Systematic Literature Review on Machine Learning Research in Education

1st Hamzah Arishi
School of Computer and Mathematical Sciences
The University of Adelaide
Adelaide, Australia
hamzah.arishi@adelaide.edu.au

2nd Nickolas Falkner
School of Computer and Mathematical Sciences
The University of Adelaide
Adelaide, Australia
nickolas.falkner@adelaide.edu.au

3rd Christoph Treude
School of Computing and Information Systems
Singapore Management University
Singapore
ctreude@smu.edu.sg

4th Thushari Attapattu
School of Computer and Mathematical Sciences
The University of Adelaide
Adelaide, Australia
thushari.atapattu@adelaide.edu.au

***Abstract*—This systematic literature review (SLR) critically examines the published literature since 2012 on the applications of machine learning (ML) in higher education. These applications include student performance classification, retention prediction, experience enhancement, and educational data mining. Machine Learning research has developed at a rapid pace to enhance different aspects of our lives. The educational sciences are at a pivotal point due to the potential influence that ML techniques have and will have on the field. We have aggregated the findings from a range of data sources, offering a thorough and objective perspective on the advancements, challenges, and future directions in this area. The volume of work that needed to be considered in the review clearly indicates the importance and relevance of this application of ML. In this SLR, we collect, analyse, and present the existing literature on ML applications in higher education. We utilise a comprehensive search strategy across multiple data sources, including 11 data sources, identifying all relevant studies published from 2012 to 2024. The inclusion criteria focused on peer-reviewed articles using ML-derived tools in areas such as student performance, retention, and experience. Exclusion criteria were rigorously applied to filter out studies that did not align with the focused domain of higher education and machine learning. The initial results yielded a pool of 16,581 articles, then refined to a substantive selection of 15,788 by removing duplicates, screening for false positives, and conducting quality assessment based on a structured quality control (QC) scoring system. The final set of papers was analysed to extract insights into ML applications' methodologies, effectiveness, and trends in higher education. The insights gained have allowed us to identify key trends, methodologies, and areas for future research. In addition, the surveying process will enable us to highlight publication patterns and critical research areas in the space.**

I. Introduction

Machine learning (ML) is an important sub-field of Artificial Intelligence. Algorithms are used to enable machines to carry out specific tasks, the most common of which is generating predictions, without human involvement. Modern civilisation benefits from the practical application of ML; in Healthcare, ML algorithms enhance disease diagnosis, personalized treatment plans, and medical imaging analysis [1]. ML in finance helps in risk assessment and fraud detection [2]. In Education, supports personalized learning, automated grading systems, and predictive analytics to enhance student outcomes [3]. For instance, electronic mailboxes are trained to identify patterns thanks to ML, saving users from having to go through many spam emails. This assumes that ML systems have been trained on appropriate data to reliably recognise the difference between useful messages and spam. ML facilitates the identification of patterns and rules in new data that resembles previously encountered data that ML has been trained to identify. The concept is that computers can be trained to identify features in the data, rather than being explicitly programmed for traditional tasks where features are static. This approach is akin to how children learn through practice to recognise different objects, such as distinguishing cars from trucks or cats from dogs [4], and computers are increasingly gaining this ability to learn from experience [5]. It is possible to automatically find patterns in existing data by using new data of the same type using ML [6]. For society to succeed, educational systems must continually evolve [7, 8]. Therefore, some areas of educational research focus on identifying and evaluating the critical factors that influence academic achievement [9]. ML is projected to be an important component of this initiative. It introduces a significant improvement in the methods available for data analysis, emphasising a data-driven approach to computational learning. Much interdisciplinary research includes ML to enhance technology capabilities. Education is a thriving field and has many potential applications for ML, especially with the pandemic encouraging many educational institutes to widen their use of technology in their education strategies. In the framework of data-driven education, this study explores how ML research will play a vital role in education research, offering researchers and educators insight into the theoretical, conceptual, and practical applications of ML in education and exploring the gaps and challenges.

The educational environment is complex; therefore, work significantly contributes to the digitalisation of educational sciences by illuminating how new ML techniques are applied successfully and dependably in research, instruction, and real-world applications. There are several applications of ML in education, some of which are self-evident. Others are overlooked, such as the validation of latent statistical models, intelligent tutoring systems (10), or prediction of student success (11). This was discussed for psychology by Yarkoni and Westfall (12), but the educational sciences have not yet fully embraced it. ML can assist educational researchers in changing the modelling culture towards a more reliable science with a stronger focus on the actual prediction of novel data and providing new analytical techniques. To develop educational science, it is essential to focus on robust models that make accurate predictions. Thus, we devote a section to the advantages of a novel model-building philosophy in educational research. Ensuring algorithmic fairness is crucial for ML applications in education, as the practical effects on people's lives can be significant. Algorithmic fairness can prevent problems with systematic discrimination, privacy, and sensitive data in ML (13). Due to the prevalence of latent phenomena in educational research, the role of ML in idea validation is essential, Kunter, Klusmann, et al., 2013 (14). In addition, in this overview we move beyond the specific topic of data mining in education, which has already generated several fascinating reviews (e.g., Ali, 2013 (15); Guleria & Sood, 2014 (16); Romero & Ventura, 2020 (17)). Fitting predictive models to data constitutes a thoroughly examined dimension of ML within educational research. The conventional framework of linear regression and the statistical inference that goes along with it are familiar to academics in the field of educational sciences. Both supervised ML and inference statistics are built on the same mathematical ideas and address the same issue of learning prediction models from data. Supervised ML is statistical modelling focused heavily on non-parametric models used to make predictions.

The adoption of these new methodologies and techniques is supported at two primary levels, namely, data gathering and data analysis (4). The growth of ML techniques positively influences educational research, fostering pivotal advancements within the domain of educational sciences (4). ML offers the potential to address various challenges present in modern academic research, including the issue of a significant proportion (4). Although the enumeration of extensively researched applications provides an understanding of the scope of ML within higher education, it is crucial to understand the underlying motivations and significance of these endeavours. Consequently, this review seeks to not only catalogue the ML applications but also shed light on why these advancements are imperative for higher education. By elucidating the research questions guiding this investigation; "What are the ML challenges and limitations in higher education research?" and "What is the significance and impact of machine learning applications in higher education?" we endeavour to provide a comprehensive framework for understanding the relevance and implications of our findings. This SLR collates and evaluates the body of ML research in higher education. Specifically, how the application of ML has helped improve the general field of higher education rather than research that focuses solely on specific courses or subject enhancement, following a structured and comprehensive approach based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (18) and (19). By adhering to a structured methodology informed by (18) and (19), this review seeks to achieve the following:

1) Identify and summarise the ML approaches and models in higher education.
2) Identify and discuss the challenges and limitations of ML in higher education.
3) Provide an overview of the research productivity of ML and education.
4) Provide insight into future research directions for ML in higher education.

The significance of this review lies in the comprehensive overview of the intersection of ML and higher education research. In the digital transformation era, educational entities must comprehend the capabilities and constraints of ML technologies. This review enriches the academic scholarly dialogue and provides actionable guidance for educators, decision makers, and researchers who seek to leverage ML to drive educational progress and enhancement. By systematically synthesising existing research, this study aims to provide a foundational framework for future investigations and implementations of ML in higher education.

The remainder of this paper is organised as follows: Starting with the research method and protocols followed in this SLR. Section 2 describes selection criteria, questions, exclusion inclusion criteria, data synthesis, and search query updates to include newly published articles. Section 3 presents the selected studies and results. Sections 4 answers to the research questions. Finally, we present our conclusion and recommendations for future work.

II. Research Methods

This Systematic Literature Review was performed based on Carrera-Rivera et al.'s (19) guidance for conducting a systematic literature review in computer science research. We also followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (20) for quality assessment of the systematic review process. Carrera-Rivera et al.'s (19) process includes planning and execution.

II-A Planning: The first step in the planning phase is to define the protocol. This involves describing the procedures involved in the review and detailing the activities to be performed. The second step is to define the PICOC (Population, Intervention, Comparison, Outcome, and Context) criteria and synonyms (19).. Population: Computer science education researchers and education researchers. Intervention: Application of ML in education. Comparison: Current literature, SLR, and what could be done better using ML in education. Outcome: Use of ML in Education, gaps, limitations, and what should be done differently. Context: Context in which the comparison takes place. In this case, our context is described in the Data

PICOC	Description
Population	Computer Science and Education Researchers
Intervention	Machine Learning Applications in Computer Science Education
Comparison	Machine Learning Applications in CS education
Outcome	The use of ML in Education, gaps, limitations, different areas to explore.
Context	Machine Learning Applications' Role in Enhancing the Student's Experience

TABLE I

PICOC (POPULATION, INTERVENTION, COMPARISON, OUTCOME, AND CONTEXT)

Synthesis section. The third step of the planning phase is to select the digital data sources. We selected 11 data sources for computer science and education research. Refer to the section Literature Search for data source details. We used the following search terms: "machine learning" OR "artificial intelligence" OR "deep learning" OR "educational data mining" OR "data science" AND "education" OR "university" OR "MOOC" OR "student performance" OR "student experience" OR "student interaction" OR "student retention" to find related articles.

II-B Research Question: The research questions focus on exploring ML in higher education, specifically examining its challenges and limitations. Reviewing the overall productivity of ML in higher education and the significance and impact of its applications on the field.

- **RQ1:** What are the ML challenges in Higher Education?
- **RQ2:** What are the ML limitations in Higher Education?
- **RQ3:** What is the overall productivity of ML research applications in Higher Education?
- **RQ4:** What is the significance and impact of ML applications in Higher Education?

II-C Literature Search: The initial search query is : ("machine learning" OR "artificial intelligence" OR "deep learning" OR "educational data mining" OR "data science" AND "education" OR "university" OR "MOOC" OR "student performance" OR "student experience" OR "student interaction" OR "student retention") yielded the following results:

1) IEEE Xplore: 2000 results
2) ACM DL: 1109 results
3) Scopus: 2000 results
4) ResearchGate: 1000 results
5) Springer Link: 1000 results
6) Web of Science: 2000 results
7) Directory of Open Access Journal: 1487 results
8) ArXiv: 1463 results
9) Science Direct: 2000 results
10) Wiley online library: 522 results
11) U.S. Department of Education: 2000 results

Total number of search results: 16,581 papers. After removing duplicated and not applicable papers: 15,788 papers.

III. STUDIES SELECTION CRITERIA

III-A Inclusion Criteria:

- Studies related to ML in higher education, including Students' performance classification, student retention prediction, student experience enhancements, and educational data mining.
- Studies using ML tools, artificial intelligence, or deep learning in the higher education domain.
- Research papers accepted and published in a peer-reviewed journal or conference.
- Papers that were published from 2012 to 2022.
- Papers written in English.

III-B Exclusion Criteria:

- Studies that do not use ML.
- Studies that are done for general education, not higher education.
- ML courses (for example, papers reviewing or proposing methods for teaching ML courses or including ML in course tools).
- Using ML to develop assessments or tools for specific courses.
- Papers that are not written in English.
- Posters, patents, already conducted reviews, Wikipedia articles, survey studies, and extended papers of already reviewed papers.
- Papers published before 2012.

III-C Study Selection: Screening and duplication: Titles and abstracts were initially screened to remove duplicates. Due to the number of selected papers and the screening required to perform a thorough systematic analysis, we followed the following steps:

- The results of each data source were uploaded to the AWS database.
- EC2 instance created.
- Each database was published five times on Amazon Mechanical Turk (Mturk) to get workers to review the title and abstracts of the papers; refer to the Amazon Mechanical Turk instruction in Section III-D.
- The mean of the five reviews was calculated to proceed with the selection.

III-D Amazon Mechanical Turk instruction: Reviewers for this study were recruited through Amazon Mechanical Turk (MTurk), a web-based platform for distributing tasks. to ensure the quality of the review process a series of steps were taken: Summary of the project, detailed instruction, two examples, short Video and daily QA check on the review process and quality, including checking the time stamp manually for each review for quality assurance purposes. The summary of the project provided the reviewers with an overview of the project and the objectives we are trying to achieve. The detailed instructions review process with examples and review submission instructions. Amazon Mechanical Turk reviewers were provided with two examples, on included papers and excluded papers. For clarity, we recorded the review process examples in a short video to be shared with the reviewers. We ran multiple daily checks to insure the review process was QA.

Quality Assessment Scoring Form:

- Are the review objectives clearly defined? Value: Not Defined: 0, Somewhat Defined: 1, Clearly Defined: 2
- Are the proposed methods well defined?
  Value: Not Defined: 0, Somewhat Defined: 1, Clearly Defined: 2
- Is the proposed accuracy measured and validated for quality assessment?
- Are the limitations of the review explicitly stated?
- What ML models are used in the study?
- What is the number of participant students?
- What type of learning environment was implemented in the study (i.e., MOOC, online, in-person or on campus)
- What data sets are used (survey, interview, social media, student interaction)?

During the review process, the results were first uploaded to the AWS database. Subsequently, an EC2 instance was created to facilitate access and management of the data. To further support data interaction, five web pages were developed for each database. The databases were each reviewed five times by reviewers from Amazon Mechanical Turk to ensure thoroughness and accuracy. Finally, complete the extraction of information form.

III-E Data Synthesis: We employed a mixed-methods protocol for data synthesis. The first step was to perform a Data Extraction Form from the selected included papers that scored more than 7 on the overall Quality Assessment Scoring Form. The Data Extraction Form helped us in identifying the following characteristics of each paper:

- Research type: theoretical research analyzes concepts and theories from literature reviews, while empirical research examines findings through scientific data or case studies.
- Research process: identifying the process type clarifies how data, models, or designs are organized.
- Framework type: identifying the employed technology provides insights into current trends, benefits, and limitations.
- Application field or domain: identifying the research domain and field.
- Gaps and challenges: reporting the current research gaps and challenges.
- Finding: research in computer science can yield different findings, for instance, an algorithm, a framework, or a new methodology.
- Evaluation method; identify the paper performance indicators.

After performing the Data Extraction Form, the last screening is a quality assurance step to ensure that the selected papers are qualified for our review.

TABLE II

SUMMARY OF UPDATED SYSTEMATIC LITERATURE REVIEW (SLR) 2023-2024

Database	New Results	Included	Excluded
IEEE Xplore	193	1	192
ACM DL	201	76	125
Scopus	254	40	214
ResearchGate	16	11	5
Springer Link	811	39	772
Web of Science	121	2	119
Directory of Open Access Journal	23	3	20
Arxive	3	1	2
Science Direct	14	6	8
Wiley Online Library	7	3	4
U.S. Department of Education	9	2	7
Total	**1652**	**184**	**1468**

III-F Search Query Update: The search was updated for the same data sources to include new results for the years 2023 and 2024, using the same search query and criteria. Table II summarises the results of the updated SLR.

IV. RESULT

After removing duplicate studies and downloading all search results, we developed a system to maintain a large number of results and systematically screen them according to the inclusion and exclusion criteria. First, we developed a database that contains the results of each data source search in separate tables. We used Amazon Web Services (AWS) cloud services. Second, an Amazon Elastic Compute Cloud EC2 was created to manage the result's database and obtain the data integrity. Third, a series of five web pages for each data source table was created. Fourth, we developed an instructional series of tasks on the Amazon Mechanical Turk crowd-sourcing marketplace; each task has details and instructions for screening [21], refer to Section III-D for Amazon Mechanical Turk instructions. To ensure the quality of the process, we performed quality assurance steps:

1) Review timer was set for 30 seconds minimum for each paper title and abstract.
2) Daily check to ensure the Amazon Mechanical Turk submission matches the updated database.
3) We used examples of false papers to filter out non-qualified responses (Red herring). If the reviewer responds negatively to the red hearing examples a pop-up window will appear on the screen telling them that they did not follow the review process instructions and the review session will end.
4) Conflict answering check, we asked the same question twice to ensure the reviewer's consistency.

Figure 1 shows the screening process of the study. Figure 2 shows the updated SLR PRISMA flow diagram. To

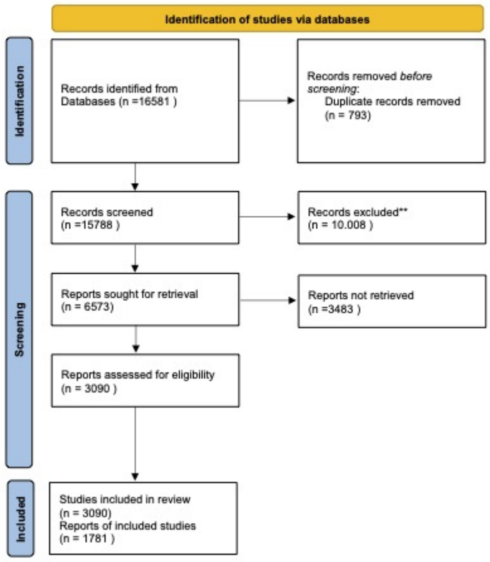

Fig. 1. Screening procedure of the studies

systematically report on the included studies we relayed on the Quality Assessment Scoring score for each database, we elected to include the papers that scored more than 70% on the Quality Assessment Scoring Form and performed Data Extraction using the Data Extraction Form in Section III-D. In the next step, we performed the Data Extraction Form using the template in the section Data Synthesis; this helped us collect more details about the included papers and helped us perform further screening for documents that did not meet the inclusion criteria. Figure 3 presents the results after performing the Data Extraction and filtering the papers. The figure indicates the data sources that yielded the most accurate results to our search criteria. ERIC, the Education Resources Information Center for the United States Department of Education, SCOPUS Elsevier's data source, and ACM the Association for Computing Machinery, produced the most accurate results as shown in Figure 3. On the contrary,

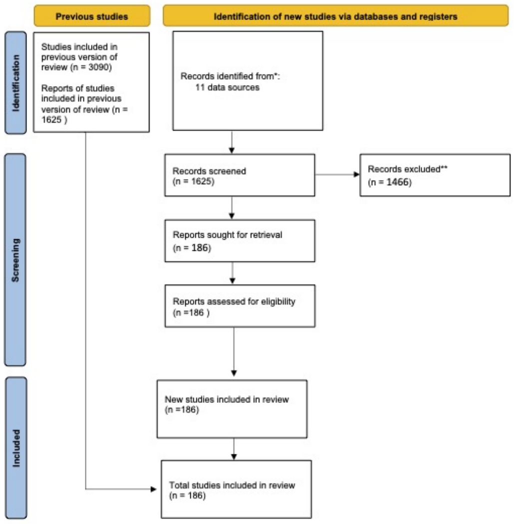

Fig. 2. Updated SLR Search

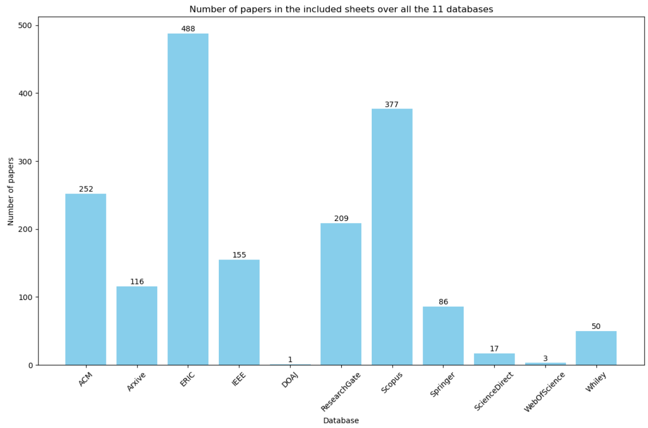

Fig. 3. Included papers count

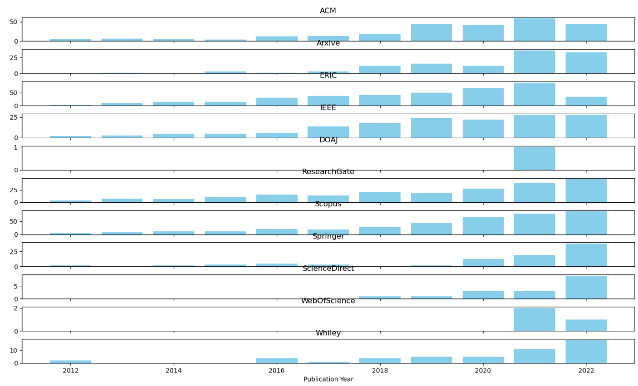

Fig. 4. Included papers by year

DOAJ the Directory of Open Access Journals, and The Web of Science data sources yielded the least accurate results. Figure 4 shows the number of publications in the data sources per year and the publication ratios for specific years. Figure 5 presents the overall categories or research methodology used in the included paper. The empirical research methodology has been used significantly in all the data sources' publications. A Thematic Analysis has been done to extract themes from the Data Extraction Form. Table III summarises the most common research gaps themes found in the data sources and provides examples from the most cited papers from each data source. Table III suggests that student engagement, Predictive Models, and Data Quality are the most common research gaps. The student's understanding of programs and their experiences represents a significant but underexplored research gap. This observation points to an emerging shift in research focus towards a more nuanced comprehension of student perceptions and experiences. The theme, lack of a comprehensive literature review and innovative learning methodologies, highlights significant research gaps by identifying innovative theoretical approaches to the current state of higher education research.

The research finding themes are summarised in Table IV. The theme of AI and ML Applications is the most frequent theme in research findings; it highlights the research community's dedication to enhancing higher education by introducing ML applications. This theme shows the significance of the ML role in higher education in the past decade. The ML predictive model is among the most common research findings, focusing on enhancing student academic performance in higher education. The theme, Interactive and engagement, focuses on improving student engagement and experience thoroughly through ML application; it shows the novel direction of enhancing learning interaction and experience. The methodology advancement theme explores new methodologies to enhance learning, testing the effectiveness of simulated methodologies and gaining insights from ML methods. The Data Science theme focuses on the data science programs and the enhancement of these programs. The research framework type themes are summarised in Table V. The Education Program Development themes show the most common framework in the extracted data, emphasising the significant amount of research done to improve educational programs with the help of ML. The application of ML in higher education and online learning provides another new and extensive research framework for this field. The Data Science theme aims to enhance educational outcomes and offer data-driven insights for improved efficiency. The theme of engagement and interaction emphasises the importance of student engagement and interaction in successful learning experiences, showing a strong focus on these factors. The Process Type followed in the extracted papers is summarised in VI. The ML theme, which focuses on ML models, refers to a growing preference for computational methods in data analysis and predictive modelling. This theme prioritises the creation of models and frameworks, which

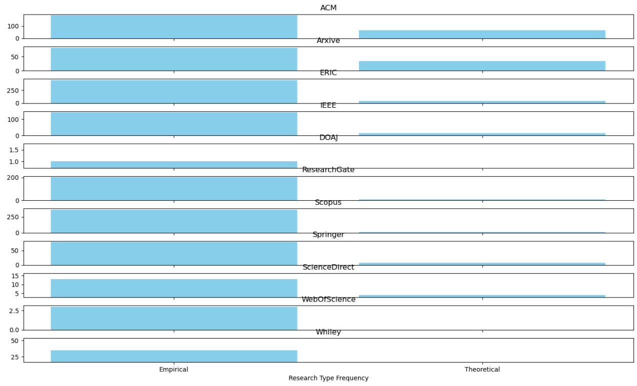

Fig. 5. Included papers by research type

are foundational for applying ML in educational settings. The Data Analysis theme greatly emphasises data analysis, which is considered a fundamental step. This involves using both qualitative and quantitative methods. The Educational Enhancement Insights theme focuses on finding elements of enhancements in the student experience, educational cutting, and curriculum enhancements. This theme shows a great dedication to enhancing the educational setting and the student experience. The Evaluation Method themes are summarised in Table VII. Regarding the ML Effectiveness theme, researchers evaluate the performance of ML models using metrics such as the F1 score, accuracy, and model evaluations. The data and Statistical Analysis theme emphasises the importance of data analysis and statistical evaluation methods in uncovering trends, patterns, and predictive capabilities within datasets. The theme of qualitative and conceptual insights focuses on qualitative analyses and conceptual frameworks, demonstrating a commitment to understanding the underlying principles and theories driving ML.

V. Answering the Research Questions

V-A The Machine Learning Challenges in Higher Education: Integrating into higher education helps improve the current state of higher education into something more promising, data-driven, and automated. However, these enhancements come with challenges; understanding these challenges and current research gaps is essential for researchers who aim to implement effective ML into higher education. Refer to Table III. These gaps are crucial for researchers to find enhancement elements within current educational settings. Data Quality: Data quality and availability are significant research gaps in higher education. In (23), Lisa Wang et al. mentioned that the rich structured data available from programming submissions needs to be exploited as complex, open-ended educational data. Utilising such data has the potential to improve data quality and enhance different aspects of educational effectiveness. In (27), Olson et al., focus on data prepossessing to improve the quality of the data science pipeline design process with ML. Moreover, concerns about data privacy and ethical use further complicate data accessibility. Engagement and Personalization: ML tools can help improve student engagement and personalise feedback. (37) explored the potential of the ML-inspired model to outperform human learners in standard assessments. (42) explored how personalised learning paths can be integrated into online laboratory practices and teaching. Personalised learning has the potential to address diverse student backgrounds, learning speeds, and styles, which are critical in remote learning environments where students may feel disconnected. They also discussed concerns about maintaining student engagement during COVID-19. Predictive Models: This research gap concerns using models to predict and find enhancement elements for student academic performance. The main concern in this gap is identifying the features that can lead to accurate prediction. (50) addressed this challenge, explicitly predicting dropout using a robust model. (56) discussed the prediction of at-risk students by utilising demographics, clickstream data, and assessment scores to predict at-risk students. There are a few other examples such as; (59),(173).

V-B Infrastructure and Analytical Challenges: This gap encompasses the identification of challenges in integrating ML into the current educational infrastructure and addresses the challenges that face the implementation of ML. One of the substantial hurdles in applying ML within higher education is the implementation of efficient, scalable infrastructures that can handle large volumes of data without compromising performance. Papers (22), (47) addressed the computational complexity challenges that face ML in educational settings and the need for more robust data handling capabilities. Understanding and Evaluation: A crucial challenge in ML applications within higher education is user understanding of how these applications function and the purpose of such programs. (53) and (54) highlight significant difficulties in users' comprehension of algorithms and their practical implementations. Another crucial aspect of this gap is the practical evaluation of these applications from the user perspective and its effect on the overall experience. (42) and (55) address the challenge of the complexities involved in assessing how students interact with and benefit from ML in educational environments. Literature and Methodological Gaps: Although there is intensive research on ML in higher education, there is still a notable gap in comprehensive literature reviews synthesising findings across diverse studies. This gap limits the researcher and practitioner's abilities to fully understand the landscape of ML applications and their impacts on higher education. (67), and (68), addressed the fragmented nature of existing reviews.

V-C The Machine Learning Limitations in Higher Education: While ML poses many potential enhancements in higher education settings, a few limitations restrict the effectiveness of the ML application and add challenges to its implementation. Technical limitations include model complexity and integrability, data dependency, and generalisation across different settings. Examples of these limitations can

TABLE III
AGGREGATED RESEARCH GAPS AND CHALLENGES IN EDUCATIONAL MACHINE LEARNING

Theme	Frequency	Key Insights and References [1]
Data Quality and Availability	271	Challenges related to the availability, consistency, and quality of data across multiple sources such as ACM (Fadhil & Gabrielli, 2017), (Wang, Sy, Liu, & Piech, 2017), (Fiebrink, 2019), (Tang et al., 2016), (Piety et al., 2014) [22] [23] [24] [25] [26]. arXiv (R. S. Olson et al.,2016) [27], (C. Brunsdon and A. Comber, 2021) [28], (P. L. Combettes and J.-C. Pesquet,2021) [29], (L. Cao et al., 2021) [30], (I. Padhi et al., 2021) [31]. And ERIC (M. J. Israel, 2015) [32], (J. Cabero-Almenara, et al., 2020) [33], (W. R. Stevenson III et al.,2015)[34], (E. Kennedy, et al.,2019)[35], (A. A. M. Alhaj et al.,2021)[36].
Engagement and Personalization	379	Focus on enhancing student engagement and personalised feedback with references from ACM (Finnie-Ansley et al., 2022)[37], (J. Pereira, 2016) [38], (M. S. Boroujeni et al., 2018)[39], (B. Shapiro et al., 2020) [40], (A. S. Imran, et al., 2019) [41] ERIC (K. A. A. Gamage et al., 2020) [42], (A. C. Graesser,,2016)[43], (A. M. Müller et al.,2016)[44], (C. Milligan et al., 2014) [45], (C. Yeager, et al., 2013) [46].
Infrastructure and Analytical Challenges	121	Issues with designing efficient, scalable infrastructures and addressing computational complexity. Seen in data from ACM (A. Fadhil et al., 2017) [22], (S. N. Liao et al., 2016) [47], (T. Robal et al., 2018) [48], (A. S. Imran et al., 201) [41], (S. S. Feger et al., 2020)[49]. IEEE, (Q. C. Zhang et al., 2019) [50], (S. D. A. Bujang et al., 2021) [51], and (V. L. Uskov et al. 2019) [52]..
Understanding and Evaluation	112	Including the challenge in users' comprehension of programs from arXiv, (B. Baumer, 2015) [53], (D. K. I. Weidele et al., 2020) [54] and evaluation of learners' experiences from ERIC, (Gamage, Kelum A et al., 2020) [42], (M "uller, Andre Matthias et al., 2021)[44], and (N. Vasilevski et al., 2022) [55].
Predictive Modeling	284	Challenges in identifying relevant features for accurate prediction, predicting performance, and academic outcomes. Highlighted in IEEE (Zhang Q C et al.,2019) [50], (M Adnan et al., 2021) [56], and (H. E. Abdelkader et al., 2022) [57]. Science Direct, (Adekitan, A. I. et al., 2019) [58], (P. M. Moreno-Marcos et al., 2019) [59], (X. Lu et al., 2020) [60]. Scopus, (H. Waheed et al.,2020) [61], (J. L. Rastrollo-Guerrero et al.,2020) [62], and (V. L. Miguéis et al., 2018) [63]. and ERIC (C. Stone et al., 2019)[64], (H. H. Yang et al., 2017)[65] and (S. Bulfin et al., 2014) [66].
Literature and Methodological Gaps	91	Lack of a comprehensive literature review and innovative learning methodologies, as mentioned in Research Gate (Y. Z. Zhang , 2022) [67], (J. Nieder et al., 2022)[68] , (L. Barik et al.,2020), [69]. and Scopus, (A. Onan, 2020) [70], (J. Rastrollo-Guerrero et al., 2020)[62], (A. Moubayed et al., 2019)[71].

TABLE IV
AGGREGATED THEMES RESEARCH FINDINGS

Theme	Frequency	Key Insights and References [2]
AI and ML Applications	517	Extensive application and exploration of AI and ML in various contexts including models accuracy prediction, pattern identification, and data-driven insights, as seen in ACM, (Wang et al., 2017)[23], (R. Fiebrink, 2019) [72], (S. Stapleton et al., 2020) [73]. ArXiv, (A. Karpatne et al., 2017) [74] note: this is the most cited paper in this SLR (739 citations), (S. Passi et al., 2018) [75], (M. Hulsebos et al., 2019) [76]. ERIC, (K. Jordan, 2014) [77](note highly cited paper 627), (Arthur C. Graesser, 2016)[43], (H. H. Yang et al., 2017) [65]. IEEE, (J. M. Kanter et al., 2015) [78], (A. Pardo et al., 2017) [79], (Q. Zhang et al., 2019)[50]. And wiley, (E. Yuzbasioglu, 2021). [80], (Z. H. Aung et al. 2022) [81], (J. Lasser etal., 2021) [82].
Data Science and Analytics	113	Focus on the development of data science systems, integration of technology, and data analytics techniques, highlighted by ArXiv, (S. Randal et al., 2016) [27], (M. Hulsebos et al., 2019) [76], (D. K. I. Weidele et al., 2020) [54]. IEEE, (J. Kanter et al., 2015) [78], (A. Pardo et al., 2017) [79], (T. Y. Yang et al., 2017) [83]. and springer, (J. G. et al., 2015) [84], (V. Aleven et al., 2016) [85], (L. Yuan, 2017) [86].
Predictive Performance and Personalization	304	Efforts towards predicting academic performance, enhancing e-learning, and proficiency measurement through personalised interventions, especially noted in ERIC, (Gamage et al., 2020) [42], (J. Mackness et al., 2013) [87], (H. Luo et al., 2013) [88]. Scopus, (S. Alturki et al., 2021) [89], (L. C. Jia et al., 2022) [90], (J. Beemer et al., 2018) [91]. and SD (S. Liu et al., 2022) [92], (X. Lu et al.,2020) [60], (M. Nilashi et al., 2022) [93] .
Educational Technology Integration	87	Challenges and insights into integrating AI in education, using EDM and ML for educational improvements, as reported by RG, (Z. S. Wilson et al., 2012) [94], (S. Lee et al., 2021) [95], (R. L. Mahler et al., 2017) [96]. Scopus, (A. Onan, 2020) [70], (A. Moubayed et al., 2019)[71], (C. F. Rodríguez-Hernández et al., 2021)[97]. arxiv, (S Biswas et al., 2022) [98], (H. Wu et al., 2018) [99], (A. Kazakçi , 2015) [100] and WOS (Y. D. Tang et al., 2015) [101].
Interactivity and Engagement	219	Investigations into patterns of interaction, the impact of engagements, and enhancing learner's experience through interpretable ML techniques, as found in ERIC; (K. Jordan, 2014) [77],(Gamage et al., 2020) [42], (A. C. Graesser,,2016)[43]. and IEEE; (J. M. Kanter et al., 2015) [78], (A. Pardo et al., 2017) (**?**), (I. Khan et al., 2019) [102] .
Methodological Advancements	152	New methodologies in examining learning transitions, effectiveness of simulated platforms, and insights from ML techniques, as explored in ERIC, (A. C. Graesser,,2016)[43], (H. H. Yang et al., 2017)[65], (N. Vasilevski et al., 2022) [55] IEEE, (J. M. Kanter et al., 2015) [78], (A. Pardo et al., 2017) [79], (Q. C. Zhang et al., 2019) [50]. and SD, (R. K. Veluri et al., 2022) [103], (E.Tebenkov et al., 2021)[104], (Y. Hu et al., 2021) [105].
Support and Community Building	71	The importance of a supportive community for success and effectiveness of tailored approaches, only discussed in RG ; (M. Hussain et al. et al., 2018) [106], (H. A. Mengash et al., 2020) [107], (M. L. Maher et al., 2015) [108].
MOOCs and Online Learning	16	Integration of learning modes through MOOCs, and challenges in MOOCs highlighted by Springer; (J. G. et al., 2015) [84], (S. Hyunjin ChaHyo-Jeong, 2020) [109], (S. Hyunjin ChaHyo-Jeong, 2020) [109]. SD ; (S. Liu et al., 2022) [92], (X. Lu et al., 2020) [60], (M. Nilashi et al., 2022) [93]. and WOS;(M. Abdullah et al., 2021)[110].

be seen in the following papers: (A. Fadhil et al., 2017) (22), (T. Robal et al., 2018). These technical limitations are related to the inheritance characteristics and challenges of models. ML complexity can make it difficult for educational institutions to adopt and effectively use these technologies without substantial investments in expertise and technology. Addressing these technical challenges is important in building effective educational ML models. Practical limitations include integration with existing educational systems and sustainability. An instance of this limitation is the student's performance, where students performance application is not transferable from one university to another without taking into account the environmental setting differences. The practical limitations are logistical and operational challenges in implementing ML within existing educational frameworks. Instances of these challenges are: (Z. S. Wilson et al., 2012) (94), (S. Lee et al., 2021) (95). Ethical limitations include algorithm bias and fairness, as well as privacy concerns. ML algorithms can perpetuate or exacerbate existing biases if they are trained on biased data or not designed to account for equity. This can lead to unfair and biased education ML applications (13).

V-D The productivity of ML research applications in higher education: Productivity in this context encompasses the effectiveness and efficiency of ML application research in higher education. The considerable volume of research, extracted from various prominent journals and conferences, showcases the scholarly commitment to enhancing educational environments through ML. This is further evidenced by the increasing number of publications from 2012 to 2024, reflecting a growing interest and a shift from theoretical models to practical applications. Additionally, the implementation of ML research findings in educational settings offers insights into its productivity, directly affecting student performance, retention rates, and personalized learning experiences. Around 2016, a notable shift in integrating deep learning techniques marked a significant milestone due to their capabilities in handling complex data (27), (22). Recent years have focused on ethical AI and fairness in ML applications, reflecting a broader understanding of the social implications of technology in educational settings (95).

V-E The significance and impact of applications in higher education: ML transformed higher education by enhancing educational settings, personalizing learning experiences, and improving student engagement and academic performance. ML integration has not only increased the accuracy of predictive models over traditional methods and human analysis but has also streamlined administrative tasks such as enrollment management, resource allocation, and student services. Additionally, ML's capacity to analyze vast data sets has facilitated the development of more effective models in higher education.

TABLE V
AGGREGATED RESEARCH FRAMEWORK TYPE THEMES ACROSS DATA SOURCES

Aggregated Theme	Frequency	Key Insights and References [3]
ML Applications	192	Widespread focus on the application of across various educational areas, with a particular emphasis on predictive modelling and artificial intelligence. Examples of this are in ACM; (Fadhil & Gabrielli, 2017) [22], (S. N. Liao et al., 2016) [47], (L. Aroyo et al., 2019)[111]. ArXiv; (M. Hulsebos et al., 2019) [76], (C. Brunsdon and A. Comber, 2021) [28], (P. L. Combettes and J.-C. Pesquet,2021) [29]. IEEE; (J. M. Kanter et al., 2015) [78], (I. Stančin and A. Jović, 2019) [112], (A. Kaur et al., 2018) [113]. Scopus; (A. Daud et al., 2017) [114], (C. C. Gray and D. Perkins, 2019) [115], (E. Wakelam et al., 2020) [116]. and Wiley; (Z. J. Zheng et al., 2022) [117], (Y. Mourdi et al., 2022)[118],(D. Mariano-Hernández et al., 2022) [119].
Education Program Development	372	Significant emphasis on developing and improving education programs, suggesting a trend towards innovation and reform in educational methodologies. Examples of this are in ACM; (Wang, Sy, Liu, & Piech, 2017) [23], (Z. Huang et al., 2019) [120], (S. Kross et al., 2019) [121]. arXiv; (R. S. Olson et al.,2016) [27], (A. Sirbu and O. Babaoglu, 2015) [122], (M.Tavakoli, et al., 2020) [123]. ERIC; (K. A. A. Gamage et al., 2020) [42], (A. C. Graesser,,2016)[43], (M. J. Israel, 2015) [32]. IEEE; (T. Y. Yang et al., 2017) [83], (M Adnan et al., 2021) [56], (I. Khan et al., 2019) [102] . Research Gate; (D. Maresch et al., 2016) [124], (A. Y. Alsobhi and K. H. Alyoubi, 2019) [125], (L. Barik et al.,2020), [69]. Sience Direct; (X. Lu et al.,2020) [60], (O. A. Olabanjo et al., 2022) [126], (Adekitan, A. I. et al., 2019) [58]. Springer; (I. Roll and R. Wylie, 2016) [127], (A. Tlili et al., 2022) [128],(V. Aleven et al., 2016) [85]. Wiely; (K. M. Stegers-Jager et al., 2012) [129], A. Cohen et al., 2019) [130], (Z. H. Zhan et al.,2022) [131].
Data Science and Analysis	176	Heavy investment in data science, analytics, and analysis techniques to improve educational outcomes and to provide data-driven insights for efficiency. Examples of this are in ACM; (C. Obaid et al., 2018) [132],(S. Kross et al., 2019) [121], (R. Fiebrink, 2019) [72]. ArXiv; (S. Randal et al., 2016) [27], (C. Brunsdon and A. Comber, 2021) [28], (P. L. Combettes and J.-C. Pesquet,2021) [29]. ERIC; (H. H. Yang et al., 2017)[65], (E. Ossiannilsson et al.,2016) [133], (D. Lee et al., 2020) [134]. IEEE; (T. Y. Yang et al., 2017) [83], (I. Stančin and A. Jović, 2019) [112], (I. Khan et al., 2019) [102] . Research Gate; (V. Gramoli et al., 2016) [135], (M. K. Kadhim and A. K. Hassanm, 2020) [136], (L. Naylor and D. Veron, 2020) [137]. Scopus; (S. A. Wartman and C. D. Combs, 2018) [138], (M. Yağcı, 2022) [139], (E. A. Amrieh et l., 2015) [140]. and Springer; (S. B. Dias et al., 2020) [141], (T. Susnjak et al., 2022) [142], (J. Samuelsen et al., 2019) [143].
Online Learning and MOOCs	193	The rise of online learning and MOOCs as a major theme reflects the ongoing digital transformation in education, focusing on student interaction and engagement. Examples of these themes are seen in: ERIC; (H. H. Yang et al., 2017) [65], (J. Loizzo et al., 2017) [144], (S. Osuna-Acedo et al., 2018) [145]. IEEE; (Y. Tan et al., 2018) [146], (W. Tenpipat et al., 2020) [147], (M. Revathy et al., 2022) [148]. science direct; (P. M. Moreno-Marcos et al., 2019) [59], (S. Liu et al., 2022) [92], (X. Lu et al., 2020) [149]. and Springer; (R. Deng and P. Benckendorff, 2022) [150],(S. Hyunjin ChaHyo-Jeong, 2020) [109], (L. Yuan, 2017) [86].
Engagement and Interaction	110	Research indicates a concentrated effort on learning engagement and student interaction, highlighting the importance of these factors in successful learning experiences. Examples of these themes are seen in: ACM; (Z. Huang et al., 2019) [120], (S. P. Shashikumar et al., 2018) [151],(S. Altaf et al., 2019) [152]. ERIC; (M. J. Israel, 2015) [32], (S. A. Iqba et al., 2022) [153], (A. Gordillo et al., 2019) [154]. Research Gate; (H. A. Mengash et al., 2020) [107], (G. Townley et al., 2013) [155], (Z. S. Wilson et al., 2012) [94].
Theoretical and Conceptual Frameworks	28	A foundational concern for solid theoretical and conceptual frameworks to underpin the use of new technologies in educational settings. Examples of these themes are seen in: Scopus; (M. Mujahid et al., 2021) [156], (M. Ciolacu et al., 2018) [157], (I. Mwalumbwe and J. S. Mtebe, 2017) [158], And Wiley; (Z. J. Zheng et al., 2022) [117], (Y. Mourdi et al., 2022)[118], (D. Mariano-Hernández et al., 2022) [119].

TABLE VI
AGGREGATED PROCESS TYPE THEMES ACROSS DATA SOURCES

Aggregated Theme	Frequency	Key Insights and References [4]
and Algorithms	485	Dominant focus on models and algorithms, suggesting a trend towards computational approaches in data analysis and predictive modelling. These themes include the development of central models and conceptual frameworks, as they support the application of and the interpretation of data in educational contexts. Examples can be seen in: ACM; (Finnie-Ansley et, al., 2022)[37], (Z. Huang et al., 2019) [120], (J. Pereira, 2016) [38]. ArXiv; (S. Passi et al., 2018) [75], (M. Hulsebos et al., 2019) [76], (L. Cao et al., 2021) [30]. ERIC; (A. M. Müller et al.,2016)[44], (J. Mackness et al., 2013) [87], (J. Cabero-Almenara, et al., 2020) [33]. IEEE; (J. M. Kanter et al., 2015) [78], (Q. C. Zhang et al., 2019) [50], (M Adnan et al., 2021) [56]. Science Direct; (S. Liu et al., 2022) [92], (M. Nilashi et al., 2022) [93], (Y. Hu et al., 2021) [105]. Springer; (S. C. Tsai et al., 2020) [159], (S. B. Dias et al., 2020) [141], (T. Susnjak et al., 2022) [142]. Wiley; (E. Yuzbasioglu et al., 2021) [80], (M. E. Sousa-Vieira et al., 2022) [160], (Z. J. Zheng et al., 2022) [117].
Data Analysis	362	High emphasise data analysis as a foundational process, including qualitative and quantitative methods across various sources. Examples of these themes are in ACM; (C. Obaid et al., 2018) [132], (A. E. Waters et al., 2015) [161], (G. Lopez et al., 2017) [162]. ArXiv;(A. Karpatne et al., 2017) [74] note: this is the most cited paper in this SLR (739 citations), (B. Baumer, 2015) [53],(C. Brunsdon and A. Comber, 2021) [28]. ERIC; (K. A. A. Gamage et al., 2020) [42], (J. Mackness et al., 2013) [87], (C. Stone et al., 2019)[64]. IEEE; (A. Tarhini et al., 2013) [163], (J. Rohde et al., 2019) [164], (D. Liu et al., 2020) [165]. Research Gate; (G. Townley et al., 2013) [155], (A. T. Angelo et al., 2021) [166],(L. Ding et al., 2016) [167]. Science Direct; (Adekitan, A. I. et al., 2019) [58], (P. M. Moreno-Marcos et al., 2019) [59], (X. Lu et al., 2020) [149]. Scopus;(H. Waheed et al.,2020) [61], (A. Onan, 2020) [70], (J. L. Rastrollo-Guerrero et al.,2020) [62]. Springer; (A. Tlili et al., 2022) [128], (S. C. Tsai et al., 2020) [159], (J. Samuelsen et al., 2019) [143]. and Wiley; (K. M. Stegers-Jager et al., 2012) [129], (E. Yuzbasioglu et al., 2021) [80], (Z. H. Zhan et al.,2022) [131].
Research Methodology	41	Considerable attention to research methodologies, such as literature reviews and theoretical analysis, reflecting the robustness of the scientific inquiry. These themes are highlighted in ACM; (J. C. Adams, 2020), [168], (R. Saini et al., 2019) [169], (P. Robe and S. K. Kuttal, 2022) [170]. ArXiv;(H. Wu et al., 2018) [99], (Y. Sun et al., 2021) [171]. IEEE; (A. Kumar et al., 2021) [172], (G. Al-Tameemi et al., 2020) [173], (Y. Wang et al., 2022) [174]. Research Gate; (J. M. Górriz et al.,2020) [175], (C. P. Rosé et al.,2019) [176]. and Science Direct; (R. K. Veluri et al., 2022) [103], (B. K. Yang et al., Yang2022Untangling) [177].
Educational Enhancement Insights	160	A focus on evaluating and enhancing student experience, educational analysis, and curriculum evaluation highlights the commitment to improving learning outcomes. Examples of these themes only seen in ERIC; (K. Jordan, 2014) [77](note highly cited paper 627), (Arthur C. Graesser, 2016)[43], (M. J. Israel, 2015) [32]. and Research Gate; (S. Hwang et al., 2019) [178], (C. A. Berry et al., 2020) [179], (E. Wong et al.,2019)[180].

TABLE VII
AGGREGATED EVALUATION METHOD THEMES ACROSS DATA SOURCES

Theme	Frequency	Key Insights and References [5]
Effectiveness	181	Evaluation of algorithms and models through performance metrics like the F1 score, accuracy, and model performance evaluations. Examples of these themes are in: ACM; (J. Pereira, 2016) [38], (Wang, Sy, Liu, & Piech, 2017) [23] , (S. Altaf et al., 2019) [152]. ERIC; (S. Altaf et al., 2019) [152], (M. Britto and S. Rush, 2013) [181], (E. Ossiannilsson et al.,2016) [133]. IEEE;(J. M. Kanter et al., 2015) [78], (T. Y. Yang et al., 2017) [83], (M Adnan et al., 2021) [56]. Scopus; (A. Salah Hashim et al., 2020) [182], (M. Arashpour et al., 2022)[183], (K. S. Rawat and I. V. Malhan, 2019) [184]. Science Direct; (P. M. Moreno-Marcos et al., 2019) [59], (S. Liu et al., 2022) [92], (R. K. Veluri et al., 2022) [103]. Wiley; (K. M. Stegers-Jager et al., 2012) [129], (T. A. Pham et al., 2021) [185].
Data and Statistical Analysis	416	A strong focus on data analysis and statistical methods to understand trends, patterns, and predictive capabilities within datasets. Comparative and Empirical Studies, Including experimental evaluation, surveys, case studies, observational, and content analysis, reflect the diversity of research designs. Examples of this are seen in: ACM; (Z. Huang et al., 2019) [120], (H. Li et al., 2020) [186], (M. Plagge, 2019) [187]. ArXiv; (P. L. Combettes and J.-C. Pesquet,2021) [29], (D. K. I. Weidele et al., 2020) [54], (S Biswas et al., 2022) [98]. ERIC; (Gamage, Kelum A et al., 2020) [42], (Arthur C. Graesser, 2016)[43], A. M. Müller et al.,2016)[44]. IEEE; (T. Y. Yang et al., 2017) [83], (M Adnan et al., 2021) [56], (I. Khan et al., 2019) [102] . Research Gate; (Z. S. Wilson et al., 2012) [94], (M. L. Maher et al., 2015) [108], (J. Wiley et al., 2017) [188]. Scopus; (D. K. Arun et al., 2021) [189], (C. A. Palacios et al., 2021) [190], (E. Wakelam et al., 2020) [116]. Springer; (V. Aleven et al., 2016) [85], (T. Susnjak et al., 2022) [142],(Y. Z. Fan et al., 2021) [191]. Wiley; (E. Yuzbasioglu et al., 2021) [80], (Z. J. Zheng et al., 2022) [117], (M. Kubsch et al., 2022) [192].
Qualitative and Conceptual Insights	80	Qualitative analyses and conceptual frameworks suggest a dedication to understanding the underlying principles and theories driving the field. Examples of these themes are seen in: ArXiv; (C. Brunsdon and A. Comber, 2021) [28], (S. Lee et al., 2020) [193], (S Biswas et al., 2022) [98]. Research Gate; (G. Townley et al., 2013) [155], (J. M. Dubinsky et al., 2019) [194], (A. T. Angelo et al., 2021) [166]. Scopus; (M. Ciolacu et al., 2018) [157], (I. Khan et al., 2021) [195], (P. Verma et al., 2017)[196]. Springer; (V. Aleven et al., 2016) [85], (J. Johnson et al., 2012) [197].

VI. CONCLUSION

In interdisciplinary research, ML is significant in improving higher education. As ML technologies continue to evolve, their potential to enhance higher education remains substantial. Although ML offers promising possibilities for higher education, its effective implementation is contingent upon overcoming significant data-related, technical, and cultural challenges. Enhancing user understanding and accurate evaluation of learner experiences are particularly critical areas that require further research and innovative solutions to ensure that ML technologies can be effectively harnessed to enhance educational outcomes and operational efficiency. Future efforts should focus on enhancing the user experience, refining these applications, improving models accessibility and fairness, assessing the impact of these applications on students and institutions, and making sure they meet the needs of both sides of these applications.

[5]Examples are from the most cited papers from each theme among the top-scoring papers in the QA system.

REFERENCES

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235–249, 2002.

[3] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," in *Advances in Neural Information Processing Systems*, 2015.

[4] S. Hilbert, S. Coors, E. Kraus, B. Bischl, A. Lindl, M. Frei, J. Wild, S. Krauss, D. Goretzko, and C. Stachl, "Machine learning for the educational sciences," *Review of Education*, vol. 9, no. 3, p. e3310, 2021.

[5] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, vol. 44, no. 1-2, pp. 207–219, 2000.

[6] Pedro Domingos, "A Few Useful Things to Know About Machine Learning," *communications of the ACM*, vol. 55, no. 10, pp. 79–88, 2012.

[7] D. W. Cheek, J. Santos, and N. Vaillant, "Education and Economic Growth," *International Encyclopedia of the Social Behavioral Sciences: Second Edition*, vol. 2, pp. 121–126, 2010.

[8] E. A. Hanushek and L. Woessmann, "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation," *Journal of Economic Growth*, vol. 17, no. 4, pp. 267–321, 2012.

[9] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, "Predicting academic performance: A systematic literature review," *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pp. 175–199, 2018.

[10] C. Conati, K. Porayska-Pomsta, and M. Mavrikis, "AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling," no. Whi, 2018.

[11] I. Živčić-Bećirević, S. Smojver-Ažić, and T. Martinac Dorčić, "Prediktori akademskog uspjeha studenata: Prospektivno istraživanje," *Drustvena Istrazivanja*, vol. 26, no. 4, pp. 457–476, 2017.

[12] T. Yarkoni and J. Westfall, "Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning," *Perspectives on Psychological Science*, vol. 12, no. 6, pp. 1100–1122, 2017.

[13] A. Chouldechova and A. Roth, "The Frontiers of Fairness in Machine Learning," pp. 1–13, 2018.

[14] M. Kunter, U. Klusmann, J. Baumert, D. Richter, T. Voss, and A. Hachfeld, "Professional competence of teachers: effects on instructional quality and student development.," *Journal of educational psychology*, vol. 105, no. 3, p. 805, 2013.

[15] Mohd Maqsood Ali, "Use of data mining in education sector," *Lecture Notes in Engineering and Computer Science*, vol. 1, no. April, pp. 513–516, 2012.

[16] P. Guleria and M. Sood, "Data mining in education: A review on the knowledge discovery perspective," *International Journal of Data Mining & Knowledge Management Process*, vol. 4, no. 5, p. 47, 2014.

[17] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1355, 2020.

[18] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, and D. Moher, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.

[19] A. Carrera-Rivera, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," nov 2022.

[20] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *International journal of surgery*, vol. 88, p. 105906, 2021.

[21] A. M. Turk, "Amazon mechanical turk," *Retrieved August*, vol. 17, p. 2012, 2012.

[22] A. Fadhil and S. Gabrielli, "Addressing challenges in promoting healthy lifestyles," *Healthy Lifestyles Promotion*, 2017.

[23] L. Wang, A. Sy, L. Liu, and C. Piech, "Deep knowledge tracing on programming exercises," in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, pp. 201–204, ACM, 2017.

[24] R. Fiebrink, "Machine learning education for artists, musicians, and other creative practitioners," *Artn*, vol. 31, 2019.

[25] S. Tang, J. C. Peterson, and Z. A. Pardos, "Deep neural networks and how they apply to sequential education data," in *L@S 2016 - Work in Progress*, 2016.

[26] P. J. Piety, D. T. Hickey, and M. Bishop, "Educational data sciences - framing emergent practices for analytics of learning organizations and systems," in *LAK '14*, ACM, 2014.

[27] R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore, "Evaluation of a tree-based pipeline optimization tool for automating data science," in *GECCO '16: Proceedings of the 2016 on Genetic and Evolutionary Computation Conference*, (Denver, CO, USA), July 2016.

[28] C. Brunsdon and A. Comber, "Opening practice: supporting reproducibility and critical spatial data science," *Journal of Geographical Systems*, vol. 23, pp. 477–496,