronkov A, and Ekins S leading in terms of citation impact. Their work collectively accounts for 25.9% of the total citations in the dataset, indicating their substantial contribution to the advancement of AI in drug discovery.

Figure 5c is a co-authorship network showcasing the most prolific authors within AI and drug discovery and

Table 2 Lotka's Law and Author Productivity Ratio (Bibliometrix & R software, 2023)

Number of articles	Number of authors	Proportion of authors
1	10,773	0.773
2	1761	0.126
3	623	0.045
4	272	0.02
5	142	0.01
6	97	0.007
7	67	0.005
8	50	0.004
9	31	0.002
10	29	0.002

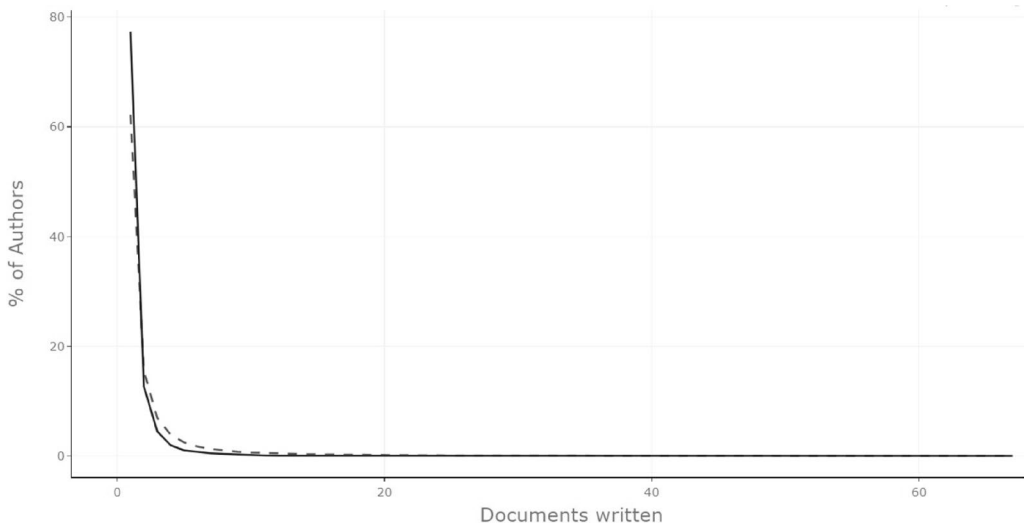

Fig. 4 Lotka's Law and Author Productivity Rate. (Bibliometrix & R software, 2023)

their collaborations. It depicts a complex web of connections, highlighting the collaborative nature of research in this domain. Authors with more connections, such as "Ekins Sean," indicate more significant research activity and collaborative engagement within the field.

We can observe several distinct clusters within the network, each representing a group of authors with a high degree of interconnectedness. For example, a large, bright blue cluster centers around "Zheng, Mingyue" and "Engkvist, Ola," suggesting a solid collaboration within a specific area of research, potentially related to machine learning applications in drug discovery. Another prominent cluster, in shades of orange, features "Hou, Tingjun," "Cao, Dongsheng," and "Wu, Zhenxing," potentially representing a group focused on developing novel AI algorithms for drug design or analysis. The cluster in green, with "Ekins, Sean" at its heart, may indicate a group concentrating on target identification or drug repurposing.

These distinct clusters suggest that the research field is multifaceted, encompassing diverse subfields or methodologies. While some research groups exhibit a more significant number of connections, suggesting stronger collaborations and a broader network within the field, others appear more focused and less collaborative. Overall, the graph highlights the significant collaborative nature of the research field, showcasing its structural characteristics by visualizing the relationships between different research groups and patterns of collaboration.

Figure 6 illustrates the key milestones in the application of AI in drug discovery from 1990 to 2023. Early studies, such as Ballester et al. [32], pioneered machine learning integration, while Lecun et al. [33] advanced deep learning techniques, significantly enhancing predictive modeling. Öztürk et al. [34] and Chen et al. [2] contributed to AI's growing role in optimizing drug-target interactions and accelerating research efficiency.

After 2015, AI applications in drug discovery saw rapid growth, peaking around 2020–2022. However, a slight decline in publications after 2022 suggests a shift toward more refined and high-quality research. AI continues to evolve, offering more sophisticated applications in pharmaceutical innovation and personalized medicine.

Articles

The top 20 most cited papers are presented in Table 3. The Article with the highest number of citations was published by Lecun Y in Nature in 2015 (Total citations = 20,779), followed by Ching T published in J R Soc Interface in 2018 (total citations = 924), and Vamatheva N J published in Nat Rev Drug Discovery in 2019 with 907 citations. Therefore, total citations are included in Table 3 to help researchers quickly identify highly cited published papers. These high-impact publications provide a quick overview of the field and broaden researchers' perspectives.

Litmaps, an advanced science discovery platform known for its visual citation navigation, was used for this study. This platform offers an interface that facilitates the discovery of scientific literature, allowing researchers to delve deep into the research area and uncover intricately linked articles within maps. Litmaps also provides convenient options for quickly importing articles through various means such as reference manager, keyword search, ORCID ID, DOI, or using a seed article [35].

Litmaps helps researchers to do literature reviews concisely and systematically. It helps to find related or

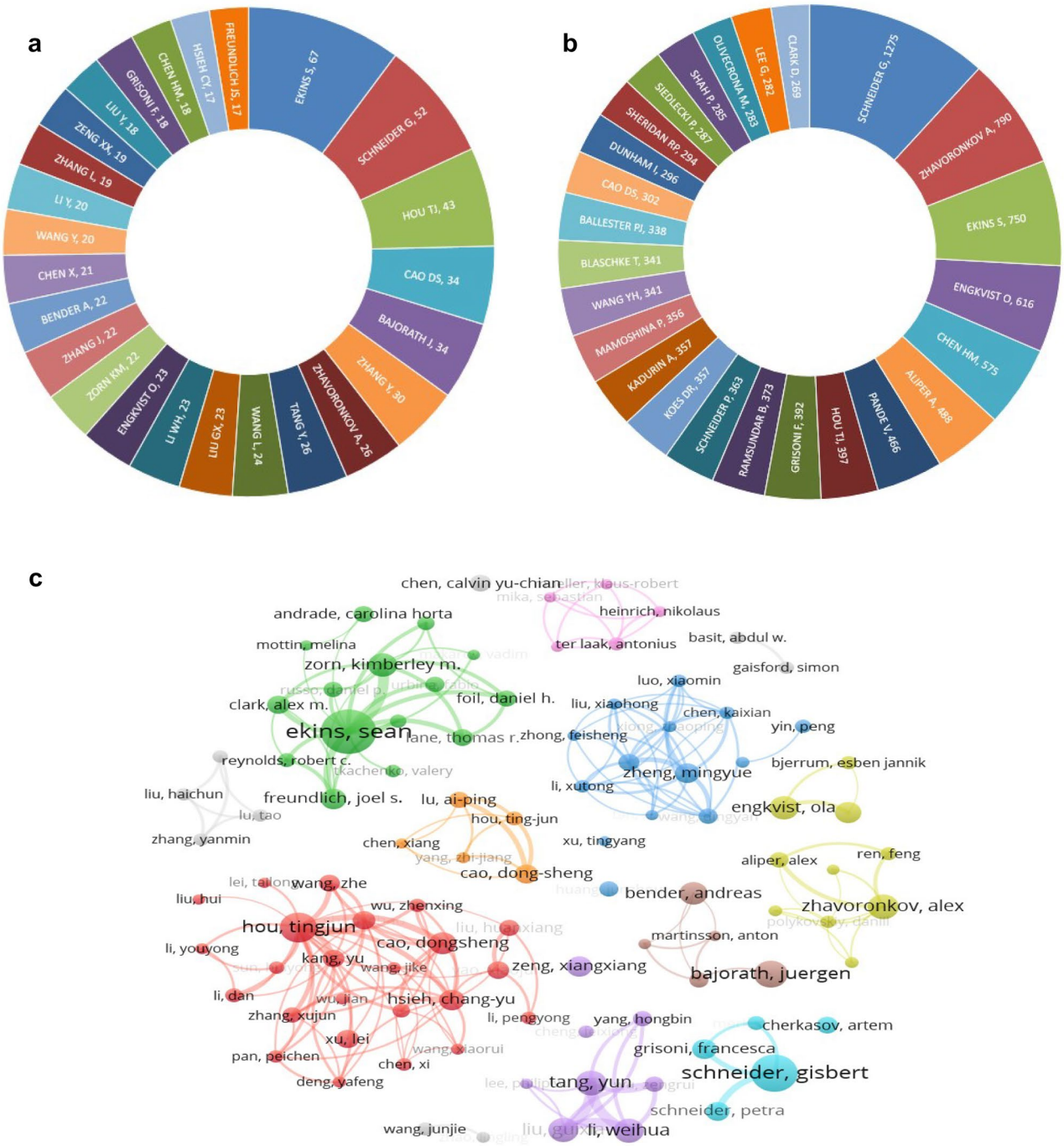

Fig. 5 **a** Top 25 authors publishing the most AI papers in Drug Discovery from 1991 to 2024. (InCites, 2024). **b** Top 25 Authors publishing the most AI citations in Drug Discovery from 1991 to 2024. (InCites, 2024). **c** Collaboration Graph of authors working in the field of Drug Discovery Using AI. (VOSviever, 2023)

relevant studies through the seed article. This will include some direct references, citations, and citations and references of your Seed Article. Litmaps provides functions for visualizing literature maps covering important papers on specific research areas through various visualization modes. In particular, papers with higher citation counts are represented by larger circles, where the node's size is proportional to the logarithm of the citation count. In Figs. 7a and b, the Seed Maps show the top 20 citations and references related to a single article.

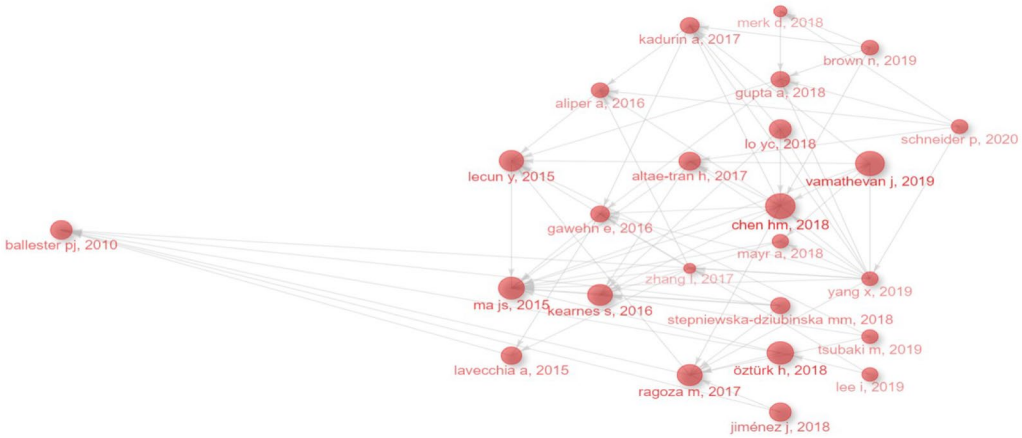

Fig. 6 History of AI articles in the field of Drug Discovery from 1990 to 2023. (Bibliometrix & R software, 2023)

According to these graphs, the change of other authors citing Ekins (2007) and Schneider (2005), who published the most AI articles in the field of Drug Discovery from 1991 to 2024, is shown.

Figure 7a shows the publication history of authors citing a specific paper by Ekins in 2007. The graph shows that the paper has been cited by numerous authors over the years, with the earliest citation dating back to 2005 and the most recent citations occurring in 2023. The graph suggests that Ekins' work has significantly impacted the field, as many authors have cited it over nearly 20 years. The graph also highlights the importance of citing previous work to ensure knowledge advancement and recognize the contributions of those who came before.

Figure 7b shows a citation network for a paper published in 2005 by Schneider [36]. The network map shows that numerous researchers across various fields have cited the paper, and it has been particularly influential in the years since its publication. The network map suggests that Schneider's paper has made a significant contribution to the field and has been cited by a wide range of authors over nearly 20 years. The graph also highlights the importance of citing previous work to ensure knowledge advancement and recognize the contributions of those who came before.

Scientific journals

Figure 8a and b show the trend of AI articles in Drug Discovery journals and reveal the decreasing or increasing trends in Drug Discovery journals over time. The results in Fig. 8a and b show the distribution of the journals with the most articles by year. The three leading journals publish the most articles on AI and Drug Discovery: Journal Of Chemical Information And Modelling, Briefings in Bioinformatics, and Journal of Cheminformatics. Although these three leading journals contributed 12% of the articles, the remaining publications are widely dispersed. This suggests that besides the top three journals, which publish most of the insights in this research area, a wide variety of other sources also contribute significantly to the literature.

When we look at the change in scientific journals over the years, the first publications in this field were published in the Journal Of Chemical Information And Modelling in 2005. After 2020, a significant increase in publications on AIapplications in drug discovery and development was observed in all journal categories. Journal of Chemical Information and Modelling, Briefing in Bioinformatics, and Journal of Cheminformatics are the journals that publish the most articles.

Figure 8b shows Scientific Journals published AI studies in Drug Discovery in 2005. This rate started to increase after 2016.

WoS CC categories

As seen in Fig. 9, according to the WoS CC database categorization, the articles are distributed in 147 scientific categories. However, about 70% of these articles are mainly covered by 10 categories. These categories are Chemistry Multidisciplinary (13%), Computer Science Interdisciplinary Applications (11.5%) Biochemistry and Molecular Biology (10.4%), Pharmacology and Pharmacy (8.9%) Medicinal Chemistry (8.7%), Mathematical and Computational Biology (8.4%), Computer Science and Information Systems (6.4%), Computer Science and Artificial Science (3.6%), Biotechnology and Applied Microbiology (3.5%).

It is important to note that there are many publications from these three disciplines in the graph to understand

Table 3 The Most Cited AI Articles in the Field of Drug Discovery from 1990 to 2023 Distribution (Bibliometrix & R software, 2023)

Paper	Title	Year	Magazine	DOI	Total Citations
LECUN Y and Others	Deep learning	2015	NATURE	10.1038/nature14539	20,779
CHING T and Others	Opportunities and obstacles for deep learning in biology and medicine	2018	J R SOC INTERFACE	10.1098/rsif.2017.0387	924
VAMATHEVAN J and Others	Applications of machine learning in drug discovery and development	2019	NAT REV DRUG DISCOV	10.1038/s41573-019-0024-5	907
BLAKEMORE DC and Others	Organic synthesis provides opportunities to transform drug discovery	2018	NAT CHEM	10.1038/s41557-018-0021-z	797
CHEN HM and Others	The rise of deep learning in drug discovery	2018	DRUG DISCOV TODAY	10.1016/j.drudis.2018.01.039	792
KEARNES S and Others	Molecular graph convolutions: moving beyond fingerprints	2016	J COMPUT AID MOL DES	10.1007/s10822-016-9938-8	789
PINZI L and Others	Molecular Docking: Shifting Paradigms in Drug Discovery	2019	INT J MOL SCI	10.3390/ijms20184331	678
MA JS and Others	Deep Neural Nets as a Method for Quantitative Structure–Activity Relationships	2015	J CHEM INF MODEL	10.1021/ci500747n	651
SCHNEIDER G and Others	Computer-based de novo design of drug-like molecules	2005	NAT REV DRUG DISCOV	10.1038/nrd1799	618
HUANG SJ and Others	Applications of Support Vector Machine (SVM) Learning in Cancer Genomics	2018	CANCER GENOME PROTEOME	10.21873/cgp.20063	604
GUPTA S and Others	In Silico Approach for Predicting Toxicity of Peptides and Proteins	2013	PLOS ONE	10.1371/journal.pone.0073957	594
THOMFORD NE and Others	Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery	2018	INT J MOL SCI	10.3390/ijms19061578	585
LO YC and Others	Machine learning in chemoinformatics and drug discovery	2018	DRUG DISCOV TODAY	10.1016/j.drudis.2018.05.010	488
CAMACHO DM and Others	Next-Generation Machine Learning for Biological Networks	2018	CELL	10.1016/j.cell.2018.05.015	470
BALLESTER PJ and Others	A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking	2010	BIOINFORMATICS	10.1093/bioinformatics/btq112	464
ÖZTÜRK H and Others	DeepDTA: deep drug-target binding affinity prediction	2018	BIOINFORMATICS	10.1093/bioinformatics/bty593	464
RAGOZA M and Others	Protein–Ligand Scoring with Convolutional Neural Networks	2017	J CHEM INF MODEL	10.1021/acs.jcim.6b00740	428
JIMÉNEZ J and Others	Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks	2018	J CHEM INF MODEL	10.1021/acs.jcim.7b00650	421
EKINS S and Others	In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling	2007	BRIT J PHARMACOL	10.1038/sj.bjp.0707305	419
CHEN X and Others	Drug-target interaction prediction: databases, web servers and computational models	2016	BRIEF BIOINFORM	10.1093/bib/bbv066	406

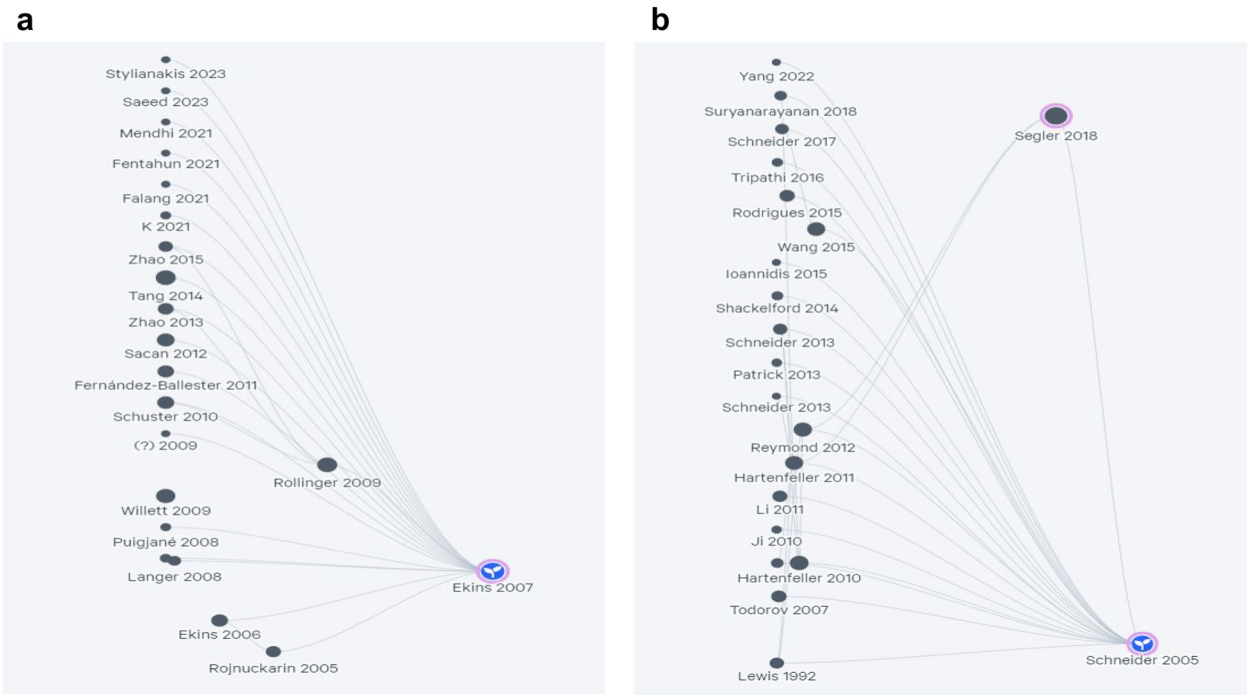

Fig. 7 **a** Ekins' Seed Maps [31]. **b:** Schneider's Seed Maps [31]

that research on topics such as "AI in Drug Discovery and Development" is conducted at the intersection of these three disciplines. It is possible to say that this research area mostly involves multidisciplinary studies, emphasizing that interdisciplinary approaches are becoming increasingly important.

Keywords

When addressing the question, "What are the most commonly used keywords?", it is essential to delve into the implications of the widespread use of terms such as "AI" and "artificial intelligence" in scientific publishing. In recent years, the hype surrounding AI has led to an increasing tendency among authors to include these terms as keywords, often to align their manuscripts with the perceived trends and scopes of high-impact journals.

For instance, some papers might emphasize AI applications in their keywords or abstracts despite limited or tangential relevance to the methodology or findings presented. Such practices risk inflating the perceived contribution of AI to the field, potentially distorting bibliometric analyses and creating a narrative more focused on the hype than on substantive advancements.

Keywords were obtained using an algorithm developed by CWTS, Leiden. These keywords are a hierarchical document-level classification system with three levels: macro-topics (10), medium-topics (326), and micro-topics (2488).

Table 4 presents the distribution of keywords used in AI and drug discovery research, categorizing them into various scientific disciplines. The dominant research areas include Chemistry, where topics like Protein Folding (775 documents) and QSAR (456 documents) highlight AI's significant role in molecular modeling and computational drug design. Similarly, Clinical & Life Sciences show a strong presence, with Gene Expression Data (377 documents) and Coronavirus-related studies (68 documents) reflecting AI's contribution to genomics and pandemic-related drug research.

Beyond these fields, AI applications extend into Electrical Engineering & Computer Science, where Deep Learning (36 documents) and Feature Selection (20 documents) demonstrate its importance in algorithm development for drug discovery. Additionally, AI is emerging in Microfluidics and Flow Chemistry, facilitating advancements in automated drug synthesis. The Physics, Mathematics, and Social Sciences categories suggest an interdisciplinary approach, where AI contributes to areas like quantum mechanics, statistical modeling, and ethical considerations in drug development.

Overall, the table illustrates AI's widespread impact across multiple disciplines, reinforcing its essential role in accelerating drug discovery. While Chemistry and Clinical Sciences dominate the landscape, emerging areas such as robotics, blockchain, and AI-driven

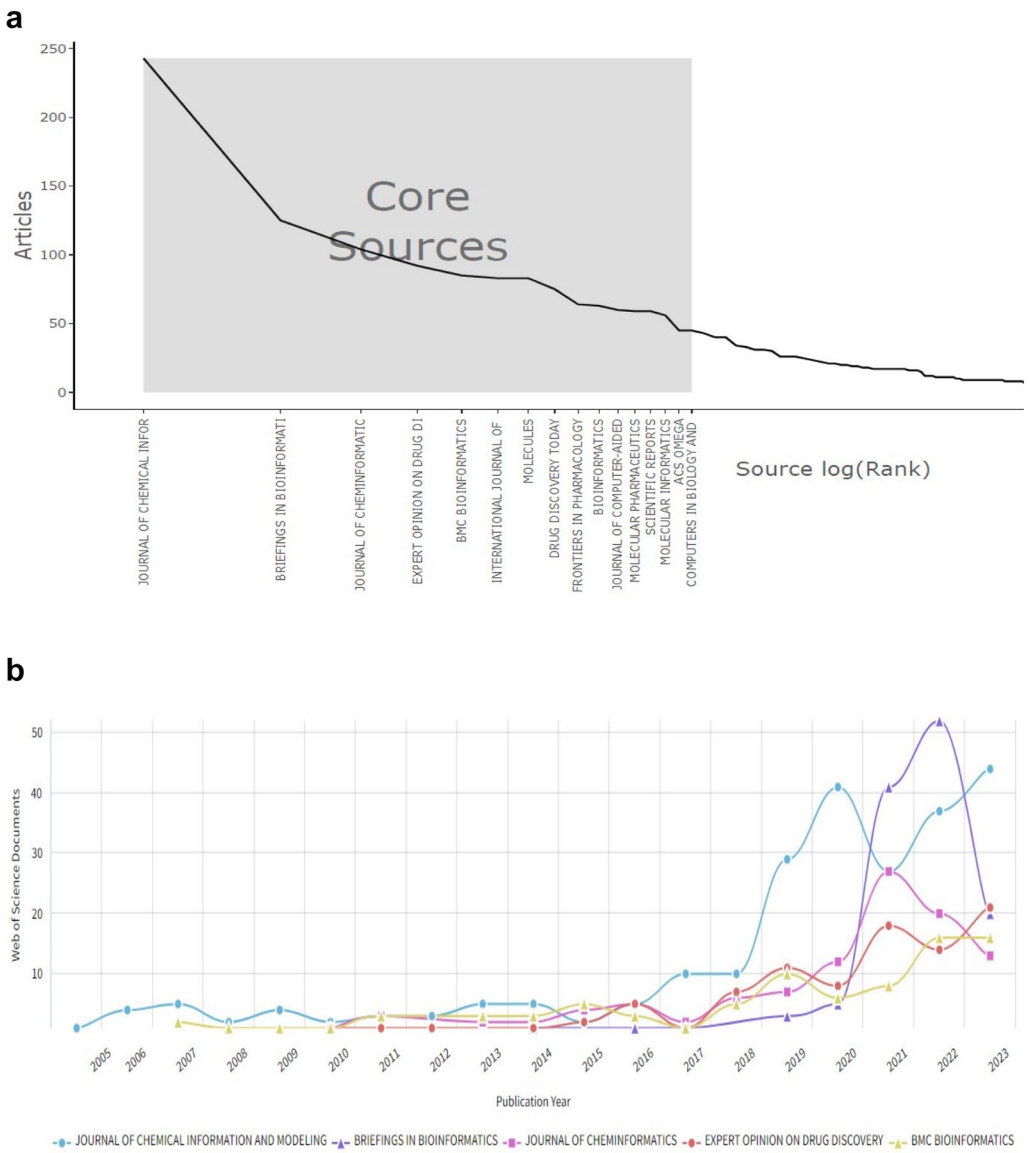

Fig. 8 **a** Distribution of scientific journals publishing the most articles in the field of AI and Drug Discovery according to Bradford's law [37]. **b** Distribution of the journals with the most articles published by year [27]

pharmacovigilance suggest new frontiers for future research and applications in pharmaceutical innovation.

In order to identify research points and future research directions in the academic field, the most frequently used author keywords were analyzed. The authors' keyword co-occurrence visualisation graph was created in this study using the Bibliometrix & R software program (Fig. 10b). The keywords used by the first 20 authors are shown. According to this, the most frequently used keywords in this field are drug discovery, prediction, design, identification, discovery, classification, docking, inhibitors, neural networks, and QSAR. The most frequent author keywords in the dataset are drug discovery, prediction, design, identification, discovery, classification, docking, and inhibitors (Fig. 10a).

Keyword co-occurrence analysis serves to understand primary themes or topics in a research area. Co-occurrence refers to the co-occurrence of two pieces of information within a dataset. Each keyword in the dataset is represented as a node, while the co-occurrence of a pair of keywords is represented as a link. The strength of this link is determined by the frequency of co-occurrence of paired keywords [38]. In this study, Author Keywords were used to perform keyword co-occurrence analysis.

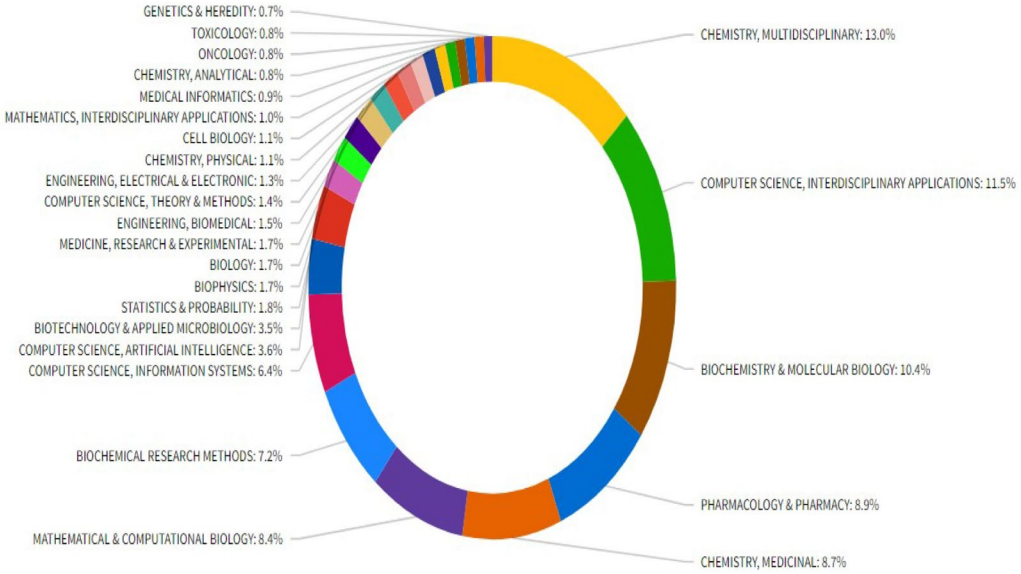

Fig. 9 Distribution of Articles According to the Top 20 WoS CC Categories. (InCites, 2024)

Author Keywords are automatically generated using a proprietary algorithm specific to Clarivate Analytics databases. Keywords associated with AIapplications in Drug Discovery research were categorized into four clusters, indicated by four colors (Fig. 10b). Each circle in the figure represents a keyword, and the lines connecting these circles show the links between keywords. Keywords sharing the same color belong to the same cluster. The size of each circle in the figure corresponds to the frequency of the keyword, with larger circles indicating higher frequency and smaller circles indicating lower frequency.

Figure 10b shows the visualization network map of the co-occurrence of author keywords. Four clusters were formed based on the occurrences. Red color indicates Cluster 1 (deep learning, neural networks, pharma, drug-target interactions, etc.); green color indicates Cluster 2 (AI, drug discovery, drug development, bioinformatics, etc.); blue color indicates Cluster 3 (machine learning, QSAR, classification, data mining, SVM, etc.); yellow color indicates Cluster 4 (artificial imaging, molecular dynamics, molecular docking, etc.).

Countries and universities

Figure 11 presents the institutional collaboration network, which visualizes the relationships and partnerships among research institutions contributing to AI-driven drug discovery. The network is composed of nodes representing institutions and edges representing collaborative ties between them. The size of each node corresponds to the institution's publication output, while the thickness of the edges reflects the strength of collaboration. Most universities are from China (see green and orange) and the USA (see blue). The leading universities are Chinese Acad Sci, Sichuan univ, and Zhejiang University. These institutions are notable for their high activity and collaborations in the field of AI papers related to Drug Discovery. The USA is the second largest cluster developed in the cooperation network. Harward Univ, Pittsburgh Univ, and the University of California are the leading universities. The Chinese Academy of Sciences and Harvard University emerge as central nodes in the network, reflecting their pivotal roles in driving research and fostering international collaborations. The network reveals strong regional collaboration patterns, particularly within China and the United States, as well as cross-regional partnerships between Europe, Asia, and North America. The institutional collaboration network highlights the importance of collaborative research in advancing AI applications in drug discovery. The strong ties between leading institutions, particularly in China and the United States, underscore the critical role of international partnerships in addressing complex challenges in drug development. Additionally, the growing contributions of institutions from emerging research hubs, such as South Korea and India, suggest a broadening geographic distribution of expertise in this field.

Figure 12 shows the top 20 institutions in terms of publications. The co-authorship analysis showed that 51 institutions published more than five articles.

Table 4 Distribution of keywords used in articles in the field of AI and Drug Discovery (Wos CC, 2024)

Fields	Total WoS CC Documents
2 Chemistry	
2.123 Protein Structure, Folding & Modelling	
2.123.13 Protein Folding	775
2,123,778 QSAR	456
2.160 Microfluidic Devices & Superhydrophobicity	
2.160.2038 Flow Chemistry	28
2.160.174 Microfluidics	7
2.190 Surfactants, Lipid Bilayers & Antimicrobial Peptides	
2.190.857 Antimicrobial Peptides	33
2.190.254 Lipid Bilayers	2
2.15 Physical Chemistry	
2.15.123 Bulk Modulus	16
2.15.3 CCSD	5
2.1 Synthesis	
2.1.1402 Click Chemistry	6
2.1.549 Chalcones	6
1 Clinical & Life Sciences	
1.54 Molecular & Cell Biology—Genetics	
1.54.79 Gene Expression Data	377
1.54.2447 Bromodomains	31
1.104 Virology—General	
1.104.1353 Coronavirus	68
1.155 Medical Ethics	
1.155.1510 Orphan Drugs	38
1.155.2316 Pharmacovigilance	29
1.117 Pharmacology & Toxicology	
1.117.179 Cytochrome P450	26
1.117.1309 Acetaminophen	17
1.25 Molecular & Cell Biology—Cancer, Autophagy & Apoptosis	
1.25.887 RAS	9
1.25.782 Proteasome	6
4 Electrical Engineering, Electronics & Computer Science	
4.17 Computer Vision & Graphics	
4.17.1901 Genome Rearrangement	38
4.17.128 Deep Learning	36
4.61 AI & Machine Learning	
4.61.145 Feature Selection	20
4.61.1335 Self Organising Maps	5
4.61.869 Clustering	5
4.48 Knowledge Engineering & Representation	
4.48.120 Complex Networks	10
4.48.1522 Big Data	1
4.116 Robotics	
4.116.1415 Human–Robot Interaction	4
4.116.862 Reinforcement Learning	4
4.187 Security Systems	
4.187.1702 Differential Privacy	5
4.187.12014 Blockchain	2

Table 4 (continued)

Fields	Total WoS CC Documents
3 Agriculture, Environment & Ecology	
3.198 Mycotoxins	
3.198.904 Streptomyces	15
3.198.926 Endophytic Fungi	2
3.16 Phytochemicals	
3.16.2062 Aloe Vera	2
3.16.2318 Silymarin	2
3.180 Microbial Biotechnology	
3.180.2188 Prolidase	4
3.180.1338 Phosphofructokinase	2
3.85 Food Science & Technology	
3.85.1687 Bioactive Peptides	4
3.85.741 Chitosan	1
3.220 Smell & Taste Science	
3.220.1242 Aspartame	4
5 Physics	
5.56 Quantum Mechanics	
5.56.1673 Geometric Phase	6
5.56.9 Entanglement	6
5.221 Nuclear Instruments	
5.221.1034 Electron Tomography	6
5.9 Particles & Fields	
5.9.746 Gravitational Waves	1
5.250 Imaging & Tomography	
5.250.1881 Photoacoustic Imaging	1
5.88 Electromagnetism	
5.88.418 X-ray Reflectivity	1
5.310 Resistive Switching	
5.310.1164 Resistive Switching	1
5.98 Geometrical Optics	
5.98.1073 Digital Holography	1
5.33 Semiconductor Physics	
5.33.632 Diluted Magnetic Semiconductors	1
5.38 Optical Electronics & Engineering	
5.38.198 Silicon Photonics	1
9 Mathematics	
9.92 Statistical Methods	
9.92.1271 Competing Risks	9
9.92.1337 Causal Inference	4
9.162 Numerical Methods	
9.162.1864 Logistic Source	3
9.143 Dynamical Systems & Time Dependence	
9.143.1163 Stochastic Resonance	1
9.50 Applied Statistics & Probability	
9.50.1564 Convex Body	1
6 Social Sciences	
6.256 Religion	
6.256.2361 Religious Education	4
6.3 Management	

Table 4 (continued)

Fields	Total WoS CC Documents
6.3.2 Knowledge Management	2
6.3.1731 Computer-supported Cooperative Work	1
6.3.2135 Sharing Economy	1
6.238 Bibliometrics, Scientometrics & Research Integrity	
6.238.166 Bibliometrics	2
6.238.1790 Plagiarism	1
6.185 Communication	
6.185.1644 Privacy	2
6.73 Social Psychology	
6.73.2034 Conditional Reasoning	1
6.73.2331 Health Locus of Control	1
6.146 Anthropology	
6.146.1728 Artisanal Mining	2
7 Engineering & Materials Science	
7.12 Metallurgical Engineering	
7.12.2236 High-Entropy Alloys	2
7.12.2224 Ductile Iron	1
7.272 Electrical—Solder & Connections	
7.272.2124 Thermal Resistance	2
7.262 Explosives	
7.262.2024 TNT	1
7.262.745 Energetic Materials	1
7.121 Concrete Science	
7.121.26 Compressive Strength	1
7.70 Thermodynamics	
7.70.1880 Falling Film	1
7.192 Testing & Maintenance	
7.192.732 Polynomial Chaos	1
7.251 Electrical—Harvesting & Discharging	
7.251.1204 Energy Harvesting	1
8 Earth Sciences	
8.8 Geochemistry, Geophysics & Geology	
8.8.2206 Compositional Data	1
10 Arts & Humanities	
10.126 Philosophy	
10.126.1568 Philosophy of Science	1

Figure 13 shows that only six clusters are formed; purple indicates Cluster 1 (France, Germany, China, etc.); yellow indicates Cluster 2 (Switzerland, China, UK, etc.); blue indicates Cluster 3 (USA, China, Germany, India, Brazil, Russia, etc.); red indicates Cluster 3 (South Korea, India, Saudi Arabia, etc.); green indicates Cluster 5 (UK, Sweden, Denmark, etc.).

Figure 14 presents the top 20 countries according to the number of articles published, categorized as Multi-Country Publication (MCP) and Single Country Publication (SCP). The multi-country publication refers to joint studies with authors from different countries and indicates international cooperation. In contrast, single-country publications refer to studies in which all authors belong to the same country and indicate domestic cooperation. American authors ranked first with 923 articles, comprising 716 single-country publications and 207 multi-country publications. They rank second in productivity with 853 articles, comprising 669 single-country publications and 184 multi-country publications, resulting in an MCP rate of 21.6 percent. See Fig. 14 and Table 5 for more detailed information.

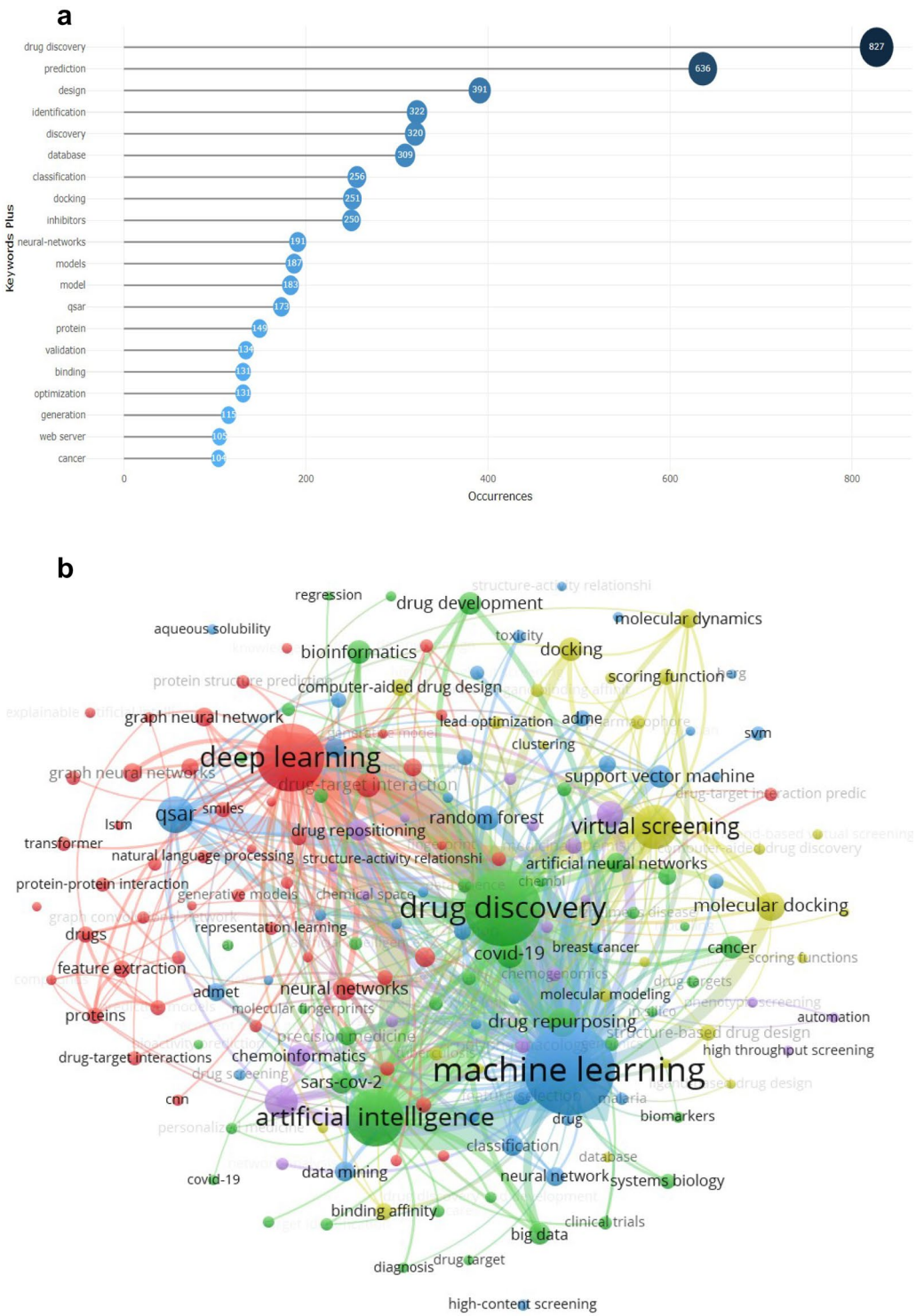

Fig. 10 **a** Most Frequently Used Author Keywords in AI and Drug Discovery Literature, 1990–2023. (Bibliometrix & R software, 2024). **b** Co-occurrence network map of author keywords occurring 10 + times in AI and Drug Discovery literature, 1990–2023. (VOSviever, 2023)

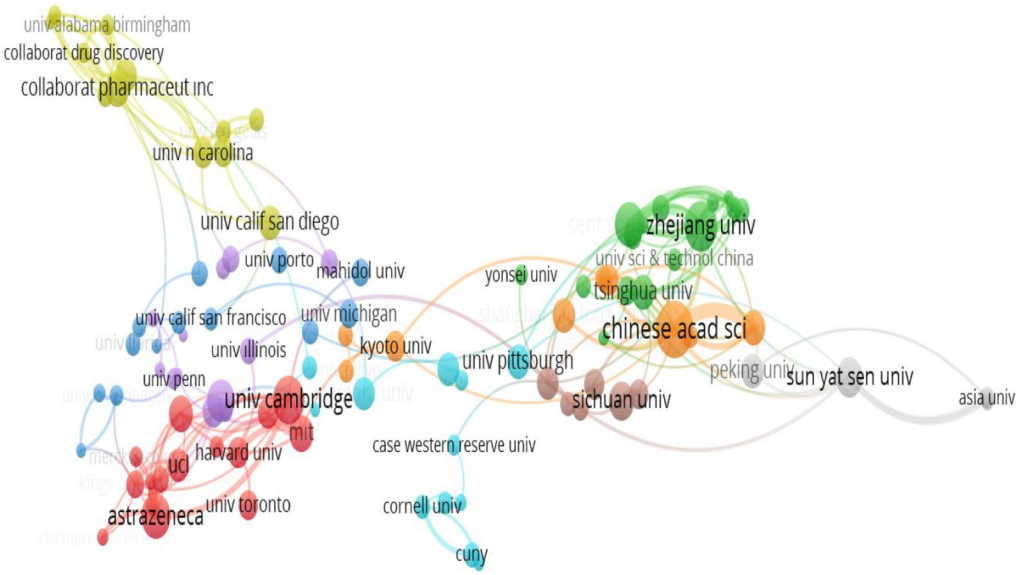

Fig. 11 Institutional network among articles whose authors are affiliated with institutions. (VOSviever, 2024)

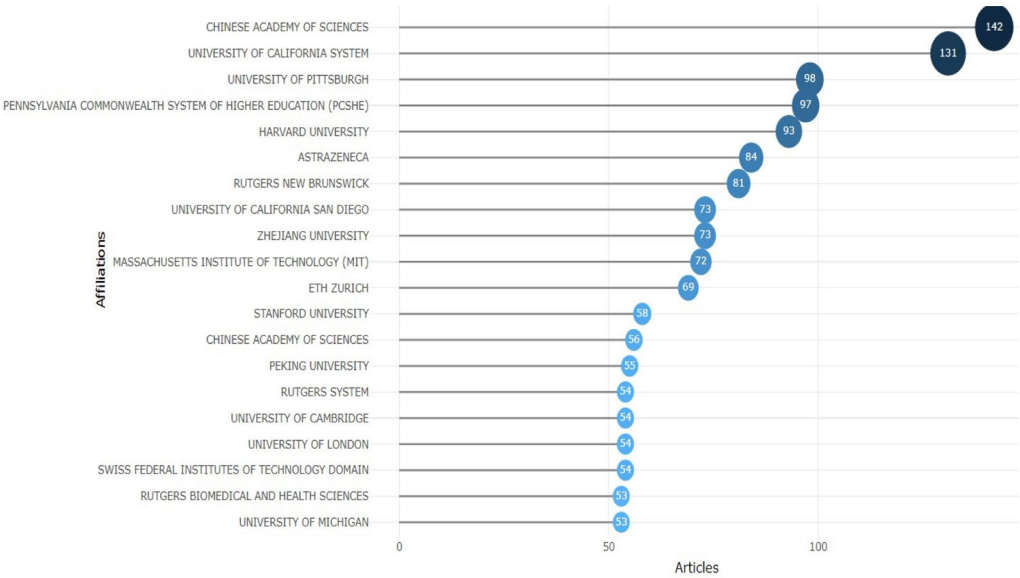

Fig. 12 Distribution of authors' affiliation with organizations according to the selected dataset (top 20 organizations). (Bibliometrix & R software, 2023)

Figure 14 also shows that individuals from the same countries author the majority of publications. This trend may be due to authors' preferences to collaborate within their research groups or with academics from the same national background.

Funding organisations

According to the results presented in Table 6, the National Natural Science Foundation of China (NSFC), with 530 studies, and the United States Department of Health & Human Services, with 378 studies, were the organizations with the highest support for the publication of scientific research on drug discovery. The National Institutes of Health (NIH)—USA, National Science Foundation (NSF), European Union (EU), UK Research & Innovation (UKRI), National Research Foundation of Korea, Spanish Government, NIH National Institute of General Medical Sciences (NIGMS) are other important

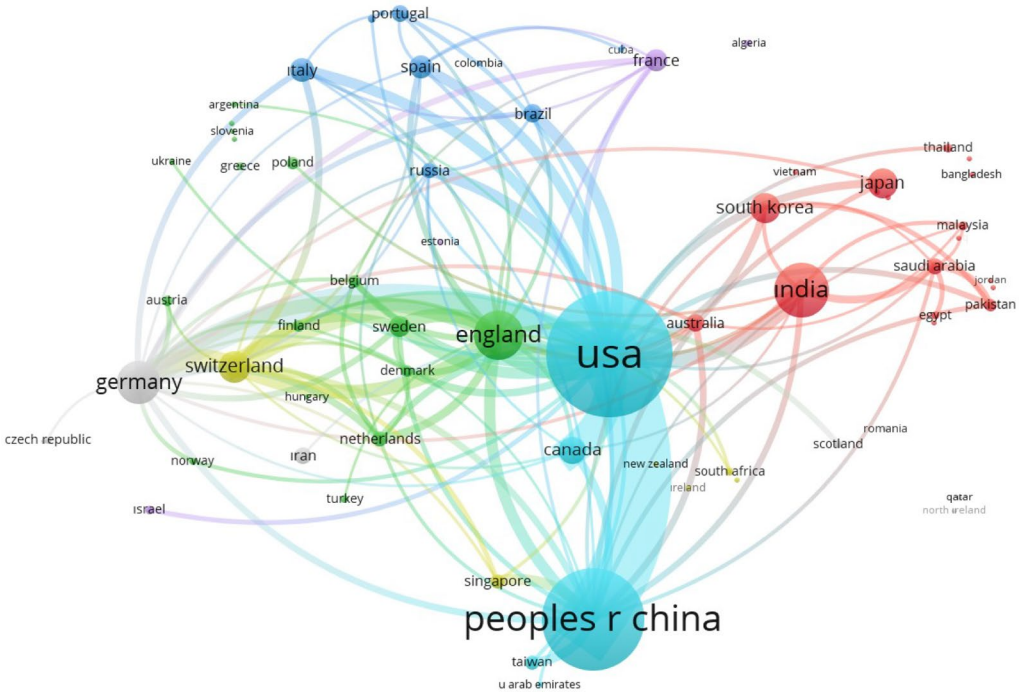

Fig. 13 Visualisation network map of co-authored articles by country. (Wosviever, 2024)

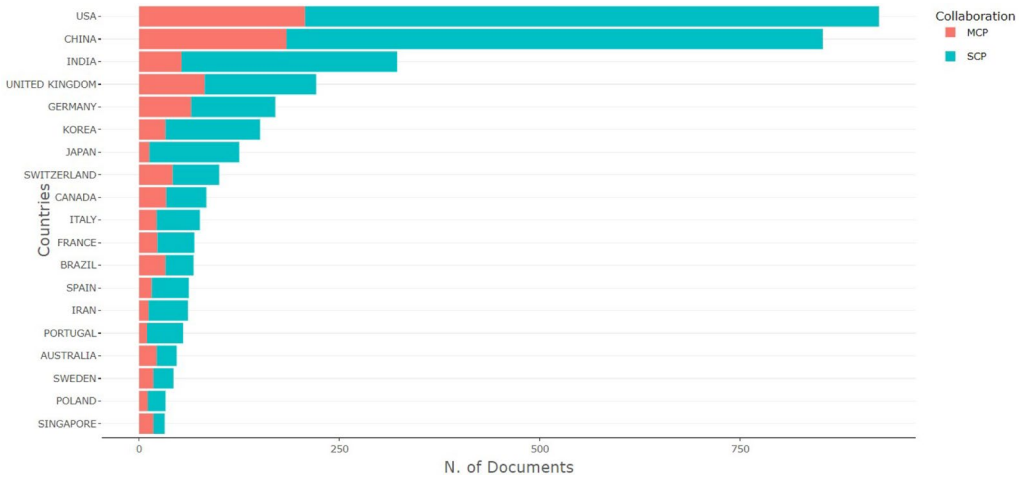

Fig. 14 Most productive countries: Single Country Publications (SCP), Multiple Country Publications (MCP). (Bibliometrix & R software, 2023)

organizations that have supported the significant development of AI for drug discovery.

As shown in Fig. 15, countries started to support AI studies in Drug Discovery in 2006. This rate started to increase significantly after 2017. The first support was provided by UK Research & Innovation (UKRI) in early 2006. Afterward, the National Institutes of Health (NIH)- USA took the lead in supporting studies in this field. After 2015, China took the lead with the National Natural Science Foundation of China (NSFC) funding institution. Figure 15 shows that NSFC supports these studies at significantly higher rates than other funding organizations. Notably, the European Union (EU) lags behind the expected support. The previous graphs and tables show how extensively NSFC has invested in this area.

Table 5 Country of the responsible author and cross-country cooperation (Bibliometrix & R software, 2023)

Country	Articles	SCP	MCP	Freq	MCP_Ratio
USA	923	716	207	0.227	0.224
CHINA	853	669	184	0.21	0.216
INDIA	322	269	53	0.079	0.165
UNITED KINGDOM	221	139	82	0.054	0.371
GERMANY	170	105	65	0.042	0.382
KOREA	151	118	33	0.037	0.219
JAPAN	125	112	13	0.031	0.104
SWITZERLAND	100	58	42	0.025	0.42
CANADA	84	50	34	0.021	0.405
ITALY	76	54	22	0.019	0.289
FRANCE	69	46	23	0.017	0.333
BRAZIL	68	35	33	0.017	0.485
SPAIN	62	46	16	0.015	0.258
IRAN	61	49	12	0.015	0.197
PORTUGAL	55	45	10	0.014	0.182
AUSTRALIA	47	25	22	0.012	0.468
SWEDEN	43	25	18	0.011	0.419
POLAND	33	22	11	0.008	0.333
SINGAPORE	32	14	18	0.008	0.563
AUSTRIA	26	12	14	0.006	0.538

Discussion

The analysis reveals a dramatic increase in publications after 2014, with a peak in 2022, followed by a slight decline. This trend aligns with the broader adoption of AI technologies, particularly deep learning and machine learning, in pharmaceutical research. The surge in publications after 2014 can be attributed to advancements in computational power, the availability of large-scale biomedical datasets, and the development of sophisticated AI algorithms capable of handling complex drug discovery tasks [2, 3]. The slight decline after 2022 may reflect a maturation of the field, where initial rapid growth is stabilizing as researchers focus on refining existing methodologies rather than exploring entirely new approaches.

The most prolific authors, such as Ekins, Schneider, Hou Tj, and Cao Ds, have significantly shaped the field through their extensive contributions. These authors are often associated with leading institutions such as the Chinese Academy of Sciences and the University of California, which dominate the collaboration networks. The prominence of these institutions underscores the importance of resource availability and interdisciplinary collaboration in advancing AI-driven drug discovery. The strong collaboration between Chinese and U.S. institutions highlights the global nature of research in this field, with both regions leveraging their unique strengths to drive innovation.

The keyword analysis identifies "protein folding," "QSAR," "gene expression data," "coronavirus," and "genome rearrangement" as the most frequently mentioned topics. These themes reflect the diverse applications of AI in drug discovery, from predicting protein structures to analyzing genomic data and accelerating drug development for emerging diseases. The prominence of "protein folding" and "QSAR" underscores the critical role of AI in addressing fundamental challenges in drug discovery, such as understanding molecular interactions and optimizing drug candidates. The focus on "coronavirus" highlights the rapid response of the scientific community to global health crises, leveraging AI to accelerate vaccine and therapeutic development during the COVID-19 pandemic.

International collaboration accounts for 28.06% of the publications, with the United States and China leading in both productivity and influence. This finding aligns with previous studies that emphasize the importance of global partnerships in addressing complex scientific challenges [10, 11]. The collaboration networks reveal distinct regional clusters, with strong intra-regional collaborations in China and the United States, as well as cross-regional partnerships between Europe, Asia, and North America. These collaborations facilitate the exchange of knowledge and resources, enabling researchers to tackle multifaceted problems in drug discovery more effectively.

Limitations

While this study provides valuable insights into the field of AI-driven drug discovery, it is not without limitations. First, the reliance on the Web of Science Core Collection (WoS CC) as the primary data source may introduce bias, as WoS CC tends to favor English-language publications and high-impact journals. The exclusion of other databases, such as Scopus and PubMed, may result in the omission of relevant studies, particularly from non-English speaking countries and emerging research hubs.

Second, the bibliometric analysis is inherently retrospective, focusing on past trends and patterns. While this provides a comprehensive overview of the field's evolution, it may not fully capture emerging trends or predict future directions. Future studies could complement bibliometric analysis with qualitative methods, such as expert interviews or case studies, to provide a more nuanced understanding of the field.

Finally, the study's focus on AI in drug discovery may overlook broader applications of AI in healthcare and biomedical research. Future research could explore the intersection of AI with other fields, such as personalized medicine, clinical trials, and healthcare delivery, to provide a more holistic view of AI's impact on the pharmaceutical industry.

Table 6 Distribution of articles according to funding international organizations [27]

Name	WoS CC Documents	Times Cited	Documents in Q1 Journals	Documents in Q2 Journals	Documents in Q3 Journals	Documents in Q4 Journals
National Natural Science Foundation of China (NSFC)	530	11,081	256	77	24	6
United States Department of Health & Human Services	378	12,810	217	63	16	4
National Institutes of Health (NIH)—USA	377	12,803	217	63	16	4
National Science Foundation (NSF)	173	26,262	80	21	5	2
European Union (EU)	100	3331	49	16	8	0
UK Research & Innovation (UKRI)	92	3078	53	16	4	1
National Research Foundation of Korea	91	1761	46	13	2	1
Spanish Government	74	2058	46	11	2	1
NIH National Institute of General Medical Sciences (NIGMS)	73	2347	42	15	2	2
Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT)	65	1377	25	12	2	1
Fundamental Research Funds for the Central Universities	62	1278	31	10	2	1
Japan Society for the Promotion of Science	59	1347	23	12	1	1
Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPQ)	55	2029	26	19	5	2
Grants-in-Aid for Scientific Research (KAKENHI)	53	1182	19	12	1	1
Engineering & Physical Sciences Research Council (EPSRC)	52	1573	36	8	2	0
Swiss National Science Foundation (SNSF)	48	2822	24	6	6	2
United States Department of Defence	47	2190	26	5	1	0
Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES)	46	1416	26	12	3	1
United States Department of Energy (DOE)	45	1219	24	9	2	0
German Research Foundation (DFG)	38	766	21	3	4	0

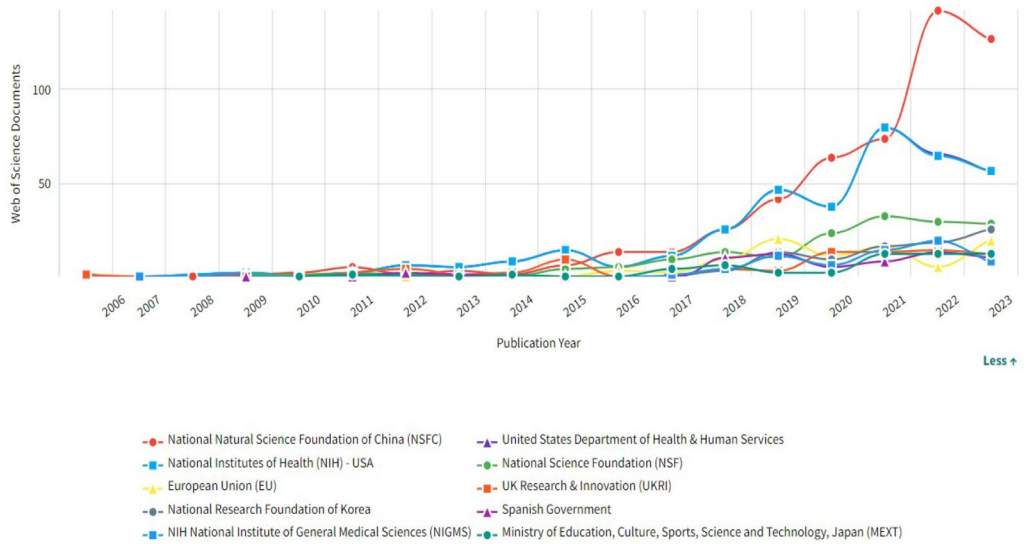

Fig. 15 Changes in the Distribution of Articles by Supporting Institutions over the Years [27]

Conclusion

This study provides a comprehensive bibliometric analysis of the application of artificial intelligence (AI) in drug discovery and development, covering the period from 1990 to 2023. By analyzing 4059 scientific publications, this research sheds light on the evolution, trends, and key contributors in this rapidly growing field. The findings reveal that the field has experienced exponential growth since 2014, driven by advancements in deep learning, machine learning, and the availability of large-scale biomedical datasets. The number of publications peaked in 2022, reflecting the increasing adoption of AI technologies in pharmaceutical research. The average number of citations per document (28.62) indicates the high impact of research in this domain, with influential works such as [29, 3] shaping the field.

The study identifies the most prolific authors, including Ekins, Schneider, Hou Tj, and Cao Ds, who have significantly contributed to the field, with their work accounting for nearly 30% of the total publications. Leading institutions such as the Chinese Academy of Sciences, Harvard University, and the University of California dominate the collaboration networks, highlighting the importance of interdisciplinary and international collaboration in advancing AI-driven drug discovery. The keyword analysis reveals that "protein folding," "QSAR," "gene expression data," "coronavirus," and "genome rearrangement" are the most frequently mentioned topics. These themes reflect the diverse applications of AI in drug discovery, from molecular modeling to genomics and pandemic response. The prominence of "coronavirus" underscores the critical role of AI in addressing global health crises, such as the COVID-19 pandemic, by accelerating drug and vaccine development.

International collaboration accounts for 28.06% of the publications, with the United States and China leading in both productivity and influence. This highlights the global nature of research in AI-driven drug discovery and the importance of cross-border partnerships. Key funding organizations, such as the National Natural Science Foundation of China (NSFC) and the United States Department of Health & Human Services, have played a pivotal role in supporting advancements in this field. The findings align with theoretical frameworks such as Lotka's Law demonstrating the concentration of productivity among a small subset of prolific authors and institutions. The study provides practical insights for researchers, industry professionals, and policymakers, emphasizing the need for continued investment in AI infrastructure, interdisciplinary collaboration, and global partnerships to drive innovation in drug discovery.

In summary, this study highlights the transformative potential of AI in revolutionizing drug discovery and development. By identifying key trends, influential contributors, and emerging research themes, this research offers a roadmap for future investigations and underscores the importance of leveraging AI technologies to address complex challenges in the pharmaceutical industry. As the field continues to evolve, further research is needed to explore new applications of AI, such as personalized medicine and clinical trial optimization, to fully realize its potential in improving global health outcomes.

Acknowledgements
The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

Institutional review board statement
Not applicable.

Author contributions
"M.K wrote the main manuscript text and prepared figures. "Z.A reviewed the manuscript".

Data availability
All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
No ethical approval was needed because this is not a human study, but only online information was used.

Research involving human participants and/or animals
None.

Informed consent
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Faculty of Medicine, Department of Medical Informatics, Baskent University, Ankara, Turkey.

Received: 12 September 2024 Accepted: 16 March 2025
Published online: 08 May 2025

References

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 9(3):203–214. https://doi.org/10.1038/nrd3078
2. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discov Today 23(6):1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039
3. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Pocock M (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18(6):463–477. https://doi.org/10.1038/s41573-019-0024-5
4. McCarthy J, Minsky ML, & ML N. (1956). Rochester, and CE Shannon (1955). Proposal for the 1956 Dartmouth Summer Research Project on Artificial Intelligence.
5. Buchanan BG, Shortliffe EH (1984) Rule based expert systems: the mycin experiments of the stanford heuristic programming project (the

Addison-Wesley series in artificial intelligence). Addison-Wesley Longman Publishing Co. Inc, Reading
6. Mak KK, Pichika MR (2019) AI in drug development: present status and future prospects. Drug Discov Today 24(3):773–780. https://doi.org/10.1016/j.drudis.2018.11.014
7. Lyu J, Wang S, Balius TE, Singh I, Levit A, Torabkhani Y, Yang X (2019) Ultra-large library docking for discovering new chemotypes. Nature 566(7743):224–229. https://doi.org/10.1038/s41586-019-0917-9
8. Topol EJ (2019) High-performance medicine: the convergence of human and AI. Nat Med 25:44–56. https://doi.org/10.1038/s41591-018-0300-7
9. Chan HS, Shan H, Dahoun T, Vogel H, Yuan S (2019) Advancing drug discovery via AI. Trends Pharmacol Sci 40(8):592–604. https://doi.org/10.1016/j.tips.2019.05.004
10. Tran BX, Nghiem S, Sahin O, Vu TM, Ha GH, Vu GT, Ho CS (2021) Artificial intelligence in drug discovery: Bibliometric analysis of its research development. J Inf 15(3):101566. https://doi.org/10.1016/j.joi.2021.101566
11. Zhang L, Zhang Y, Chen X, Zhang H (2022) A bibliometric analysis of artificial intelligence in drug discovery and development from 2000 to 2021. Expert Opin Drug Discov 17(8):813–824. https://doi.org/10.1080/17460441.2022.2096016
12. Aria M, Cuccurullo C (2017) bibliometrix: an R tool for comprehensive science mapping analysis. J Informet 11(4):959–975. https://doi.org/10.1016/j.joi.2017.08.007
13. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct bibliometric analysis? An overview and guidelines. J Bus Res 133:285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
14. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. J Informet 5(1):146–166. https://doi.org/10.1016/j.joi.2010.10.002
15. Russell SJ, Norvig P (2022) AI: a modern approach (4th ed. Global edition. ed.). Pearson Education Limited.
16. Koçak M, Akçalı Z. Development trends and knowledge framework of artificial intelligence (AI) applications in oncology by years: a bibliometric analysis from 1992 to 2022. *Discov Onc* 15, 566 (2024). https://doi.org/10.1007/s12672-024-01415-0
17. Pacurari AC, Bhattarai S, Muhammad A, Avram C, Mederle AO, Rosca O, Mavrea A (2023) Diagnostic accuracy of machine learning AI architectures in the detection and classification of lung cancer: a systematic review. Diagnostics. https://doi.org/10.3390/diagnostics13132145
18. Huang MH, Rust RT (2018) AI in service. J Serv Res 21(2):155–172. https://doi.org/10.1177/1094670517752459
19. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260. https://doi.org/10.1126/science.aaa8415
20. Yang X, Wang Y, Byrne R, Schneider G, Yang S (2020) The landscape of artificial intelligence in drug development: bibliometric analysis of trends in 2014–2019. Front Pharmacol 11:1551. https://doi.org/10.3389/fphar.2020.01551
21. Snyder H (2019) Literature review as a research methodology: An overview and guidelines. J Bus Res 104:333–339
22. Hood WW, Wilson CS (2001) Bibliometrics, scientometrics and informetrics literature [Review]. Scientometrics 52(2):291–314. https://doi.org/10.1023/a:1017919924342
23. Broadus RN (1987) Towards a definition of bibliometrics. Scientometrics 12(5–6):373–379. https://doi.org/10.1007/bf02016680
24. Andres A (2009) Measuring academic research: how to conduct a bibliometric study. Measuring academic research: how to conduct a bibliometric study. Chandos Publishing, Kidlington, pp 1–169
25. Martín-Martín A, Orduna-Malea E, López-Cózar ED (2018) A new method to delineate academic disciplines through Google Scholar Citations: the case of bibliometrics. Scientometrics 114(3):1251–1273. https://doi.org/10.1007/s11192-017-2587-4
26. WoS CC (2022). https://www.webofscience.com/wos/woscc/summary/d395368a-6372-4770-ba65-01c2a114dea3-b5fc6c6e/times-cited-descending/1. Accessed 21 Nov 2023.
27. InCites (2023). InCites. https://incites.clarivate.com/. Accessed 21 Nov 2023.
28. Bibliometrix (2022). Bibliometrix. https://bibliometrix.org/biblioshiny/biblioshiny1.html. Accessed 21 Nov 2023.
29. Van Eck NJ, Waltman L (2010) Software research: VOSviewer, a computer programme for bibliometric mapping. Scientometrics 84(2):523–538. https://doi.org/10.1007/s11192-009-0146-3
30. Laskaris R (2015) AI: a modern approach 3rd ed. Libr J 140(6):45–45
31. Rowlands I (2005) Emerald authorship data, Lotka's law and research productivity. ASLIB Proc 57(1):5–10. https://doi.org/10.1108/00012530510579039
32. Ballester PJ, Mitchell JB, Boobis AR. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics, 26(9), 1169–1175. https://doi.org/10.1093/bioinformatics/btq112
33. LeCun Y, Bengio Y, Hinton G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
34. Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: Deep drug-target binding affinity prediction. Bioinformatics, 34(17), i821–i829. https://doi.org/10.1093/bioinformatics/bty593
35. Kaur A, Gulati S, Sharma R, Sinhababu A, Chakravarty R (2022) Visual citation navigation of open educational resources using Litmaps. Libr Hi Tech News 39(5):7–11
36. Schneider G. (2005). Computer-based de novo design of drug-like molecules. Nature Reviews Drug Discovery, 4(8), 649–663. https://doi.org/10.1038/nrd1799
37. Garfield E (1980) Bradford's law and related statistical patterns. Curr Contents 19:5–12
38. Radhakrishnan S, Erbis S, Isaacs JA, Kamarthi S (2017) Novel keyword co-occurrence network-based methods for promoting systematic reviews of scientific literature. PLoS ONE 12(3):e0172778. https://doi.org/10.1371/journal.pone.0172778

Publisher's Note