

Research

Analyzing the evolution of machine learning integration in educational research: a bibliometric perspective

Musa Adekunle Ayanwale¹  · Rethabile Rosemary Molefi²  · Saheed Oyeniran³ 

Received: 6 January 2024 / Accepted: 8 April 2024

Published online: 08 May 2024

© The Author(s) 2024 [OPEN](#)

Abstract

Machine learning, a subset of artificial intelligence, has experienced rapid advancements and applications across various domains. In education, its integration holds great potential to revolutionize teaching, learning, and educational outcomes. Despite the growing interest, there needs to be more comprehensive bibliometric analyses that track the trajectory of machine learning's integration into educational research. This study addresses this gap by providing a nuanced perspective derived from bibliometric insights. Using a dataset from 1986 to 2022, consisting of 449 documents from 145 sources retrieved from the Web of Science (WoS), the research employs network analysis to unveil collaborative clusters and identify influential authors. A temporal analysis of annual research output sheds light on evolving trends, while a thematic content analysis explores prevalent research themes through keyword frequency. The findings reveal that co-authorship network analysis exposes distinct clusters and influential figures shaping the landscape of machine learning in educational research. Scientific production over time reveals a significant surge in research output, indicating the field's maturation. The co-occurrence analysis emphasizes a collective focus on student-centric outcomes and technology integration, with terms like "online" and "analytics" prevailing. This study provides a nuanced understanding of the collaborative and thematic fabric characterizing machine learning in educational research. The implications derived from the findings guide strategic collaborations, emphasizing the importance of cross-disciplinary engagement. Recommendations include investing in technological infrastructure and prioritizing student-centric research. The study contributes foundational insights to inform future endeavors in this ever-evolving field.

Keywords Artificial intelligence · Bibliometric analysis · Machine learning · Educational research · Performance analysis · Science mapping

Abbreviations

SSCI	Social Sciences Citation Index
ESCI	Emerging Sources Citation Index
CPCI-SSH	Proceedings Citation Index-Social Science & Humanities
BKCI-SSH	Book Citation Index-Social Sciences & Humanities
CI	Collaboration index
AU	Author
CR	Cited References

✉ Musa Adekunle Ayanwale, ayanwalea@uj.ac.za; Rethabile Rosemary Molefi, rethabile68@gmail.com; Saheed Oyeniran, oyeniran@unilorin.edu.ng | ¹Department of Science and Technology Education, Faculty of Education, University of Johannesburg, Johannesburg, South Africa. ²Department of Languages and Social Education, Faculty of Education, National University of Lesotho, Roma, Lesotho. ³Department of Educational Management, Faculty of Education, University of Ilorin, Ilorin, Nigeria.



DT	Document type
TC	Total citation
RP	Corresponding author
DE	Keywords
WC	Science Categories
JCR-IF	Journal citation report impact factor
MCP	Multiple-country publications
SCP	Single-country publications

1 Introduction

The fusion of education and technology has ushered in a new era of learning, where traditional pedagogical methods intersect with cutting-edge advancements in machine learning (ML) [29]. As classrooms transform into digital spaces and learning materials become increasingly digitized, integrating ML algorithms into educational frameworks holds the promise of reshaping the essence of teaching and learning [42]. This paradigm shift prompts a critical examination of the trajectory of ML integration in educational research globally, a subject of paramount importance in the contemporary educational landscape. Historically, education has been a cornerstone of societal progress, serving as the bedrock upon which future generations are molded [48]. However, the challenges faced by modern education are complex and multifaceted. Conventional teaching methods, although time-tested, often need help to cater to the diverse needs of today's learners [28]. Simultaneously, the proliferation of digital technologies has opened new avenues for education, paving the way for innovative and interactive learning experiences. A study conducted by Onyekwere and Enamul Hoquei [35] highlight a positive correlation between the flexibility of digital learning and the attitudes of students towards distance learning programs. Additionally, a beneficial association is observed between technological support and students' perceptions of distance learning programs. This suggests that students having a positive attitude towards digital technologies act as an added advantage since they could easily adopt machine learning in their learning.

Global machine learning (ML) education research has witnessed a surge in attention, reflecting the field's escalating prominence across industries [49]. Universities and online platforms now offer diverse ML courses catering learners with varying backgrounds [22]. The interdisciplinary integration of ML with other disciplines, such as computer science and business, underscores the holistic approach to educational programs. Collaboration between educational institutions and industry has strengthened, ensuring alignment with workforce demands. However, challenges persist, including the need to reform and have up-to-date curricula [49] and a shortage of qualified instructors [36]. Ongoing research focuses on pedagogical approaches and effective teaching methods, aiming to equip students with foundational knowledge and the adaptability required in this rapidly evolving field. ML education research globally reflects a dynamic landscape, addressing emerging challenges and aligning educational offerings with the evolving needs of learners and industries [37].

On the other hand, within Sub-Saharan Africa, where traditional educational approaches have struggled to bridge the gap between educational demand and supply, the integration of machine learning holds the potential to revolutionize education delivery and access [16]. However, access to quality education in Sub-Saharan Africa has often been marred by systemic issues such as limited resources, inadequate infrastructure, and vast disparities in educational opportunities [5, 28, 30]. This implies that in this context, the convergence of education and ML is a beacon of hope, promising tailor-made educational journeys and data-driven insights into student performance.

Machine learning, a subset of artificial intelligence, empowers educational systems to learn from vast datasets, identifying patterns and refining approaches in response to real-time feedback [45]. From personalized learning algorithms that adapt content based on individual student progress to predictive analytics that foresees academic challenges, ML technologies offer transformative potential [44]. Additionally, intelligent tutoring systems, driven by ML, provide students with personalized guidance, bridging the gap between traditional classroom instruction and individual learning needs [38, 43]. This suggests that these advancements signify a shift from the one-size-fits-all approach to an era of customized, adaptive education. Amidst this transformative landscape, the rationale behind examining the trajectory of ML integration in global educational research becomes apparent. This study is propelled by the need to comprehensively map the evolution of ML's role in shaping educational research worldwide. By employing mapping analysis techniques, the study endeavors to identify established research themes, elucidate geographical patterns in ML integration and

unravel the collaborative networks that underpin innovative educational research. Furthermore, this research explores the implications of ML integration for educational policy, teacher practices, and learner experiences.

Although one could argue that teaching the importance of ML methods and concepts is common in higher education institutions, numerous studies have also attempted to review the literature on ML concepts in K-12. These studies include, for example Sanusi et al. [43, 44], which explores the systematic review of teaching and learning machine learning in K-12. Belmonte et al. [8] also carried out a bibliometric review with Scientific Mapping in Web of Science on Machine Learning and Big Data in the Impact Literature, while Zawacki-Richter et al. [51] presented a systematic review of research on artificial intelligence applications in higher education. Ezugwu et al. [15] also conducted a 30-year overview with a Bibliometric Analysis Review on Machine Learning Research Trends in Africa. On the other hand, this study differs from previous studies because it presents a global trajectory of machine learning in education. The importance of this research extends far beyond academia. Insights from this study are poised to inform educational policies, enabling policymakers to make evidence-based decisions about technology integration in classrooms. Armed with a nuanced understanding of ML applications, educators can adapt their teaching methods to foster more engaging and effective learning environments. Additionally, this research is expected to guide future scholars and researchers, directing their focus toward unexplored avenues and uncharted territories within ML-integrated education. Through meticulous analysis and thoughtful interpretation, this study endeavors to illuminate the path forward, shaping the future of education in an increasingly interconnected and technologically driven world. While the overarching goal is to answer the following questions: What are the annual scientific article production trends on ML in educational research from 1986 to 2022? How have the average total citations per article per year changed over time, and what can we learn from these patterns? What is the global citation impact of prominent authors, and how does their work contribute to the scholarly discourse on ML in educational research? Which affiliations and countries have significantly contributed to ML in educational research, and what are their collaborative dynamics? What are the main themes and trends within ML, as shown by word cloud analysis? And how are author collaboration networks structured, and what can we learn from their collaborative patterns?

This paper is structured as follows: The subsequent section will provide an extensive review of the relevant literature on ML in education and reviews on machine learning in K-12 and higher learning. The core of this paper will be dedicated to exploring machine learning applications globally in education and the challenges and barriers they face. The methodology section will detail the research methods, data sources, and analytical tools employed in the scientific mapping and bibliometric analysis. The findings from the bibliometric analysis will be presented, followed by a discussion of their implications and actionable recommendations for stakeholders in education.

2 Review of related literature

2.1 Machine learning applications in education

One of the applications offered by ML in education is personalized Learning and adaptive systems. Adaptive learning systems continuously analyze individual student progress. Advanced algorithms assess correct and incorrect answers and the steps taken to arrive at solutions [16]. This granular analysis enables these systems to adapt the content in real-time, providing additional challenges for advanced students and targeted support for struggling learners. Furthermore, ML algorithms identify learning plateaus, ensuring that each student's content remains engaging and stimulating [26]. For example, ML applications could be more employed in distance learning program which enables students to engage in online education at their own convenience and in a location of their choosing [35]. This educational approach effectively supports students in cultivating a sense of autonomy, allowing them to tailor their learning experience based on their individual backgrounds, skills, and areas of interest. This implies as students advance through this mode of learning, they may cultivate a more favorable outlook towards Distance Learning (ODL).

Furthermore, educational institutions harness predictive analytics and ML algorithms to analyze vast datasets encompassing student performance, attendance records, and socio-economic factors. Predictive models built on this data facilitate nuanced interventions [26]. For instance, by identifying students at risk of dropping out based on attendance patterns and academic performance, schools can deploy targeted support systems, ranging from counseling to additional tutoring, ensuring these students receive the necessary assistance.

Additionally, language learning applications leverage ML to comprehend learners' language proficiency levels. These platforms utilize adaptive algorithms that assess the accuracy and speed of responses, thereby tailoring subsequent challenges accordingly [45]. ML algorithms employ neural networks and deep learning techniques to decipher context

in translation, resulting in remarkably accurate translations [38]. This implies that technology not only aids language learners but also enables seamless communication across diverse linguistic backgrounds. It is an undeniable fact that automated assessment tools use machine learning for nuanced evaluation. These systems employ algorithms that analyze the correctness of answers and the reasoning and problem-solving strategies utilized [21, 26]. This portrays that by recognizing patterns in student responses, these tools provide actionable feedback, guiding students toward better understanding and reinforcing specific concepts where needed. Finally, ML-powered tutoring systems are designed to assist both students and educators. These systems analyze student responses to identify common misconceptions and learning gaps. Educators receive detailed reports outlining areas where their students struggle the most, enabling targeted classroom interventions [26]. Furthermore, teacher support platforms employ natural language processing to analyze educators' queries and provide curated resources, aiding continuous professional development.

2.2 Machine learning challenges and barriers to education

Machine learning (ML) holds immense promise in revolutionizing education by personalizing learning experiences, improving educational outcomes, and enhancing administrative efficiency. However, the widespread adoption of ML in education is hindered by several challenges and barriers ranging from technological limitations to policy and regulatory issues. One of the primary challenges in implementing machine learning in education is the need for adequate infrastructure [34]. Many educational institutions, especially in developing countries, need more computing power and network infrastructure to support sophisticated ML algorithms. With robust hardware and high-speed internet connectivity, implementing machine learning applications becomes more accessible, limiting the scope and effectiveness of educational interventions [23]. In addition, data privacy concerns pose a significant barrier to the widespread adoption of machine learning in education. ML algorithms rely heavily on vast amounts of data to make accurate predictions and recommendations, and educational data, including student records, learning patterns, and assessments, are sensitive and confidential [2, 11]. This suggests that ensuring the privacy and security of this data is crucial to gaining the trust of students, parents, and educators. Striking a balance between utilizing data for educational improvement and protecting individual privacy rights is a complex challenge that education institutions must navigate.

Further, implementing machine learning technologies in education necessitates a skilled workforce capable of developing, deploying, and maintaining these systems. This highlights that a shortage of professionals with expertise in education and machine learning could hinder the effective adoption of ML in education [34]. Bridging this skill gap requires substantial investments in training programs, workshops, and educational initiatives to empower educators with the knowledge and skills needed to integrate machine learning into the curriculum effectively. Another line of argument is socioeconomic disparities among students which create a digital divide, hindering equal access to machine learning-powered educational tools [9]. Students from low-income families may lack access to personal computers, smartphones, or a stable internet connection, putting them at a disadvantage in utilizing online ML-based learning resources. This is again depicted by Ayanwale et al. [5], Makumane [28] and Molefi and Ayanwale [30] that most students in Sub-Saharan countries come from disadvantaged families and cannot afford the abovementioned gadgets. This indicates that addressing this challenge requires targeted interventions such as providing subsidized or free devices and internet access to economically disadvantaged students, ensuring equitable educational opportunities.

To sum up, policy and regulatory frameworks also play a crucial role in shaping the implementation of machine learning in education. Ambiguities in regulations regarding data usage, privacy, and intellectual property rights create uncertainty for educational institutions and technology developers. Clear and comprehensive policies must be formulated to guide the ethical use of machine learning in educational settings. Moreover, policymakers should collaborate with experts to establish standards and guidelines that ensure responsible AI development and deployment in education.

2.3 Machine learning deployment in education

Integrating machine learning into pedagogy and curriculum has been an important concept many scholars have worked on. Hilbert et al. [18] researched machine learning for educational sciences. This study shows that ML can offer fresh perspectives on data analysis and establish a novel benchmark for statistical models that can also be applied in classical research settings with smaller data samples. In addition to its capability to handle numerous variables and model intricate relationships, utilizing ML techniques in the educational sciences can enhance the scientific field. Zawacki-Richter et al. [51] also explored a systematic review of research on artificial intelligence applications in higher education, specifically on educators. Their review has yielded a remarkable outcome, which is the conspicuous absence of critical reflection

on the pedagogical and ethical implications and the risks involved in implementing AI applications in higher education. The other survey was conducted by Kucak et al. [21] and examined the use of machine learning in education. The findings demonstrate that machine learning algorithms have the potential to assist educational institutions in reaching out to students and providing them with the necessary support to achieve success at the earliest opportunity. Again, Kučak et al. [21] revealed that one of the primary advantages of machine learning, based on abundant research in scientific databases, is its ability to predict student performance. By comprehending each student's unique characteristics, this technology can identify areas of weakness and offer recommendations for improvement. Another review on how machine learning is transforming higher education highlighted that the most widely researched application of ML in higher education is related to predicting students' academic performance and employability [37]. A review on machine learning in education conducted by Jagwani [20] pointed out that machine learning will be a game changer in education in the future since it will bring many new opportunities to uphold management and lessen the effort and learning gaps between the students and teachers. A review by Balaji et al. [7] on the contributions of machine learning models toward student academic performance prediction shows that machine learning methods can envisage the students' performance based on specified features as categorized and can be used by students as well as academic institutions.

3 Materials and methods

3.1 Category of bibliometric approach

According to Donthu, Kumar, Mukherjee, et al. [12], there are two categories of bibliometric analysis techniques: performance analysis (PA) and science mapping (SM). Science mapping comprises various methods such as citation analysis, co-citation analysis, bibliographic coupling, co-word analysis, and co-authorship analysis. These techniques, when combined with network analysis, effectively illustrate the bibliometric structure and intellectual structure of the research field [6]. On the other hand, performance analysis (PA) is commonly used in reviews to assess the contributions and performance of different research constituents, such as authors, institutions, countries, and journals, within the field. It is similar to the background or profile of participants typically presented in empirical research, but with a more analytical approach [10, 12, 13, 39]. Figure 1 below provides a detailed breakdown of the components of these two categories.

Importantly, the complexities of research methodologies and theoretical frameworks have increased collaboration among scholars [1, 32]. This collaborative approach has been proven to enhance the quality of research outcomes by incorporating diverse perspectives and expertise, leading to greater clarity and deeper insights [47]. In this study, we employed various bibliometric techniques to understand research dynamics comprehensively such as co-citations, scientific output over time, co-occurrence and co-authorship. Co-citation analysis is a commonly used technique to map the scientific landscape [19]. This method is based on the notion that publications frequently cited together share similar themes. Through co-citation analysis, researchers can uncover the underlying intellectual structure of a research field, revealing thematic patterns and connections among publications. Notably, Rossetto et al. [40] and Liu et al. [25] have demonstrated the effectiveness of this approach in identifying prevalent themes within a research domain. Additionally, analysis of scientific output, which examines the trajectory of scientific production over time, considering predetermined variables to assess its evolution, co-authorship analysis, which evaluates the extent of collaboration among authors and research groups, shedding light on collaborative networks within the field; and word set analysis, which identifies key descriptors or keywords within analyzed articles, offering insights into the central themes and content of research publications [12–14, 27]. When integrated, these bibliometric techniques provide a multifaceted perspective on research trends and dynamics, enabling researchers to discern patterns, collaborations, and thematic foci within a field.

3.2 Data collection procedure

Bibliometrics has gained widespread adoption for assessing the contemporary state of research across various disciplines [24]; [31]). It involves employing meta-analysis techniques to evaluate scientific output [17], setting forth principles and standards for analyzing the progression of publications both quantitatively and descriptively. In this study, it was employed to assess the number of global studies related to ML in education from 1986 to 2022. The study primarily relied on the Web of Science (WoS) database provided by Clarivate Analytics, which is known for its uniqueness, reliability, and comprehensiveness. WoS hosts millions of studies from over 12,000 journals with impressive impact factors, making it the preferred database for global scientific research assessment [3]; [24]. The

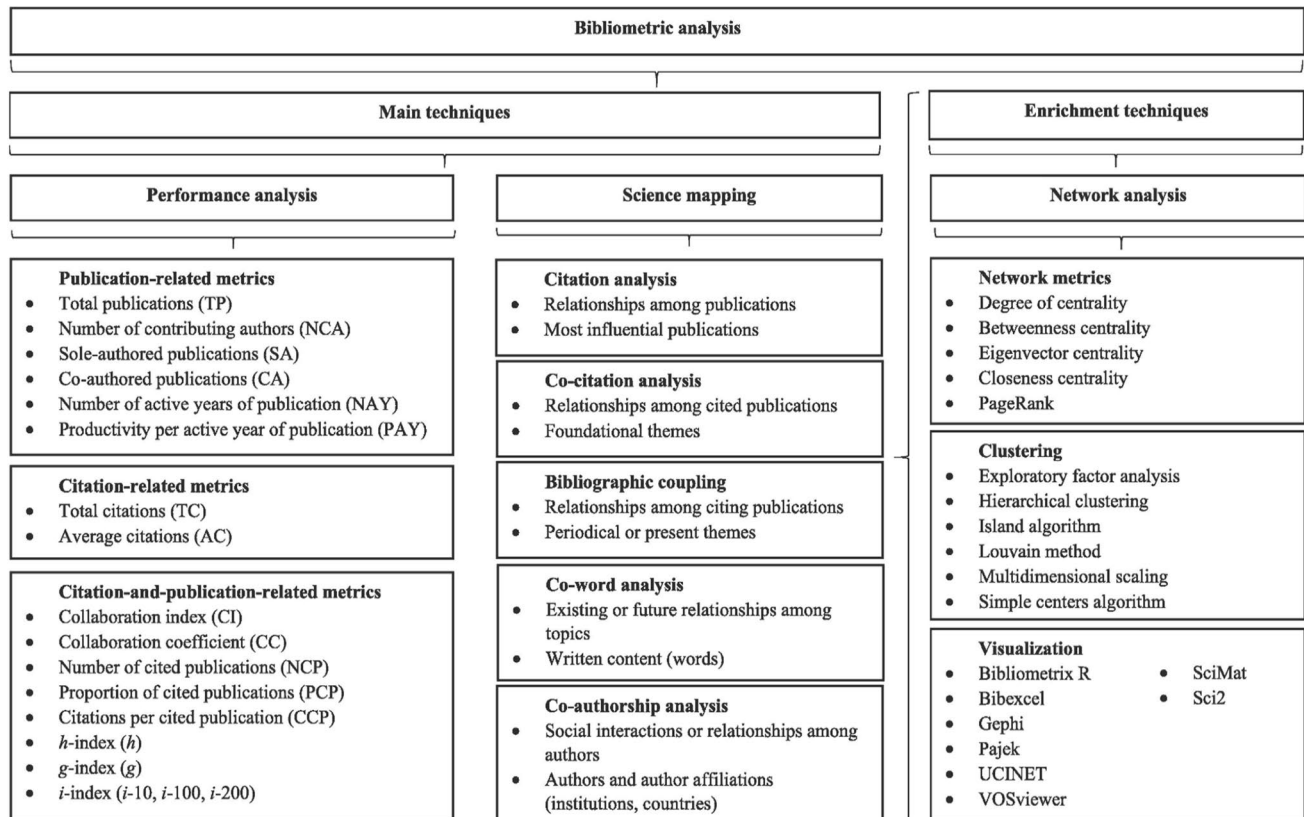


Fig. 1 Schematic representation of PA and SM [12]

data for this study was collected from the Social Sciences Citation Index (SSCI), Emerging Sources Citation Index (ESCI), Proceedings Citation Index—Social Science & Humanities (CPCI-SSH), and Book Citation Index – Social Sciences & Humanities (BKCI-SSH) within the WoS platform, which has a proven track record of data retrieval [50]. In the search process, the authors opted for a title search, which offers several advantages, including eliminating irrelevant documents and reducing loss of sensitivity and specificity, as highlighted by [33]. The search involved a list of specific machine learning-related keywords from January 1, 1986, to December 31, 2022. We ran our search on October 18, 2023, at 17 h:58 using the following search strings—TS = (“Machine learning” OR “Machine learning algorithm” OR “Deep learning” OR “Learning analytics” OR “Artificial intelligence” OR “Data mining” OR “Machine learning in K-12 education” OR “Machine learning in higher education” OR “Machine learning in primary education”).

The initial search across ten indices (Science Citation Index Expanded (SCI-EXPANDED), Emerging Sources Citation Index (ESCI), Social Sciences Citation Index (SSCI), Conference Proceedings Citation Index—Science (CPCI-S), Arts & Humanities Citation Index (A&HCI), Book Citation Index—Science (BKCI-S), Conference Proceedings Citation Index—Social Science & Humanities (CPCI-SSH), Index Chemicus (IC); Book Citation Index—Social Sciences & Humanities (BKCI-SSH), and Current Chemical Reactions (CCR-EXPANDED)) of the WoS Core Collection yielded a total of 256,096 results (see Appendix 1 for the PRISMA flowchart illustrating the included data). Subsequently, the authors refined the search results by focusing on SSCI, ESCI, CPCI-SSH, and BKCI-SSH, which reduced the count to 21,966. Further refinement included excluding non-English language documents and removing those published in 2023 and 2024, resulting in 17,586 documents. Lastly, the documents were filtered based on the research area, specifically educational research, resulting in 449 documents used for subsequent analyses. These filtered data were meticulously cleaned and validated to ensure their relevance to the study’s focus. They were then downloaded in BibTeX format and processed using the Biblioshiny package in RStudio (see Fig. 2 for the workflow).

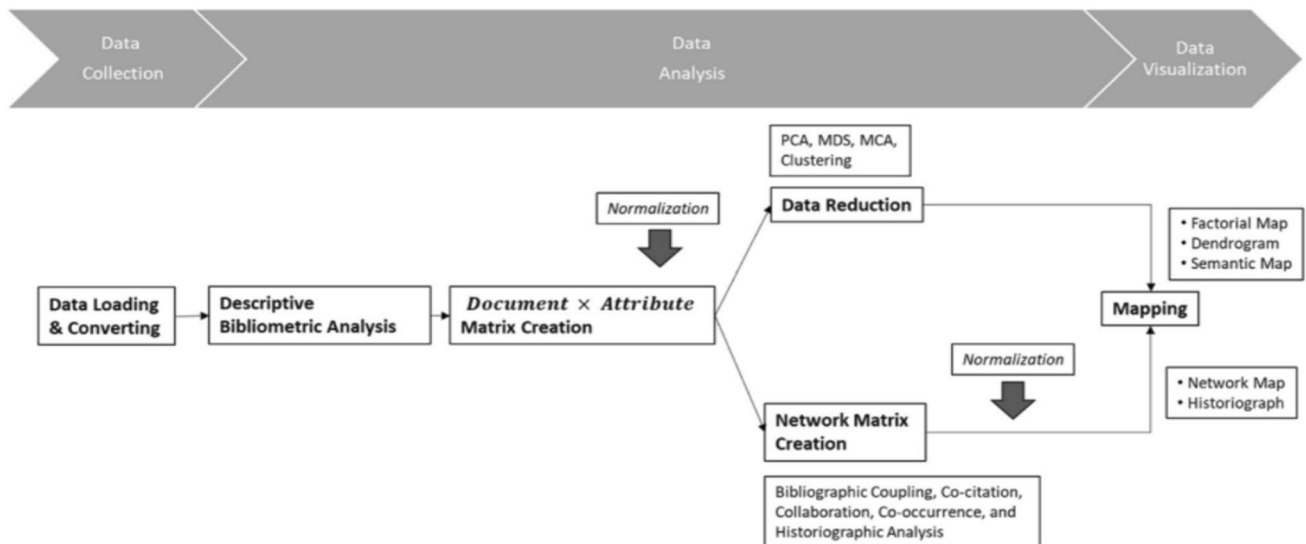


Fig. 2 Bibliometrix science mapping workflow in R language

3.3 Method of data analysis

Data analysis was performed using RStudio software version 4.3.1 [41] in conjunction with the “bibliometrix” package. The data was effectively imported into the “biblioshiny” tool, following the procedures outlined by [4]. Also, the authors employed various bibliometric parameters to visualize, tabulate, and analyze aspects such as author collaboration, yearly scientific production, top authors, annual citations by country, countries of corresponding authors, top publications by citations, publication impact, significant organizations, keywords, and keyword combinations [33].

4 Results

From 1986 to 2022, a comprehensive exploration into the intersection of ML in educational research has unfolded, as presented in Table 1. Drawing from a diverse array of 145 sources, including journals and books, this investigation analyzed 449 documents. The annual growth rate of 15.04% indicates a consistent and notable expansion in research output, underscoring the increasing significance and interest in this interdisciplinary field. Also, documents with dataset of an average age of 4.09 years, signifying a contemporary focus in the research, with scholars actively contributing to current advancements.

Furthermore, the average of 11.46 citations per document suggests that the research in this field is not only recent but also widely recognized and referenced by other scholars, indicating its impact and relevance. The rich network of interconnected research is highlighted by 12,771 references, showcasing the depth and breadth of the literature in this field. A diverse range of 471 keywords reflects various topics and concepts explored within machine learning in educational research. Nine hundred eighty-nine authors have contributed to this work, signaling a broad and diverse community of researchers exploring machine learning in educational contexts. Collaboration is evident, with 64 single-authored documents and an average of 3.04 co-authors per document, emphasizing the interdisciplinary nature of the research in this field. Our findings reinforce the idea that the successful exploration of ML in educational research often requires diverse skill sets and perspectives.

International collaboration is notable, with 16.04% of documents featuring co-authors worldwide. This global collaboration underscores the significance of diverse perspectives in advancing the understanding of ML applications in education. Document types range from articles (397) to other forms such as early access articles, proceedings papers, book reviews, corrections, editorials, letters, and reviews. This diversity in document types suggests a multidimensional approach to exploring machine learning’s implications in education. Our findings reflect scholars’

Table 1 Primary information about the research studies on ML in educational research

Description	Results
<i>Main information about the data</i>	
Timespan	1986:2022
Sources (Journals, Books, etc.)	145
Documents	449
Annual growth rate %	15,04
Document average age	4,09
Average citations per doc	11,46
References	12,771
Document contents	
Keywords plus (Id)	471
Author's keywords (De)	1087
Authors	
Authors	989
Authors of single-authored docs	49
Authors collaboration	
Single-authored docs	64
Co-authors per doc	3,04
International co-authorships %	16,04
Document types	
Article	397
Article: early access	9
Article; proceedings paper	2
Book review	9
Correction	3
Editorial material	11
Letter	1
Review	17

comprehensive approach to understanding and applying ML techniques across various facets of education. Consequently, the trends and characteristics outlined in this analysis paint a picture of a vibrant and expanding field of research at the confluence of machine learning and educational research. The collaborative and international nature of these efforts, coupled with a focus on diverse topics, underscores the complexity and richness of the ongoing exploration in this dynamic intersection.

Moreover, the annual scientific production data from 1986 to 2022 (see Fig. 3) provides a compelling narrative of the evolution and trends within machine learning in educational research. In the initial years, from 1986 to 1995, there is a minimal presence of articles suggesting a nascent stage in the exploration of machine learning applications in education, with limited attention and interest during this period. From 1996 to 2004, there is a gradual increase in the number of articles, with notable peaks in 1998 and 2002. This phase may indicate a growing awareness of the potential applications of machine learning in education, prompting increased research interest. The years 2005 to 2015 witnessed a substantial rise in the number of articles, signifying a maturation of the field. This growth suggests a heightened recognition of the importance of machine learning in addressing educational challenges, possibly driven by technological advancements and increased access to data. The most notable observation is the remarkable increase in scientific production from 2016 onwards. The numbers surge, reaching a peak in 2022 with 155 articles. This period aligns with the era where machine learning technologies, such as deep learning, gained prominence and found diverse applications. The surge in research output may signify an increasing urgency to understand and harness the potential of machine learning in educational settings.

Further, Table 2 presents data on annual total citations per article and year for machine learning in educational research, revealing nuanced patterns that provide insights into the evolving impact and recognition of scholarly works in this interdisciplinary field. In the early years, from 1986 to 1995, the mean total citations per article and mean total citations per year were relatively modest, indicating a period when the impact and recognition of machine learning in

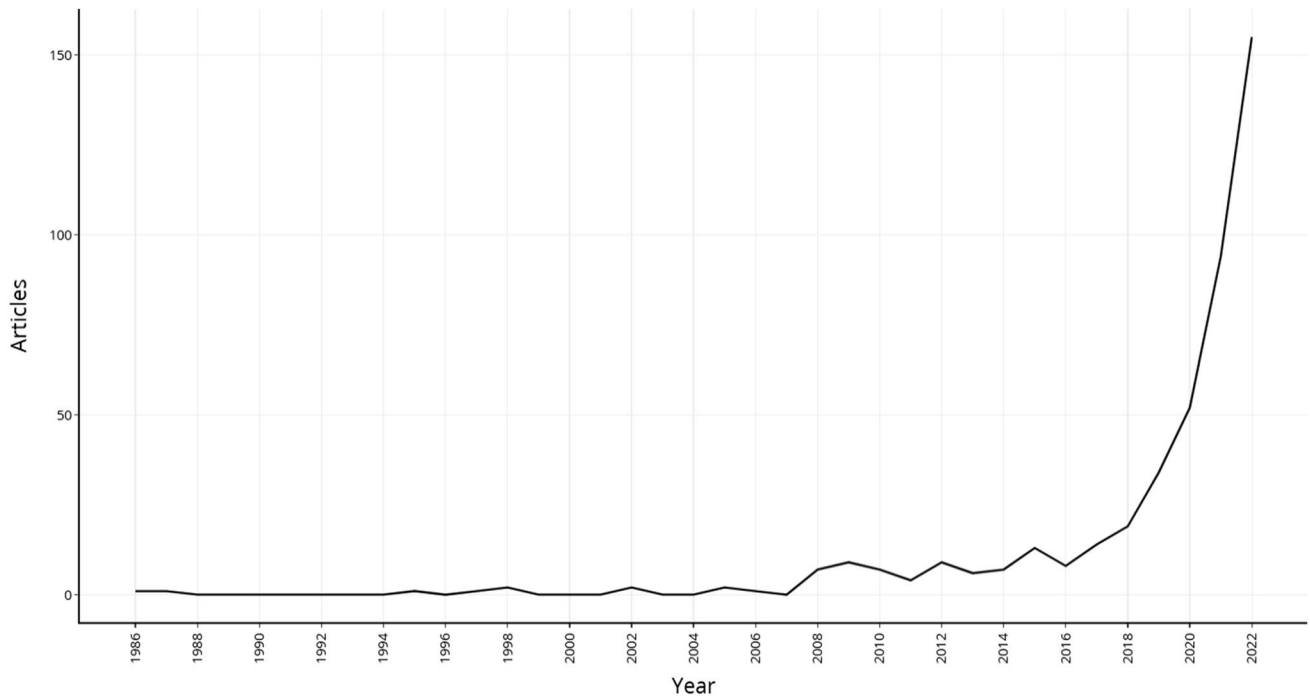


Fig. 3 Plot showing annual scientific production of articles

Table 2 Summary of mean citations per article and year on ML studies in educational research

Year	MeanTCperArticle	N	MeanTCperYear
1986	9	1,00	0,24
1987	1	1,00	0,03
1995	4	1,00	0,14
1997	29	1,00	1,07
1998	14	2,00	0,54
2002	4	2,00	0,18
2005	93,5	2,00	4,92
2006	6	1,00	0,33
2008	46,86	7,00	2,93
2009	30,78	9,00	2,05
2010	26,71	7,00	1,91
2011	19,75	4,00	1,52
2012	29,22	9,00	2,44
2013	22,33	6,00	2,03
2014	16,57	7,00	1,66
2015	17,62	13,00	1,96
2016	9	8,00	1,12
2017	6,36	14,00	0,91
2018	19,16	19,00	3,19
2019	20,41	34,00	4,08
2020	15,94	52,00	3,98
2021	9,38	94,00	3,13
2022	2,13	155,00	1,06

educational research were in their formative stages. The turning point occurred in 1997, where a notable increase in mean total citations per article (29) and mean total citations per year (1.07) suggests the emergence of influential research. This period may have laid the foundation for subsequent studies, marking a phase of growing interest and recognition within the academic community. From 2008 to 2015, the impact experienced a significant surge, peaking in 2005 with a remarkably high mean total citations per article of 93.5. This surge may signify a period of pivotal advancements or seminal works that garnered widespread attention and recognition. Subsequent years maintain a relatively high level of impact, indicating sustained interest and recognition in the field. The years from 2016 to 2022 exhibit fluctuations in mean total citations per article and mean total citations per year, suggesting a dynamic and evolving landscape. Notably, 2019 is a particularly impactful year with a mean total citations per article of 20.41 and a mean total citations per year of 4.08, indicating a concentration of high-impact studies during that period.

Recent trends, specifically in 2021 and 2022, show a decrease in mean total citations per article and mean total per year. This may prompt researchers to explore emerging trends, adopt new methodologies, or delve into novel research directions to maintain the visibility and impact of ML in educational research. The trend of these findings implies that researchers should be attuned to the dynamic nature of the field, recognizing influential periods and adjusting strategies to sustain impact. The sustained high impact from 2008 to 2015 underscores the enduring relevance of ML in educational research during that period. However, the recent fluctuations suggest the need for ongoing innovation and exploration to keep the field vibrant and impactful.

The global citation of an author signifies the frequency with which their work has been referenced, as recorded in the WoS database, prior to being accessed or retrieved [46]. This metric does not merely reflect the author's popularity among researchers but rather quantifies the instances their work has been cited, thereby indicating the perceived value and excellence of both the citing individuals and documents. Figure 4 showcases the 15 authors with the highest global citations in the study, providing a snapshot of the considerable impact and recognition garnered by their contributions within the scholarly community. Among the identified authors, Zhai X stands out as the most prolific contributor with 11 articles, indicating a sustained and impactful presence. The fractionalized value of 4.43 underscores the significant influence of Zhai X's work, suggesting a central role in shaping discussions within the field. Following closely, Nehm RH has contributed substantially with 7 articles and a fractionalized value of 2.07. This indicates a considerable impact, highlighting Nehm RH's consistent engagement in the research landscape. Musso MF, with 6 articles and a fractionalized value of 2.00, also demonstrates a significant contribution, contributing to the foundational knowledge in machine learning

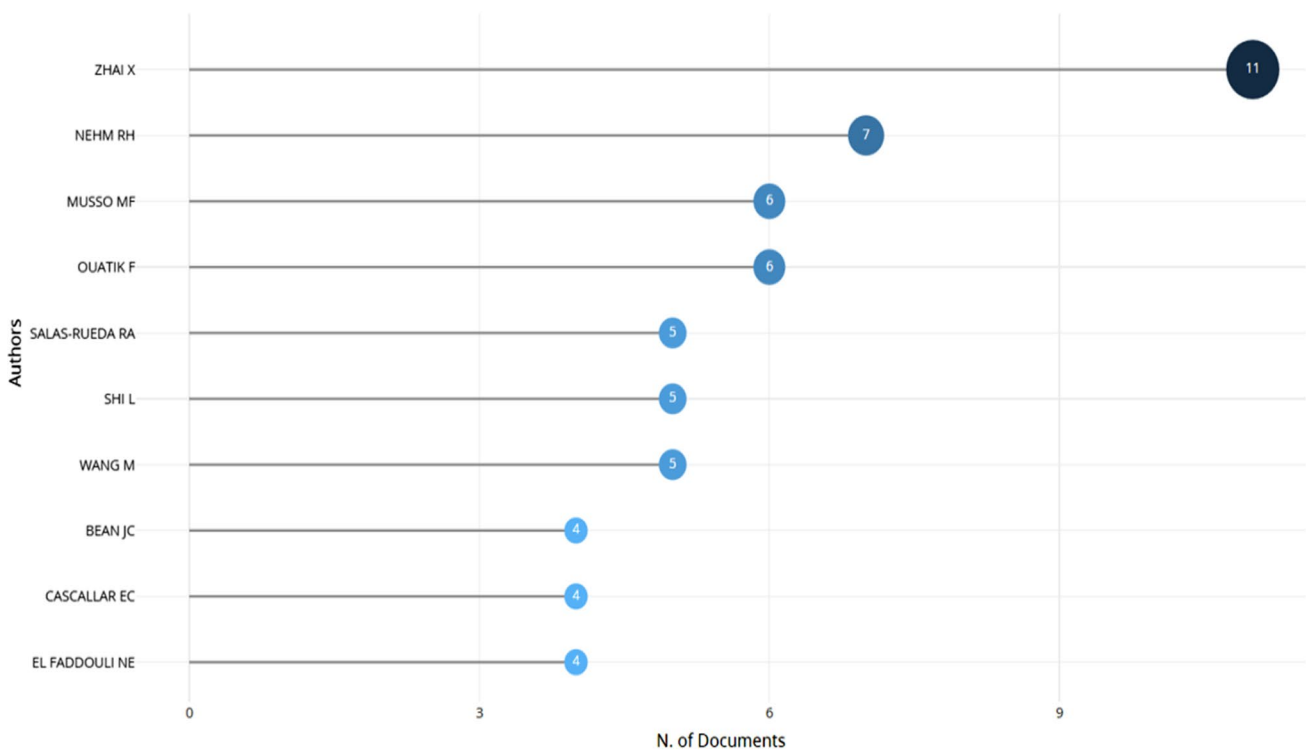


Fig. 4 Most relevant authors on machine learning studies in educational research

applications in education. Other authors, including Salas-Rueda RA, Shi L, Wang M, Bean JC, Cascallar EC, El Faddouli NE, Green GP, Ha M, Huang NF, and Kilic AF, have made noteworthy contributions with varying numbers of articles. The fractionalized values provide insights into the relative impact of their work, showcasing a diverse research landscape.

These findings imply that prolific authors like Zhai X, Nehm RH, and Musso MF likely influence the field through their research, methodologies, and agenda-setting. The diversity of authors underscores a multifaceted research landscape, crucial for a comprehensive understanding of the field. Moreover, prolific authors can serve as knowledge disseminators, bridging the gap between research and practice. Their influence may extend to educators, policymakers, and fellow researchers, contributing to the practical application of machine learning techniques in educational contexts. The identification of key contributors also opens avenues for potential research collaborations, fostering interdisciplinary approaches and innovative solutions. The fractionalized values offer a nuanced assessment of impact, considering both the quantity and perceived significance of articles. This nuanced approach helps stakeholders, including funding agencies and institutions, gauge the influence of specific researchers in the broader landscape of machine learning in educational research.

As presented in Fig. 5, the distribution of articles across different affiliations in the field of ML in educational research offers insightful patterns and implications for the landscape of academic contributions. The University of Georgia emerges as a standout contributor, leading with 26 articles. This signifies a sustained and significant commitment to research in machine learning for educational applications. The institution's prominence suggests a central role in shaping the discourse and advancements within the field. Michigan State University and Purdue University share the second position, each contributing 16 articles. This indicates active engagement from these institutions in research endeavors related to machine learning in education. The comparable numbers suggest a shared dedication to advancing knowledge in the field. Ohio State University—14 articles, University of Hong Kong—14 articles, University of West Attica—14 articles. These affiliations share the third position, each with 14 articles. This suggests a global distribution of research efforts, highlighting the diverse perspectives and approaches to machine learning applications in educational research from different regions. Several other affiliations, including Velagapudi Ramakrishna Siddhartha Engineering College, University of Granada, Adiyaman University, National Tsing Hua University, Seattle University, St. Peters Engineering College, University Federal of Piaui, and Aarhus University, also make substantial contributions with varying numbers of articles. This finding implies that the University of Georgia's leading position implies institutional leadership in the exploration of machine learning in educational research. This may be indicative of robust research programs, collaborations, and a supportive academic environment fostering innovation. The inclusion of institutions from diverse regions signifies the global nature of research in machine learning for educational applications. This diversity enriches the field with a variety of perspectives, methodologies, and contextual considerations. Also, affiliations with similar article counts, such as

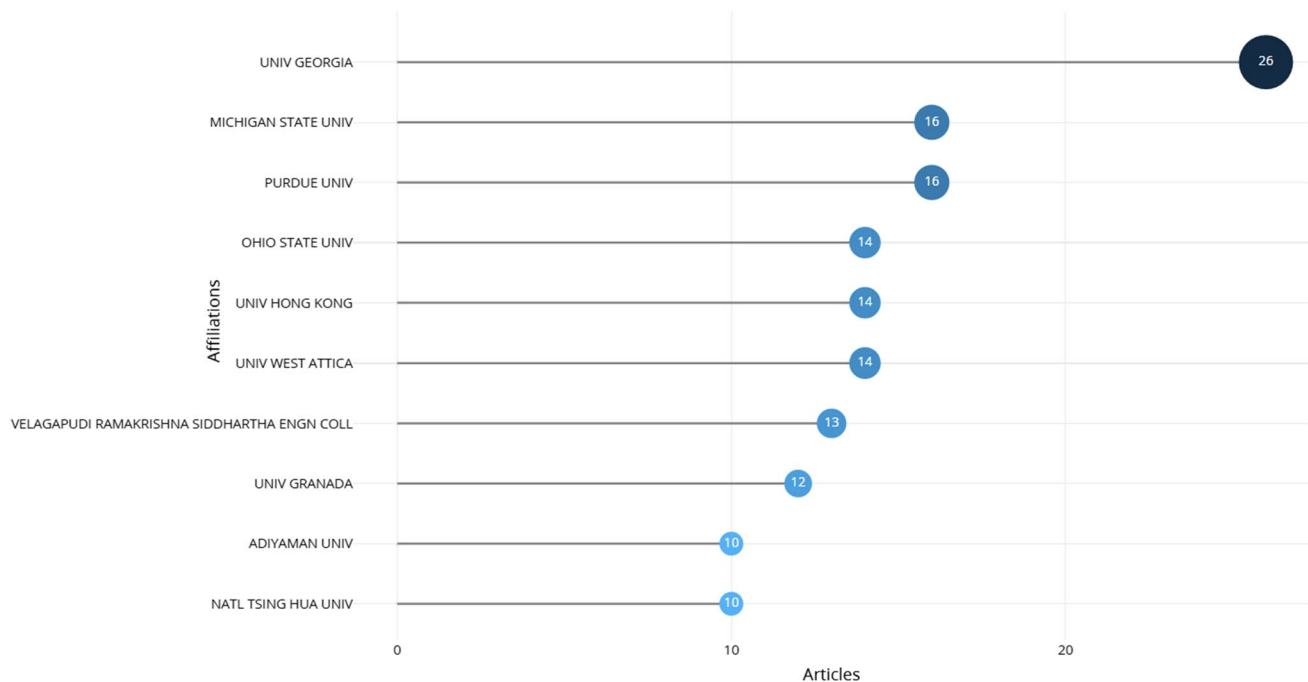


Fig. 5 Most relevant affiliations on ML studies in educational research

Michigan State University and Purdue University, may find opportunities for collaboration. Joint initiatives could lead to synergies in research, the sharing of expertise, and the development of comprehensive solutions to challenges in the application of machine learning in education.

Figure 6 presents the top 15 countries for educational research focusing on ML, ranked by the number of articles published with a corresponding author. Certain trends emerge from this analysis. The United States leads in both the quantity and independent production of research articles, with a substantial portion of its publications being Single Country Publications (SCPs), indicating a strong domestic research presence. However, it maintains a moderate Multiple Country Publication (MCP) Ratio of 0.135, suggesting a balanced mix of domestic and international collaboration. India closely follows the USA regarding research output, with a notable emphasis on SCPs, indicating a robust domestic research landscape. However, its relatively low MCP Ratio of 0.056 suggests limited engagement in international collaborative efforts. While China exhibits a slightly lower total number of articles compared to the USA and India, it demonstrates a significant level of independent research activity, as evidenced by its high proportion of SCPs. Moreover, its MCP Ratio of 0.203 indicates a strong inclination towards international collaboration, potentially contributing to the global advancement of machine learning in education. The United Kingdom maintains a moderate research output, a balanced distribution between SCPs and MCPs, and a relatively high MCP Ratio of 0.161, suggesting active engagement in international collaboration to drive progress in the field. Among the other countries in the top 15, including Australia, Morocco, Turkey, the Netherlands, Spain, Canada, Germany, Malaysia, Mexico, Denmark, and France, there are varying levels of research output and collaboration. Some prioritize domestic research efforts, while others exhibit a strong propensity for international collaboration, as indicated by their respective MCP Ratios.

The globally cited documents listed in Table 3 cover various educational research topics and have been highly cited, indicating their significant contributions to advancing knowledge and practice in the field. Although not all of them explicitly focus on ML, their relevance and implications for the evolution of machine learning in educational research are found in their potential intersections with machine learning applications, such as personalized learning, learning analytics, and adaptive instruction. These papers serve as a foundation for future research in educational machine learning, addressing emerging challenges and opportunities in the field.

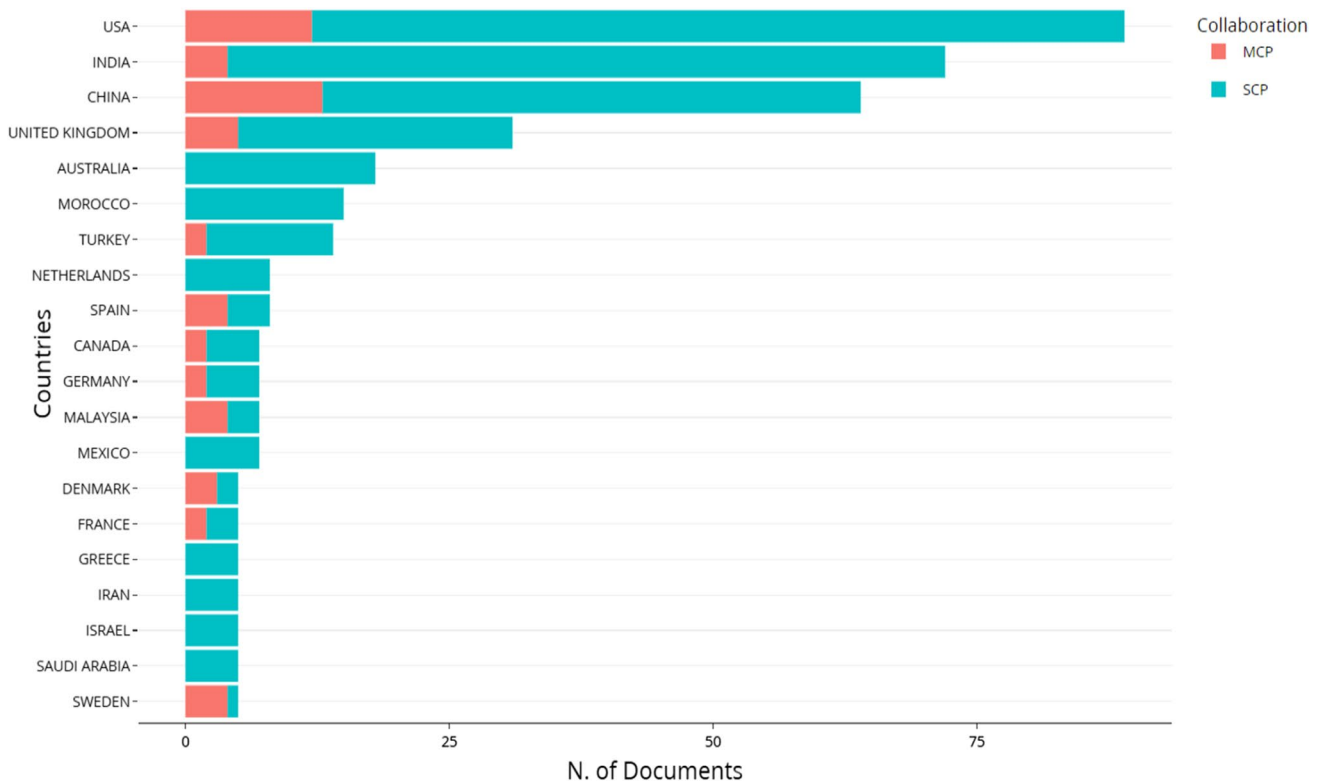
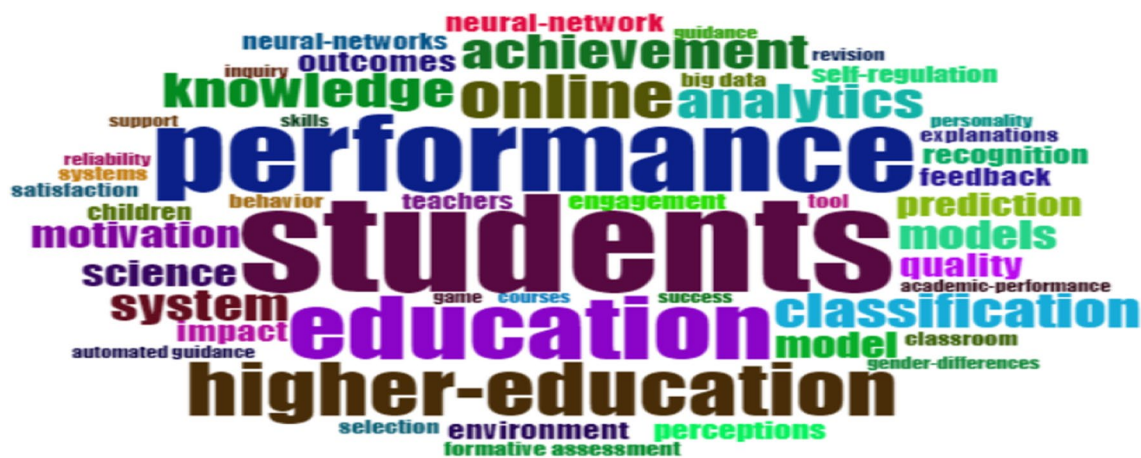


Fig. 6 Most relevant countries by corresponding author

Table 3 Most global cited documents

Paper	DOI	Total citations	TC per Year	Normalized TC
ONAN A, 2020, COMPUT APPL ENG EDUC	10.1002/cae.22179	155	38,75	9,72
RUSHTON A, 2005, MED TEACH	10.1080/01421590500129159	124	6,53	1,33
XING W, 2019, J EDUC COMPUT RES	10.1177/0735633118757015	113	22,60	5,54
KE F, 2009, INTERNET HIGH EDUC	10.1016/j.iheduc.2009.08.001	95	6,33	3,09
DUMMER TJB, 2008, J GEOGR HIGH EDUC	10.1080/03098260701728484	94	5,88	2,01
PANG B, 2020, J EDUC BEHAV STAT	10.3102/1076998619872761	81	20,25	5,08
NEHM RH, 2012, J SCI EDUC TECHNOL	10.1007/s10956-011-9300-9	75	6,25	2,57
GRAY CC, 2019, COMPUT EDUC	10.1016/j.compedu.2018.12.006	74	14,80	3,63
SPIKOL D, 2018, J COMPUT ASSIST LEARN	10.1111/jcal.12263	73	12,17	3,81
KHANAL SS, 2020, EDUC INF TECHNOL	10.1007/s10639-019-10063-9	72	18,00	4,52
DRAPER SW, 2009, BR J EDUC TECHNOL	10.1111/j.1467-8535.2008.00920.x	72	4,80	2,34
CHAPMAN C, 2005, INNOV EDUC TEACH INT	10.1080/01587910500167910	63	3,32	0,67
ZHAI X, 2020, STUD SCI EDUC	10.1080/03057267.2020.1735757	60	15,00	3,76
D'MELLO SK, 2009, INT J LEARN TECHNOL	10.1504/IJLT.2009.028805	60	4,00	1,95
SHIELD M, 2013, EDUC STUD MATH	10.1007/s10649-012-9415-9	58	5,27	2,60
TAN M, 2015, INT J EMERG TECHNOL LEARN	10.3991/ijet.v10i1.4189	58	6,44	3,29

**Fig. 7** Word cloud relating to studies on ML in the educational research

The analysis of frequently occurring terms in machine learning in educational research reveals key thematic focuses and trends (see Fig. 7). The predominant emphasis is on students, performance, education, and higher education, reflecting a comprehensive exploration of technology's role in enhancing learning experiences and outcomes. Notably, there is a strong interest in online education, analytics, and predictive modeling, signaling a growing use of technology to personalize and optimize the learning experience. The terms also highlight considerations for student engagement, recognition, and perceptions, indicating a nuanced approach that encompasses both technological and human aspects within the educational context. Overall, the findings underscore a multifaceted application of ML techniques to advance educational practices, improve student outcomes, and address challenges in the field.

Figure 8 illustrates the presence of keywords in scientific research on education. It offers valuable insights into the relationships and importance of these keywords. Clustered keywords like "Online," "Impact," "Feedback," "Tool," "Explanations," and "Satisfaction" demonstrate a strong connection among various aspects of online education, including its impact, feedback mechanisms, tools used, the validity of findings, explanations provided, and user satisfaction. The prominence of the keyword "Online" emphasizes the increasing significance of digital learning environments and related factors in machine learning within educational research. Another cluster consists of keywords such as "Classification," "Analytics," "System," "Prediction," "Quality," "Recognition," and "Big Data." Additionally, keywords related to student outcomes and

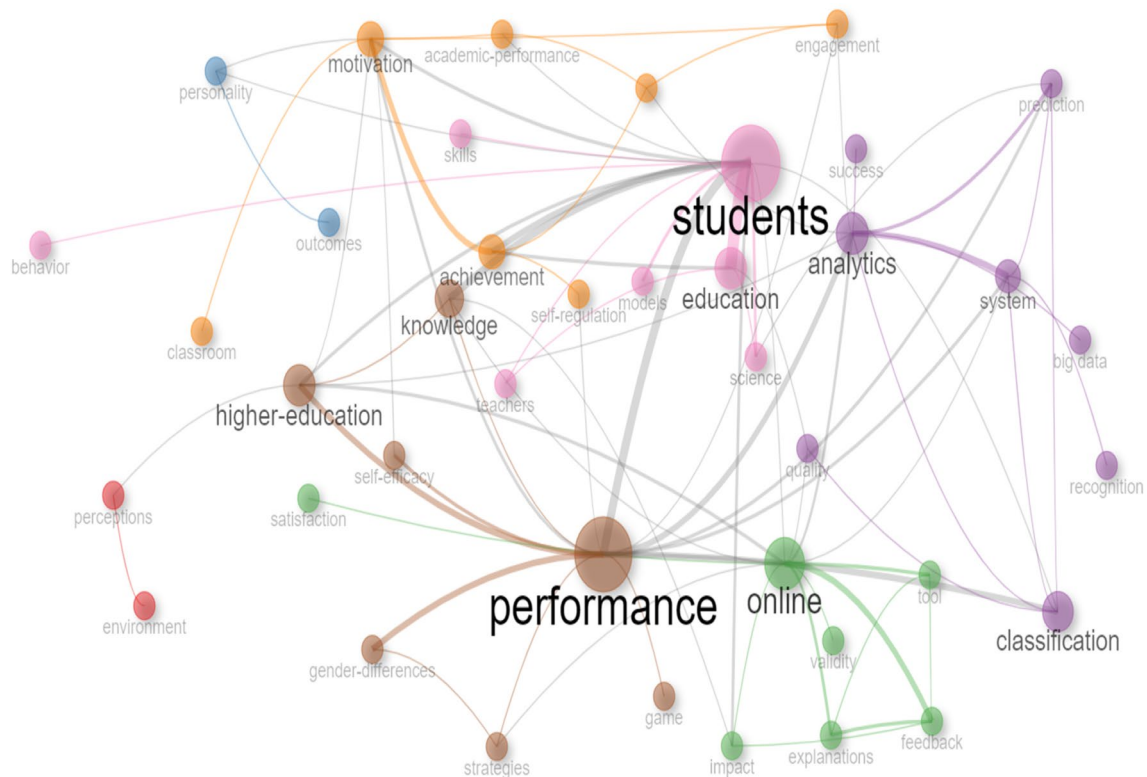


Fig. 8 Co-occurrence of keywords in scientific production

behaviors, such as “Achievement,” “Motivation,” “Engagement,” “Self-regulation,” and “Academic performance,” form another cluster. This suggests a comprehensive approach to understanding and promoting student success in educational settings. The co-occurrence of keywords associated with online education, feedback mechanisms, and satisfaction indicates a focus on improving digital learning environments and enhancing user experiences. ML algorithms can optimize online platforms, customize learning content, and enhance user engagement and satisfaction. Furthermore, clustering keywords related to personality, motivation, engagement, and self-regulation implies a growing emphasis on personalized learning experiences and adaptive interventions. The prominence of analytics, prediction, and quality assessment keywords highlights the importance of data-driven decision-making in educational contexts.

The co-citation results provided in Fig. 4 offer insights into the relationships between authors within the machine learning domain of educational research. These authors, namely Anonymous, Breiman L, Romero C, Hastie T, Kotsiantis SB, Xing WL, Ravenscroft A, Wu JY, and Boekaerts M, are clustered together based on their co-citation frequencies. The presence of “Anonymous” in the cluster suggests that some citations might not be attributed to specific authors, possibly indicating references to collective works or contributions without individual attribution. Additionally, the Organization for Economic Cooperation and Development (OECD), while not an individual author, is included in the co-citation results. Its presence alongside individual authors suggests that OECD publications are frequently cited in conjunction with the works of these authors. OECD reports and studies often provide valuable insights and data on educational policies, practices, and trends, which may influence the research agenda and discussions on machine learning in education.

The co-authorship network analysis, as presented in Fig. 6, reveals distinct clusters of authors within the field of machine learning in educational research, each representing potential collaborative groups or subfields. Cluster 1 (Zhai X, Nehm RH, Shi L, Ha M, Krajcik J, Pellegrino JW, Urban-Lurain M). This cluster exhibits a close-knit, collaborative group with varying centrality measures. Zhai X emerges with the highest betweenness, indicating a potential bridge between other researchers, while Pellegrino JW holds the highest PageRank, signifying overall importance in the network. Cluster 2 (Kilic AF, Koyuncu I). Kilic AF and Koyuncu I form a distinct cluster with high Closeness and PageRank, suggesting central and influential roles within their collaborative network. Also, Musso MF and Cascallar EC represent another cluster, showcasing high Closeness and PageRank, indicating prominence within their collaborative network. Tal T and Tsashu M form a small cluster with high Closeness and PageRank, indicating centrality within their collaborative group. Wang

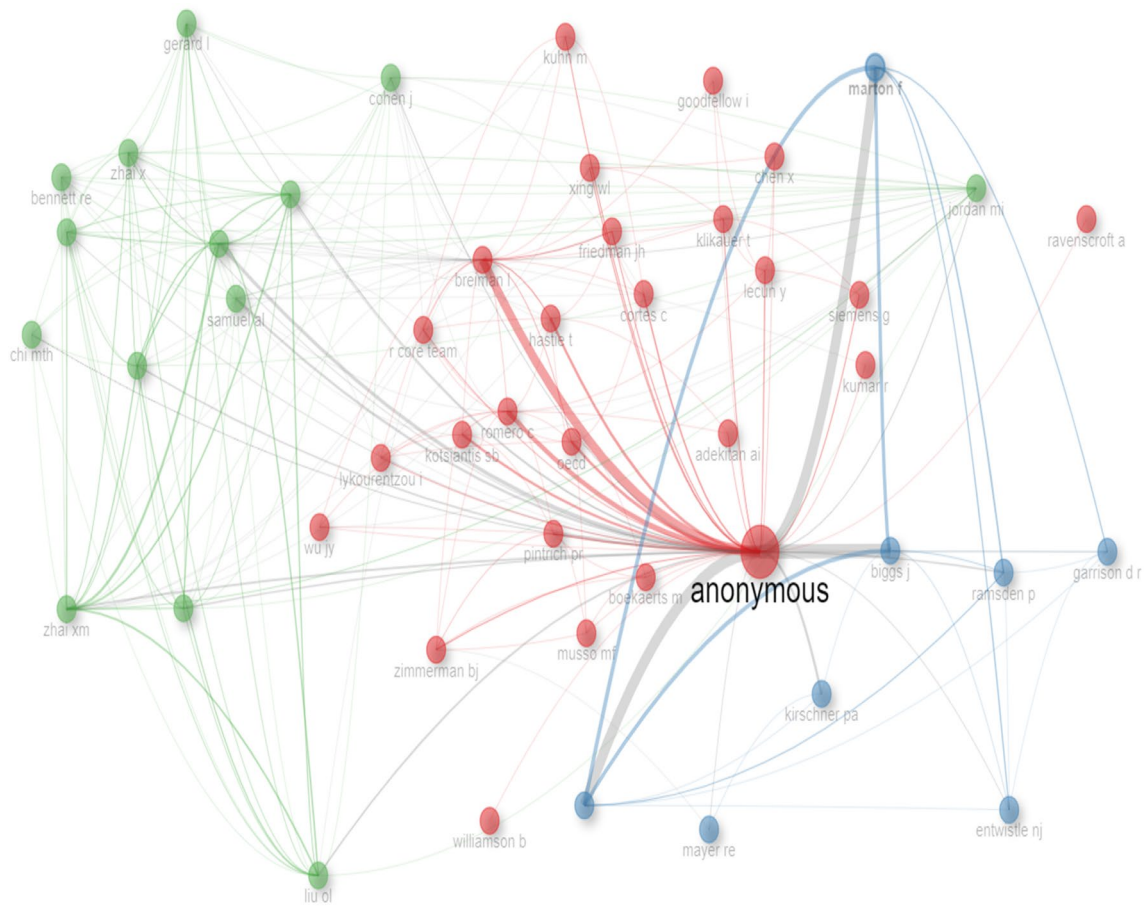


Fig. 9 Co-citation with “authors” as a unit of analysis

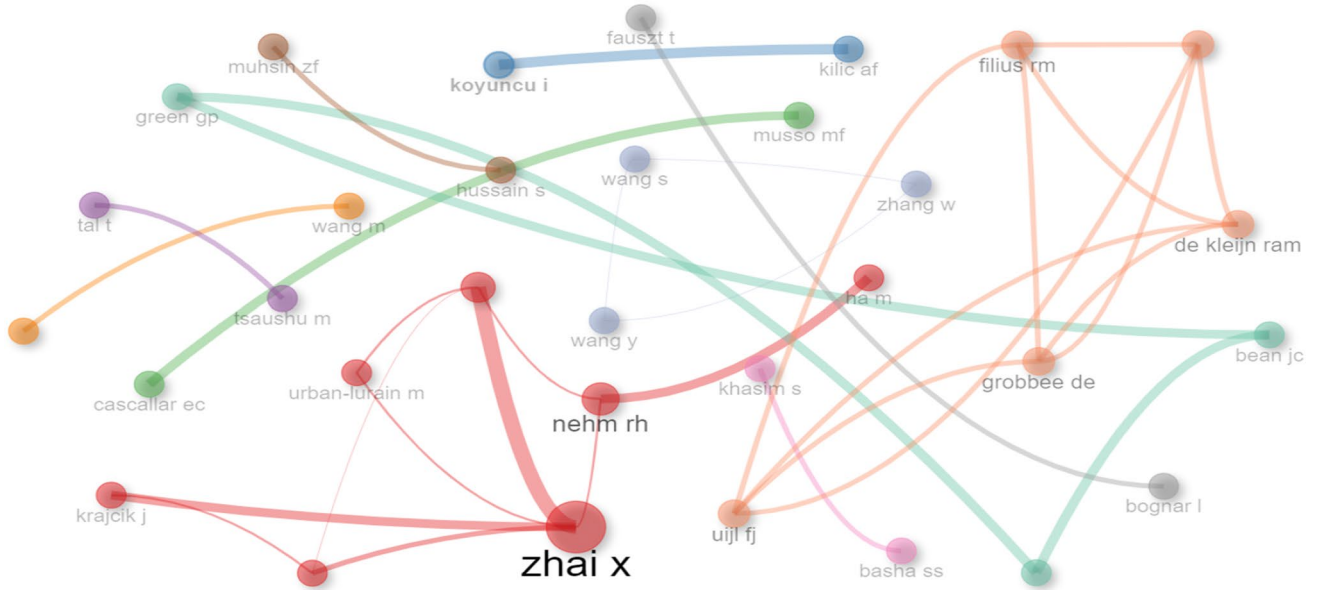


Fig. 10 Co-authorship among the prominent authors on machine learning in educational research

M and Peng J constitute a cluster with equal centrality measures, suggesting a cohesive collaborative subgroup. Husain S and Muhsin ZF form a cluster with high Closeness and PageRank, indicating central and influential roles within their collaborative network. Cluster 7 (Basha SS, Khasim S). Basha SS and Khasim S represent a small cluster, potentially focusing on a specific thematic area within the broader field (Figs. 9 and 10).

Cluster 8 (Bognar L, Fauszt T): Bognar L and Fauszt T constitute another cluster, suggesting collaboration in a distinct research theme. Bean JC, Green GP, and Peterson DJ form a collaborative group with moderate centrality measures, indicating influence within their cluster. De Kleijn RAM stands as a separate cluster with low centrality measures, playing a unique role within this collaborative subgroup. The collaboration network showcases diverse collaborative clusters, each potentially contributing to specific aspects or subfields within machine learning in educational research. Authors with high centrality measures, such as PageRank, may be considered influential figures within their collaborative networks, indicating leadership or key contributions. Clusters with high Closeness suggest tightly-knit groups with strong internal connections, facilitating effective communication and collaboration. Nodes with high Betweenness may act as bridges between different clusters, fostering potential interdisciplinary collaboration. Consequently, the analysis provides insights into the collaborative structure and dynamics within the realm of machine learning in educational research, identifying influential authors and collaborative clusters contributing to the advancement of knowledge in this field.

5 Discussions

Based on the results of this study, machine learning is ubiquitous in educational research, which emphasizes the interdisciplinary nature of this field. Researchers are increasingly recognizing the potential of ML techniques in educational contexts, as reflected in the consistent growth in research output over time. As noted by scholars like Sanusi et al. [42], ML is being leveraged for educational purposes in-line with global trends. There is great recognition and acceptance in the scholarly community for the research in this field based on the large number of citations per document. The impact of this research cannot only be quantified, but also emphasized in terms of its relevance to advancing our understanding of ML applications in education. As a result of the collaborative and international nature of the research efforts, ML in educational contexts can be explored to incorporate diverse perspectives and cultural nuances. In fact, Ezugwu et al.'s study [15] emphasizes how important ML research is to broader scientific discourse and its importance in the area of education. As a result of the diversity of document types and keywords, the multidimensional approach highlights the complexity of studying machine learning's implications in education. Through the lens of machine learning, researchers can analyze various aspects of educational practices and challenges, contributing to a better understanding of its potential applications. It is also evident that the machine learning (ML) research landscape within the educational domain is dynamic and evolving. The implications of this are that educational research and practice need to expand and become more integrated. It is important to understand that this trend has multiple implications.

Research and practitioners are increasingly recognizing the relevance and applicability of machine learning in educational settings, which suggests that the field is maturing. Articles have surged in the last few years because of rapid advancements in machine learning technology. This indicates that researchers are leveraging new techniques and tools to address complex educational challenges. Moreover, there has been a significant increase in scientific production, particularly in recent years. This indicates a growing interest from academia and industry in the intersection of machine learning and education. This heightened attention could lead to increased investment, collaboration, and practical implementations in educational settings. This finding aligns with the submission of, as the rising number of articles suggests fertile ground for innovation. The field may be on the brink of discovering novel applications, methodologies, and solutions that can revolutionize education by integrating machine learning. As the field experiences this surge in research output, there may be a growing need for synthesizing findings and integrating diverse perspectives. This could stimulate collaborative efforts to create comprehensive frameworks and guidelines for implementing machine learning in education. Also, findings from authors in relevant countries underscore the dynamic and collaborative nature of ML research in educational contexts, with implications for local innovation and global cooperation in addressing complex educational challenges. It is worth noting that while African countries may have lower levels of research output compared to other regions, there is a growing interest in educational research, including machine learning. This presents an opportunity for these countries to invest in building research capacity, enhancing infrastructure, and fostering collaborations with international partners to strengthen their research capabilities. Partnering with researchers and institutions from other regions facilitates

knowledge exchange, access to resources, and opportunities for skill development [10, 13]. Such collaborations can address the unique educational challenges that African nations face and foster the development of contextually relevant solutions.

Furthermore, the clustering highlights the shared relevance and interconnected nature of these topics within educational research, particularly data analytics and management. It suggests a significant emphasis on the use of data-driven approaches and technologies, such as ML, for analyzing educational data, predicting outcomes, designing systems, ensuring quality, and effectively utilizing big data. Machine learning algorithms have the potential to enable personalized learning paths, adaptive feedback mechanisms, and targeted interventions that are tailored to the specific needs of each student. The clustering of these authors indicates that they are often cited together in educational research literature, implying a shared relevance or common themes in their work. This suggests that their works are influential in the intersection of ML and educational research. Their collective contributions may represent foundational concepts, methodologies, or machine learning applications in educational contexts. Analyzing the works of these authors could provide valuable insights into the historical development, current trends, and future directions of machine learning in educational research. This finding supports the idea that co-citation analysis focuses primarily on highly-cited publications, potentially leaving out recent or niche publications from its thematic clusters [25].

6 Implications and recommendations

The findings of this study carry significant implications for the field of machine learning in educational research. The identified collaborative clusters and influential authors indicate a dynamic and interconnected research community. Recognizing and understanding these collaborative dynamics is vital for fostering cross-disciplinary collaboration and leveraging diverse expertise to address complex challenges. The global representation of authors and affiliations also reflects a widespread interest in the application of machine learning in education. Researchers from diverse countries contribute to a rich and varied research landscape, showcasing a global commitment to advancing educational technology. The observed increase in research output over the years signals a growing interest and recognition of the potential impact of machine learning on education. This highlights the evolving nature of the field and underscores the need for continued exploration of innovative approaches and technologies. Additionally, the recurring emphasis on terms like “students,” “performance,” and “achievement” indicates a strong focus on student-centric outcomes. This underscores a collective effort to harness machine learning to enhance the learning experiences and academic achievements of students. Terms such as “online,” “analytics,” and “prediction” point to a significant focus on integrating technology into educational practices. This trend suggests a growing interest in leveraging machine learning for adaptive learning platforms, personalized content delivery, and predictive modeling to optimize educational outcomes. Understanding the structure of the collaboration network and identifying influential authors provides valuable insights for strategic collaborations. Researchers and institutions can benefit from fostering collaborations with influential figures to amplify the impact of their work and contribute to the advancement of the field.

Based on the implications derived from the study, several recommendations can guide future research and practice. Facilitate interdisciplinary collaboration among researchers from machine learning, education, and related fields. Encouraging cross-disciplinary partnerships can lead to innovative solutions and a more comprehensive understanding of educational challenges. Identify and support emerging researchers who show promise in contributing to the intersection of machine learning and education. Mentorship programs, collaborative workshops, and research grants can nurture the next generation of scholars in this evolving field. Foster international collaborations to incorporate diverse perspectives and approaches. Collaborative projects involving researchers from different countries can lead to a more holistic understanding of educational challenges and the development of globally applicable solutions. Institutions and policymakers should consider investing in technological infrastructure to support the integration of machine learning in educational settings. This includes the development of adaptive learning platforms, data analytics systems, and tools for personalized education. Prioritize studies that directly impact student learning experiences and outcomes. Investigations into the effectiveness of machine learning applications on student performance, engagement, and satisfaction can guide the development of impactful educational technologies. By considering these implications and recommendations, stakeholders in the field can contribute to the continued growth and positive impact of machine learning on education globally.

7 Conclusion

In conclusion, this extensive study spanning the period from 1986 to 2022 has yielded insightful observations and comprehensive analyses regarding the intersection of machine learning and educational research. The exploration covered various facets, including collaboration networks, temporal trends, citation patterns, authorship characteristics, and thematic content, collectively painting a nuanced picture of the evolving landscape at the intersection of these two domains. The findings regarding collaboration networks have unveiled the intricate web of interactions among researchers, emphasizing the collaborative nature of knowledge creation and dissemination in the interdisciplinary field of machine learning in education. The temporal analysis has revealed a substantial growth in research output, signifying the increasing importance and interest in applying machine learning technologies to educational contexts. This growth trajectory suggests a maturation of the field with expanding opportunities and challenges. Examining citation patterns has indicated not only an increase in the quantity of scholarly contributions but also a growing recognition within the academic community, reflecting the impact of machine learning research in education. The analysis of authorship characteristics has showcased a diverse and globally distributed authorship landscape, with influential authors and collaborative clusters identified. This provides valuable insights for future collaboration and mentorship initiatives.

The examination of thematic content, through the frequency analysis of terms, has shed light on the prevailing research focuses. Terms such as “students,” “performance,” and “online” underscore a collective effort within the research community to address student-centric outcomes and integrate technology into educational practices. In light of these findings, several implications and recommendations have emerged. Strategic collaborations and a global engagement approach are recommended to foster cross-disciplinary collaborations and enrich the global discourse on machine learning in education. Acknowledging influential authors and collaborative clusters provides guidance for strategic collaboration initiatives, amplifying the impact of research in the field. The thematic emphasis on terms like “online” and “analytics” encourages institutions and policymakers to invest in technological infrastructure, supporting the integration of machine learning in educational settings. Moreover, the emphasis on student-centric terms underscores the importance of prioritizing research that directly impacts student learning experiences. Future research endeavors should aim to enhance educational outcomes through the application of machine learning. The call for longitudinal studies and impact assessments advocates for evidence-based decision-making, contributing to the field’s maturity and guiding the implementation of effective educational technologies. In essence, this study contributes significantly to the ongoing discourse on machine learning and education. The findings and recommendations presented here serve as a foundational resource for future research initiatives, collaborations, and policy decisions, shaping the trajectory of machine learning in educational research for the foreseeable future.

8 Limitations and future work

One limitation of our study is that we relied only on the Web of Science database to generate the datasets. This approach may have overlooked relevant publications from other sources, resulting in an incomplete view of ML in educational research. Moreover, our findings may have been influenced by publication bias, as studies with significant results or positive outcomes are more likely to be published. This could have led to a skewed representation of the data. To overcome these limitations and further advance the understanding of ML in educational research, there are several avenues for future research. First and foremost, efforts should be made to enhance data collection by incorporating a wider range of sources. This could include unpublished studies and international databases, which would provide a more comprehensive perspective. Additionally, addressing publication bias through techniques such as funnel plot analysis and promoting the pre-registration of studies could help minimize biases and improve the accuracy of our findings.

Author contributions Musa Adekunle Ayanwale: Conceptualization, Writing- Original draft preparation, Methodology, Data Curation, Validation, Formal analysis, Writing—Review & Editing, Visualization, Supervision. Rethabile Rosemary Molefi: Writing—Original draft preparation, Writing Review, Formatting, Investigation. References alignment, Editing Saheed Oyeniran: Data generation through WoS database, Writing—Review & Editing.

Funding No funding received.

Data availability The authors obtained data for this study from the Web of Science (WoS) database and are available online for verification.

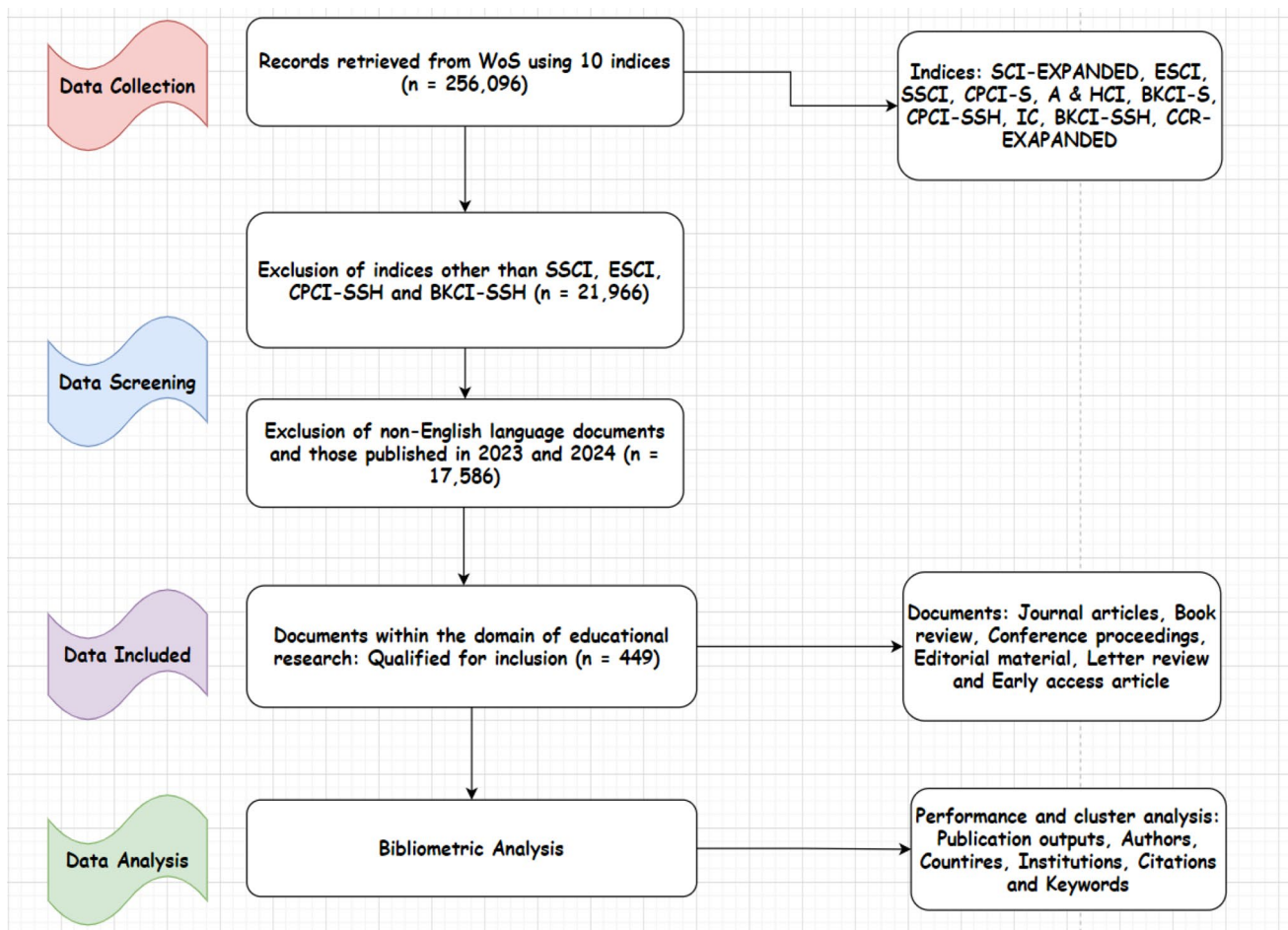
Declarations

Competing interests The authors wish to declare that there is no competing of interests regarding the publication of this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1

Flow chart of the screened publications based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.



References

1. Acedo FJ, Barroso C, Casanueva C, Galan JL. Co-authorship in management and organizational studies: an empirical and network analysis. *J Manage Stud.* 2006;43(5):957–83. <https://doi.org/10.1111/j.1467-6486.2006.00625.x>.

2. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. *Appl Sci*. 2023;13(12):7082. <https://doi.org/10.3390/app13127082>.
3. Aleixandre-Tudó JL, Castelló-Cogollos L, Aleixandre JL, & Aleixandre-Benaven, R. Emerging topics in scientific research on global water-use efficiency. *J Agric Sci*. 2019;157(6):480–92. <https://doi.org/10.1017/S0021859619000789>
4. Aria M, Cuccurullo C. Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informet*. 2017;11(4):959–75. <https://doi.org/10.1016/j.joi.2017.08.007>
5. Ayanwale MA, Mosia PA, Molefi RR, Shata L. Reliability components of online teaching and learning tools in lesotho higher education institutions: a systematic review. *Pertan J Sci Technol*. 2022;31(1):595–614. <https://doi.org/10.47836/pjst.31.1.34>.
6. Baker HK, Pandey N, Kumar S, Haldar A. A bibliometric analysis of board diversity: current status, development, and future research directions. *J Bus Res*. 2020;108:232–46. <https://doi.org/10.1016/j.jbusres.2019.11.025>.
7. Balaji P, Alelyani S, Qahmash A, Mohana M. Contributions of machine learning models towards student academic performance prediction: a systematic review. *Appl Sci*. 2021;11(21):10007. <https://doi.org/10.3390/APP112110007>.
8. Belmonte JL, Segura-Robles A, Moreno-Guerrero AJ, Parra-González ME. Machine learning and big data in the impact literature. A bibliometric review with scientific mapping in web of science. *Symmetry*. 2020. <https://doi.org/10.3390/SYM12040495>.
9. Bozic V. *Risks of Digital Divide in Using Artificial Intelligence (Ai)*. May. 2023. <https://doi.org/10.13140/RG.2.2.18156.13443>
10. Cobo MJ, L'opez-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field. *J Inf*. 2011;5(1):146–66. <https://doi.org/10.1016/j.joi.2010.10.002>.
11. Combrink HME, Marivate V, Masikisiki B. Technology-enhanced learning, data sharing, and machine learning challenges in South African education. *Educ Sci*. 2023;13(5):438–438. <https://doi.org/10.3390/EDUCSCI13050438>.
12. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. *J Bus Res*. 2021;133:285–96. <https://doi.org/10.1016/j.jbusres.2021.04.070>.
13. Donthu N, Gremler DD, Kumar S, Pattnaik D. Mapping of journal of service research themes: A 22-year review. *J Serv Res*. 2020. <https://doi.org/10.1177/1094670520977670032>.
14. Donthu N, Reinartz W, Kumar S, Pattnaik D. A retrospective review of the first 35 years of the international journal of research in marketing. *Int J Res Mark*. 2020. <https://doi.org/10.1016/j.ijresmar.2020.10.006>.
15. Ezugwu AE, Oyelade ON, Ikotun AM, Agushaka JO, Ho Y-S. Machine learning research trends in Africa: a 30 years overview with bibliometric analysis review. *Arch Comput Methods in Eng*. 2023;30(7):4177–207. <https://doi.org/10.1007/s11831-023-09930-z>.
16. Fomunyan KG. Theorising machine learning as an alternative pathway for higher education in Africa. *Int J Educ Pract*. 2020;8(2):268–77. <https://doi.org/10.18488/journal.61.2020.82.268.277>.
17. Gonzalez E, Colomo E, Cívico A. Quality education as a sustainable development goal in the context of 2030 Agenda: bibliometric approach. *Sustainability*. 2020;12(15): e5884. <https://doi.org/10.3390/su12155884>.
18. Hilbert S, Coors S, Kraus E, Bischl B, Lindl A, Frei M, Wild J, Krauss S, Goretzko D, Stachl C. Machine learning for the educational sciences. *Rev Educ*. 2021. <https://doi.org/10.1002/rev3.3310>.
19. Hjørland B. Facet analysis: the logical approach to knowledge organization. *Inf Process Manage*. 2013;49(2):545–57.
20. Jagwani A. A review of machine learning in education. *J Emerg Technol Innov Res*. 2019;6(5):384–6.
21. Kučak D, Juričić V, Đambić G. Machine learning in education—A survey of current research trends. *Ann DAAAM Proc Int DAAAM Symp*. 2018;29(1):0406–10. <https://doi.org/10.2507/29th.daaam.proceedings.059>.
22. Latif G, Abdelhamid SE, Fawagreh KS, Brahim GB, Alghazo R. Machine learning in higher education: students' performance assessment considering online activity logs. *IEEE Access*. 2023;11:69586–600. <https://doi.org/10.1109/ACCESS.2023.3287972>.
23. Lu D-N, Le H-Q, Vu T-H. the factors affecting acceptance of e-learning: a machine learning algorithm approach. *Educ Sci*. 2020;10(10):270. <https://doi.org/10.3390/educsci10100270>.
24. Liu W, Zheng J, Wang Z, Li R, Wu, T. A bibliometric review of ecological research on the Qinghai-Tibet Plateau, 1990–2019. *Ecol Inform*. 2021;64:101337. <https://doi.org/10.1016/j.ecoinf.2021.101337>.
25. Liu Z, Yin Y, Liu W, Dunford M. Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis. *Scientometrics*. 2015;103(1):135–58.
26. Lynch M. 6 Ways Machine Learning Will Revolutionize the Education Sector—The Tech Edvocate. *The Tech Edvocate*. 2019. <https://www.thetechedvocate.org/6-ways-machine-learning-will-revolutionize-the-education-sector/>. Accessed 26 Sep 2023
27. Magana EC, Ariza AC, Poletti G. Analysis of scientific production on pedagogy of death in the scopus. *Electron J Res Educ Psychol*. 2023;21(59):223–50.
28. Makumane MA. Students' perceptions on the use of LMS at a Lesotho university amidst the COVID-19 pandemic. *Afr Ident*. 2021;00(00):1–18. <https://doi.org/10.1080/14725843.2021.1898930>.
29. Matthew U, Kazaure J, Okafor N. Contemporary development in E-learning education, cloud computing technology and internet of things. *EAI Endors Trans Cloud Syst*. 2018;7(20):169173. <https://doi.org/10.4108/eai.31-3-2021.169173>.
30. Molefi RR, Ayanwale MA. Using composite structural equation modeling to examine high school teachers' acceptance of E-learning after Covid-19. *New Trends Issues Proc Humanit Soc Sci*. 2023;10(1):01–11. <https://doi.org/10.18844/prosoc.v10i1.8837>.
31. Moral-Munoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. *Prof de la inf/Inf Prof*. 2020. <https://doi.org/10.3145/epi.2020.ene.03>.
32. Mukherjee D, Lim WM, Kumar S, Donthu N. Guidelines for advancing theory and practice through bibliometric research. *J Bus Res*. 2022;148:101–15.
33. Okaiyeto K, Oguntibeju OO. Trends in diabetes research outputs in South Africa over 30 years from 2010 to 2019: a bibliometric analysis. *Saudi J Biol Sci* 2021;28:2914–24. <https://doi.org/10.1016/j.sjbs.2021.02.025>
34. Okolo CT, Aruleba K, Obaido G. Responsible AI in Africa—Challenges and Opportunities. Cham: Springer International Publishing; 2023. p. 35–64. https://doi.org/10.1007/978-3-031-08215-3_3.
35. Onyekwere J, EnamulHoquei K. Relationship between technological change, digitization, and students' attitudes toward distance learning in lagos higher education institutes. *Innoeduca Int J Technol Educ Innov*. 2023;9(1):126–42. <https://doi.org/10.24310/innoeduca.2023.v9i1.15286>.

36. Pedro F. Artificial intelligence in education: challenges and opportunities for sustainable development. Paris: Unesco; 2019.
37. Pinto AS, Abreu A, Costa E, Paiva J. How machine learning (ML) is transforming higher education: a systematic literature review. *J Inf Syst Eng Manag.* 2023;8(2):21168. <https://doi.org/10.55267/iadt.07.13227>.
38. Pugliese R, Regondi S, Marini R. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Sci Manag.* 2021;4(December):19–29. <https://doi.org/10.1016/j.dsm.2021.12.002>.
39. Ramos-Rodrigue AR, Ruiz-Navarro J. Changes in the intellectual structure of strategic management research: a bibliometric study of the strategic management journal, 1980–2000. *Strateg Manag J.* 2004;25(10):981–1004. <https://doi.org/10.1002/smj.397>.
40. Rossetto DE, Bernardes RC, Borini FM, Gattaz CC. Structure and evolution of innovation research in the last 60 years: review and future trends in the field of business through the citations and co-citations analysis. *Scientometrics.* 2018;115(3):1329–63. <https://doi.org/10.1007/s11192-018-2709-7>.
41. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. 2024. <https://www.R-project.org/>.
42. Sanusi IT, Oyelere SS, Omidiora JO. Exploring teachers' preconceptions of teaching machine learning in high school: a preliminary insight from Africa. *Comput Educ Open.* 2022;3:100072. <https://doi.org/10.1016/j.caeo.2021.100072>.
43. Sanusi IT, Oyelere SS, Vartiainen H, Suhonen J, Tukiainen M. A systematic review of teaching and learning machine learning in K-12 education. *Educ Inf Technol.* 2023. <https://doi.org/10.1007/s10639-022-11416-7>.
44. Sanusi IT, Oyelere SS, Vartiainen H, Suhonen J, Tukiainen M. Developing middle school students' understanding of machine learning in an African school. *Comput Educ Artif Intell.* 2023. <https://doi.org/10.1016/j.caeai.2023.100155>.
45. Sharma N, Sharma R, Jindal N. Machine learning and deep learning applications-A Vision. *Global Transit Proc.* 2021;2(1):24–8. <https://doi.org/10.1016/j.gltp.2021.01.004>.
46. Sivankalai S, Virumandi A, Sivasekaran K, Bala SB, Balamurugan B, Sharmila M, Kaladevi P. Scientometric analysis and visualization of Astrovirus based on R-packages. *Library Philos Pract.* 2021;3(1):1–15. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=9946&context=libphilprac>. Accessed 05 Oct 2023
47. Tahamtan I, Safipour Afshar A, Ahamdzadeh K. Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics.* 2016;107(3):1195–225. <https://doi.org/10.1007/s11192-016-1889-2>.
48. Tokmak A. Historical Transformation of the Ideal Citizen Profile in Turkey (1789–1938). *Int Online J Educ Sci.* 2023;15(4):799–812. <https://doi.org/10.15345/iojes.2023.04.014>.
49. Webb ME, Fluck A, Magenheimer J, Malyn-Smith J, Waters J, Deschênes M, Zagami J. Machine learning for human learners: opportunities, issues, tensions and threats. *Educ Tech Res Dev.* 2021;69(4):2109–30. <https://doi.org/10.1007/s11423-020-09858-2>.
50. Wei X, Yang F, Chen D, Li J, Shi X, Li B, Zhang C. Analyzing nanoparticle-induced neurotoxicity: a bibliometric analysis. *Toxicol Ind Health.* 2020;36(1):22–9. <https://doi.org/10.1177/0748233719900844>.
51. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education—where are the educators? *Int J Educ Technol High Educ.* 2019;16(1):39. <https://doi.org/10.1186/s41239-019-0171-0>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.