https://doi.org/10.53656/math2026-5-6-mlm

MACHINE LEARNING IN MATHEMATICS AND COMPUTER SCIENCE EDUCATION: A SYSTEMATIC THEMATIC REVIEW OF DISCIPLINARY LEARNING SUPPORT

Eyal Sadeh
South-West University "Neofit Rilski" (Bulgaria)

Abstract. Machine learning (ML) is being increasingly used in mathematics and computer science education, but the educational value of machine learning is not theorized or assessed evenly. The current studies focus on predictive accuracy, automation, and personalization, which raises the question of whether such applications are relevant to supporting higher-order disciplinary learning. This gap is filled by the current study, which is a qualitative systematic review based on the PRISMA framework of open-access and Q1/Q2 peer-reviewed articles published between 2014 and 2025. Through reflexive thematic analysis, sixteen studies were synthesised in four dimensions of analysis, namely application types, evaluation practices, cognitive support, and pedagogical integration. The results show that there is a systematic predominance of predictive analytics and performance-based assessment, but little overt scaffolding in problem solving, abstraction, and conceptual knowledge. Notably, ML systems are frequently used as back-end analytics and not as a learning-focused instructional tool. This review is novel in that it is a discipline-specific, cognition-based synthesis, shifting the ML paradigm from outcome prediction to pedagogically grounded cognitive scaffolding. Such a point of view provides practical design and assessment guidelines to narrow the divide between ML innovation and disciplinary learning in the education of mathematics and computer science.

Keywords: Machine Learning; Mathematics Education; Computer Science Education; Learning Analytics; Educational Technology; Cognitive Scaffolding; PRISMA; Reflexive Thematic Analysis.

1. Introduction

1.1. Contextual background

Machine learning (ML) has evolved rapidly, and many areas have been transformed by these advances, and education is one of the areas that has been significantly affected by increased access to learner information and the development of computational power. Baker and Inventado (2014) demonstrated that a key role of educational data mining has been taken by ML techniques, which allow the analysis of behaviours of learners, assessment data, and instructional interactions on a large scale. Similarly, Romero and Ventura (2020) established that classification, clustering, and neural networks are ML-based techniques that are commonly used to aid automated assessment, adaptive learning processes, and customised feedback systems. In mathematics and computer science teaching, the structure and the intensive data content of the subjects have motivated these developments, which are easily modelled algorithmically.

A large amount of literature on the topic of learning analytics has shown that ML models are effective in their ability to anticipate academic outcomes and to recognize at-risk students who are liable to failure or dropout. Siemens and Long (2011) stated that predictive analytics can offer actionable insights in institutional decision-making and learner support, and that Peña-Ayala (2014) presented a systematic demonstration of the fact that supervised learning algorithms are especially effective in predicting the outcomes of a student in a variety of educational settings. Lakkaraju et al. (2015) found that in discipline-specific contexts, predictive models using ML methods could be used to effectively predict student achievement in STEM courses, such as mathematics, but Ihantola et al. (2015) and Tarek et al. (2022) found that automated assessment systems operated by ML methods could be effectively used to assess programming assignments at scale. Taken together, these studies point to the technical soundness, scalability, and efficiency of the application of ML-based approaches to mathematics and computer science education.

Although these benefits are proven, an emerging body of critical literature argues that much of the current research is focused on predictive accuracy and optimization of systems rather than pedagogical and cognitive factors. Holmes et al. (2019) believed that in the majority of cases, AI and ML in education are created with a strong technological focus, and little is based on learning theory or instructional design principles. In a similar manner, Williamson and Eynon (2020) warned that educational technologies that use data are likely to emphasize learning as an outcome more than a complex thinking process. This is particularly concerning in mathematics and computer science education, where higher-order cognitive and disciplinary abilities, including abstraction, problem solving, algorithmic thinking, and conceptual understanding, are the focus of learning achievement (Wing, 2006; Schoenfeld, 2016).

This imbalance is further emphasised by the recent reviews of ML in education. The researchers demonstrated that the majority of empirical investigations are based on system performance, prediction, and administrative efficiency, and less attention is paid to how ML tools can assist with the learning processes on the cognitive or disciplinary level (Zawacki-Richter et al., 2019). Chen et al. (2020) also discovered that personalization and analytics are the most popular in the literature, but explicit pedagogical and learning impact studies are underrepresented. Because of this, synthesized evidence on whether and how ML applications in mathematics and computer science education support disciplinary learning processes other than analytics and personalization is sparse. This gap emphasizes the necessity of a narrow and systematic thematic analysis of the use of MLs in these areas, with a specific emphasis on whether they align with cognitive and pedagogical learning goals.

1.2. Statement of purpose

Most applications of machine learning in mathematics and computer science education are structured and assessed based on predictive analytics and personalization results, with a relatively low level of conceptual and empirical focus on systematically scaffolding the higher-order cognitive functions and disciplinary learning of students, including problem solving, abstraction, and conceptual learning.

1.3. Research Questions

These are the main sub-questions that will guide the central purpose of this paper, which will translate the purpose of the statement into actionable research questions:

a) What are the most common machine learning applications reported in the mathematics and computer science education research?

b) Which outcomes and evaluation criteria of machine learning applications in mathematics and computer science education are most likely to be used?
c) How do current machine learning applications explicitly facilitate or support higher-order thinking, e.g., problem solving, abstraction, and conceptual knowledge, in mathematics and computer science learning?
d) What are the pedagogical ways to implement machine learning in teaching and learning designs in math and computer science education practices?

2. Literature review

2.1 Research design and method

This study adopted a qualitative systematic review design guided by the PRISMA 2020 framework and employed reflexive thematic analysis to synthesize evidence on the application of machine learning in mathematics and computer science education. The PRISMA approach was used to ensure transparent identification, screening, eligibility assessment, and inclusion of studies, while reflexive thematic analysis provided a structured yet interpretive approach for examining patterns across the selected literature (Page et al., 2021; Braun & Clarke, 2019).

The review focused on peer-reviewed studies published between 2014 and 2025 that examined machine learning applications within mathematics or computer science education contexts. Following the screening process, the final corpus of studies was imported into a structured extraction framework and reviewed multiple times to facilitate familiarity with study aims, methodologies, machine learning applications, evaluation approaches, and reported educational outcomes.

The thematic analysis followed a hybrid deductive–inductive coding strategy. Deductive coding was initially guided by the four research questions, resulting in four parent analytical domains: (1) machine learning application types, (2) evaluation practices and outcomes, (3) cognitive and disciplinary learning support, and (4) pedagogical integration. During repeated reading of the included studies, inductive coding was subsequently used to identify emerging concepts, patterns, and relationships that were not predetermined by the research questions. Examples of inductively generated codes included predictive monitoring, automated assessment, performance-oriented evaluation, cognitive scaffolding, teacher agency, and pedagogical transparency.

Codes were iteratively reviewed, compared, and grouped into candidate themes. Themes were refined through repeated examination of conceptual coherence, internal consistency, and alignment with the objectives of the review. Following the principles of reflexive thematic analysis, themes were treated as patterns of shared meaning rather than frequency-based categories and were interpreted in relation to the broader educational and disciplinary context (Braun & Clarke, 2006; Braun & Clarke, 2019).

To enhance methodological transparency and trustworthiness, analytic decisions were documented throughout the review process using an audit trail. Coding definitions, theme revisions, and interpretive decisions were recorded systematically to ensure consistency and reflexivity. A reflexive approach was maintained throughout the analysis, acknowledging the active role of the researcher in theme development and interpretation. These procedures contributed to the credibility, transparency, and methodological rigor of the thematic synthesis (Byrne, 2022; Nowell et al., 2017).

2.2 Data collection and search strategy plan

A PRISMA-guided systematic review, with the help of which the data collection and search strategy were applied, was used. This study adopts a qualitative systematic review design guided by PRISMA 2020 and employs reflexive thematic analysis to synthesize peer-reviewed literature. The search was performed in Scopus, Web of Science Core Collection, ERIC, IEEE Xplore, and the ACM Digital Library to provide multidisciplinary coverage.

2.3 Data analysis plan

Reflexive thematic analysis was applied to the final corpus of included studies in accordance with the six-phase framework suggested by Braun and Clarke (2006, 2019). This method was chosen because it is appropriate to synthesize qualitative evidence and focus on its interpretive dimensions, reflexivity, and conceptual consistency rather than straightforward quantitative measures. The analysis has been carried out through repeated familiarisation with the studies, approaches, assessment standards, and documented learning effects. The first codes were generated using a hybrid deductive-inductive approach, whereby deductive codes were produced based on the four research questions and inductive codes were produced based on the emergent patterns that were not envisaged before.

Codes were then grouped into candidate themes in accordance with each parent research question, with specific consideration given to the framing of machine learning applications in relation to predictive analytics, cognitive support, and pedagogical integration. Themes were revisited and narrowed down so as to provide internal consistency and conceptual specificity, and were defined and given names that reflected their key organizing concepts. Analytic decisions through the process were recorded in order to have transparency and reflexivity. Based on this thematic structure, the synthesis in the Findings section makes it possible to conduct a critical analysis of the prevailing trends, gaps, and unexploited dimensions in the literature (Braun & Clarke, 2019; Byrne, 2022).

2.4 Restrictions on the research methodology

There are a number of limitations to be expected. First, the restriction to open-access studies can introduce availability bias, meaning that high-impact but paywalled studies can be underrepresented. Second, the use of Q1/Q2 journals only enhances rigour, but can miss novel or emerging work in niche or unranked journals, which can impact comprehensiveness. Third, the review is restricted to English-language studies in other non-English settings. Fourth, due to the wide range of designs and evaluation standards, in research on ML-in-education, heterogeneity prevents direct comparability between studies; this issue is resolved by not employing a meta-analysis but relying on a thematic synthesis. Lastly, reflexive thematic analysis is associated with interpretive judgement, despite the increased dependability of systematic procedures and audit trails; themes can still be biased by the analytic perspective of the reviewers (Braun & Clarke, 2019).

2.5 Ethical consideration and prejudice elimination strategy

The study has no human participants and interventions or the gathering of personally identifiable information, so formal human-subject ethical approval is usually not necessary. Nevertheless, ethical practice will involve proper claims, interpretation, and citation practices of the study. Bias will be mitigated at various levels. Multi-database searching will help minimize search and screening bias, and PRISMA-based transparent documentation of decisions and exclusions will be used (Page et al., 2021).

Predefined criteria will be used to diminish selection bias, and the reasons for excluding full texts will be documented. The presence of analytic bias will be minimized through keeping a reflexive log that will indicate the development and revision of interpretations according to the principles of reflexive thematic analysis (Braun & Clarke, 2019; Byrne, 2022). Reporting clarity and transparency will be based on trustworthiness criteria (Lincoln & Guba, 1985; Nowell et al., 2017).

3. Findings

Table 1 provides an overview of the sixteen studies included in the review. The evidence base comprised a combination of empirical studies and review articles published between 2014 and 2025, with a primary focus on machine learning applications in mathematics, computer science, STEM, and broader educational contexts. The reviewed studies addressed predictive modelling, learning analytics, automated assessment, adaptive learning systems, and educational data mining.

Table 1. Summary of the sixteen studies

Author(s)	*Year*	*Study Type*	*Educational Context*	*Primary ML Focus*
Chen et al.	2020	Empirical	STEM/Higher Education	Predictive modelling
Lakkaraju et al.	2015	Empirical	Higher Education	Academic risk prediction
Ihantola et al.	2015	Empirical	Computer Science Education	Automated assessment
Romero & Ventura	2020	Review	Educational Data Mining	Learning analytics and ML applications
Zawacki-Richter et al.	2019	Review	Higher Education	AI and ML trends in education
Schoenfeld	2016	Empirical/Theoretical	Mathematics Education	Mathematical problem-solving and cognition
Valencia-Arias et al.	2025	Empirical	Virtual Learning Environments	Engagement analytics and prediction

Ibarra-Vazquez et al.	2023	Review	Open Education	Competency prediction using ML
Khan & Al-Lawati	2025	Empirical	Mathematics and Engineering Education	Academic resilience and proficiency prediction
Nguyen & Pham	2025	Review	Mathematics Education	AI integration trends
Huang et al.	2025	Review	Educational Process Analytics	ML applications in educational data
Ersozlu et al.	2024	Empirical	Educational Data Mining	Learner performance analytics
Ximenes	2025	Review	Mathematics Education	AI opportunities and pedagogical implications
Sanusi et al.	2022	Review	K–12 Education	Teaching and learning machine learning
Gaitantzi & Kazanidis	2025	Review	Computer Science Education	AI applications in database instruction
Peña-Ayala	2014*	Review	Educational Data Mining	Educational data mining applications

3.1. Theme 1: The Preeminent Types of Machine Learning Applications in Mathematics and Computer Science Education

Machine learning is often framed as an instrument of prediction, usually aimed at predicting student performance, academic risk, and other outcomes related to engagement instead of facilitating disciplinary learning. The most recent meta-analyses reveal that predictive modelling (decision trees, random forests, support vector machines, and neural networks) is prevalent in the application of ML in education (Zawacki-Richter et al., 2019; Romero & Ventura, 2020). In the field of mathematics and computer science education, ML models are usually trained on past assessment data sets, interaction history, or submission history to predict end grades, pass-fail results, or likelihood of dropping out (Lakkaraju et al., 2015; Chen et al., 2020). Corpus empirical studies often have high predictive accuracy, often above 80 – 90% especially when predicting academic risk or predicting course success. Nonetheless, predictive success is commonly viewed as an end in itself, following of the work of Siemens and Long (2011) and Romero and Ventura (2020). Interventions Model outputs (e.g., risk scores or probability estimates) are hardly converted into instructional interventions aimed at enhancing mathematical reasoning or computing abstraction. Instead, predictive analytics are used mainly for institutional monitoring purposes (such as early warning systems and administrative decision-making) as opposed to learner-focused cognitive assistance.

The targets of prediction are different in different studies, but they are still focused on the result. Others look at the grades and test scores, especially in undergraduate mathematics and introductory programming classes (Lakkaraju et al., 2015). Others focus on predicting dropouts and retention, and they view ML as an approach to the early detection of risky learners (Chen et al., 2020). The involvement-based prediction based on the clickstream or log data is widespread as well, whereas the concept of engagement is usually operationalized as the frequency of activities instead of the depth of the conceptual processing (Valencia-Arias et al., 2025). In a minor portion of the literature, prediction drives resource recommendation or personalisation, though, even in this case, predictive modelling tends to precede pedagogical design rather than arise from it.

The second dominant application is automated assessment and grading, especially in computer science education. Code submissions, error type classification, and scoring on a large scale are some of the most common uses of ML-based systems (Ihantola et al., 2015; Romero & Ventura, 2020). Although these systems are more efficient, the feedback that is generated is usually correctness-based and focuses on right-wrong decisions and not on strategy-based advice. Therefore, automated evaluation is likely to support the optimization of performance without the explicit scaffolding of problem-solving heuristics, abstract thinking, or conceptual debugging.

Another obvious type of application is personalization and adaptive recommendation systems. These systems propose materials, order of activities, or change the difficulty according to expected performance (Zawacki-Richter et al., 2019; Nguyen &Pham, 2025). Nevertheless, personalization is often viewed as a technical optimization challenge, where the learning model is poorly represented and articulated. Personalization, unless specifically pedagogically underpinned, may end up as adaptive pacing, as opposed to cognitive scaffolding.

Among the studies reviewed, relatively few focused explicitly on facilitating mathematical proving, conceptual change, abstraction development, or structured problem-solving, and little is based on established learning theories like the problem-solving framework by Schoenfeld or the models of computational thinking. This lack supports the main argument of this review that, even with all their technical sophistication, the applications of ML in mathematics and computer science education are largely still focused on prediction, automation, and personalization, with little systematic support for higher-order disciplinary learning practices.

3.2. Theme 2: Evaluation Outcomes and Evaluation Criteria applied in research on ML-based mathematics and computer science education

In the reviewed articles, technical performance measures, in particular classification accuracy, area under the curve (AUC), precision-recall metrics, and F1 scores, prevail in the evaluation practices. The most frequent interpretations of empirical studies using ML models on educational data are predictive performance of approximately 75–90 percent, especially on predicting final grades or identifying at-risk students (Lakkaraju et al., 2015; Romero & Ventura, 2020). Comparative model benchmarking is the major contribution in a variety of instances, where various algorithms are tested to see which is the most accurate predictor (Chen et al., 2020).

Despite the fact that such metrics are necessary in terms of validating model reliability, they are predominant indicators of the evaluation focus on model performance, but not the impact on learning. The high level of predictive accuracy proves that a system can make accurate predictions, yet it does not prove that the system enhances the mathematical reasoning or students' understanding of mathematical concepts. According to Zawacki-Richter et al. (2019), numerous AI-in-education studies implicitly compare predictive success and educational value even though they offer little evidence that predictions are translated into instructional interventions that can modify the learning tracks.

A second pattern is a reduction of educational achievement to the proxies of the quantitative outcomes, most often grades, examination grades, pass-fail status, or course completion. In mathematics education, the most common type of dependent variable is final exam or cumulative assessment percentage (Lakkaraju et al., 2015), whereas in computer science education, and in introductory programming courses, success is often defined by assignment completion or correct code (Ihantola et al., 2015). These tests provide limited information about the development of a conceptual foundation, the ability to use abstraction, or the strategies of solving problems. Conceptual outcomes are not central when reported but are usually weakly operationalized, like short self-report measures or implied through performance indicators (Valencia-Arias et al., 2025).

Some of them contain deep learning process logs, such as clickstream records, step-by-step sequences of solving problems, or code execution logs, especially in Web-based learning (Romero & Ventura, 2020). Nonetheless, these data are mostly utilized as the input features to enhance the predictive effectiveness, but not as analytical tools to comprehend the reasoning of learners. Process data is hardly used to model problem-solving strategies, conceptual transitions, or debugging behaviour, even in cases where they are available. In this respect, process data is often considered a source of features, rather than a window into cognition.

Learning-theory-based evaluation methods, including those addressing metacognition, conceptual change, and abstraction, are still rare. Although these constructs are conceptually mentioned in some studies, very few of them are operationalized in the context of evaluation (Nguyen & Pham, 2025). This evaluation-light orientation restricts interpretability and transferability, and it is hard to determine whether genuine learning is achieved or whether the effects are merely short-term optimization effects.

Lastly, little is given to the ethical and practical evaluation aspects, such as bias, fairness, and interpretability. Even though these issues are recognized in the review studies (Zawacki-Richter et al., 2019), the empirical assessment of the differentiated effect on groups of learners is seldom studied. In education in mathematics and computer science, which place a premium on reasoning and transparency, the use of opaque, accuracy-based measurement may eventually erode pedagogical practice as well as equity.

3.3. Theme 3: Impact on advanced Cognitive and Disciplinary Learning assistance

In the reviewed literature, assertions that machine learning aids learning are widespread, but, on closer look, the support is often implied rather than explicitly intended. In most situations, the outputs of ML systems are predictions, classifications, or suggestions that are believed to be educationally valuable without an explicit statement of how such outputs support problem solving, abstraction, or conceptual knowledge. As noted by Romero and Ventura (2020), learning analytics and ML applications usually end with the identification of patterns within the data of learners, yet the pedagogical application of these patterns is not detailed. In this regard, a predictor does not educate, and a recommender does not necessarily construct reasoning. The literature focuses much on learning about performance-risk scores, probabilities, rankings, etc., as opposed to learning how to think in the field of mathematical or computational analysis.

Across the studies included in this review, explicit cognitive problem-solving scaffolding appeared relatively uncommon, although the limited number of studies examining this issue means this finding should be interpreted cautiously. A very limited fraction of the studies reported using structured hints, worked-example scaffolds, or prompts that remind students to consider the plan of action. Such scaffolding is usually task-oriented rather than based on transferable problem-solving heuristics, even when it does exist. An example of successful prediction of students at risk in STEM courses is Lakkaraju et al. (2015), which demonstrated a strong predictive capacity, but the logic of the intervention is aimed at alerting the instructor instead of helping students to develop mathematical reasoning. Consequently, the relationship between the outputs of the ML and disciplinary cognition is usually indirect.

Another version of this trend can be seen in computer science education, where systems based on ML often evaluate the outputs of code, rates of errors, or success. The automated assessment systems proposed by Ihantola et al. (2015) are reliable for measuring correctness at scale, but they seldom instruct learners on how to generate correct code by following specific reasoning processes. There is a lack of support for abstraction, algorithmic thinking, decomposition, and debugging strategies, and many systems support outcome validation as opposed to cognitive advice. As a result, learners get feedback on what not to work on and not why or how to think differently.

Within the reviewed studies, conceptual knowledge was infrequently operationalized through direct measures, with most studies relying on performance-based indicators such as grades or task completion. Most of the studies included in the corpus do not directly measure changes in conceptual understanding or the development of abstraction, but use performance proxies like scores or completion rates. According to Valencia-Arias et al. (2025), conceptual constructs that appear tend to be hidden through engagement measures or measures of achievement, which undermines the argument of disciplinary learning. Claims of improvements in learning with the help of ML systems are hard to prove without definite conceptual indices. Effectively, the majority of studies show performance forecasting but not conceptualization.

Significantly, there are few studies that are positive exceptions. These works are distinguished by the fact that the applications of ML are based on learning theory, which combines process-oriented feedback and positions systems as part of instructional processes instead of viewing them as independent analytics (Nguyen & Pham, 2025). Nevertheless, these methods are not common. Their limited number highlights the main point of the given review: being highly technical, the evidence reviewed suggests that many ML applications prioritize prediction, automation, and performance optimization, while comparatively fewer studies explicitly address higher-order cognitive and disciplinary learning, indicating that there is still a notable gap between analytics-driven innovation and learning-oriented design.

3.4. Theme 4: Pedagogical Implementations of machine learning to Teaching and Learning designs

In all the studies considered, machine learning is most frequently introduced into education as a component of the learning infrastructure and not as a direct part of the learning design. Across many of the reviewed studies, ML systems were implemented primarily as dashboards, risk indicators, or automated grading tools that supported instructional decision-making rather than directly shaping learning activities (Romero & Ventura, 2020). The main role of students and teachers in such implementations is to provide data, whereas the outputs of ML are intended to be sent to the administrative monitoring or decision-making in institutions. Such a setup makes ML an efficiency-focused technology, which can facilitate oversight and scalability, but with little instructional usefulness in disciplinary learning of mathematics and computer science.

This trend is also supported by the assigned role of teachers. In the majority of the incorporated research, teachers are placed as end-users who get prediction or analytics results, or as executors who will act on the system's recommendations (Zawacki-Richter et al., 2019). A few studies characterize teachers as co-designers in the process of developing pedagogical processes based on ML insights. The lack of agency among teachers is significant: in the case of instructional choices predetermined by predictive models, the correspondence to the goals of curriculum, classroom practice, and disciplinary epistemologies is diluted. This tends to lead to limited uptake or only superficial use of ML tools, especially where there are mathematical or computationally intensive tasks.

Another gap is with regard to classroom implementation. A great number of these studies justify ML systems with retrospective datasets or controlled experiments, limiting their analysis to model performance but not considering the performance of systems in real-world classrooms (Chen et al., 2020). Consequently, there is very little practical consideration of the constraints that exist, e.g., curriculum pacing, assessment policies, and workload of teachers. This gap between laboratory and classroom inhibits the interpretability and transferability of results, because technically sound models might not be applicable in the context of the instructional setting to create meaningful results.

Another essential weakness is pedagogical transparency. Only a limited number of reviewed studies reported the use of explainable feedback mechanisms that explicitly communicated the rationale behind system recommendations explaining why a system gave a specific recommendation or prediction. Opaque ML outputs can be incompatible with epistemic standards of reasoning and justification, which are two main disciplinary values in mathematics and computer science education. Without transparency, the learners would be placed as passive receivers of recommendations as opposed to being active participants in the sense-making processes.

4. Discussion

The aim of this review was to critically discuss the conceptualization, evaluation, and pedagogical incorporation machine learning applications into mathematics and computer science instruction. In all four of these themes, it seems that a similar trend is present, that is, ML systems are largely developed and evaluated as prediction, automation, and personalization tools, with limited focus on higher-order cognitive and disciplinary learning processes. Although predictive accuracy and scalability are often shown (Lakkaraju et al., 2015; Romero & Ventura, 2020), these technical successes are not often reflected in instructional frameworks that explicitly help students solve problems, make abstractions, or develop conceptual understanding.

The results demonstrate that the disciplinary learning goals and the priorities of ML application were still out of balance. Education in mathematics and computer science focuses on reasoning, justification, and conceptual transfer, but most ML systems concentrate on outcome proxies like grades, completion, or risk classification (Zawacki-Richter et al., 2019). Even with rich process data, these are usually instrumentalized to enhance prediction, but not to support cognition as scaffolding, thereby reinforcing a performance-oriented conception of learning. Consequently, the evidence indicates that many ML applications exhibit a strong analysis capability but a low pedagogical foundation.

Notably, this review is not an indication that ML does not have educational potential. Instead, there is evidence indicating that the design needs to be reoriented. Potential options include studies that incorporate ML into pedagogical practice, i.e., embedding knowledge in the lesson planning process, aiding in teacher decision-making, providing explainable feedback in a form of process-oriented feedback, etc. (Nguyen & Pham, 2025). Such strategies position ML as a cognitive scaffolding collaborator, and not a back-end analytics engine.

5. Recommendations

These findings suggest that the key priorities for future research should be: (a) theory-based design of ML systems that are based on domain-specific learning frameworks; (b) the evaluation strategies that can measure conceptual knowledge and logical processes as well as performance results; and (c) participatory design models that will make teachers co-designers of ML-based instruction. These changes are necessary to align ML innovation with the epistemic goals of mathematics and computer science education.

6. Conclusion

This thematic review was a systematic review that studied machine learning use in mathematics and computer science education based on the factors of application type, evaluation practice, cognitive support, and pedagogical integration. Based on the results within the reviewed corpus, ML applications were predominantly centred on prediction, automation, and personalization with little systematic assistance of higher-order disciplinary learning. Practices used in evaluation heavily focus on predictive performance, and conceptual understanding, abstraction, and problem-solving skills are seldom operationalised and explicitly measured.

This review reveals an important gap between ML-driven analytics and learning-focused educational design, suggesting a need for greater integration between predictive technologies and pedagogically grounded instructional approaches. To fill this gap, it is necessary that there is a transition from outcome-optimized to pedagogically more informed ML systems. To ensure machine learning innovation can meaningfully improve mathematics and computer science education, it is important to align machine learning innovation with disciplinary learning theory and instruction practice rather than merely measuring its performance.

REFERENCES

Baker, R.S. and Inventado, P.S. (2014). Educational Data Mining and Learning Analytics. *Learning Analytics*, 61 – 75. https://doi.org/10.1007/978-1-4614-3305-7_4.

Braun, V. and Clarke, V. (2006). Using Thematic Analysis in Psychology, *Qualitative Research in Psychology*, *3*(2), 77 – 101. https://doi.org/10.1191/1478088706qp063oa.

Braun, V. and Clarke, V. (2019). Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), pp. 589–597. https://doi.org/10.1080/2159676X.2019.1628806.

Byrne, D. (2021). A Worked Example of Braun and Clarke's Approach to Reflexive Thematic Analysis. *Quality & Quantity*, *56*(1), 1391 – 1412. https://doi.org/10.1007/s11135-021-01182-y.

Chen, L., Chen, P. and Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*(8), 75264 – 75278. https://doi.org/10.1109/ACCESS.2020.2988510.

Ersozlu, Z., Taheri, S. and Koch, I. (2024). A review of machine learning methods used for educational data. *Education and information technologies* [Preprint]. https://doi.org/10.1007/s10639-024-12704-0.

Gaitantzi, A. and Kazanidis, I. (2025). The Role of Artificial Intelligence in Computer Science Education: A Systematic Review with a Focus on Database Instruction. *Applied Sciences*, *15*(7), 3960. https://doi.org/10.3390/app15073960.

Holmes, W., Bialik, M. and Fadel, C. (2019). *Artificial Intelligence in Education Promises and Implications for Teaching and Learning.* Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-Excerpt-CCR.pdf.

Huang, J., Xin, Y.P. and Chang, H.H. (2025). The Application of Machine Learning to Educational Process Data Analysis: A Systematic Review. *Education Sciences*, *15*(7), 888. https://doi.org/10.3390/educsci15070888.

Ibarra-Vazquez, G. et al. (2023). Predicting open education competency level: A machine learning approach, *Heliyon*, *9*(11), e20597. https://doi.org/10.1016/j.heliyon.2023.e20597.

Ihantola, P. et al. (2015). Educational Data Mining and Learning Analytics in Programming. *Proceedings of the 2015 ITiCSE on Working Group Reports – ITICSE-WGR '15* [Preprint]. https://doi.org/10.1145/2858796.2858798.

Jonas Dreyøe Herfort et al. (2023). Twenty years of research on technology in mathematics education at CERME: a literature review based on a data science approach. *Educational Studies in Mathematics*, *112*(2), 309 – 336. https://doi.org/10.1007/s10649-022-10202-z.

Khan, G.D. and Al-Lawati, H. (2025). Machine learning insights into mathematics proficiency, academic resilience, and engineering success amid pandemic disruptions. *Social Sciences & Humanities Open*, *12*, 101936 – 101936. https://doi.org/10.1016/j.ssaho.2025.101936.

Lakkaraju, H. et al. (2015). A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [Preprint]. https://doi.org/10.1145/2783258.2788620.

Lincoln, Y.S. and Guba, E.G. (1985). Naturalistic inquiry. *International Journal of Intercultural Relations*, *9*(4), 289 – 331. http://dx.doi.org/10.1016/0147-1767(85)90062-8.

Long, P.D. and Siemens, G. (2014). Penetrare la nebbia: tecniche di analisi per l'apprendimento. *Italian Journal of Educational Technology*, *22*(3), 132 – 137. https://doi.org/10.17471/2499-4324/195.

Nguyen, D.T. and Pham, Q.V. (2025). The evolving landscape of AI integration in mathematics education: A systematic review of trends (2015 – 2025). *Eurasia Journal of Mathematics Science and Technology Education*, *21*(10), em2714–em2714. https://doi.org/10.29333/ejmste/17078.

Nowell, L.S. et al. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*(1). https://doi.org/10.1177/1609406917733847.

Page, M.J. et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*(71). https://doi.org/10.1136/bmj.n71.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, *41*(4), 1432 – 1462. https://doi.org/10.1016/j.eswa.2013.08.042.

PRISMA (2020). *PRISMA 2020 Flow Diagram*, *PRISMA*. https://www.prisma-statement.org/prisma-2020-flow-diagram.

Rodríguez-Ortiz, M.Á., Santana-Mancilla, P.C. and Anido-Rifón, L.E. (2025). Machine Learning and Generative AI in Learning Analytics for Higher Education: A Systematic Review of Models, Trends, and Challenges. *Applied Sciences*, *15*(15), 8679. https://doi.org/10.3390/app15158679.

Romero, C. and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3). https://doi.org/10.1002/widm.1355.

Santoso, P.H. et al. (2022). Thematic Analysis of Indonesian Physics Education Research Literature Using Machine Learning. *Data*, *7*(11), 147. https://doi.org/10.3390/data7110147.

Sanusi, I.T. et al. (2022). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, *28*. https://doi.org/10.1007/s10639-022-11416-7.

Schoenfeld, A.H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Journal of Education*, *196*(2), 1 – 38. https://doi.org/10.1177/002205741619600202.

Stamatoglou, M., Moirogianni, K. and Roussou, M. (2025). Play as an Inclusive Pedagogical Tool: Early Childhood Educators' Perspectives on Supporting Young Children with Special Educational Needs. *International Research in Education*, *13*(2), 1. https://doi.org/10.5296/ire.v13i2.23055.

Tan, B., Jin, H.-Y. and Cutumisu, M. (2023). The applications of machine learning in computational thinking assessments: a scoping review. *Computer Science Education*, *34*(2), 193 – 221. https://doi.org/10.1080/08993408.2023.2245687.

Tarek, M. et al. (2022). Review of Programming Assignments Automated Assessment Systems. *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 230 – 237. https://doi.org/10.1109/miucc55081.2022.9781736.

Valencia-Arias, A. et al. (2025). Use of machine learning in virtual learning environments: A bibliometric review. *Eurasia Journal of Mathematics Science and Technology Education*, *21*(9), em2695 – em2695. https://doi.org/10.29333/ejmste/16806.

Williamson, B. and Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, *45*(3), 1 – 13. https://doi.org/10.1080/17439884.2020.1798995.

Wing, J.M. (2006). Computational Thinking. *Communications of the ACM*, *49*(3), 33 – 35. https://doi.org/10.1145/1118178.1118215.

Ximenes, S. (2025). Artificial Intelligence in Mathematics Education: A Systematic Review of Opportunities, Challenges, and Pedagogical Implications. *Journal of Education Method and Learning Strategy*, *3*(03), 517 – 531. https://doi.org/10.59653/jemls.v3i03.1871.

Zawacki-Richter, O. et al. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators?. *International Journal of Educational Technology in Higher Education*, *16*(1), 1 – 27. https://doi.org/10.1186/s41239-019-0171-0.

✉ **Mr. Eyal Sadeh, M.A.**
ORCID iD: 0009-0005-6083-9187
Department of Computer Science
South-West University "Neofit Rilski"
Blagoevgrad, Bulgaria
E-mail: eyal1983@gmail.com