The Current Application Status of Big Data Information Technology in Higher Education: A Bibliometric Analysis from 2013 to 2023

Hailong Niu

Accounting and Finance, Shandong Business Institute, sdbi, Yantai, China

`haroldniu@163.com`

Abstract. For a more comprehensive and clear understanding of the historical development and current status of big data information technology in higher education, as well as a more precise judgment of future development directions, this paper conducts bibliometric and visualization analyses of research on big data informationization in higher education from 2013 to 2023 using VOSviewer and CiteSpace software. Combining the strengths of VOSviewer and CiteSpace in quantitative analysis, the paper generates visualizations such as author analysis tables, journal analysis tables, and keyword network visualization graphs. These analyses serve as a foundational reference for future researchers in this field.

Keywords: higher Education; big data; information technology.

1   Introduction

In today's era of rapid development in information technology, our society has entered a new epoch characterized by digitization and big data. The prominent features of this era include the rapid generation of data, a vast array of types, and the increasingly complex nature of its properties. Faced with such a scenario, higher education has had to make corresponding adjustments and adaptations. Advances in educational technology, such as online learning platforms and learning management systems, have not only accumulated a massive amount of data on student learning behaviors and academic performance but have also provided rich materials for big data analysis. These analyses can assist higher education institutions in gaining deeper insights into the teaching process and students' learning outcomes, thereby effectively enhancing the quality and effectiveness of education. The application of big data allows us to make real-time adjustments to course design and teaching methods to better cater to the needs of students. Some scholars have also proposed that big data can be applied not only to assist students and teachers but also to predict and prevent the occurrence of student dangers, ensuring student safety [1].

Higher education institutions can more accurately grasp students' learning dynamics and needs, optimizing teaching programs to enhance the overall quality of education,

Personalized education services based on big data can provide tailored learning support for each student, fully tapping into their potential and enhancing learning efficiency. Additionally, big data analysis provides scientific decision-making tools for university management, enabling them to accurately understand the school's operational status, allocate resources reasonably, and improve management efficiency. Furthermore, big data technology has brought abundant data resources to the field of educational research, facilitating in-depth exploration of educational phenomena and patterns, and promoting the depth of educational research. The integration of educational models with big data technology can cultivate professionals who are more adaptable to future societal employment demands, enhancing students' competitiveness in the workplace. Some scholars also indicate that big data brings risks of privacy disclosure [2]. Other scholars suggest that in the face of the development of information and big data, teachers and students must master relevant technologies, which may pose a learning burden on them [3].

2 Methodology

2.1 Research Methods

This study adopts a combined approach of bibliometrics and scientometrics for research. Bibliometrics focuses on the quantitative analysis of literature, such as the quantity and citation frequency, to assess the current status of a field, a method widely recognized by many scholars [4]. This approach transcends disciplinary boundaries, revealing interactions and collaborations across different domains. Simultaneously, it scrutinizes the research activities, collaboration relationships, and academic impact of authors and institutions, providing a basis for evaluating their academic standing [5]. The strength of bibliometrics lies in its objectivity, as it reduces subjective bias through quantitative analysis while handling large amounts of data, comprehensively revealing research dynamics [6].

On the other hand, scientometrics emphasizes citation analysis, exploring citation patterns among documents to evaluate the influence of research. It focuses on the quality and significance of academic achievements, employing indicators like impact factor to assess the academic standing of journals. Scientometric analysis helps understand the hotspots and frontiers of research fields, providing objective grounds for the evaluation of academic quality, selection of research directions, journal assessments, as well as research investment and policy formulation. In summary, bibliometrics and scientometrics, as research tools based on data analysis, offer an objective perspective for the academic community to understand research progress and impact. However, they are not flawless evaluation systems and may not delve deeply into single-article studies that have a significant impact on a specific research area.

2.2 Research Materials

To ensure the comprehensiveness and reliability of the data analysis in this study, Web of Science was chosen as the data source, specifically from Science Citation Index

Expanded (SCI-EXPANDED) and Social Sciences Citation Index (SSCI). The selected index keywords and strategy were TS = ("higher education" and ("big data" or "information technology")), with a time range from January 1, 2013, to December 1, 2023. The document types included in the search were Articles and Review Articles, and the language chosen was English. The search yielded a total of 892 results. To ensure the precision of this study, a data cleansing process was performed, excluding articles that did not align with the research theme and eliminating redundant review articles.

Ultimately, the researchers selected a total of 334 valid articles from the initial pool of 892 articles, resulting in a selection rate of 37.4%. The papers used in this study on the subject of big data informationization in higher education came from 71 countries, 545 organizations, involved 967 authors, and were published in 123 different journal sources. These selected articles collectively cited 16,233 references.

3   SCIENTOMETRICS and VISUAL ANALYSIS

3.1   Publication Analysis of Big Data Informationization in Higher Education

Table 1. Time of trend of the publications on Big Data Information Technology in Higher Education

Year of publication	Number of Documents
2013	2
2014	7
2015	10
2016	11
2017	20
2018	18
2019	22
2020	38
2021	58
2022	99
2023	50

Firstly, this paper conducts a year-by-year publication analysis of the 334 selected papers. As indicated in Table I, it is evident that the research field of big data informationization in higher education attracted attention starting from 2017 and subsequently experienced rapid development. The year 2022 emerged as the most prominent in the past decade for this field. It is noteworthy that although the data collection for this paper concludes on December 1, 2023, the current trends in research activity in the field for the current year appear to be comparable to those observed in 2021.

3.2   Bibliometric Analysis of Authors

Analyzing authors in the field of big data informationization in higher education using VOSviewer software, the criteria for selecting key authors in this domain include

publishing three or more articles. A total of four authors meeting these criteria were identified.

Table 2. Key Authors in the Field of Big Data Informationization in Higher Education Research

Autuor Name	Documents	Citations	Average Citation/Publicafion
alyoussef,ibrahim youssef	3	121	40.33
salloum,said a.	3	29	9.67
tsai,sang-bing	3	5	1.67
wang,jing	3	1	0.33

From Table II, it can be observed that there is not much difference in the number of publications among authors in this research field. It is noteworthy that the top-ranked author, Ibrahim Youssef Alyoussef, has an average citation rate per article of 40.33, significantly higher than the other three authors. Further investigation into this author reveals a primary research focus on issues related to Massive Open Online Courses (MOOCs) [7].

3.3   Bibliometric Analysis of Journals

Conducting an analysis of journal inclusion in the research on big data informationization in higher education using VOSviewer software, a selection was made among the targeted 123 journals. The criterion for selection was set at seven or more publications. As a result, a total of nine journals met the specified criteria.

Table 3. The Top 9 Journals in Big Data Informationization in Higher Education

Source	Documents	Citations	Average Citation
education and information technologies	34	743	21.85
leee access	10	95	9.5
computers in human behavior	7	894	127.71
sustainability	29	266	9.17
journal of computing in higher education	7	615	87.85
international journal of educational technology in higher education	9	232	25.77
international review of research in open and distributed learning	7	499	71.28
mobile information systems	15	4	0.267
wireless communications &mobile computing	14	8	0.57

Observing from Table III, it can be noted that among the selected nine journals, the one with the most publications is "Education and Information Technologies," totaling 34 articles with an average citation count of 21.85. Of particular interest is the journal

ranked third, "Computers in Human Behavior," which, despite having only 7 articles, boasts an exceptionally high average citation count of 127.71 per article, indicating a superior quality, considerable authority, and significant research reference value compared to other journals.

3.4   Bibliometric Analysis of Countries

Using VOSviewer to analyze the nationalities of authors in the papers, a minimum publication threshold of 15 articles was set, resulting in the identification of five countries meeting this criterion. These five countries are all actively engaged in in-depth research on big data informationization in higher education.

Table 4. The top 5 countries in the field of big data informationization in higher education research

Country	Documents	Citations	Average Citation
Usa	44	960	21.81
China	125	1441	11.52
England	23	715	31.08
Saudi Arabia	30	929	30.97
Spain	16	526	32.87

Table IV shows that among the top 5 countries, Chinese scholars have the highest number of publications, with 125 articles. This is attributed to the large number of Chinese universities in recent years and the widespread application of big data informationization. The main research areas are still concentrated in mobile teaching and network-related topics. However, the average citation rate for Chinese scholars is not high, at 11.51 per article, indicating that there is room for improvement in the depth of their research. Spanish scholars, despite having the lowest number of publications at 16 articles, rank first in average citation count with 32.87.

Fig. 1. Analysis of collaboration among authors from different countries

Using CiteSpace software to analyze the collaboration among authors from different countries, Figure 1 depicts a network graph of international collaboration among authors. The thickness of the lines connecting country labels represents the degree of collaboration between the two countries. Upon observation, it is evident that China collaborates with the United States and many other countries' authors. Authors from Spain, on the other hand, only collaborate with authors from the United States
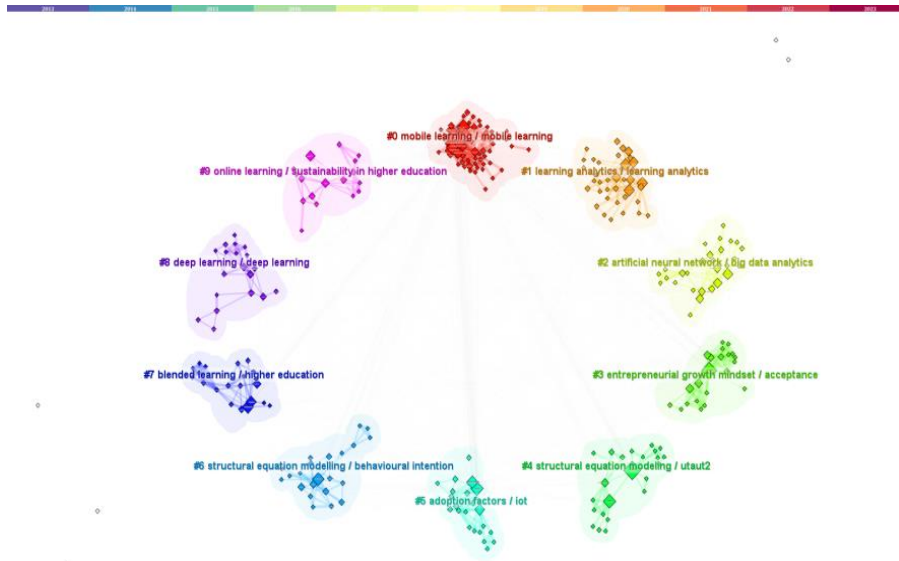
3.5   Network visualization

Fig. 2. Network visualization in the Field of Instructional Design

Using CiteSpace software for Network Visualization analysis of paper keywords is a highly expressive visual analysis tool. It can cluster keywords to discover the structural distribution of research hotspots, explore research hotspots, and understand the development process of a research field. As shown in Figure 2, the analysis of keywords reveals 10 clusters such as 'mobile learning,' 'learning analytics,' 'online learning,' etc. Each cluster is represented by a distinct color, and next to the main label of each cluster, there are small connected nodes, representing the sub-branches of research around that cluster. The larger and denser the cluster area, the more mature and strongly connected the research in that area is, as seen with 'mobile learning.' On the other hand, the color of the clusters also indicates the time when the research hotspots emerged. For example, the blue cluster appeared in 2013, where scholars proposed an innovative educational and teaching approach, achieving a deeper integration of teaching and learning through the fusion of learning and research [8]. The red cluster research emerged in recent years, influenced by the outbreak of COVID-19, where a large number of students had to learn through mobile devices. Scholars proposed that attitude is the most important influencing factor in the intention to use mobile learning [9]. Network Visualization analysis can assist researchers in finding research directions.

4   CONCLUSION AND PROSPECT

4.1   Conclusion

Research on the Informatization of Big Data in Higher Education, as an important field within educational studies, is based on the use of two software tools, VOSviewer and CiteSpace, employing bibliometric research methods. This study comprehensively reviews the development process and technological evolution in the global field of big data in higher education over the past decade. Through the use of big data technology, higher education institutions can more accurately understand students' learning dynamics and needs, optimize teaching plans, thereby enhancing the overall quality of education and ensuring that students can solidly grasp necessary knowledge and skills. Personalized educational services based on big data can provide tailored learning support for each student, fully tapping into their potential and improving learning efficiency. Additionally, big data analysis provides scientific decision-making tools for higher education management, enabling them to more accurately grasp the operational status of the institution, allocate resources rationally, and improve management efficiency. Furthermore, big data technology brings rich data resources to the field of educational research, facilitating in-depth exploration of educational phenomena and patterns, and advancing educational research in depth.

The integration of big data technology into educational models can cultivate professionals better suited to future employment demands, enhancing students' competitiveness in the workplace. The study also conducted bibliometric analysis on 334 articles written by 967 authors from 71 countries in this field, with the following overall summaries:

Research on the Informatization of Big Data in Higher Education entered a period of rapid growth in 2018.

Nine journals on the subject were introduced, with the highest-quality articles found in "Computers in Human Behavior." Keywords such as "students," "science," "performance," and "technology" have become hot topics in the field of instructional design in recent years.

Chinese authors contributed the most articles in this research field, but in terms of average citations per article, scholars of Spanish nationality received the highest recognition.

Research on the Informatization of Big Data in Higher Education has made significant progress over the past decade. In the early stages, researchers mainly focused on basic concepts of big data, technological frameworks, and potential applications in the education field. As big data technology matured, researchers began exploring how to integrate big data with educational informatization to enhance education quality and efficiency. This includes research on data mining, learning analytics, and educational data management.

4.2   Contribution and outlook

In the new era of the Internet and artificial intelligence, the application of big data informatization in higher education is attracting increasing attention. This study conducts an in-depth analysis of the development history and current status of this field over the past decade. For new scholars entering this field, the study not only provides valuable perspectives and methodologies but also guides them in their research directions.

For scholars who wish to initiate research on the informatization of big data in higher education in the future, as well as those aiming to quickly establish a clear cognitive framework for their research area, this study will be of great assistance. It allows scholars to understand the research dynamics and development trends in the field, laying a solid foundation for their academic exploration.

Through in-depth analysis of authors and journals, this study assists scholars in identifying core authors and reference journals for big data informatization technology in higher education. It helps scholars quickly select research references.

The study also provides a detailed clustering of the development and research of big data informatization technology in higher education. This will help scholars quickly find specific research directions, providing a solid foundation for their research, and offering constructive advice for paper submissions.

Looking ahead, we have reason to believe that, on the one hand, ensuring the positive impact of big data informatization technology in higher education requires balancing technological innovation with sustainable social development. Attention must also be given to ethical, privacy, and fairness issues. Educational institutions and decision-makers should cautiously apply big data technology to safeguard student rights and social justice.On the other hand, higher education has significant potential in the future with the support of big data and information technology. Therefore, this research project has profound value and significance. In future research, scholars can attempt to use multiple literature databases for resource collection and employ more bibliometric software for analysis to form a more comprehensive understanding of the research field.

References

1. Cano, E. M., Ruiz, J. G., & Garcia, I. A. (2015). Integrating a learning constructionist environment and the instructional design approach into the definition of a basic course for embedded systems design. Computer Applications in Engineering Education, 23(1), 36-53.
2. Korhonen, V., & Rautopuro, J. (2019). Identifying problematic study progression and "at-risk" students in higher education in Finland. Scandinavian Journal of Educational Research, 63(7), 1056-1069.
3. García-Vélez, R. A., López-Nores, M., González-Fernández, G., Robles-Bykbaev, V. E., Wallace, M., Pazos-Arias, J. J., & Gil-Solla, A. (2019). On data protection regulations, big data and sledgehammers in Higher Education. Applied Sciences, 9(15), 3084.
4. Aldraiweesh, A., & Alturki, U. (2023). Exploring Factors Influencing the Acceptance of E-Learning and Students' Cooperation Skills in Higher Education. Sustainability, 15(12), 9363.

5. Yang, Y., Wang, C., &Lai,M. (2012). Using bibliometric analysis to explore research trend of electronic word-of-mouth from 1999 to 2011. International Journal of Innovation, Management and Technology, 3,337-342.

6. Van Doren, C. S. V., & Heit, M. J. (1973). Where it's at: A content analysis and appraisal of the "journal of leisure research". Journal ofLeisure Research, 5,67-73.

7. Lesnikowski, A., Belfer, E., Rodman, E., Smith, J., Biesbroek, R., Wilkerson, J. D., … Berrang-Ford, L. (2019). Frontiers in data analytics for adaptation research: topic modeling. Wiley Interdisciplinary Reviews: Climate Change, 10, e576.

8. Alyoussef, I. Y. (2021). Massive open online course (MOOCs) acceptance: The role of task-technology fit (TTF) for higher education sustainability. Sustainability, 13(13), 7374.

9. Alhumaid, K., Habes, M., & Salloum, S. A. (2021). Examining the factors influencing the mobile learning usage during COVID-19 Pandemic: An Integrated SEM-ANN Method. Ieee Access, 9, 102567-102578.