Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 14 No. 7 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijep.v14i7.50261

PAPER

Towards a Literature Review Methodology: A Practical Guide in the Context of Using Artificial Intelligence in Education

Mohamed El Jihaoui[1](✉), Oum El Kheir Abra[2], Khalifa Mansouri[1], Moulay El Houssine Ech-Chhibat[1]

[1]Laboratory: Modeling and Simulation of Intelligent Industrial Systems (M2S2I), ENSET Mohammedia, University Hassan II of Casablanca, Casablanca, Morocco

[2]Regional Center for Education and Training Professions, Rabat, Morocco

mohamed.eljihaoui-etu@etu.univh2c.ma

ABSTRACT

Scientific research plays an essential role in the development of mankind through its ability to provide innovative solutions and technological advances in various fields. In this perpetual quest, researchers must be active members of the scientific community, sharing their experiences and publishing their work and discoveries. All research work generally begins with a literature review (LR), which involves delving into previous work to examine the current state of knowledge, situate a research topic, and thus identify opportunities and explore new avenues of research. This important step can be arduous and complicated for new researchers. This paper aims to address the challenge faced by new researchers in conducting LRs, particularly in formulating search queries, identifying and selecting relevant studies, and extracting data from each paper, by proposing a methodological approach and applying it through a practical example in the context of using artificial intelligence in education. The approach is guided by the PRISMA (preferred reporting items for systematic reviews and meta-analyses) framework and focuses on recent studies conducted in 2021, 2022, and 2023. The proposed approach identified 52 out of 336 relevant studies. 65% of these studies were deemed to be of high quality (Q1 and Q2 rankings), and 40% of the articles were published in high-impact academic journals (Q1). This approach is versatile and can be adapted to different fields.

KEYWORDS

artificial intelligence (AI), case study, education, literature review (LR), practical guide, preferred reporting items for systematic reviews and meta-analyses (PRISMA)

1 INTRODUCTION

Scientific research is the pillar on which advances in knowledge and understanding of phenomena in many fields rest, offering a wide range of perspectives for resolving them. In this context, researchers play a very important role in the ongoing

El Jihaoui, M., Abra, O.E.K., Mansouri, K., Ech-Chhibat, M.E.H. (2024). Towards a Literature Review Methodology: A Practical Guide in the Context of Using Artificial Intelligence in Education. *International Journal of Engineering Pedagogy (iJEP)*, 14(7), pp. 119–145. https://doi.org/10.3991/ijep.v14i7.50261

Article submitted 2024-06-05. Revision uploaded 2024-07-11. Final acceptance 2024-07-11.

quest for knowledge, making them key players in the scientific community, where their contribution helps to advance the world around us. This contribution often takes the form of the publication of new discoveries in specialized journals, thereby enriching the knowledge base. The scientific research process usually begins with a state-of-the-art analysis or literature review (LR), which can involve a large number of publications [1] and, according to Chen and Zhuge [2], is a more time-consuming process than reading the news, as readers have to research and read numerous citations.

In this dynamic process of scientific research, which requires the examination of previous work, new researchers are often faced with the challenge of carrying out a LR. This is a crucial and arduous task, requiring more perseverance and time, as well as a clear methodology of organization and research, and tools and techniques that facilitate the identification, selection, extraction of data and information, and analysis of studies already carried out in a field. All this must be brought together to clearly define the scope of the subject. This crucial stage needs to be approached with a rigorous, considered approach that forces researchers to identify trends, discrepancies, and gaps in existing research.

Literature review is an initial step in any research project. During this crucial stage of a research project, the researcher must identify opportunities and explore new avenues of research to effectively develop and position the research topic. To achieve this goal, researchers must immerse themselves in existing studies on the subject, then analyze them to explore current trends in the field. This analysis is necessary for all scientists (researchers and academics) wishing to pursue their professional careers, as confirmed by Cuschieri et al. [3]. A LR is a scientific document that synthesizes the knowledge of a body of previous research in its entirety [4]. This document consists of identifying, collecting, analyzing, and discussing existing publications on a given topic in a methodical and well-structured way [5]. According to Shah et al. [6] and Socolofsky [7], its elaboration is a daunting task for new researchers. At this stage, beginners are faced with the challenge of selecting, evaluating, and organizing relevant sources, extracting useful information and knowledge, analyzing and criticizing it, overcoming prejudices, and managing their time efficiently.

Over time, various types of LR have emerged, including narrative, systematic, meta-analysis, general, descriptive, field, and critical LR. The authors Taherdoost [8] and Rocco et al. [9] review this taxonomy, highlighting the advantages and disadvantages of each type and examining their scope and applicability. Performing a LR essentially involves identifying and collecting studies on the subject under review and extracting data and information useful for analysis and writing. These two steps will be discussed and developed in detail throughout this paper.

In this paper, we aim to contribute to this field by proposing a methodological approach to conducting a LR. The proposed contribution consists in enriching the LR and populating the knowledge base by focusing on the various stages involved in carrying out a LR, including the stage of formulating the search query, the identification of relevant studies, and the phase of extracting data and information from each paper. The proposed approach will enable novice researchers to conduct a LR with greater ease and fewer methodological concerns. We will illustrate this approach with a practical example of the use of artificial intelligence (AI) in education.

Integrating AI into education enhances learning experiences, provides personalized feedback, and improves educational outcomes, making it an essential tool for tackling contemporary educational challenges and ensuring a sustainable future. It offers vast opportunities for innovative research and practical applications, as well as educational benefits [10], [11], [12].

For example, AI is playing a crucial role in the democratization of education by enhancing learning in massive open online courses (MOOCs) [13]. This is achieved by offering personalized content and interactive features, making high-quality resources more accessible to a wider audience. AI can be integrated into intelligent systems that can help education managers make data-driven decisions [14], such as the early identification of at-risk students [15], whether in terms of academic performance [16] or at risk of dropping out of school [17]. In addition, AI can analyze virtual social networks [18] to provide insight into students' social skills, helping educators to develop essential social skills. In addition, AI analysis enhances platforms such as YouTube [19] for asynchronous teaching, offering insights into student engagement and learning outcomes.

In the post-pandemic COVID-19 era [20], AI provides scalable and adaptive learning solutions that ensure continuity and quality of education while focusing on digital literacy and inclusion. The integration of AI into immersive technologies [21] enhances adaptive learning environments and real-time feedback, resulting in more engaging and effective pedagogical tools. AI also has the ability to analyze student interactions and provide personalized feedback and support by incorporating signal processing techniques [22]. In this way, AI can personalize educational experiences to meet individual needs, leading to improved learning outcomes.

The other sections of this document are organized as follows: Section 2 presents related work and analyzes recent and relevant studies of previous work related to our topic. Section 3 details our methodology and the steps involved in the LR, explained with the help of a concrete, step-by-step example. Section 4, Results, presents the conclusions drawn from our analysis, highlighting key ideas and significant patterns observed. In Section 5, we interpret and explain the results in the context of our research questions, providing explanations and discussing the implications of our findings. Finally, Section 6 summarizes the main points of this paper, highlights its contributions, and suggests potential directions and considerations for future research.

2 RELATED WORK

In the academic context, mastering a LR process is essential in scientific research, especially for new researchers learning how to conduct a LR independently. This topic has aroused the interest of many researchers, and several articles have been published on conducting a LR [23], [24], [25], [26], [27], [28]. To explore what has been achieved recently and present, as well as the current state of accumulated knowledge on this subject, we will use the process shown in Figure 1 to demonstrate our selection of recent and relevant studies for discussion.

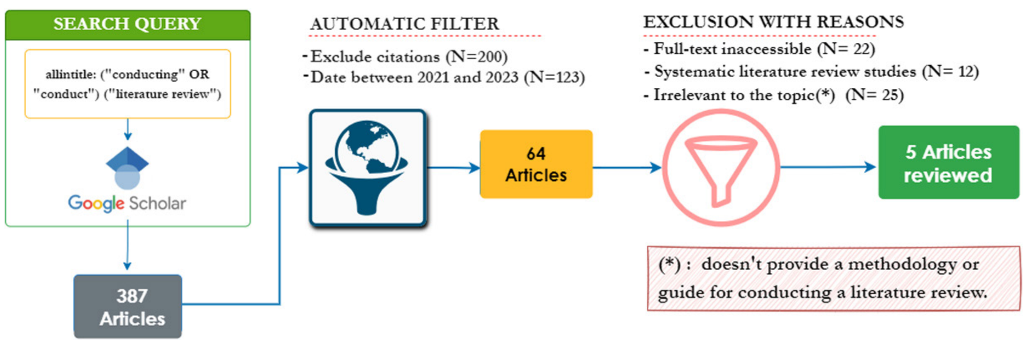

Fig. 1. Article selection process

This selection process uses Google Scholar as a search engine to target articles related to the conduct of LRs. Initially, 387 articles were identified. After filtering by date and excluding citations, 64 articles were evaluated. Of these, 59 were excluded for various reasons: 22 of their full texts were inaccessible, 12 were systematic review studies, and 25 articles were deemed irrelevant because they did not provide methodology or guidance for performing a LR. At the end of this process, we selected 5 studies that were considered relevant for conducting LR. These studies mainly describe and explain the methods and steps to follow in order to conduct a LR effectively [8], [9], [29], [30], and a state-of-the-art LR [31].

In this context, Fernandez et al. [29] propose a five-step protocol for conducting a scoping LR: (1) identify research questions; (2) identify relevant studies; (3) study selection; (4) chart the data; (5) summarize, and report. However, in this protocol, a crucial step that seems to be missing is the data extraction stage, which enables us to extract and collect data and information needed to prepare a synthesis of the studies included in a literature review.

The authors in [30] have designed seven activities for writing a LR: (1) Selecting a reviewed theme; (2) identifying keywords; (3) searching the source of research articles; (4) generating a reading list; (5) evaluate the research articles; (6) structuring the research article; (7) writing a LR, but without providing more practical details to elucidate these activities for new researchers.

In the studies proposed by [29] and [30], we noted the absence of a detailed section on the steps involved in constructing the search query, including the generation of the script to be executed on the chosen electronic databases.

In [8], the author offers step-by-step tips for conducting an LR and presents a detailed taxonomy of LRs. However, the in-depth review steps have not been practically clarified.

Barry et al. [31] outline a six-step process for conducting a state-of-the-art LR. The aim of this process is to develop a three-part argument regarding the state of knowledge about a specific phenomenon: *This is where we are now*. *This is how we got here*. *This is where we could go next*. However, this process fails to provide practical guidance on the methodological steps needed to conduct a LR and summarize the existing knowledge on a particular subject.

The previous analysis mainly revealed the need to explain in detail and contextualize the data extraction stage for the articles to be included in the LR. Given the importance of this part, it forms the basis of a robust analysis and guides the answers to research questions related to a topic. This analysis also revealed the need to further develop the search query construction stage, particularly the part linked to the script that will be executed on the electronic databases. Hence, there is a need for a more detailed approach to fill these gaps.

This paper, which is a continuation of previous work [8], [9], [29], [30], [31], we propose to contribute to this work by first detailing the search query script design stage, which aims to identify relevant studies related to the subject under review and further developing the data extraction phase. Our approach will be explained and illustrated step-by-step through a case study on using AI in education.

Through our exploration of previous research, this paper will address the following questions:

Q1: What are the essential steps for conducting a successful literature review?
Q2: How can the PRISMA framework be effectively utilized in the study selection process related to a research subject?

Q3: How effective are the keywords used in the search query?
Q4: How can we measure the approach's effectiveness in selecting relevant articles?
Q5: How can we organize and summarize the findings from the selected studies?

3 MATERIALS AND METHODS

This section will detail the research methodology used to conduct a LR, particularly the process of identifying articles and studies already carried out about our subject. In addition, we will propose a strategy for extracting useful information from each article. To carry out this study and ensure a complete and accurate analysis based on the studies [8], [9], [29], [30], [31], we propose an approach for conducting a LR as depicted in Figure 2.

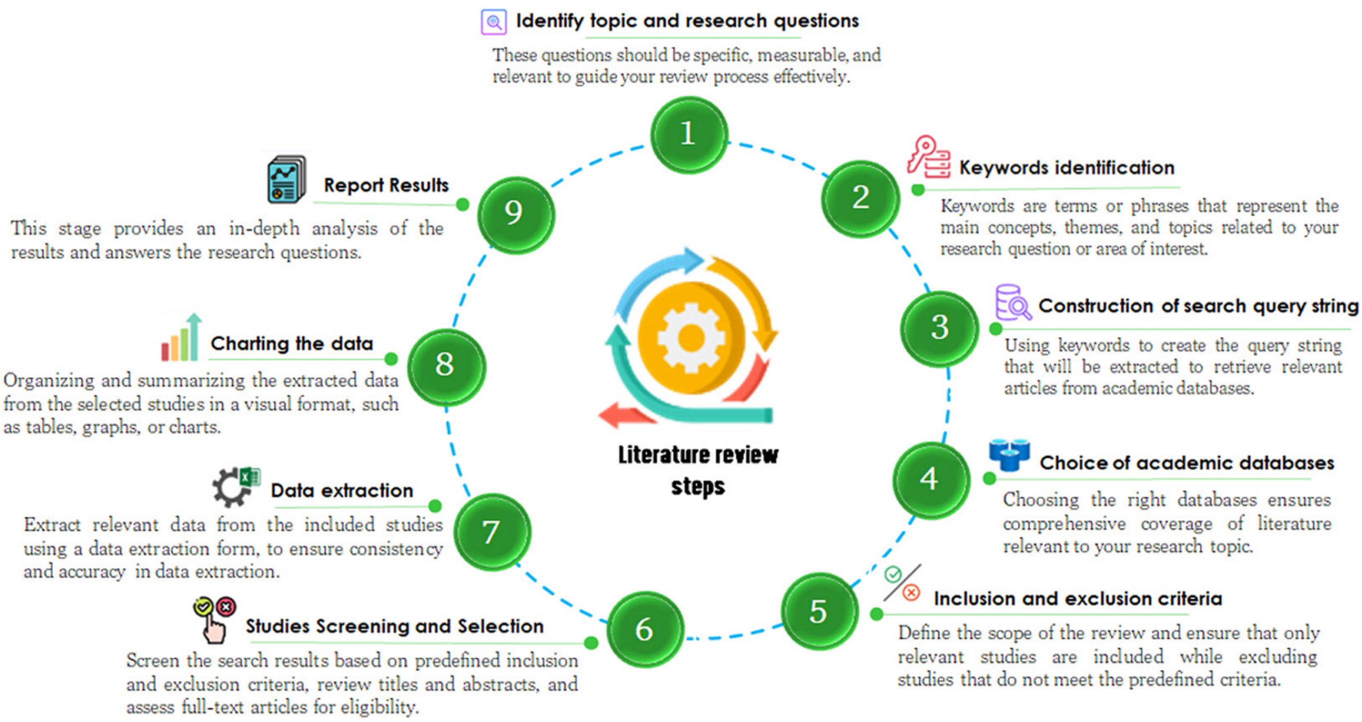

Fig. 2. Key steps for conducting a literature review (own elaboration)

3.1 Identify topic and research questions

To provide a concrete illustration of this process, we will explain our approach using an example: a case study on the topic of "Student retention in the education system through predictive analysis and educational data mining." In order to gather and analyze all relevant scientific articles on this study topic, we will use the following main research questions (RQ) to guide our approach:

RQ1: What are the most effective data mining and analysis methods for identifying at-risk students?
RQ2: What indicators can identify whether a student is at risk of dropping out?

RQ3: How can these indicators monitor and improve student retention in an education system?
RQ4: How is educational data used to improve student retention?
RQ5: What variables or characteristics are often used by predictive models to identify at-risk students in an education system?
RQ6: What are the ethical considerations and challenges of using educational data?

The RQs provided here are relevant to the subject we are using as an example for demonstration purposes. Although this paper won't delve into them, we will utilize these questions to identify keywords for our research query and to guide our paper selection process.

We have selected this specific subject for our approach because addressing problems using AI follows a similar procedure and utilizes comparable tools. This makes the data extraction phase, as described in Figure 6, more convenient. Initially, our focus will be on the process of selecting relevant studies and extracting valuable data and information from these chosen articles. Subsequently, we will conduct a thorough analysis of these studies to uncover insights and trends related to the research topic.

3.2 Keywords identification

The main concepts unveiled based on the keyword in the topic title and research questions are as follows: the first is related to retention, the second is related to the students or the sample we will be studying, the third is predictive analytics, often involving machine learning (a branch of AI), and the fourth concept is about educational data mining. We will identify the relevant terms related to these four concepts in Table 1 to construct our search query.

Table 1. Terms relevant to the construction of our search query

Concept	Related Concepts	Count
Retention	student retention, retention rate, student performance, dropping out, dropout, graduation rate, student at risk, student support, student engagement	9
Student	student, education, educational, school, college, university, higher education	7
Educational Data Mining	Data Mining, Educational Data Mining, or EDM, or Learning Analytics	3
Predictive Analytics	AI, artificial intelligence, predictive models, machine learning, *supervised learning, deep learning, predictive analytics, explainable AI	8
Total		27

For the sake of organization, we're going to classify keywords into three groups:

- **Technical aspect:** Terms related to the techniques used.
- **Topic:** On-the-spot terms.
- **Population or sample:** Terms that delineate the sample being studied.

All the keywords used in our search are listed and organized in Figure 3.

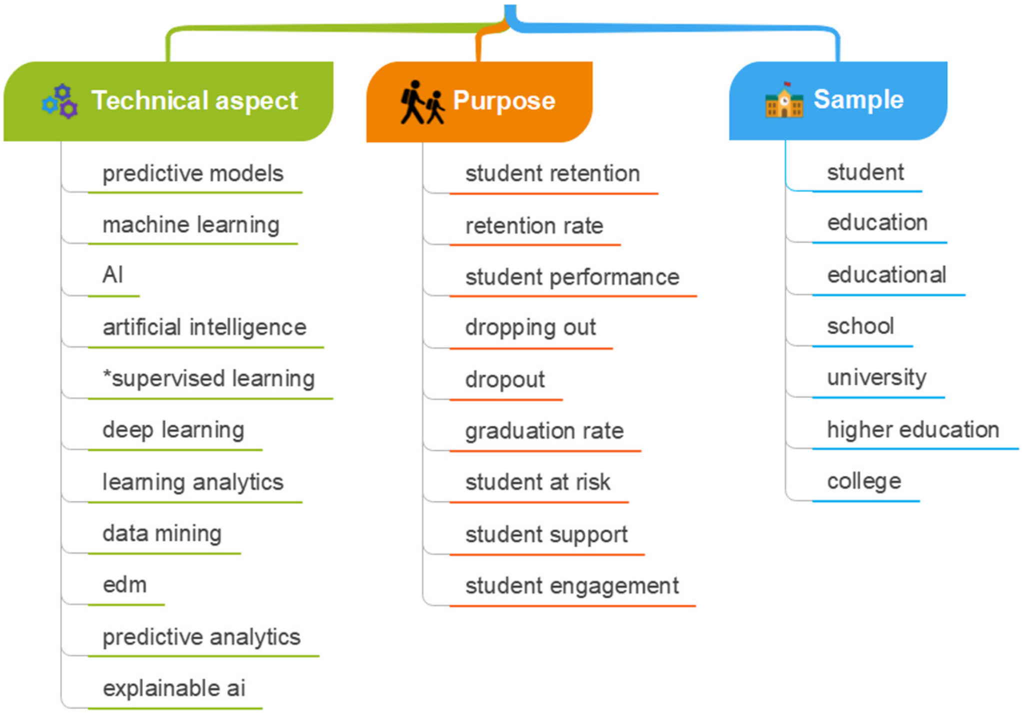

Fig. 3. Keywords used in our search

3.3 Search query string construction

After identifying the keywords (see Figure 3), we will construct our search query using logical operators and then develop the script for execution on electronic databases. Table 2 provides a summary of the various components of our request.

Table 2. Script for our search query

#ID	Keywords	Query Script
Req1	**Technical aspect**	("predictive models" OR "machine learning" OR "AI" OR "artificial intelligence" OR "*supervised learning" OR "deep learning" OR "learning analytics" OR "data mining" OR "edm" OR "predictive analytics" OR "explainable ai")
Req2	**Subject**	("student retention" OR "retention rate" OR "student performance" OR "dropping out" OR "dropout" OR "graduation rate" OR "student at risk" OR "student support" OR "student engagement")
Req3	**Sample**	("education" OR "educational" OR "school" OR "college" OR "university" OR "higher education")

Thus, the complete query results in the intersection of different parties cited in Table 2.

Search Query = **Req1** $\cap$ **Req2** $\cap$ **Req3**

3.4 Choice of academic databases

We will execute our script on the Scopus, Web of Science, and IEEE Xplore databases. We chose these databases for our literature research because of the wide range of available sources they provide. Moreover, the advanced search tools these databases offer will enable us to conduct precise and targeted searches by supporting the execution of complex and composed scripts. The complete script to be executed on the various databases is outlined in Table A1 in the Appendix.

Figure 4 displays the initial result obtained after the query script is executed on the various databases.

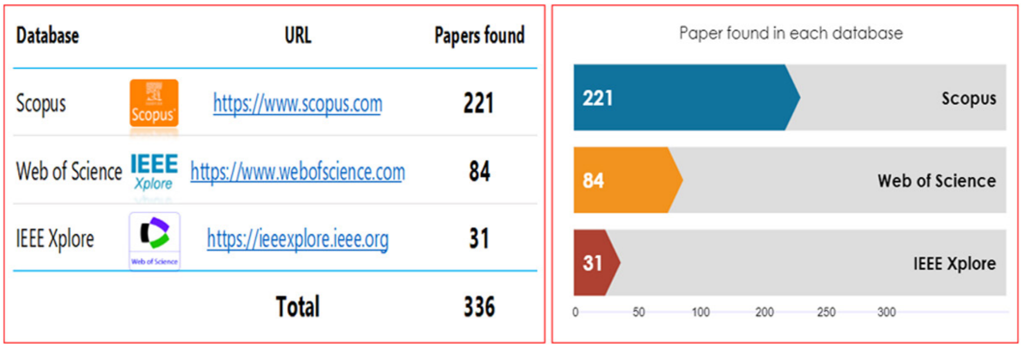

Fig. 4. Articles found in three databases

3.5 Inclusion and exclusion criteria

It is important to specify the inclusion and exclusion criteria for studies, i.e., the criteria based on which studies will or will not be included in a synthesis. This ensures that the list of articles included in our LR is accurate to our research topic. These criteria are used to define the research parameters and ensure the quality of the included studies. Clear, well-defined criteria guarantee that other researchers will reproduce the same work. Table 3 lists all the criteria used.

Table 3. Inclusion and exclusion criteria

Method	#ID	Criterion	Inclusion Criteria	Exclusion Criteria
Automatic	1	Publication Date	– Articles published in 2021, 2022, 2023	– Articles older than 2021
	2	Language	– Articles written in English	– Articles not written in English
	3	Document Type	– Journal Articles – Conference papers – Book Chapters – Review	– Erratum – Preprints
Semi-automatic	4	Remove Duplicate records	– No Duplicate records	– Duplicate records
Human intervention	5	Content	– Relevant to topic	– Irrelevant to topic
	6	Approach	– Education Data Mining or Machine learning Approaches	– Not an EDM or ML approaches

These criteria ensure that the LR focuses on recent and relevant approaches and techniques most widely used in the context of AI applications in education, namely machine learning (ML) and educational data mining (EDM) approaches:

- Publication Date: Articles from 2021 to 2023 are included to ensure the study incorporates the latest research.
- Language: The most popular language for publishing scientific articles within the international academic community.
- The document type criterion includes peer-reviewed articles, conference papers, book chapters, and reviews to ensure high credibility and reliability. Errata and preprints are excluded because errata are corrections to previous work and do not represent complete studies, while preprints are unvetted research that has not yet undergone peer review. This ensures the integrity and quality of the literature review.
- Content: Only articles relevant to the topic are included to maintain the clarity and focus of the review.
- Approach: studies using EDM or ML approaches are included to ensure the review is specific to the chosen methodologies and consistent in examining these techniques in educational contexts.

These criteria collectively ensure that the review is based on high-quality, recent, and relevant studies, providing a solid foundation for examining the application of EDM and ML in education.

After applying the criteria (1), (2), (3), and (4), we identified 117 items. Following a comprehensive review of each article's title, abstract, problem statement, methodology, and contribution, we established additional exclusion criteria, as outlined in Table 4.

Table 4. Additional exclusion criteria

#ID	Reason for Exclusion	Excluded Papers
7	Not a Machine learning Or Education Data Mining Approaches	**10**
8	Content written in the Spanish language	**4**
Total		**14**

3.6 Studies screening and selection

The PRISMA statement provides guidelines to help authors enhance the reporting of systematic reviews and meta-analyses. These guidelines ensure transparency and clarity in the reporting process. In our study, we will focus on the paper selection process by incorporating a vital component [32] of this framework, the PRISMA flow diagram. This diagram, shown in Figure 5, comprises four phases: **identification**, **screening**, **eligibility**, and **included** [33]. It depicts the process of selecting articles for our review, including the number of studies identified, included, and excluded, as well as the reasons for these exclusions.

A brief summary of each phase:

- **Identification:** Conduct comprehensive searches across multiple databases and other sources and remove duplicates.
- **Screening:** Review titles and abstracts to exclude studies that do not meet inclusion criteria.
- **Eligibility:** Assess the full text of potentially eligible studies and document reasons for exclusions.
- **Included:** Confirm eligible studies and extract relevant data for analysis.

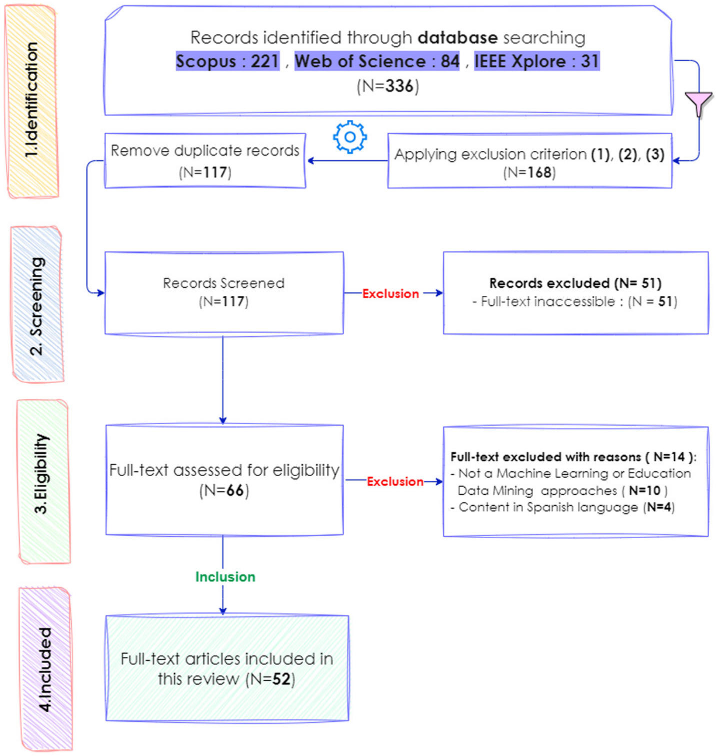

Fig. 5. PRISMA flow diagram adapted from [33]

The diagram depicted in Figure 5 illustrates the step-by-step approach used for selecting articles in a LR. Initially, a total of 336 records were found through database searches, including 221 from Scopus, 84 from Web of Science, and 31 from IEEE Xplore. After applying systematic exclusion filters (1), (2), and (3), as cited in Table 3, and removing duplicates, 117 articles remained for a screening process. This process involved evaluating titles, abstracts, problems, methodology, and the contribution of each paper. Through this process, 51 articles were excluded due to inaccessible full texts. The next phase involved a full-text eligibility assessment of the remaining 66 articles, among which 10 articles were excluded for not clearly applying a machine learning approach or educational data mining, and four articles were in Spanish. Finally, 52 complete articles met the inclusion criteria and were selected for the review.

The list of studies identified through this process has been categorized into four topics: student performance, student dropout, student retention, and student engagement, as shown in Table 5.

Table 5. Selected papers

Topic	AI Techniques	References	Count
Student performance	Supervised learning	[34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54]	21
	Unsupervised Learning	[48], [49], [50], [51], [52], [53]	6
	Deep Learning	[16], [55], [56]	3
	Semi-supervised learning	[57], [54]	2

(Continued)

Table 5. Selected papers *(Continued)*

Topic	AI Techniques	References	Count
Student Dropout	Supervised learning	[58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [17], [73], [74], [75], [76], [77]	21
	Unsupervised Learning	[78], [75], [76], [77]	4
	Deep Learning	[58]	1
Student Retention	Supervised learning	[79], [80], [81]	3
	Unsupervised Learning	[79]	1
	Deep Learning	[82]	1
Student Engagement	Unsupervised Learning	[83]	1

3.7 Data extraction strategy

By reviewing existing work, we can understand the current state of knowledge in our field of study and identify gaps in previous research. These gaps can steer our research toward unresolved questions or lesser-explored aspects of the topic. For this reason, we propose the model illustrated in Figure 6 to extract valuable information and insight from each paper. This approach can be used as a spreadsheet to simplify the capitalization of knowledge and the analysis of the data collected, which will be used to answer research questions and prepare a synthesis.

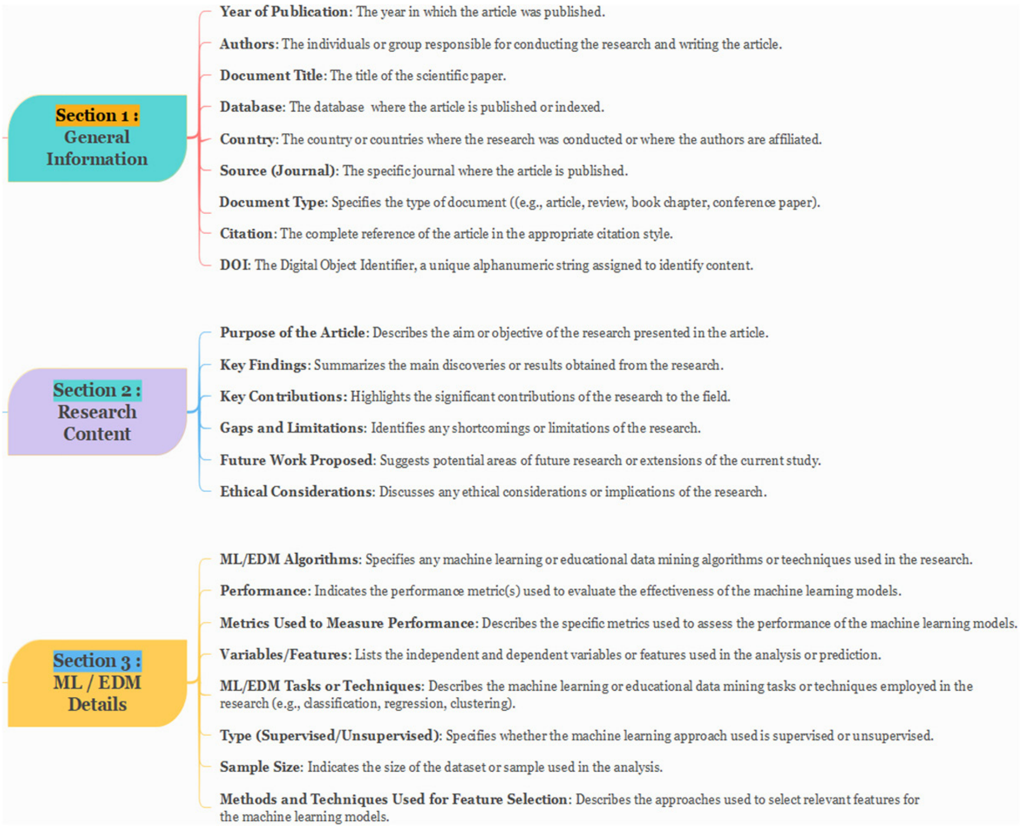

Fig. 6. Data extraction strategy model (own development)

In Table 6, we show an example of how the data extraction model was used to extract data and useful information for the paper "*The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates*" (Kiss et al., 2022).

Table 6. A practical example of how to use the data extraction strategy

Reference	**Kiss et al. (2022)**
Document Title	The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates
Database	Scopus
Document Type	Article
Country	United States
Quotation	Kiss, V., Maldonado, E., & Segall, M. (2022). The Use of Semester Course Data for Machine Learning Prediction of College Dropout Rates. Journal of Higher Education Theory and Practice, 22(4), 64–74. https://doi.org/10.33423/jhetp.v22i4.5130
Source (Journal)	Journal of Higher Education Theory and Practice
Purpose of the Article	Predict the likelihood of college students dropping out by analyzing demographic information and course performance overall enrolled semesters using machine learning (ML) techniques.
Key Findings	Logistic regression was found to be the most accurate ML technique among the four evaluated (including decision trees, neural networks, and support vector machines), with an accuracy of **84.8%** for predicting dropout rates using semester course performance data.
Key Contributions	– Compares multiple ML techniques for dropout prediction and demonstrates the effectiveness of logistic regression for this purpose. – Provides a model that can assist educational institutions in identifying and supporting at-risk students, thereby addressing student retention proactively.
Methodological Approach	The research employed a quantitative analysis, comparing four different ML techniques using demographic and course performance data across multiple semesters to predict dropout rates.
ML Algorithms	**Neural Networks**, **Decision Trees**, **Logistic Regression**, **Support Vector Machines**.
Performance	– Logistic regression method had the best accuracy of **84.8%**. – Decision trees (**82.2%**), – Neural networks (**80.8%**), – Vector machine support (**72.5%**).
Metrics used to measure performance	Positive Rate (TPR), False Positive Rate (FPR), Precision, and Accuracy, ROC CURVE
Variables or Features used for prediction	**Demographic predictors:** Gender (Binary), Race (Categorical), Age (Numerical) Major (Categorical), Student Type (Categorical), Transferred GPA (Numerical) **Performance Metrics:** Grade at Every Course (Categorical)
ML tasks	**Classification**
Types of ML techniques	**Supervised Learning**
School Level	**Higher Education**
Sample Size	**29,282** Students
Future Work Proposed	Future research directions include developing models based on more sophisticated algorithms like deep learning and applying the current model to other colleges within the institution.

3.8 Charting the data

Generally, this section is part of the "**RESULTS**" and involves organizing and summarizing data from selected studies in tables, graphs, or diagrams to make it easier to understand.

3.9 Report results

In this section, we will discuss the findings of the studies included in the LR, address the research questions, and provide an in-depth analysis of the results and conclusions of previous research, emphasizing the relevant data in a logical manner. We will not delve further into this section in this paper, as the answers to the research questions mentioned earlier in the section "**IDENTIFY THE TOPIC AND RESEARCH QUESTIONS**" regarding this case study will be published shortly in another article.

4 MAIN RESULTS

After applying the process described in the PRISMA flow diagram (see Figure 5) concerning the selection of articles to be studied and analyzed and applying all the inclusion and exclusion criteria mentioned in Table 3, we will summarize the results obtained in Figures 7–15.

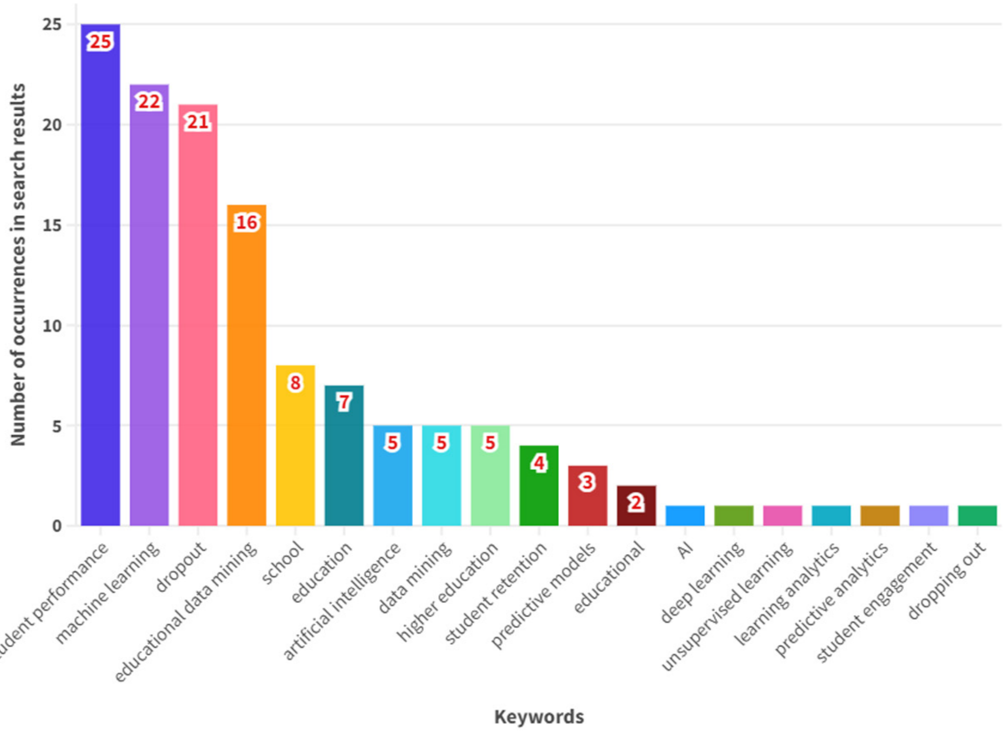

Fig. 7. Keyword occurrences in search results

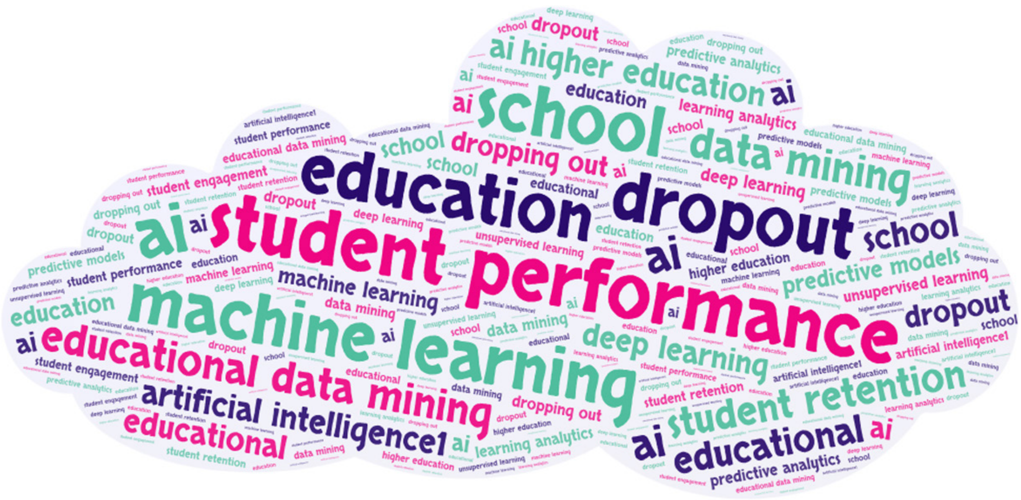

Fig. 8. Word cloud representation of the most important and frequently mentioned keywords in search results

These two figures (see Figures 7 and 8) clearly show that among the keywords used in our research, *'student performance,' 'machine learning,' 'dropout,' 'educational data mining,' 'school,' and 'education'* are the most common keywords that appear in the search results (in the titles of 52 article titles). This shows that there is a strong relationship between these keywords and the topic of school retention.

	artificial intelligence	deep learning	dropout	dropping out	education	educational	educational data mining	higher education	learning analytics	machine learning	predictive analytics	predictive models	school	student engagement	student performance	student retention	unsupervised learning	Total
AI					1			1										2
artificial intelligence					2	1							1					4
data mining			1		3	1		2		1			1		2	1		12
deep learning					1													1
dropout	1	1			4			3		14		3	6				1	33
educational data mining									1		1		1	1				4
machine learning				1	5	1	1	5					5		3			21
predictive models					1			1										2
student performance	3				4	1	9	2					1					20
student retention	1				2		1	2		2	1							9
Total	5	1	1	1	23	4	11	16	1	17	2	3	15	1	5	1	1	108

Fig. 9. Co-occurrence matrix

The co-occurrence matrix is used to capture the frequency with which these keywords appear together in search results (the titles of the 52 articles).

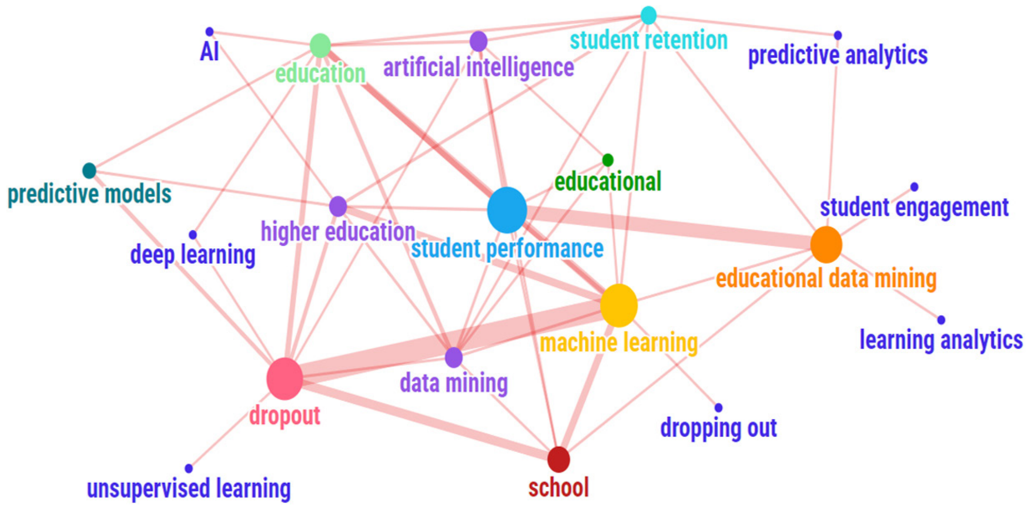

Fig. 10. Keyword co-occurrences as a network

This network representation makes it possible to visualize and analyze the relationships and interactions between different keywords used. Features are represented by nodes, and relationships between them are represented by links. The size of each node depends on the total occurrence of the keyword in the search results (see Figure 7), and the thickness of the link between two keywords is related to the frequency with which these keywords appear together; these frequencies are represented in the matrix (see Figure 9).

Database	Number of Studies	Percentage
Scopus	27	52%
Web of Science	20	38%
IEEE Xplore	5	10%
Total	52	

Fig. 11. Number of studies based on recognized scientific databases

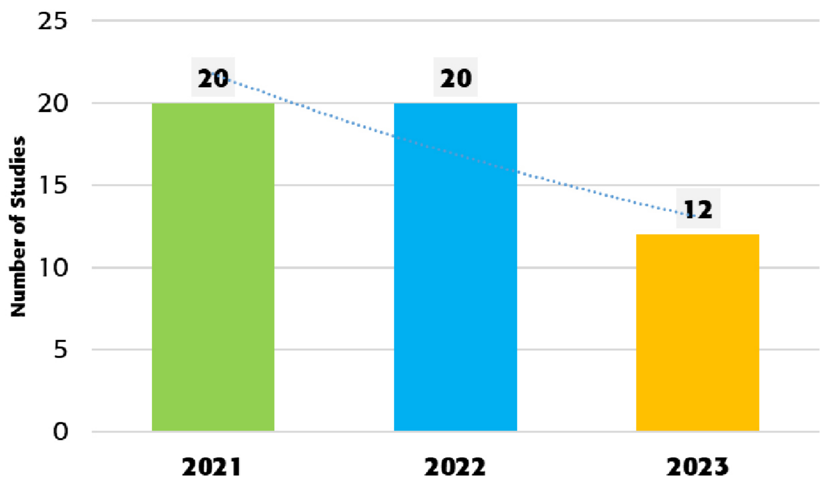

Fig. 12. Number of studies published in 2021, 2022, and 2023

This figure shows the number of articles published over the selected period. In 2021 and 2022, the number of articles published remained the same. The number of articles published fell slightly in 2023, but this is justified by the fact that the search for these articles was carried out before the end of the year, exactly on 09/11/2023.

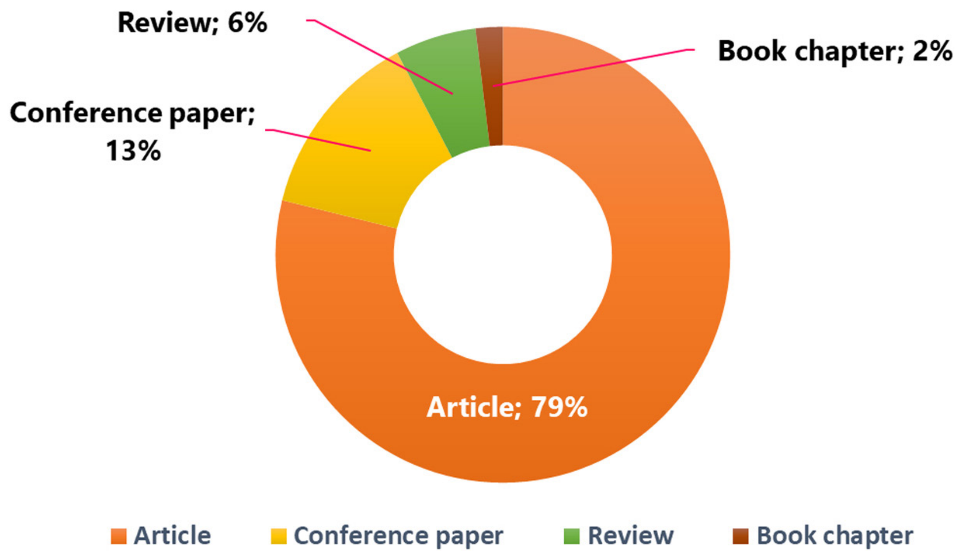

Fig. 13. Breakdown of studies based on type of document

In terms of the type of papers identified and studied, journal articles (41 studies) were the main type of papers, accounting for 79% of all papers identified by our approach.

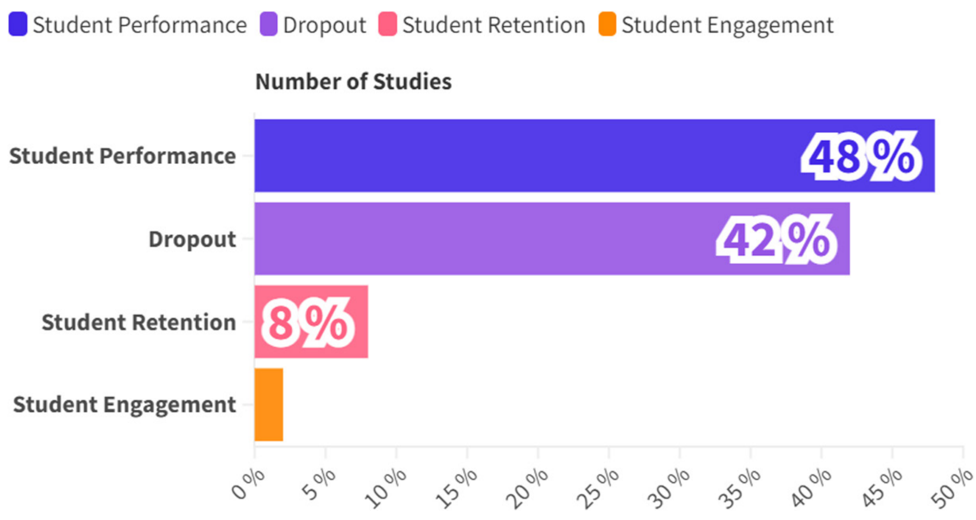

Fig. 14. Breakdown of studies based on theme

This figure illustrates the distribution of studies according to four distinct themes, which also shows the predominance of studies related to "student performance" (48%) and "student dropout" (42%). These two aspects are intimately related to the subject of school retention.

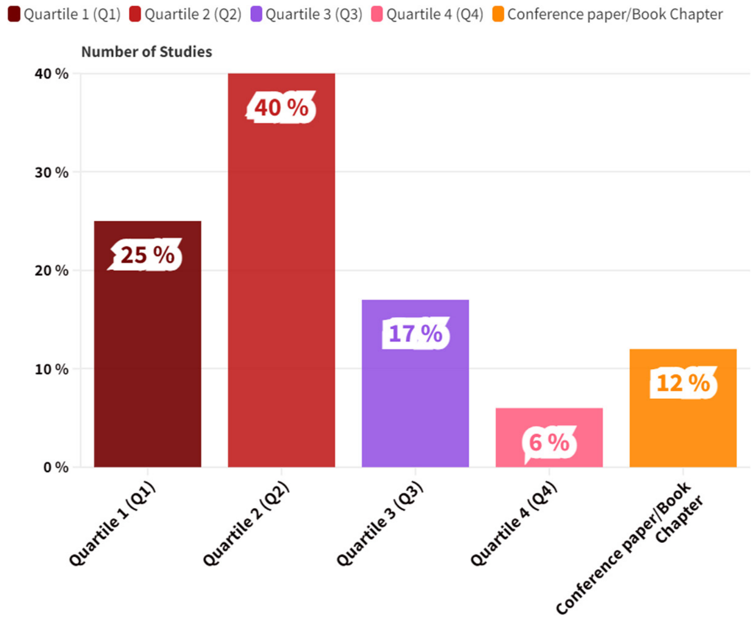

Fig. 15. Classification of studies based on the ranking of the journal

Figure 16 shows the distribution of studies according to the ranking of the academic journals in which they were published according to the SJR indicator [84]. This ranking (Q1, Q2, Q3, Q4) shows the quality and influence of these studies in their field of research.

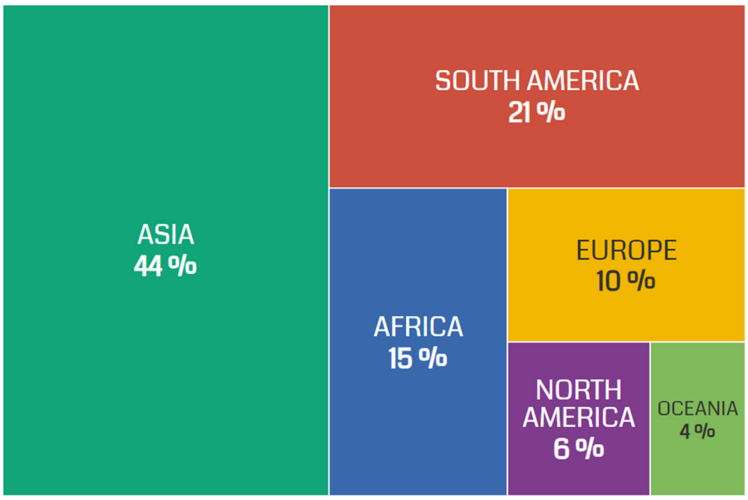

Fig. 16. Breakdown of studies based on region

5 DISCUSSION

There are a number of steps involved in the successful completion of a LR. These steps are described in Figure 2 and will be explained throughout this paper. In addition, a concrete example will be provided to demonstrate the process across multiple databases to gather a wide range of relevant studies. The use of the PRISMA framework can enhance this process by providing a structured approach to creating a search strategy, documenting the selection process with a flow chart (see Figure 5), systematically assessing study eligibility, and ensuring transparent reporting. The selection process involves the application of predetermined inclusion and exclusion criteria (refer to Table 3) to eliminate irrelevant studies and ensure that only relevant literature is considered.

Figure 6 provides a summary model for extracting useful knowledge and information from the selected studies that involves in-depth analysis to identify common themes and patterns, critical evaluation of methodologies and findings, synthesis of results to provide a comprehensive understanding, identification of gaps in the existing literature, and discussion of practical implications and applications of the findings.

Based on the results obtained after applying our methodological approach, 52 articles that meet the inclusion and exclusion criteria were selected from among the 336 items identified in the initial phase. Most of these studies focused on the topic of student performance (48%) and student dropout (42%), with less attention paid to student retention (8%) and engagement (2%) (see Figure 14), and most of these studies were indexed in the academic article databases Scopus (52%) and Web of Science (38%) (see Figure 11). The effectiveness of this approach can assess the impact factor and quality of journals in Figure 15, which shows that, among the 52 articles identified, 65% of publications are of high quality (Q1 and Q2 rankings), and 40% of these articles are published in journals with the highest impact factor (Q1) in the field of academic research.

The search query's effectiveness is evaluated based on the presence of the keywords used in the search results. In our case, out of the 27 keywords employed, 19 keywords (70%) appeared in the search results, as illustrated in Figure 7. This indicates a relatively high effectiveness of the chosen keywords, suggesting that most of the keywords were relevant and well-targeted for the research topic. However, the fact that eight keywords did not appear in the search results highlights the need for further refinement. This could involve adjusting or replacing fewer effective keywords to enhance the comprehensiveness and relevance of the search, ensuring that a broader and more accurate range of studies is captured. In our literature search, Figures 7 and 8 clearly show that among the keywords, 'student performance,' 'machine learning,' 'dropout,' 'educational data mining,' 'school,' and 'education' are the most frequent keywords in the search results (in the titles of 52 articles). This proves that there is an intersection and complementarity between these keywords and the topic of school retention. In addition, Figures 9 and 10 show a strong correlation between keyword pairs (*machine learning*, *dropout*) and (*educational data mining*, *student performance*). These keyword pairs often appear together in search results.

Although this paper presents a clear and concise approach that aims to equip junior researchers with practical knowledge, methodology, and tools (refer to Table A2) to conduct a LR easily, our study has some limitations: It depends on specific databases for academic papers; this could have omitted relevant studies not indexed in Scopus, Web of Science, or IEEE Xplore. In addition, the period limited to the last three years (2021, 2022, and 2023) may exclude relevant work before this time frame, which could have provided important additional information. Another gap that needs to be addressed is the method of selecting keywords (refer to Table 1), which should be systematic and well-argued. Indeed, of the 27 keywords used in the search query, 19 keywords (70%) appear in the search results (see Figure 7).

6 CONCLUSIONS

In this paper, we have addressed the issue of conducting a LR by proposing a methodological approach guided by the PRISMA framework, developed and illustrated by a practical example, a case study on the use of AI in education, and covering studies conducted over the last three years (2021, 2022, and 2023).

The steps in this approach provide a methodological guide for conducting and carrying out a LR on topics related to the use of AI and machine learning in education. This is confirmed by the proposed data extraction strategy model, which is perfectly suited to this type of topic. It can also be applied to topics in other fields, but with the simple challenge of adapting the data extraction strategy to meet the requirements of this field of study, in particular Section 3 of Figure 6, which concerns the technical aspects and tools of artificial intelligence.

This work is about equipping novice researchers with a solid arsenal that needs to be effectively exploited in order to clarify the steps, methods, and strategies for conducting an LR with the ultimate goal of publishing an article, which remains the dream of every researcher, especially a novice one. Our contribution serves not only as a practical guide for novice researchers but also as a demonstration of how to approach a research topic, as it provides a clear framework enriched with concrete examples that aim to enlighten novice researchers on the methods and strategies to follow in order to discover and detect untapped opportunities and innovative research avenues. This coincides with the ultimate goal of effectively and meaningfully defining and delimiting the research topic by capitalizing on the knowledge already accumulated in the field. Our approach presents a practical framework for synthesizing existing knowledge and identifying research gaps.

In conclusion, our work presents a methodological approach to conducting a LR and opens the way to ongoing reflection and future developments to perfect and generalize it.

It would be worth applying this methodology to different domains in order to assess its robustness and applicability in various contexts. In addition, an improvement that could be beneficial would be to automate the task of choosing and determining keywords by automatically generating them from the subject; this would avoid choosing keywords that will never appear in search results. Furthermore, the integration of AI techniques, such as the automatic analysis of article content (PDF format), could enrich our approach by enabling a more exhaustive and rapid exploration of existing literature.

7 ACKNOWLEDGMENT

We thank Professor Hala EL MOUTAOUKIL for his invaluable help in proofreading this paper. His mastery of the English language greatly improved the quality of this work.

8 REFERENCES

[1] N. I. Altmami and M. E. B. Menai, "Automatic summarization of scientific articles: A survey," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 4, pp. 1011–1028, 2022. https://doi.org/10.1016/j.jksuci.2020.04.020

[2] J. Chen and H. Zhuge, "Summarization of scientific documents by detecting common facts in citations," *Future Generation Computer Systems*, vol. 32, pp. 246–252, 2014. https://doi.org/10.1016/j.future.2013.07.018

[3] S. Cuschieri, V. Grech, and C. Savona-Ventura, "WASP (Write a Scientific Paper): Structuring a scientific paper," *Early Human Development*, vol. 128, pp. 114–117, 2019. https://doi.org/10.1016/j.earlhumdev.2018.09.011

[4] C. Okoli, "A guide to conducting a standalone systematic literature review," *CAIS*, vol. 37, 2015. https://doi.org/10.17705/1CAIS.03743

[5] M. Gottlieb and M. C. Utesch, "Publish or Perish: A scientific blueprint for a journal article," *Int. J. Eng. Ped. (iJEP)*, vol. 12, no. 3, pp. 171–177, 2022. https://doi.org/10.3991/ijep.v12i3.28253

[6] J. Shah, A. Shah, and R. Pietrobon, "Scientific writing of novice researchers: What difficulties and encouragements do they encounter?" *Academic Medicine*, vol. 84, no. 4, pp. 511–516, 2009. https://doi.org/10.1097/ACM.0b013e31819a8c3c

[7] S. A. Socolofsky, "How to write a research journal article in engineering and science," pp. 1–17, 2004. [Online]. Available: http://people.math.sfu.ca/~stockie/research/cfdgroup/Socolofsky2004.pdf

[8] H. Taherdoost, "Towards nuts and bolts of conducting literature review: A typology of literature review," *Electronics*, vol. 12, no. 4, p. 800, 2023. https://doi.org/10.3390/electronics12040800

[9] T. S. Rocco, M. S. Plakhotnik, C. M. McGill, D. Huyler, and J. C. Collins, "Conducting and writing a structured literature review in human resource development," *Human Resource Development Review*, vol. 22, no. 1, pp. 104–125, 2023. https://doi.org/10.1177/15344843221141515

[10] A. Harry, "Role of AI in education," *Injurity*, vol. 2, no. 3, pp. 260–268, 2023. https://doi.org/10.58631/injurity.v2i3.52

[11] N. D. Nguyen, "Exploring the role of AI in education," *London Journal of Social Sciences (LJSS)*, no. 6, pp. 84–95, 2023. https://doi.org/10.31039/ljss.2023.6.108

[12] A. Ezzaim, F. Kharroubi, A. Dahbi, A. Aqqal, and A. Haidine, "Artificial intelligence in education – State of the art," *International Journal of Computer Engineering and Data Science (IJCEDS)*, vol. 2, no. 2, 2022. https://www.ijceds.com/ijceds/article/view/37

[13] S. Sharov, S. Tereshchuk, A. Tereshchuk, V. Kolmakova, and N. Yankova, "Using MOOC to learn the Python programming language," *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 2, pp. 17–32, 2023. https://doi.org/10.3991/ijet.v18i02.36431

[14] Y. Wang, "When artificial intelligence meets educational leaders' data-informed decision-making: A cautionary tale," *Studies in Educational Evaluation*, vol. 69, p. 100872, 2021. https://doi.org/10.1016/j.stueduc.2020.100872

[15] R. Ravikumar, F. Aljanahi, A. Rajan, and V. Akre, "Early alert system for detection of at-risk students," in *2018 Fifth HCT Information Technology Trends (ITT)*, 2018, pp. 138–142. https://doi.org/10.1109/CTIT.2018.8649508

[16] Z. Chen, G. Cen, Y. Wei, and Z. Li, "Student performance prediction approach based on educational data mining," *IEEE Access*, vol. 11, pp. 131260–131272, 2023. https://doi.org/10.1109/ACCESS.2023.3335985

[17] E. Melo, I. Silva, D. G. Costa, C. M. D. Viegas, and T. M. Barros, "On the use of explainable artificial intelligence to evaluate school dropout," *Education Sciences*, vol. 12, no. 12, p. 845, 2022. https://doi.org/10.3390/educsci12120845

[18] S. Sharov, M. Vorovka, T. Sharova, and A. Zemlianska, "The impact of social networks on the development of students' social competence," *Int. J. Eng. Ped.*, vol. 11, no. 3, pp. 84–98, 2021. https://doi.org/10.3991/ijep.v11i3.20491

[19] Z. Kanetaki *et al.*, "Acquiring, analyzing and interpreting knowledge data for sustainable engineering education: An experimental study using YouTube," *Electronics*, vol. 11, no. 14, p. 2210, 2022. https://doi.org/10.3390/electronics11142210

[20] S. Jacques, A. Ouahabi, and Z. Kanetaki, "Post-COVID-19 education for a sustainable future: Challenges, emerging technologies and trends," *Sustainability*, vol. 15, no. 8, p. 6487, 2023. https://doi.org/10.3390/su15086487

[21] H. Chemerys, A. Vynogradova, H. Briantseva, and S. Sharov, "Strategy for implementing immersive technologies in the professional training process of future designers," *J. Phys.: Conf. Ser.*, vol. 1933, pp. 1–8, 2021. https://doi.org/10.1088/1742-6596/1933/1/012046

[22] S. Femmam, N. K. M'Sirdi, and A. Ouahabi, "Perception and characterization of materials using signal processing techniques," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 5, pp. 1203–1211, 2001. https://doi.org/10.1109/19.963184

[23] C. Cooper, A. Booth, J. Varley-Campbell, N. Britten, and R. Garside, "Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies," *BMC Med Res Methodol*, vol. 18, 2018. https://doi.org/10.1186/s12874-018-0545-3

[24] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, pp. 333–339, 2019. https://doi.org/10.1016/j.jbusres.2019.07.039

[25] A. S. Denney and R. Tewksbury, "How to write a literature review," *Journal of Criminal Justice Education*, vol. 24, no. 2, pp. 218–234, 2013. https://doi.org/10.1080/10511253.2012.730617

[26] J. Randolph, "A guide to writing the dissertation literature review," *Practical Assessment, Research, and Evaluation*, vol. 14, no. 1, p. 13, 2009. https://doi.org/10.7275/b0az-8t74

[27] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002. [Online]. Available: http://www.jstor.org/stable/4132319

[28] R. Ferrari, "Writing narrative style literature reviews," *Medical Writing*, vol. 24, no. 4, pp. 230–235, 2015. https://doi.org/10.1179/2047480615Z.000000000329

[29] K. Fernandez, A. Buhler, and S. Rivera-Jimenez, "Methods for conducting a scoping literature review on institutional culture and transformational change in engineering education," in *2023 ASEE Annual Conference & Exposition Proceedings*, Baltimore, Maryland: ASEE Conferences, 2023. https://doi.org/10.18260/1-2--43639

[30] D. Haris, Chairunisah, and D. Armanto, "Writing a literature review: Activities on research in mathematics education to support students in conducting research," *J. Phys.: Conf. Ser.*, vol. 1819, pp. 1–6, 2021. https://doi.org/10.1088/1742-6596/1819/1/012028

[31] E. S. Barry, J. Merkebu, and L. Varpio, "How to conduct a state-of-the-art literature review," *Journal of Graduate Medical Education*, vol. 14, no. 6, pp. 663–665, 2022. https://doi.org/10.4300/JGME-D-22-00704.1

[32] S. Patole, Ed., *Principles and Practice of Systematic Reviews and Meta-Analysis.* Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-71921-0

[33] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *International Journal of Surgery*, vol. 8, no. 5, pp. 336–341, 2010. https://doi.org/10.1016/j.ijsu.2010.02.007

[34] M. S. Ahmad, A. H. Asad, and A. Mohammed, "A machine learning based approach for student performance evaluation in educational data mining," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2021, pp. 187–192. https://doi.org/10.1109/MIUCC52538.2021.9447602

[35] M. N. Yakubu and A. M. Abubakar, "Applying machine learning approach to predict students' performance in higher educational institutions," *Kybernetes*, vol. 51, no. 2, pp. 916–934, 2022. https://doi.org/10.1108/K-12-2020-0865

[36] V. Siafis and M. Rangoussi, "Educational data mining-based visualization and early prediction of student performance: A synergistic approach," in *Proceedings of the 26th Pan-Hellenic Conference on Informatics (PCI '22)*, 2022, pp. 246–253. https://doi.org/10.1145/3575879.3576000

[37] M. Ragab, A. M. K. Abdel Aal, A. O. Jifri, and N. F. Omran, "Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, pp. 1–9, 2021. https://doi.org/10.1155/2021/6241676

[38] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. M. S. Islam, "Mining educational data to predict students' performance: A comparative study of data mining techniques," *Educ. Inf. Technol.*, vol. 26, pp. 6051–6067, 2021. https://doi.org/10.1007/s10639-021-10575-3

[39] D. M. Ahmed, A. M. Abdulazeez, D. Q. Zeebaree, and F. Y. H. Ahmed, "Predicting university's students performance based on machine learning techniques," in *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, 2021, pp. 276–281. https://doi.org/10.1109/I2CACIS52118.2021.9495862

[40] J. R. Asor, G. M. B. Catedrilla, C. A. C. Buama, M. E. Malabayabas, and C. E. Malabayabas, "Prediction of senior high school students' performance in a state university: An educational data mining approach," *International Journal of Information and Education Technology (IJIET)*, vol. 13, no. 6, pp. 925–931, 2023. https://doi.org/10.18178/ijiet.2023.13.6.1888

[41] N. Z. Salih and W. Khalaf, "Prediction of student's performance through educational data mining techniques," *IJEECS*, vol. 22, no. 3, pp. 1708–1715, 2021. https://doi.org/10.11591/ijeecs.v22.i3.pp1708-1715

[42] S. Sengupta, "Towards finding a minimal set of features for predicting students' performance using educational data mining," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 15, no. 3, pp. 44–54, 2023. https://doi.org/10.5815/ijmecs.2023.03.04

[43] M. Zafari, A. Sadeghi-Niaraki, S.-M. Choi, and A. Esmaeily, "A practical model for the evaluation of high school student performance based on machine learning," *Applied Sciences*, vol. 11, no. 23, p. 11534, 2021. https://doi.org/10.3390/app112311534

[44] G. Latif, S. E. Abdelhamid, K. S. Fawagreh, G. B. Brahim, and R. Alghazo, "Machine learning in higher education: Students' performance assessment considering online activity logs," *IEEE Access*, vol. 11, pp. 69586–69600, 2023. https://doi.org/10.1109/ACCESS.2023.3287972

[45] A. Jokhan, A. A. Chand, V. Singh, and K. A. Mamun, "Increased digital resource consumption in higher educational institutions and the artificial intelligence role in informing decisions related to student performance," *Sustainability*, vol. 14, no. 4, p. 2377, 2022. https://doi.org/10.3390/su14042377

[46] H. Pallathadka, B. Sonia, D. T. Sanchez, J. V. De Vera, J. A. T. Godinez, and M. T. Pepito, "Investigating the impact of artificial intelligence in education sector by predicting student performance," in *Materials Today: Proceedings*, vol. 51, 2022, pp. 2264–2267. https://doi.org/10.1016/j.matpr.2021.11.395

[47] J. Malini and Y. Kalpana, "Investigation of factors affecting student performance evaluation using education materials data mining technique," in *Materials Today: Proceedings*, vol. 47, 2021, pp. 6105–6110. https://doi.org/10.1016/j.matpr.2021.05.026

[48] D. Moonsamy, N. Naicker, T. T. Adeliyi, and R. E. Ogunsakin, "A meta-analysis of educational data mining for predicting students performance in programming," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 2, 2021. https://doi.org/10.14569/IJACSA.2021.0120213

[49] M. A. Saleh, S. Palaniappan, N. A. A. Abdallah, and M. M. B. Baroud, "Predicting student performance using data mining and learning analysis technique in Libyan higher education," *PEN*, vol. 9, no. 3, pp. 509–519, 2021. https://doi.org/10.21533/pen.v9i3.2212

[50] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Educ Inf Technol*, vol. 26, pp. 205–240, 2021. https://doi.org/10.1007/s10639-020-10230-3

[51] A. Dhankhar, K. Solanki, S. Dalal, and Omdev, "Predicting students performance using educational data mining and learning analytics: A systematic literature review," in *Innovative Data Communication Technologies and Application*. in Lecture Notes on Data Engineering and Communications Technologies, J. S. Raj, A. M. Iliyasu, R. Bestak, and Z. A. Baig, Eds., Springer, Singapore, vol. 59, 2021, pp. 127–140. https://doi.org/10.1007/978-981-15-9651-3_11

[52] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational data mining techniques for student performance prediction: Method review and comparison analysis," *Front. Psychol.*, vol. 12, 2021. https://doi.org/10.3389/fpsyg.2021.698490

[53] M. H. Bin Roslan and C. J. Chen, "Educational data mining for student performance prediction: A systematic literature review (2015–2021)," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 5, pp. 147–179, 2022. https://doi.org/10.3991/ijet.v17i05.27685

[54] W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students' performance," *Engineering Reports*, vol. 4, no. 5, p. e12482, 2022. https://doi.org/10.1002/eng2.12482

[55] R. Josphineleela, V. Sundararajan, K. Meenakshi, A. Maruthi Varaprasad, P. Kumar Yadavalli, and D. Praveenadevi, "Ai based structural equation modelling to classify the students' performance in higher education institutions," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 4s, pp. 203–212, 2023. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/2647

[56] A. Bressane *et al.*, "Fuzzy artificial intelligence—Based model proposal to forecast student performance and retention risk in engineering education: An alternative for handling with small data," *Sustainability*, vol. 14, no. 21, p. 14071, 2022. https://doi.org/10.3390/su142114071

[57] R. H. Ali, "Educational data mining for predicting academic student performance using active classification," *Iraqi Journal of Science*, vol. 63, no. 9, pp. 3954–3965, 2022. https://doi.org/10.24996/ijs.2022.63.9.27

[58] Y.-T. Shiao, C.-H. Chen, K.-F. Wu, B.-L. Chen, Y.-H. Chou, and T.-N. Wu, "Reducing dropout rate through a deep learning model for sustainable education: Long-term tracking of learning outcomes of an undergraduate cohort from 2018 to 2021," *Smart Learn. Environ.*, vol. 10, 2023. https://doi.org/10.1186/s40561-023-00274-6

[59] A. J. Fernandez-Garcia, J. C. Preciado, F. Melchor, R. Rodriguez-Echeverria, J. M. Conejero, and F. Sanchez-Figueroa, "A real-life machine learning experience for predicting university dropout at different stages using academic data," *IEEE Access*, vol. 9, pp. 133076–133090, 2021. https://doi.org/10.1109/ACCESS.2021.3115851

[60] L. Cañete-Sifuentes, V. Robles, E. Menasalvas, and R. M. Borja, "Comparing automated machine learning against an off-the-shelf pattern-based classifier in a class imbalance problem: Predicting university dropout," *IEEE Access*, vol. 11, pp. 139147–139156, 2023. https://doi.org/10.1109/ACCESS.2023.3336596

[61] T. I. Pehlivanova and V. I. Nedeva, "Attributes selection using machine learning for analysing students' dropping out of university: A case study," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1031, no. 1, p. 012055, 2021. https://doi.org/10.1088/1757-899X/1031/1/012055

[62] H. S. Park and S. J. Yoo, "Early dropout prediction in online learning of university using machine learning," *JOIV: Int. J. Inform. Visualization*, vol. 5, no. 4, pp. 347–353, 2021. https://doi.org/10.30630/joiv.5.4.732

[63] J. Leo, "Early-warning dropout visualization tool for secondary schools: Using machine learning, QR code, GIS and mobile application techniques," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 11, 2022. https://doi.org/10.14569/IJACSA.2022.0131176

[64] Y. N. Mnyawami, H. H. Maziku, and J. C. Mushi, "Enhanced model for predicting student dropouts in developing countries using automated machine learning approach: A case of Tanzanian's secondary schools," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022. https://doi.org/10.1080/08839514.2022.2071406

[65] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100066, 2022. https://doi.org/10.1016/j.caeai.2022.100066

[66] K. S. Selim and S. S. Rezk, "On predicting school dropouts in Egypt: A machine learning approach," *Educ. Inf. Technol.*, vol. 28, pp. 9235–9266, 2023. https://doi.org/10.1007/s10639-022-11571-x

[67] A. B. Urbina-Nájera and L. A. Méndez-Ortega, "Predictive model for taking decision to prevent university dropout," *International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI)*, vol. 7, no. 4, pp. 205–213, 2022. https://doi.org/10.9781/ijimai.2022.01.006

[68] J. K. Hoyos Osorio and G. Daza Santacoloma, "Predictive model to identify college students with high dropout rates," *Revista Electrónica de Investigación Educativa*, vol. 25, 2023. https://doi.org/10.24320/redie.2023.25.e13.5398

[69] V. Kiss, E. Maldonado, and M. Segall, "The use of semester course data for machine learning prediction of college dropout rates," *Journal of Higher Education Theory and Practice*, vol. 22, no. 4, pp. 64–74, 2022. https://doi.org/10.33423/jhetp.v22i4.5130

[70] D. Opazo, S. Moreno, E. Álvarez-Miranda, and J. Pereira, "Analysis of first-year university student dropout through machine learning models: A comparison between universities," *Mathematics*, vol. 9, no. 20, p. 2599, 2021. https://doi.org/10.3390/math9202599

[71] V. Flores, S. Heras, and V. Julian, "Comparison of predictive models with balanced classes using the SMOTE method for the forecast of student dropout in higher education," *Electronics*, vol. 11, no. 3, p. 457, 2022. https://doi.org/10.3390/electronics11030457

[72] M. Segura, J. Mello, and A. Hernández, "Machine learning prediction of university student dropout: Does preference play a key role?" *Mathematics*, vol. 10, no. 18, p. 3359, 2022. https://doi.org/10.3390/math10183359

[73] D. A. Gutierrez-Pachas, G. Garcia-Zanabria, E. Cuadros-Vargas, G. Camara-Chavez, and E. Gomez-Nieto, "Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the Latin American context," *Education Sciences*, vol. 13, no. 2, p. 154, 2023. https://doi.org/10.3390/educsci13020154

[74] J. Kabathova and M. Drlik, "Towards predicting student's dropout in university courses using different machine learning techniques," *Applied Sciences*, vol. 11, no. 7, p. 3130, 2021. https://doi.org/10.3390/app11073130

[75] P. Rodríguez, A. Villanueva, L. Dombrovskaia, and J. P. Valenzuela, "A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: The case of Chile," *Educ. Inf. Technol.*, vol. 28, pp. 10103–10149, 2023. https://doi.org/10.1007/s10639-022-11515-5

[76] K. Oqaidi, S. Aouhassi, and K. Mansouri, "Towards a students' dropout prediction model in higher education institutions using machine learning algorithms," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 18, pp. 103–117, 2022. https://doi.org/10.3991/ijet.v17i18.25567

[77] L. Rondado De Sousa, V. Oliveira De Carvalho, B. Penteado, and F. Affonso, "A systematic mapping on the use of data mining for the face-to-face school dropout problem," in *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU)*, vol. 1, 2021, pp. 36–47. https://doi.org/10.5220/0010476300360047

[78] M. A. Valles-Coral *et al.*, “Density-based unsupervised learning algorithm to categorize college students into dropout risk levels,” *Data*, vol. 7, no. 11, p. 165, 2022. https://doi.org/10.3390/data7110165
[79] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, “Student retention using educational data mining and predictive analytics: A systematic literature review,” *IEEE Access*, vol. 10, pp. 72480–72503, 2022. https://doi.org/10.1109/ACCESS.2022.3188767
[80] W. Villegas-Ch, J. Govea, and S. Revelo-Tapia, “Improving student retention in institutions of higher education through machine learning: A sustainable approach,” *Sustainability*, vol. 15, no. 19, p. 14512, 2023. https://doi.org/10.3390/su151914512
[81] C. A. Palacios, J. A. Reyes-Suárez, L. A. Bearzotti, V. Leiva, and C. Marchant, “Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile,” *Entropy*, vol. 23, no. 4, p. 485, 2021. https://doi.org/10.3390/e23040485
[82] S. M. Arqawi, E. A. Zitawi, A. H. Rabaya, B. S. Abunasser, and S. S. Abu-Naser, “Predicting university student retention using artificial intelligence,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 9, 2022. https://doi.org/10.14569/IJACSA.2022.0130937
[83] L. M. Nkomo and M. Nat, “Student engagement patterns in a blended learning environment: An educational data mining approach,” *TechTrends*, vol. 65, pp. 808–817, 2021. https://doi.org/10.1007/s11528-021-00638-0
[84] “Scimago Journal & Country Rank.” Accessed: Apr. 21, 2024. [Online]. Available: https://www.scimagojr.com/

9 APPENDICES

Table A1. Complete script of our search query for the three databases

Database	Query Script
Scopus	TITLE (“student retention” OR “retention rate” OR “student performance” OR “dropping out” OR “dropout” OR “graduation rate” OR “student at risk” OR “student support” OR “student engagement”) AND TITLE (“predictive models” OR “machine learning” OR “AI” OR “artificial intelligence” OR “*supervised learning” OR “deep learning” OR “learning analytics” OR “data mining” OR “edm” OR “predictive analytics” OR “explainable ai”) AND TITLE (“education” OR “educational” OR “school” OR “college” OR “university” OR “higher education”)
IEEE Xplore	((“Document Title”:”predictive models” OR “Document Title”:”machine learning” OR “Document Title”:”AI” OR “Document Title”:”artificial intelligence” OR “Document Title”:”*supervised learning” OR “Document Title”:”deep learning” OR “Document Title”:”learning analytics” OR “Document Title”:”data mining” OR “Document Title”:”edm” OR “Document Title”:”predictive analytics” OR “Document Title”:”explainable ai”) AND (“Document Title”:”student retention” OR “Document Title”:”retention rate” OR “Document Title”:”student performance” OR “Document Title”:”dropping out” OR “Document Title”:”dropout” OR “Document Title”:”graduation rate” OR “Document Title”:”student at risk” OR “Document Title”:”student support” OR “Document Title”:”student engagement”) AND (“Document Title”:”education” OR “Document Title”:”educational” OR “Document Title”:”school” OR “Document Title”:”college” OR “Document Title”:”university” OR “Document Title”:”higher education”))
Web of Science	TI=(“predictive models” OR “machine learning” OR “AI” OR “artificial intelligence” OR “*supervised learning” OR “deep learning” OR “learning analytics” OR “data mining” OR “edm” OR “predictive analytics” OR “explainable ai”) AND TI=(“student retention” OR “retention rate” OR “student performance” OR “dropping out” OR “dropout” OR “graduation rate” OR “student at risk” OR “student support” OR “student engagement”) AND TI=(“education” OR “educational” OR “school” OR “college” OR “university” OR “higher education”)

Table A2. Useful tools for bibliographic organization and diagrams

Tool	Description
https://www.zotero.org/	**Zotero** is a free and easy-to-use tool for bibliographic management and organization. Its ability to integrate with browsers and word processing software (such as Word) makes it easier to manage academic sources, saving valuable time and ensuring accurate citations.
EdrawMind https://www.edrawsoft.com/	**EdrawMind** is a mind mapping and brainstorming software. It offers a wide range of features for mind-mapping creation.
draw.io https://app.diagrams.net/	**Draw.io** offers robust diagramming tools, which are essential for creating clear and informative visual representations of relationships and processes. These visual aids can significantly improve the understanding and presentation of results.
WordArt https://wordart.com/	WordArt.com is an online word cloud generator that enables you to create amazing and unique word clouds with ease even for users with no prior knowledge of graphic design. This word cloud generator allows us to create a visual representation of text that gives a higher rank to words that appear more frequently.
Flourish https://app.flourish.studio/	Flourish is a user-friendly platform for creating interactive data visualizations and storytelling experiences. It offers customizable templates, interactivity features, storytelling tools, and easy sharing options. With Flourish, you can import data, create engaging visuals, add narratives, and share your work seamlessly.

10 AUTHORS

Mohamed El Jihaoui holds a bachelor's degree in mathematical and computer sciences and a master's in educational technologies and pedagogical innovations. He is currently a PhD student in the modeling and simulation of intelligent industrial systems (M2S2I) research laboratory at ENSET of Mohammedia, Hassan II University of Casablanca. His research focuses on information and communication technology for education management (E-mail: mohamed.eljihaoui-etu@etu.univh2c.ma).

Oum El Kheir Abra is an Associate Professor of computer science and Deputy Director of the Regional Center of Educational and Training Professions, Rabat Sale Kenitra region, Morocco. Her research interests include geographic information systems (GIS), multi-criteria decision-making methods (MCDM), and information and communication technologies for education management (E-mail: oumelkheir.abra@taalim.ma).

Khalifa Mansouri is currently a researcher-professor in computer science, Training Director, and Director of the M2S2I Research Laboratory at ENSET of Mohammedia, Hassan II University of Casablanca. His research interests include information systems, e-learning systems, real-time systems, artificial intelligence, and industrial systems (modeling, optimization, numerical computation). Graduated from ENSET of Mohammedia in 1991, CEA in 1992, and PhD (Computation and Optimization of Structures) in 1994, HDR in 2010, and National PhD (in Computer Science) in 2016. He is the author of 10 books in computer science, a scientific book with the publisher Springer, 441 research papers, including 248 in the Scopus library, and supervised 36 defended doctoral theses (E-mail: khalifa.mansouri@enset-media.ac.ma).

Moulay El Houssine Ech-Chhibat is an Associate Professor and Doctor of Mechanical Engineering. He has the Accreditation to Direct Research (ADR) in the Laboratory of Modeling and Simulation of Intelligent Industrial Systems at the Higher Normal School of Technical Education, Hassan II University of Casablanca, Morocco. He is interested in maintenance, reliability, robotics, automatics, Industry 4.0, and artificial intelligence applications (E-mail: echchhibate@gmail.com).