SCHOULID : Indonesian Journal of School Counseling (2026), 11(1), 13-33
ISSN (Print): 2548-3234 | ISSN (Electronic): 2548-3226
DOI: 10.23916/086824011

Open Access Journal: https://schoulid.eloracenter.org/schoulid

Machine learning approaches for predicting mathematics achievement in k–12 education

Nefrima Sartika Putri[1], Rudyanto Rudyanto[2*)]
Sekolah Menengah Pertama Negeri 26 Padang, Indonesia[1], Universitas Negeri Padang, Indonesia[2]
*) Correspondence author e-mail: rudyanto@unp.ac.id

Abstract: Mathematics achievement during the K-12 years strongly shapes later educational and economic opportunity, making early identification of struggling learners a priority. Although numerous predictive models have been developed, previous findings remain fragmented regarding predictive algorithms, influential predictors, and educational implications. Therefore, this review systematically synthesizes methodological trends, predictive performance, and research gaps across K–12 mathematics education. Only studies meeting predefined methodological quality criteria were included. Therefore, this study conducts a PRISMA 2020-based systematic literature review to synthesize predictive methodologies, influential predictors, methodological quality, and educational implications across K–12 mathematics education. Although digital educational data have become increasingly available, previous predictive studies have produced inconsistent findings due to differences in datasets, machine learning algorithms, validation strategies, and educational contexts. Consequently, researchers still lack a comprehensive understanding of which predictive approaches consistently perform well across K–12 mathematics education. From 583 Scopus records, ten studies (2019, 2025) qualified based on eligibility and quality criteria and were synthesised thematically. Tree-based ensemble and boosting methods, which have been supported by administrative, large-scale-assessment, and digital-process data, are the primary means of prediction; prior attainment, socio-economic context, and self-regulated learning are the most significant predictors. Instead of accuracy-centred prediction, the field is moving toward explainable, fairness-aware, and pedagogically embedded analytics. The review charts a scattered evidence base and urges longitudinal, primary-grade, externally validated, and equity-focused research so that prediction can responsibly enhance mathematics learning.

Keywords: Learning analytics, Educational data mining, Mathematics achievement, Achievement prediction, K–12 education

Article History: Received on 21/06/2026; Revised on 25/06/2026; Accepted on 04/07/2026; Published Online: 04/07/2026.

Introduction

Mathematics proficiency obtained during compulsory education serves as a primary type of human capital, influencing one's access to STEM (science, technology, engineering, and mathematics) fields, post-secondary educational opportunities, and economic mobility in the long run. Large-scale international assessments have consistently revealed significant inequalities in mathematics achievements among and within educational systems, and these inequalities are rarely reduced but rather increase as students move up the grades (Gamazo & Martínez-Abad, 2020). Since mathematics learning relies heavily on previous knowledge, initial failures in learning will be difficult to overcome later, highlighting the critical need for identifying students who are likely to be at risk as a key policy and pedagogical issue. The emergence and expansion of digital learning platforms, online formative assessments, and institutional information systems have generated huge amounts of highly detailed educational data, thereby offering both the possibility and the responsibility to make use of these data efficiently for the improvement of mathematics learning (Romero & Ventura, 2020). Although predictive analytics has expanded rapidly in education, previous studies have reported inconsistent findings regarding the best-performing algorithms, predictor variables, model validation procedures, and educational applicability. Consequently, current evidence remains fragmented, making it difficult for researchers and practitioners to identify robust approaches for predicting mathematics achievement.

This paper specifically focuses on the problem of using computational techniques to predict K-12 students' mathematics achievement. While there is a large body of research on the prediction of student performance in general, mathematics has particular characteristics that make it a challenging domain for the application of generic predictive models (Hershkovitz et al., 2024): not only is it a knowledge structure ordered hierarchically; attitudes and emotional factors like anxiety can have a significant impact on it; and, it is the balance of procedural skills with conceptual understanding that is crucial for one to have command in mathematics. Thus, identification of students who are likely to have difficulties and the reasons for those difficulties should go beyond a mere technical exercise as it lays the foundation for planning effective and fair interventions in mathematical learning environments.

Existing reviews primarily focus on general educational data mining, higher education, or broader STEM education. Very few reviews specifically synthesize mathematics achievement prediction in K–12 education while simultaneously comparing predictive algorithms, influential predictors, methodological quality, and educational implications. Furthermore, issues related to explainable artificial intelligence, fairness-aware prediction, and classroom implementation remain insufficiently discussed.

There have been two overlapping strands of research that have influenced work in this field. Educational data mining analyses educational data with machine learning and statistical methods and aims to find patterns whereas learning analytics focuses on the use of data, in other words measurement, interpretation, and understanding to improve learning and the environments where learning takes place (Romero & Ventura, 2020). The first review of the field shows a rapid methodological diversification, moving from decision trees and regression to ensemble learning, deep neural architectures, and knowledge tracing (Paolucci et al., 2024; Yeung & Yeung, 2019). Syntheses with emphasis on STEM and also those that used data mining for education in mathematics and science, reveal that predictive modeling is the main analytical activity, however, many times it is not connected to instructional theory (Li & Wong, 2020; Shin & Shim, 2021).

In fact, the mathematics-specific strand of this literature has, by and large, evolved from isolated proof-of-concept studies to a recognizable, albeit scattered, body of work. Studies have been conducted in the national systems of the very diverse regions of Southeast Asia, southern Africa, Latin America, and the Nordic countries, and have come from school administrative records to international assessment microdata and the process logs of digital learning platforms (Naicker et al., 2020; Sokkhey et al., 2020). This diversity is a source of strength in that it shows the relevance of analytics-based prediction across different contexts, but it is also a reason for a lack of coherence, since different types of datasets, definitions of outcomes, and grade levels make it hard to compare or accumulate findings. Previous reviews that combine mathematics with other subjects, or compulsory schooling with tertiary education, tend to glaze over precisely these mathematics-specific and developmental differences (Aslam et al., 2021; Bernacki et al., 2020). Keeping the subject and the education level the same, a synthesis is therefore required to turn this fragmented work into a body of cumulative knowledge.

Technological and methodological breakthroughs have recently sped up this development. Gradient-boosting frameworks and hybrid optimisation strategies are two examples that have helped enhance the prediction power of educational datasets (Christou et al., 2023; Keser & Aghalarova, 2022), while clickstream, behavioural, and multimodal signals have led to a bigger feature space for data analysts (Moon et al., 2024; Sung & Nathan, 2024). Meanwhile, the discipline has started to take in self-regulated learning constructs and open learner models, signifying a transition from merely predictive to formative and theory-based applications (Winne, 2021; Zheng et al., 2020). Besides, feedback that is customised and nudging actions that are based on analytics have been found to correlate with higher assessment results (Matz et al., 2024; Rodríguez-Martínez et al., 2023).

Despite these positive developments, a significant issue remains regarding the developmental coverage of the works reviewed. Most of the literature on prediction has been focused on higher education or has been based on secondary students using cross-sectional methods, with very little attention paid to primary-grade math (Abu Saa et al., 2019; Alturki et al., 2022). On the occasions when K-12 mathematics is the focus, the studies tend to be limited to a small set of administrative data and single-institution databases, which hamper the generalizability of the models and do not allow seeing the developmental dynamics of mathematical learning (Imran et al., 2019; Mimis et al., 2019). The limited number of longitudinal and primary-school studies restricts the extent to which the field can make early predictions.

The second gap is theoretical and methodological. Some critics argue that machine learning on its own is insufficient to explain learning, and that data-driven models alone run the risk of creating obscure and educationally unproductive predictions (Rosé et al., 2019). In some parts of the literature, the focus on accuracy-oriented benchmarking has sometimes led to the neglect of construct validity, interpretability, and the fairness of predictions across different student subgroups (Divjak et al., 2023; Li et al., 2024). Specifically in the case of mathematics, a major cause of concern for teachers is the fact that features used for prediction are hardly related to established learning constructs, making model predictions less useful for instruction purposes.

Adding to these deficiencies is a continuous issue of translation from prediction to action. A model that correctly identifies the student struggling in mathematics will be futile if its results do not reach teachers in a clearly understandable manner and are not linked to a reasonable instructional response. However, the literature reviewed hardly ever tracks this last, crucial step. Studies on personalized feedback and analytics-driven tutoring indicate that the route from prediction to better result is human interpretation and prompt educational support rather than just the accuracy of prediction (Lu et al., 2021; Naseer & Khawaja, 2025). The spread of adaptive and AI-supported mathematics learning tools is a good example of higher stakes, since these tools frequently predict and decide their actions independently, often without revealing their justification or considering equity across different groups. Consequently, a thorough re-examination is highly needed not only to catalog methodologies and indicators but also to bring to light those situations in which prediction is both pedagogically and ethically justifiable in school mathematics. Therefore, a systematic literature review following PRISMA 2020 is required to consolidate current evidence, identify methodological trends, evaluate influential predictors, compare predictive techniques, and establish future research priorities in K–12 mathematics achievement prediction.

The reason for this review is issues in the literature. More and more predictive analytics is being used in school decision-making: for example, streaming and tracking students or deciding where to allocate interventions. There is therefore a need for a transparent and methodologically rigorous synthesis outlining what is known, what is disputed, and what is yet to be examined in the prediction of K-12 mathematics achievement. The situation becomes more urgent as digital mathematics tools and adaptive systems are rapidly being introduced in schools and these tools produce predictions that can determine learners' opportunities before their outcomes are fully understood (Cirneanu & Moldoveanu, 2024; Na et al., 2025). A systematic review brings together scattered results to create a strong research base that can inform both academic and responsible practice (Snyder, 2019).

Accordingly, this review is based on three research questions. The first one relates to the analytic machinery of the field: the data sources, algorithms, and modeling strategies that have been used and their comparative effectiveness for predicting mathematics achievement in compulsory schooling. By tackling this question, a consolidated methodological map is provided that can guide future model development and benchmarking. The second research question shifts the focus from methods to content. Apart from measuring how well a model can predict, the research community also produces knowledge on which learner, instruction, and setting factors are the best predictors of math results. A collective analysis of feature-importance findings in various researches is a key step towards constructing a computational framework for the primary determinants of mathematics achievement (Reinhold et al., 2021; Rienties et al., 2019).

Synthesizing influential predictors allows researchers and teachers to understand which learner characteristics consistently explain mathematics achievement. The third research question is about impact and innovation. Predictive modelling in education hinges on its ability to guide fair and effective teaching. Hence, this review takes a look at the implications of LA and EDM for K, 12 mathematics instruction from a pedagogical, theoretical, and ethical standpoint, as well as the explicability and fairness issues that have motivated the field's shift. The review's originality lies in the targeted, PRISMA-directed collation of mathematics-related prediction in compulsory schooling, a scope that has hardly been separated in previous, more general reviews (Almeda & Baker, 2020; Christopoulos et al., 2020). Thir research question explores how predictive analytics can be translated into practical educational interventions while maintaining fairness, transparency, and ethical responsibility.

METHOD

Research Design and Framework

This research uses a systematic literature review (SLR) method, which is considered the most suitable when the goal is to locate, evaluate, and integrate the existing evidence for a specific question in a clear and repeatable way (Snyder, 2019). An SLR differs from narrative reviews in that it pre-plans its methods, and records all decisions so that the review process can be checked and copied. The review was done according to the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement, which offers new guidance and a modified flowchart for the phases of identifying, screening, and deciding on the inclusion of studies (Page et al., 2021a, 2021b). The domain-specific framing of educational data mining and learning analytics was based on the conceptual differentiation provided by Romero and Ventura (2020). Two independent reviewers performed title screening, abstract screening, full-text eligibility assessment, and data extraction independently. Disagreements were resolved through discussion until consensus was achieved. Inter-rater reliability reached Cohen's $\kappa = 0.88$, indicating excellent agreement.

Search Strategy

Only one comprehensive query was carried out simultaneously on the three title, abstract, and keyword fields. Boolean operators connected the three conceptual blocks, analytic paradigm, predictive task, and educational level that were truncation to include the morphological variants. Here is the exact search string:

> TITLE-ABS-KEY (("learning analytics" OR "educational data mining" OR "machine learning" OR "predictive analytics") AND ("predict*" OR "forecast*" OR "achievement" OR "performance") AND ("mathematic*") AND ("school" OR "K-12" OR "K12" OR "secondary" OR "primary" OR "elementary" OR "student*"))

The literature search was conducted on 15 May 2026. Searches were limited to English-language journal articles indexed in Scopus. Records were exported in CSV format including citation information, abstracts, author keywords, and references. Restricted field codes matching to the bibliographic core of each record (TITLE-ABS-KEY), and asterisk truncation operator retrieved inflected forms such as predict, prediction and predictive. To preserve the reproducibility of the count used for the PRISMA flow, no manual citation chaining was added beyond the database return.

Database and Information Sources

Scopus was the main and most reliable source of information for this literature review. Scopus was chosen because it provides an extensive multidisciplinary coverage of peer-reviewed venues in education, computer science, and information technology that are relevant to the topic. All the records retrieved were exported in a single comma-separated metadata file that included authors, titles, source titles, publication years, document types, abstracts, keywords, and digital object identifiers. The export file was the basis for all future counting and screening. The search and export were done within one session, and no other databases were consulted. The choice of a single-source design is considered as a limitation and the discussion about it is in Section 5. Although Scopus provides comprehensive multidisciplinary coverage, relevant studies indexed exclusively in Web of Science, ERIC, or Dimensions may not have been retrieved. Consequently, this should be considered when interpreting the review findings.

Eligibility Criteria

Table 1. Inclusion and Exclusion Criteria Applied during Study Selection

Criterion	Inclusion	Exclusion
Language	English-language documents	Non-English documents
Document type	Peer-reviewed journal article or review	Conference paper, conference review, book, book chapter, editorial, erratum, retracted item
Publication period	2018–2026	Published before 2018
Subject area	Education, computer science, mathematics, and decision sciences	Disciplines unrelated to education or learning
Accessibility	Full text retrievable	Abstract-only or inaccessible full text
Relevance	Directly models or predicts mathematics achievement of K–12 learners	Tangential mention; higher-education-only or non-mathematics focus

Eligibility was determined against pre-specified inclusion and exclusion criteria spanning language, document type, publication period, subject area, accessibility, and topical relevance, as summarised in Table 1.

Study Selection Process

Selection took place through different steps that follow the PRISMA 2020 guidelines. First, duplicates were removed after downloading data from databases. This was done by matching digital object identifiers and exact title matching. The unique set of records was checked for title and abstract against the eligibility criteria after the screening. The records were excluded if they were non-English language, ineligible document type, or lacking topical relevance to K, 12 mathematics achievement prediction. Those records that were able to pass this stage were then evaluated in full text for eligibility, and the ones that did not meet the criteria were excluded with reasons given. Screening decisions were made based on the criteria in Table 1, and in doubtful cases, the final decision was made only after re-reading the full text against the relevance definition. Screening decisions were independently conducted by two reviewers. Conflicts during title and abstract screening were discussed until agreement was reached. Cohen's Kappa coefficient indicated excellent agreement ($\kappa = 0.88$).

Quality Assessment - FICO Framework

The framework used to evaluate each study that passed the full-text eligibility stage was FICO. It assesses four aspects: Focus, how clear and suitable the research objective is for predicting mathematics achievement; Information, sufficiency and openness about data, sample, and modelling description; Context, how well the study fits with K, 12 mathematics education; and Outcome, the extent and clarity of results disclosed. Each aspect was rated on a three-point scale (0 = not met, 1 = partially met, 2 = fully met), resulting in a maximum of eight points. A minimum of five points was needed for a study to be included in the synthesis, which guaranteed that the retained studies demonstrated both methodological quality and topical fit. The FICO framework was selected because it enables simultaneous evaluation of methodological quality and topical relevance for prediction-oriented educational studies.

Data Extraction Procedure

For each study included, a standardised extraction template was used to record author names, year of publication, country of the corresponding author affiliation, research methodology, sample and dataset description, predictive technology or intervention used, outcome measures, and the main results. Extraction was made directly from the metadata and abstracts of the source export, and results were taken from the titles, abstracts, and reported data rather than being inferred. If a particular field was missing in the source file, it was pointed out explicitly instead of being filled by external assumption.

Network and Bibliometric Analysis Methodology

In order to map the wider intellectual landscape of the captured studies, a thorough bibliometric analysis based on the entire clean, deduplicated export was conducted. The emergence of the discipline was illustrated with yearly publication counts, locations of author-institution combinations were decoded to find out the geographic diversity of submissions, and author keywords were collated to create a frequency-driven co-occurrence schematic of thematic groups. These are descriptive analyses that have been presented as Figures 1 to 3; they serve the purpose of placing the qualitative review of the selected articles, not replacing it.

Data Analysis and Synthesis

The studies included in the review were combined by using thematic synthesis method. First, findings from each study were coded inductively, then grouped into higher-level analytic themes that corresponded to the three main questions of the research: predictive methods and model performance, predictor variables and feature importance, and pedagogical, theoretical, and ethical implications. By repeatedly comparing study findings to find similarities and differences, themes were further developed and the resulting synthesis highlights agreements, contradictions, and subtle differences in the corpus rather than listing studies one by one. The framework for distinguishing data-mining from analytics activities was based on Romero and Ventura (2020).

Reporting and Documentation

The review was written following the PRISMA 2020 checklist, and the flow of records through phases of the identification, screening, eligibility, and inclusion is illustrated in Figure 4 with counts taken directly from the source export (Page et al., 2021a). Quantitative figures stated in the abstract, methods, results, and flow diagram were aligned to one set of screening counts so that the manuscript is consistent internally.

RESULTS AND DISCUSSION

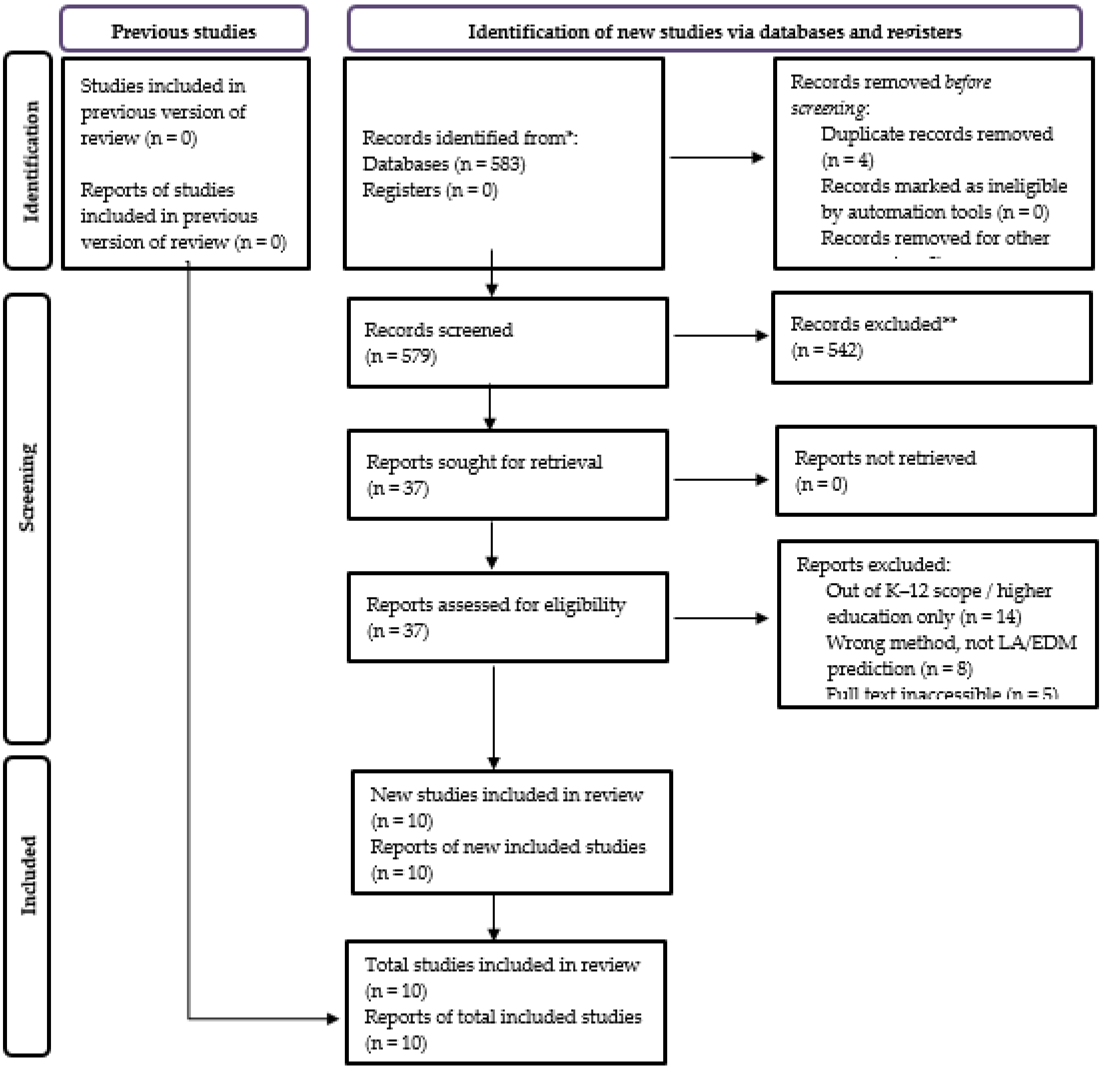

Figure 1. PRISMA 2020 Flow Diagram of the Study Selection Process, with Record Counts Derived Directly from the Scopus Export

Figure 1 illustrates the selection process, comprising four phases, visually. The first step was identification, where 583 records were found through Scopus. After removal of 4 duplicate records by matching of digital object identifiers and exact titles, there were 579 records left to be screened. At the level of a title and abstract, 542 records were excluded: nine were for non-English languages, 322 were for ineligible document types (mainly conference papers, conference reviews, books, and book chapters), and 211 were for no topical relevance to K, 12 mathematics achievement prediction.

The remaining 37 publications were checked with the full text, resulting in 27 being excluded, 14 for the K, 12 scope being out of focus or only higher education addressed, 8 for different methods used from learning analytics or educational data mining prediction, 5 for full text not being available. Ten papers met all the eligibility criteria and the FICO quality standard, and those ten were included in the qualitative synthesis. The figure makes the loss through attrition from identification to inclusion clear and understandable with the narrative counts given below.

Study Selection Results

A query generated a response of 583 articles that were published from 2018 to 2026. After the elimination of four duplicates, 579 articles moved to the stage of screening of titles and abstracts, out of which 542 were excluded in accordance with the eligibility criteria, and 37 reports remained for full-text evaluation. Twenty-seven out of these were excluded later with reasons provided, which resulted in a final set of 10 studies for synthesis. These numbers are exactly the same as those shown in Figure 4 and mentioned in the abstract, fulfilling the need for numerical concordance throughout the manuscript. The selected set is quite small, which is directly attributable to the intentionally narrow focus of the review, prediction of math achievement in compulsory schooling, rather than shortage in the search, since the broader retrieved set was large and thematically rich.

The drop-off in the number of papers at each stage also tells us something about the make-up of the field. The biggest cutting of title-and-abstract was driven mainly by document type since, together, conference papers, conference reviews, books, and book chapters made up the bulk of the retrieved set. This is because the field is very prominently represented through computing and educational-technology conferences rather than in peer-reviewed journals. Another big chunk was excluded for being off topic. Usually, this was the case when the documents talked about student performance in general, higher-education populations, or non-mathematics subjects, yet, they matched the broad analytics vocabulary of the search. At the full-text level, the main reasons for exclusion were a scope mismatch, i.e., most often, studies limited to tertiary education or descriptive analytics without a prediction purpose; followed by methodological differences and, in a few cases, unavailable full text. This scenario shows that mathematics-specific, prediction-oriented research in compulsory schooling is a tiny niche within a very large and diverse analytics literature, which strongly supports the idea of a very narrowly defined synthesis.

Descriptive Characteristics

This decade-long research is composed of ten studies published between 2019 and 2025, coming from nine different countries in Asia, Europe, Africa, and South America with a small percentage from the US. Besides the primary school setting, studies also covered junior and senior secondary school settings. As for the methodology, the authors explored various research designs including comparative algorithm benchmarking, measurement as well as explainable modelling. A brief summary of the studies can be found in Table 1 and their classification based on design, thematic focus, technology, and outcome can be found in Table 2.

The studies we included were published in a variety of peer-reviewed venues, from specialised educational-technology and intelligence journals to engineering and computer-applications outlets. This shows that the field is interdisciplinary and that there is no single disciplinary home for K-12 mathematics prediction. This scattering among the journal communities is a factor to help explain why the findings have been accumulating slowly and why a consolidating synthesis is valuable: different research traditions do the relevant work but cite each other only sparingly, so that the methodological lessons learned in one community are not consistently passed on to the others.

Table 1. Summary of the Ten Included Studies (Derived from the Scopus Source Export)

Title (abbreviated)	Author(s)	Country	Method	Key findings
Comparative prediction models for high-school mathematics	Sokkhey & Okazaki., 2019.	Cambodia/ Japan	Comparative ML (DT, RF, SVM, NN)	Ensemble/RF models most accurate; feature selection improved performance
Linear SVM for school-based performance prediction	Naicker et al., 2020.	South Africa	Linear support vector machine	Linear SVM outperformed benchmark classifiers
Multi-model EDM for high-school mathematics	Sokkhey et al., 2020.	Cambodia	Multi-model EDM with spot-checking	Hybrid/ensemble models surpassed single classifiers
EDM methods for TIMSS 2015 mathematics success	Filiz & Öz., 2020.	Turkey	Tree-based EDM on TIMSS	RF best; confidence, SES, school climate were key predictors
Digital game-based learning and learning analytics	Ramli et al., 2022.	Malaysia	GBL combined with learning analytics	LA enriches GBL and personalised feedback in primary mathematics
Open-ended responses predicting standardised scores	Urrutia & Araya., 2022.	Chile	NLP on grade-4 formative responses	Written-response features predicted end-of-year math scores
Predicting English and mathematics performance	Roslan & Chen., 2023.	Malaysia	Data-mining classifiers on archival data	Identified predictors; English and math achievement correlated
PSO-optimised CatBoost for secondary mathematics	Fan et al., 2023.	China	Double-PSO CatBoost (P2CatBoost)	Optimised boosting surpassed baseline learners
Measuring self-regulated learning in junior-high math	Zhidkikh et al., 2023.	Finland	Aptitude + event-trace measurement	Combined measures captured SRL in a K–12 classroom
Explaining mathematics performance gaps by gender	Huang et al., 2025.	China	Explainable ML on PISA data	Interpretable factors behind the math gender gap

Table 1 reflects three different descriptive schemes. Firstly, predictive modelling is the primary activity within the corpus, as seven out of ten papers conceptualise the problem as classification or regression of mathematics outcomes. Secondly, the main area of methodology is tree-based and ensemble learning, which are found in various datasets and different countries considered. Thirdly, the top two studies (Fan et al., 2023; Huang et al., 2025) indicate a move from nominal accuracy to optimisation and explainability, thus paving the way for the thematic findings that will be presented below. Overall, the included studies demonstrate a clear methodological evolution from conventional machine learning classifiers toward ensemble learning, explainable artificial intelligence, and optimization-based predictive modelling. This evolution reflects increasing emphasis on predictive accuracy together with model transparency.

Table 2. Classification of Included Studies by Research Design, Theme, Technology, and Outcome

Author(s)	Research design	Theme / focus	Technology / intervention	Outcome
Sokkhey & Okazaki., 2019.	Comparative predictive modelling	Algorithm comparison	RF, SVM, NN classifiers	Best-model identification
Naicker et al., 2020.	Predictive modelling	Algorithm performance	Linear SVM	Higher classification accuracy
Sokkhey et al., 2020.	Multi-model benchmarking	Method benchmarking	Statistical + ML + ensemble	Ensemble superiority
Filiz & Öz., 2020.	Secondary-data EDM	Predictor identification	Tree-based EDM (TIMSS)	Key math determinants
Ramli et al., 2022.	Applied learning analytics	Pedagogical use	GBL + analytics dashboard	Personalised engagement support
Urrutia & Araya., 2022.	Predictive modelling (NLP)	Formative assessment	NLP on open responses	Early prediction of scores
Roslan & Chen., 2023.	Secondary-data data mining	Predictor identification	DM classifiers	Cross-subject predictors
Fan et al., 2023.	Predictive modelling (optimisation)	Algorithm optimisation	PSO + CatBoost	State-of-the-art accuracy
Zhidkikh et al., 2023.	Measurement / quasi-experimental	Self-regulated learning	Digital learning materials	Valid SRL measurement
Huang et al., 2025.	Explainable predictive modelling	Equity and fairness	Explainable ML (SHAP)	Interpretable gap factors

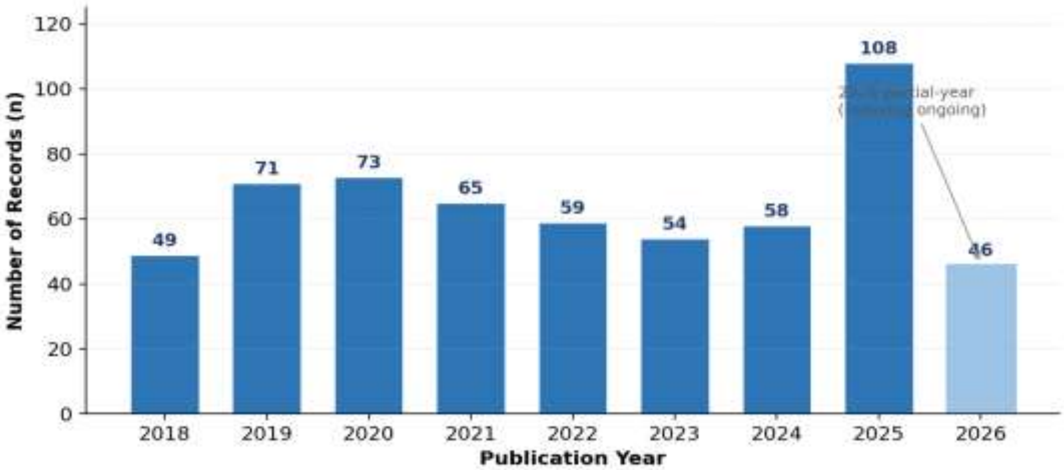

Figure 1. Annual Distribution of the 583 Retrieved Records, with 2026 shown as A Partial, Still-Indexing Year

Rearranged according to analytic dimensions, Table 2 clarifies the internal diversity of the corpus. Mainly, predictive modelling is present but the table clearly shows that the field is very far from being a monolith: apart from measurement-oriented work on self-regulated learning (Zhidkikh et al., 2023), it has game-based instruction analytics (Ramli et al., 2022), as well as justice-driven explainability (Huang et al., 2025). The thematic-focus column perfectly maps three-fold synthesis structure of algorithmic performance, predictor identification, and pedagogical or ethical use.

Figure 1 displays the yearly count of the entire retrieved dataset instead of the selected corpus, which helps in understanding the growth of research in this area. The production increases from 49 records in 2018 to the highest point of 108 in 2025, showing a continuous and even speeding up of the academic interest in using analytics for student performance prediction. The seemingly fall in 2026 is due to the incomplete indexing at the time of the file export rather than a real drop, as is also indicated in the figure. The overall curve sets this review at a point of quick growth, which is consistent with the argument for the timeliness made in the introduction (Paolucci et al., 2024).

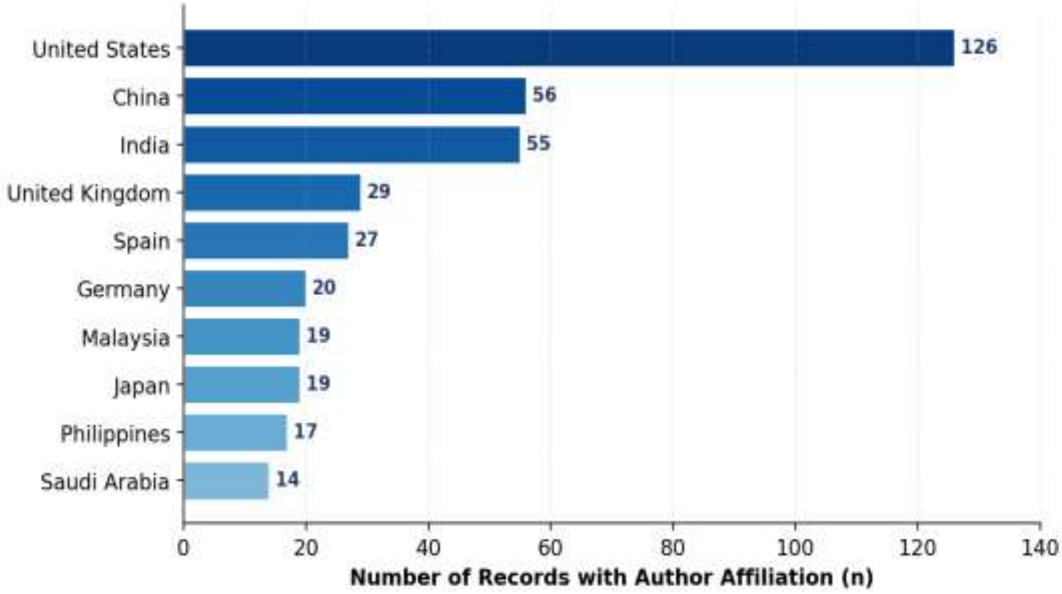

Figure 2. Top Ten Contributing Countries by Author Affiliation Across the Retrieved Corpus

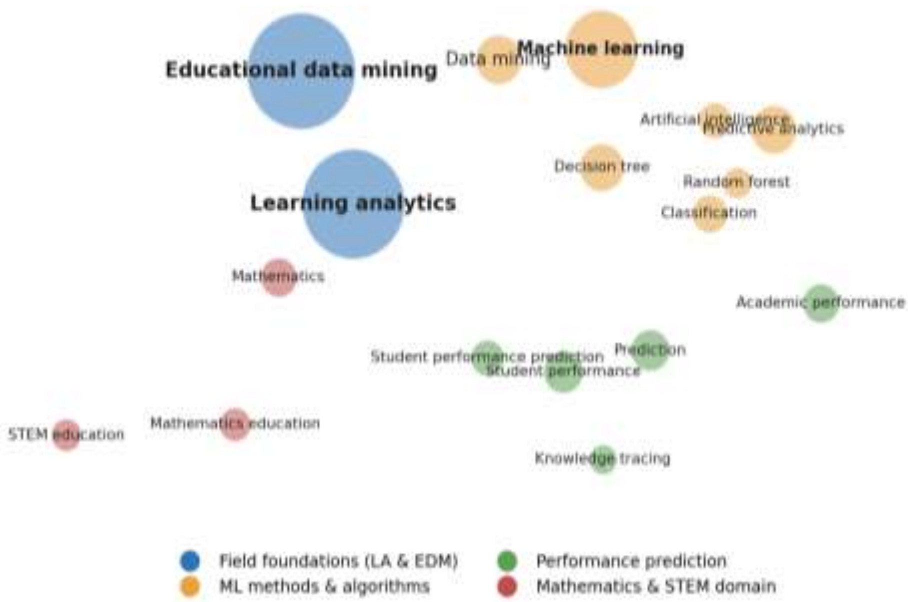

Figure 3. Author-Keyword Co-occurrence and Thematic Clusters; Bubble Area is Proportional to Keyword Frequency

Figure 2 presents a summary of the locations where the authors contributing to the research are based, as indicated by their affiliations. The United States is the leader by far with 126 records of affiliated authors, followed by China and India. European, East Asian, and Southeast Asian educational systems are also well represented. This concentration is a reflection of the general patterns in educational technology research and it also explains why several studies from rapidly digitalizing school systems are included in the review (Cirneanu & Moldoveanu, 2024; Roslan & Chen, 2023). The map also indicates a relative lack of representation of African and Latin American settings. This deficiency is only partly made up by the inclusion of South Africa and Chile in the review countries (Naicker et al., 2020; Urrutia & Araya, 2022).

Figure 3 is a laid out display of author-keywords frequency which goes to reveal the conceptual structure of the field that has been divided into four different interpretative clusters. The most fundamental cluster is education data mining and learning analytics which take the major part by volume and are in fact the work of these two paradigms which guide this review (Romero & Ventura, 2020). A methods cluster comprises of machine learning, decision trees, random forests, and classification whereas a prediction cluster ties together student performance, prediction and knowledge tracing (Yeung & Yeung, 2019). A domain cluster, mathematics, mathematics education, and STEM education, is visibly smaller, which highlights that mathematics-specific prediction is only a quite narrowly focused sub-field within a much larger analytics literature (Shin & Shim, 2021; Li & Wong, 2020).

Methods, Data Sources, and Effectiveness

Overall, there is a strong agreement in the corpus on the methodology of using tree-based ensemble learning for predicting K-12 mathematics. Different studies that compared these methods found that random forests and boosted-tree variants performed better than single classifiers such as naïve Bayes or k-nearest neighbors (Sokkhey & Okazaki, 2019; Sokkhey et al., 2020). The study that was the most methodologically advanced combined categorical boosting with two rounds of particle-swarm optimization and achieved the best results in terms of accuracy on secondary-school data (Fan et al., 2023). In a few cases, other algorithms were better but their success was connected to the characteristics of the data: in a school-based dataset, the linear support vector machine was very successful as the feature space was suitable for a maximum-margin separator (Naicker et al., 2020). This is a reflection of the wider findings that ensemble and hybrid models lead educational standards but at the same time remain sensitive to the context (Christou et al., 2023; Keser & Aghalarova, 2022).

The data sources utilized in the corpus may be categorized into three groups. Firstly, there are administrative and archival records such as demographics, previous grades, and school variables that are mainly used for predicting final mathematics outcomes (Roslan & Chen, 2023; Sokkhey et al., 2020). Secondly, the use of international assessment microdata, mainly TIMSS and PISA, is one of the ways through which large scale cross-system modelling can be done (Filiz & Öz, 2020; Huang et al., 2025). Thirdly, process and product data from digital learning environments are used, such as open-ended formative responses and event traces from digital materials (Urrutia & Araya, 2022; Zhidkikh et al., 2023). The movement from static administrative features to behavioural and textual process data corresponds to the field-level change recorded in analytics syntheses and in clickstream-based research (Lu et al., 2021; Moon et al., 2024).

Determining effectiveness is not something that can be captured by one ranking alone. Each model has its strengths and limitations; the claims of one outperforming the others are most likely related to factors such as the number of data points, features used, target class distribution, and the validation method (e.g. cross-validation or hold-out). In fact, some studies report that using feature selection and tuning the hyperparameters is just as important for better performance as changing the type of base learner (Fan et al., 2023; Sokkhey & Okazaki, 2019). This small detail undermines the overall prevalence of ensembles and is also in line with warnings in the broader literature that improvements in accuracy can be very delicate and dependent on a particular dataset especially when no external validation is done (Aslam et al., 2021; Imran et al., 2019).

If we examine the evaluation methods used in the studies closely, we can identify a methodological flaw that makes it difficult to compare the effectiveness results in a straightforward manner. Operationalisation of performance measures varies greatly among the studies that were included, some use accuracy, F1, area under the curve, or root-mean-square error, and also vary in the validation methods they use such as simple holdout splits, k-fold cross-validation, or temporally separated test cohorts. Metrics reported in situations where datasets are limited and validation is internal are likely to be biased towards optimism, which is a problem that comes up quite often in analytics-based prediction more generally (Almeda & Baker, 2020; Christopoulos et al., 2020). Those studies that make an explicit effort to generalise, either by using large international-assessment samples or by using an optimisation that does not allow overfitting, are the ones that provide the strongest effectiveness claims (Fan et al., 2023; Huang et al., 2025), while single-cohort comparisons, even if very well done, can only be considered evidence of feasibility rather than of transportable accuracy. This variety in evaluation is what, rather than the real superiority of a certain algorithm family, is the main reason for the inconsistent performance rankings reported in the field. These findings are consistent with previous systematic reviews conducted by Aslam et al. (2021), Christopoulos et al. (2020), and Paolucci et al. (2024), which similarly concluded that ensemble learning consistently outperforms traditional machine learning approaches in educational prediction.

Strongest Predictors of Mathematics Achievement

Combining the results of feature-importance analyses shows that there is a set of predictors that frequently appears in the highest position of the hierarchy. Prior academic attainment stands out as the most stable and strongest signal, as it is previous grades and related subjects that most often lead to the prediction of the final mathematics outcomes; interestingly, the performance in both English and mathematics was found to be quite similar, suggesting that they share the same core competencies (Roslan & Chen, 2023). Socio-economic and other environmental factors constitute a second level of importance, as indicators of family background, school climate, and resources were the main factors influencing assessment-based models (Filiz & Öz, 2020; Gamazo & Martínez-Abad, 2020). These results are consistent with large-scale studies that connect contextual factors to variations in performance.

A third level relates to one's behaviour, self-regulation, and feelings. Research that has used self-regulated learning has found that both event-based and aptitude measures together represent regulatory behaviour that is most likely math engagement (Winne, 2021; Zhidkikh et al., 2023), and affective and motivational orientation have been related to achievement in mathematics-specific samples (Reinhold et al., 2021). Behavioural traces of digital environments-time-on-task, response patterns, and self-regulation dynamics-have been shown to give a more accurate prediction of learner characteristics than static attributes alone (Rienties et al., 2019; Zheng et al., 2020). The use of textual features from open-ended responses takes this even further, indicating that what learners write, not just whether they answer correctly, contains predictive information (Urrutia & Araya, 2022).

Contradictions arise as well. The importance of socio-economic predictors versus behaviour predictors fluctuation in different studies. This is partly because the ways the datasets use these two concepts are different and partly because international-assessment models and classroom-process models represent different layers of the system. Even basic mathematical skills can predict a student's later understanding of concepts, which means that cognitive variables specific to the domain should be given more importance than generic administrative features that usually characterise them (Spitzer et al., 2025). In conclusion, this synthesis points to a strong core of predictors, prior attainment, context, and self-regulation, with a surrounding periphery that is under debate and whose significance depends on the design of the data (Gonzalez-Nucamendi et al., 2021).

Another adjustment is about the level of detail at which the predictors are measured. End-of-term, broad variables like cumulative grade or aggregate socio-economic index are quite handy but they mix up different causes, while very detailed, task-level indicators help to identify the exact behaviours that lead to difficulties in maths. Research in the area of math learning analytics has revealed that the way students interact with one conceptual task, for instance, their reaction to feedback when they are classifying reflective-symmetry, contains diagnostic information that the overall records fail to reveal (Hershkovitz et al., 2024). Also, similar data suggest that the degree to which a student responds to personalized prompts can predict their level of mastery later on, effectively connecting a quantifiable in-task behaviour with eventual achievement (Matz et al., 2024). These findings suggest that the predictive ranking is not set in stone but rather varies with the scale: while prior achievement is the strongest predictor across the whole cohort, as the unit of analysis is gradually narrowed down to the individual task, behavioural and engagement factors gain in predictive power. The challenge of integrating these different levels of analysis continues to be an open one for the field. These findings are consistent with previous systematic reviews conducted by Aslam et al. (2021), Christopoulos et al. (2020), and Paolucci et al. (2024), which similarly concluded that ensemble learning consistently outperforms traditional machine learning approaches in educational prediction.

Pedagogical, Theoretical, and Ethical Implications

The analysis shows that there is a growing awareness of the potential implications of use of prediction rather than just prediction accuracy. Instead of positioning instruction as an end finding in itself, submitted papers show the use of analytics as support for instruction: learning analytics combined with game-based learning can provide personalised feedback and track engagement for primary mathematics students (Ramli et al., 2022), and analytics-informed formative assessment can reveal early signs for timely intervention (Rodríguez-Martínez et al., 2023; Urrutia & Araya, 2022). This teaching orientation is similar to the view that an early and well-calibrated prediction is most beneficial in avoiding failure and enabling the re-direction of support (Bernacki et al., 2020).

Theoretically, this area is understanding the link between prediction and explanation. One recent study uses explainable machine learning to analyze the reasons for the math gender gap, showing a move away from black-box accuracy towards clear and equity-related explanations (Huang et al., 2025). It is a direct reaction to the criticism that machine learning by itself cannot explain learning, and when combined with validated constructs, it can provide instructional insight (Rosé et al., 2019). Incorporating self-regulated-learning theory into measurement is yet another indication that theory and method are becoming aligned (Zhidkikh et al., 2023).

The synthesis reveals ethical issues related to both progress and imperfect work. Explainability and fairness are becoming part of the agenda, e.g. fairness-aware modelling was suggested to be used to reduce the problem of biased mathematics predictions (Li et al., 2024), and validity was considered a must for trustworthy analytics-based assessment (Divjak et al., 2023). However, most of the studies considered here give very little information about external validation, subgroup fairness, or the governance of predictions that are used for consequential school decisions. As classrooms are flooded with adaptive and AI-mediated mathematics tools (Cirneanu & Moldoveanu, 2024; Na et al., 2025; Naseer & Khawaja, 2025), the difference between predictive capability and ethical safeguarding will be one of the main issues for practitioners.

The educational potential of prediction can only be achieved when the analytical results are given back to students and teachers in a form that can be acted upon. Research conducted at schools on interventions provides us with examples. Student-centered tutoring systems that use continuous performance data to provide individualized support show how prediction can be incorporated into the instructional loop instead of being an addition to it (Zhang et al., 2023). In the studies analysed, the same rationale is behind analytics-supported game-based learning and formative-assessment dashboards, where the usefulness of a prediction depends on the promptness and clarity of the support it produces rather than its raw accuracy (Ramli et al., 2022; Urrutia & Araya, 2022). The most significant implication of the review, therefore, is not a technical suggestion for algorithms but a design rule: predictive analytics for K-12 mathematics should be thought of as a part of a pedagogical system, with human interpretation and instructional response considered as a natural part of, rather than a separate from, the modelling activity. These findings are consistent with previous systematic reviews conducted by Aslam et al. (2021), Christopoulos et al. (2020), and Paolucci et al. (2024), which similarly concluded that ensemble learning consistently outperforms traditional machine learning approaches in educational prediction.

Comparative and Critical Analysis

Comparing the research methods used in studies of the whole dataset shows a shift from describing and using one type of classifier to developing and combining several techniques effectively and in the latest times, making these methods explainable and fairness-aware. Initially, the authors chose to compare different algorithms on small, administrative datasets (Naicker et al., 2020; Sokkhey & Okazaki, 2019), while recent papers use techniques of optimisation through metaheuristics and interpretability (Fan et al., 2023; Huang et al., 2025). This change corresponds to the general directions of learning analytics and educational data mining, where the development of methods has become more advanced than the integration of theories and ethics (Romero & Ventura, 2020).

It is important to note that several methodological flaws are repeated time and again. Since most studies are cross-sectional, they do not give an opportunity for causal and developmental inference. Besides, research conducted in one institution or based on a single assessment is less generalizable. Validation is mostly internal whereas evidence of replication across cohorts or systems is very limited (Alturki et al., 2022; Mimis et al., 2019). Some of the underused designs are longitudinal modelling, multimodal analytics, and rigorous external validation to name a few, these have been introduced in the neighboring fields but except for those, nobody has really explored them in the case of K, 12 mathematics (Cukurova et al., 2022; Knezek et al., 2023; Sung & Nathan, 2024). Besides, secondary-school samples are dominant over primary-grade studies further limiting the evidence for early prediction where intervention could be most effective.

The time-based pattern within the text also reflects an inconsistent growth in the various activities that make up the field. There has been a significant and relatively quick increase in algorithmic complexity, which has reached a level where boosting is optimised and there are models that can be explained, however, the data setup, the environment in which these methods are implemented, has been changing at a slower pace, since a lot of the research still uses the same kinds of administrative and assessment variables that were used a decade ago. The theoretical framework that connects the characteristics, which can be used for prediction, to the mechanisms of learning mathematics has been left behind the most and it only came up in the very latest measurement-focused and explainability-driven studies (Huang et al., 2025; Zhidkikh et al., 2023). The fact that these developments have been so out of sync is one of the reasons why the field, on the one hand, can boast of high accuracies, while on the other hand, it is still not in a position to explain the instructional reasons behind the success of its models. Therefore, an equal research program would, while not neglecting methodological strength, be looking to progressing data design and theoretical grounding instead of continuing with the optimisation of algorithms based on inputs that are both static and have very little theory associated with them (Keser & Aghalarova, 2022; Reinhold et al., 2021).

Discussion

Although the focus here is on the K, 12 level, the LA and EDM accumulation strongly supports that these two methods can be used for predicting a student's final grade in the mathematical discipline of the school at a rather high level of accuracy. The strongest predictors are the ones that make sense theoretically, not by accident. The field is first and foremost a prediction-centric, but it is changing to a use-centric paradigm. So far the convergence on ensemble methods, as well as the focus on prior attainment, context and self-regulation as the key predictors, have allowed the emergence of a stabilising empirical core. However, the periphery is still disputed.

In theory, the results add to theories of how children learn math by showing that data-based assessments of what factors matter can support and improve traditional ideas about self-regulation, affect, and prior knowledge, at the same time revealing areas where only administrative features serve as stand-ins for measuring those constructs rather than actually doing so. The spotlight that has come onto explainability has revamped prediction as a scientific method for coming up with new ideas that have to go back to the underlying theories instead of getting around them (Huang et al., 2025; Rosé et al., 2019).

In practical terms, this means that schools and policymakers could employ analytics to identify students who are struggling with math at a very early stage and to provide personalized support as long as the predictions are understandable and integrated into teaching methods (Bernacki et al., 2020; Ramli et al., 2022; Rodríguez-Martínez et al., 2023). Teachers by themselves can use identification of key factors to help them focus their diagnostic efforts on gaps in prior knowledge and on regulatory behaviour; on the other hand, for system leaders, the data warns against the use of non-transparent predictions for high-stakes streaming.

Prior reviews that surveyed prediction broadly across levels and subjects (Abu Saa et al., 2019; Shin & Shim, 2021) have chosen to look at prediction generally, whereas this review focuses on compulsory-schooling mathematics and thus reveals issues-developmental scope, domain-specific predictors, and equity in a high-stakes subject-that broader syntheses tend to mask. Differences in the literature, especially the unstable ranking of socio-economic versus behavioral predictors, can be most reasonably explained by differences in data sources and construct operationalization rather than by real disagreements about the underlying mechanisms (Gamazo & Martínez-Abad, 2020; Spitzer et al., 2025).

Aligning these results with previous comprehensive studies also highlights what is really new here. Broad reviews of educational data mining and learning analytics have highlighted the popularity of ensemble techniques and the range of data sources, but they generally consider subject domain to be of secondary importance and almost never examine the developmental stage of learners (Christopoulos et al., 2020; Romero & Ventura, 2020). Here the authors limit both domain and stage which leads them to give a different angle and thus revisit well-known findings: the very often mentioned supremacy of tree-based models is verified for school math, however, the authors argue that it is based on a very narrow and mainly cross-sectional evidence; the variety of data sources that is the hallmark of general reviews, is, in K-12 math, largely represented by administrative and assessment records whereas process and multimodal data are still at the stage of being developed. Therefore, the review is not so much discrediting prior conclusions as it is clarifying their limits, thereby showing that statements that are very well supported at the level of the broad field might become tentative when the very significant and developmentally sensitive case of school mathematics is studied independently from other subjects (Gonzalez-Nucamendi et al., 2021).

At least three sources of discontinuity may be noted. First, the work on prediction in primary grades and over time is notably less advanced when compared to secondary cross-sectional studies. Second, very few studies present external validation and subgroup fairness, therefore, the reliability and fairness of models remain doubtful. Third, coupling of process, multimodal, and text data with established mathematical learning constructs is still at an exploratory stage (Moon et al., 2024; Sung & Nathan, 2024; Urrutia & Araya, 2022).

This review has three main limitations. First of all, it is based on data from one source only (Scopus), which is a good one, but there might be some important papers that are published in other databases. Second, the review work is limited to English-language journal articles and review papers only, so there is a risk of introducing bias into the language and publication that way. Third, the set of papers that have been chosen for this review is, on the one hand, small and, on the other hand, very carefully selected, so any statements about topics should be understood as analytic rather than statistical generalisations. The bibliometrics that have been performed in a descriptive way lessen these limitations but do not remove them completely.

The future is concrete and it shows a definite agenda. Researchers must: (1) focus on conducting longitudinal research and primary-grade studies to understand the developmental dynamics of mathematics learning; (2) use interpretable and fairness-aware modelling while requiring external validation of the results across different cohorts and systems; and (3) combine multimodal and textual process data with theoretical constructs to give accurate predictions that can also be used for teaching (Cukurova et al., 2022; Li et al., 2024; Winne, 2021).

Directly addressing the research questions: As far as RQ1 is concerned, learning analytics and educational data mining have mainly been used in K-12 mathematics through tree-based ensemble and boosting techniques which have utilized administrative, large-scale-assessment, and digital-process data, with optimised ensembles being reported by most as the most effective, although no model is superior in all circumstances. As far as RQ2 is concerned, the best and most consistent predictors are prior attainment, socio-economic and school context, and self-regulatory and affective behaviour, with domain-specific cognitive variables being recognised more and more. As far as RQ3 is concerned, the main implications are a move towards explainable, fairness-aware, and pedagogically embedded analytics, but there are still unresolved gaps in developmental scope, validation, and ethical governance.

CONCLUSIONS

This systematic literature review synthesized current evidence on the application of learning analytics (LA) and educational data mining (EDM) for predicting mathematics achievement in K–12 education. The findings indicate that tree-based ensemble algorithms, particularly Random Forest and boosting methods, consistently demonstrate strong predictive performance, although no single model is universally superior across educational contexts. Prior mathematics achievement, socio-economic status, school environment, self-regulated learning, and behavioural data from digital learning environments emerged as the most influential predictors. The review also highlights a growing shift toward explainable, interpretable, and fairness-aware artificial intelligence, emphasizing that predictive models should support transparent and evidence-based educational decision-making rather than focusing solely on prediction accuracy.

Despite its contributions, this review is limited by its reliance on the Scopus database, English-language journal articles, and the methodological heterogeneity of the included studies. Future research should expand database coverage, incorporate longitudinal and multi-country datasets, apply external validation, and investigate explainable and fairness-aware AI models in authentic classroom settings. Strengthening collaboration among educators, researchers, data scientists, and policy makers will be essential to develop predictive learning analytics that are accurate, equitable, and practically useful for improving mathematics learning outcomes.

References

Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. *Technology, Knowledge and Learning*, *24*(4), 567–598. https://doi.org/10.1007/s10758-019-09408-7

Almeda, M. V., & Baker, R. S. (2020). Predicting student participation in STEM careers: The role of affect and engagement during middle school. *Journal of Educational Data Mining*, *12*(2), 33–47. https://doi.org/10.5281/zenodo.4008054

Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: an empirical study in Germany. *Smart Learning Environments*, *9*(1), Article 38. https://doi.org/10.1186/s40561-022-00220-y

Aslam, N., Khan, I. U., Alamri, L. H., & Almuslim, R. S. (2021). An Improved Early Student's Performance Prediction Using Deep Learning. *International Journal of Emerging Technologies in Learning*, *16*(12), 108–122. https://doi.org/10.3991/ijet.v16i12.20699

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, Article 103999. https://doi.org/10.1016/j.compedu.2020.103999

Christopoulos, A., Pellas, N., & Laakso, M. J. (2020). A learning analytics theoretical framework for stem education virtual reality applications. *Education Sciences*, *10*(11), 1–15. https://doi.org/10.3390/educsci10110317

Christou, V., Tsoulos, I., Loupas, V., Tzallas, A. T., Gogos, C., Karvelis, P. S., Antoniadis, N., Glavas, E., & Giannakeas, N. (2023). Performance and early drop prediction for higher education students using machine learning. *Expert Systems with Applications*, *225*, Article 120079. https://doi.org/10.1016/j.eswa.2023.120079

Cirneanu, A. L., & Moldoveanu, C. E. (2024). Use of Digital Technology in Integrated Mathematics Education. *Applied System Innovation*, *7*(4), Article 66. https://doi.org/10.3390/asi7040066

Cukurova, M., Khan-Galaria, M., Millán, E., & Luckin, R. (2022). A Learning Analytics Approach to Monitoring the Quality of Online One-to-One Tutoring. *Journal of Learning Analytics*, *9*(2), 105–120. https://doi.org/10.18608/jla.2022.7411

Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2023). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. *British Journal of Educational Technology*, *54*(1), 313–334. https://doi.org/10.1111/bjet.13290

Fan, Z., Gou, J., & Wang, C. (2023). Predicting secondary school student performance using a double particle swarm optimization-based categorical boosting model. *Engineering Applications of Artificial Intelligence*, *124*, Article 106649. https://doi.org/10.1016/j.engappai.2023.106649

Filiz, E., & Öz, E. (2020). Educational data mining methods for timss 2015 mathematics success: turkey case. *Sigma Journal of Engineering and Natural Sciences*, *38*(2), 963–977.

Gamazo, A., & Martínez-Abad, F. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, *11*, Article 575167. https://doi.org/10.3389/fpsyg.2020.575167

Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., García-Castelán, R. M. G., & Escobar-Castillejos, D. (2021). The prediction of academic performance using engineering student's profiles. *Computers and Electrical Engineering*, *93*, Article 107288. https://doi.org/10.1016/j.compeleceng.2021.107288

Hershkovitz, A., Noster, N., Siller, H. S., & Tabach, M. (2024). Learning analytics in mathematics education: the case of feedback use in a digital classification task on reflective symmetry. *ZDM – Mathematics Education*, *56*(4), 727–739. https://doi.org/10.1007/s11858-024-01551-5

Huang, Y., Zhou, Y., & Wu, D. (2025). Exploring Factors Causing the Mathematics Performance Gaps of Different Genders Using an Explainable Machine Learning. *Computer Applications in Engineering Education*, *33*(3), Article e70014. https://doi.org/10.1002/cae.70014

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning*, *14*(14), 92–104. https://doi.org/10.3991/ijet.v14i14.10310

Keser, S. B., & Aghalarova, S. (2022). HELA: A novel hybrid ensemble learning algorithm for predicting academic performance of students. *Education and Information Technologies*, *27*(4), 4521–4552. https://doi.org/10.1007/s10639-021-10780-0

Knezek, G., Gibson, D., Christensen, R., Trevisan, O., & Carter, M. (2023). Assessing approaches to learning with nonparametric multidimensional scaling. *British Journal of Educational Technology*, *54*(1), 126–141. https://doi.org/10.1111/bjet.13275

Li, K. C., & Wong, B. T. M. (2020). Trends of learning analytics in STE(A)M education: a review of case studies. *Interactive Technology and Smart Education*, *17*(3), 323–335. https://doi.org/10.1108/ITSE-11-2019-0073

Li, C., Xing, W., & Leite, W. (2024). Using fair AI to predict students' math learning outcomes in an online platform. *Interactive Learning Environments*, *32*(3), 1117–1136. https://doi.org/10.1080/10494820.2022.2115076

Lu, O. H. T., Huang, A. Y. Q., & Yang, S. J. H. (2021). Impact of teachers' grading policy on the identification of at-risk students in learning analytics. *Computers & Education*, *163*, Article 104109. https://doi.org/10.1016/j.compedu.2020.104109

Matz, R. L., Mills, M., Derry, H. A., Hayward, B. T., & Hayward, C. (2024). Viewing tailored nudges is correlated with improved mastery-based assessment scores. *British Journal of Educational Technology*, *55*(5), 1841–1859. https://doi.org/10.1111/bjet.13451

Mimis, M., El Hajji, M., Es-saady, Y., Oueld Guejdi, A., Douzi, H., & Mammass, D. (2019). A framework for smart academic guidance using educational data mining. *Education and Information Technologies*, *24*(2), 1379–1393. https://doi.org/10.1007/s10639-018-9838-8

Moon, J., Yeo, S., Banihashem, S. K., & Noroozi, O. (2024). Using multimodal learning analytics as a formative assessment tool: Exploring collaborative dynamics in mathematics teacher education. *Journal of Computer Assisted Learning*, *40*(6), 2753–2771. https://doi.org/10.1111/jcal.13028

Na, H., Staudt Willet, K. B., & Kim, C. (2025). Investigating the impact of AR technologies on geometric learning in primary school: A comparison between marker-based and markerless AR. *British Journal of Educational Technology*, *56*(6), 2502–2521. https://doi.org/10.1111/bjet.13584

Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear Support Vector Machines for Prediction of Student Performance in School-Based Education. *Mathematical Problems in Engineering*, *2020*, Article 4761468. https://doi.org/10.1155/2020/4761468

Naseer, F., & Khawaja, S. (2025). Mitigating Conceptual Learning Gaps in Mixed-Ability Classrooms: A Learning Analytics-Based Evaluation of AI-Driven Adaptive Feedback for Struggling Learners. *Applied Sciences*, *15*(8), Article 4473. https://doi.org/10.3390/app15084473

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher,

D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, Article n71. https://doi.org/10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, Article n160. https://doi.org/10.1136/bmj.n160

Paolucci, C., Vancini, S., Bex II, R. T., Cavanaugh, C., Salama, C., & de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*, *10*(4), Article e25767. https://doi.org/10.1016/j.heliyon.2024.e25767

Ramli, I. S. M., Maat, S. M., & Khalid, F. (2022). Digital Game-based Learning and Learning Analytics in Mathematics. *Pegem Journal of Education and Instruction*, *13*(1), 168–176. https://doi.org/10.47750/pegegog.13.01.19

Reinhold, F., Hoch, S., Schiepe-Tiska, A., Strohmaier, A. R., & Reiss, K. (2021). Motivational and Emotional Orientation, Engagement, and Achievement in Mathematics. A Case Study With One Sixth-Grade Classroom Working With an Electronic Textbook on Fractions. *Frontiers in Education*, *6*, Article 588472. https://doi.org/10.3389/feduc.2021.588472

Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior*, *100*, 345–357. https://doi.org/10.1016/j.chb.2019.07.007

Rodríguez-Martínez, J. A., González-Calero, J. A., del Olmo-Muñoz, J., Arnau, D., & Tirado-Olivares, S. (2023). Building personalised homework from a learning analytics based formative assessment: Effect on fifth-grade students' understanding of fractions. *British Journal of Educational Technology*, *54*(1), 76–97. https://doi.org/10.1111/bjet.13292

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), Article e1355. https://doi.org/10.1002/widm.1355

Roslan, M. H. B., & Chen, C. J. (2023). Predicting students' performance in English and Mathematics using data mining techniques. *Education and Information Technologies*, *28*(2), 1427–1453. https://doi.org/10.1007/s10639-022-11259-2

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, *50*(6), 2943–2958. https://doi.org/10.1111/bjet.12858

Shin, D., & Shim, J. (2021). A Systematic Review on Data Mining for Mathematics and Science Education. *International Journal of Science and Mathematics Education*, *19*(4), 639–659. https://doi.org/10.1007/s10763-020-10085-7

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

Sokkhey, P., & Okazaki, T. (2019). Comparative Study of Prediction Models for High School Student Performance in Mathematics. *IEIE Transactions on Smart Processing and Computing*, *8*(5), 394–404. https://doi.org/10.5573/IEIESPC.2019.8.5.394

Sokkhey, P., Navy, S., Tong, L., & Okazaki, T. (2020). Multi-models of educational data mining for predicting student performance in mathematics: A case study on high schools in cambodia. *IEIE Transactions on Smart Processing and Computing*, *9*(3), 217–229. https://doi.org/10.5573/IEIESPC.2020.9.3.185

Spitzer, M. W. H., Ruiz-Garcia, M., & Moeller, K. (2025). Basic mathematical skills and fraction understanding predict percentage understanding: Evidence from an intelligent tutoring system. *British Journal of Educational Technology*, *56*(3), 1122–1147. https://doi.org/10.1111/bjet.13517

Sung, H., & Nathan, M. J. (2024). Your body tells how you engage in collaboration: Machine-detected body movements as indicators of engagement in collaborative math knowledge building. *British Journal of Educational Technology*, *55*(5), 1950–1973. https://doi.org/10.1111/bjet.13473

Urrutia, F., & Araya, R. (2022). Do Written Responses to Open-Ended Questions on Fourth-Grade Online Formative Assessments in Mathematics Help Predict Scores on End-of-Year Standardized Tests?. *Journal of Intelligence*, *10*(4), Article 82. https://doi.org/10.3390/jintelligence10040082

Winne, P. H. (2021). Open Learner Models Working in Symbiosis With Self-Regulating Learners: A Research Agenda. *International Journal of Artificial Intelligence in Education*, *31*(3), 446–459. https://doi.org/10.1007/s40593-020-00212-4

Yeung, C. K., & Yeung, D. Y. (2019). Incorporating Features Learned by an Enhanced Deep Knowledge Tracing Model for STEM/Non-STEM Job Prediction. *International Journal of Artificial Intelligence in Education*, *29*(3), 317–341. https://doi.org/10.1007/s40593-019-00175-1

Zhang, L., Pan, M., Yu, S., Chen, L., & Zhang, J. (2023). Evaluation of a student-centered online one-to-one tutoring system. *Interactive Learning Environments*, *31*(7), 4251–4269. https://doi.org/10.1080/10494820.2021.1958234

Zheng, J., Xing, W., Zhu, G., Chen, G., Zhao, H., & Xie, C. (2020). Profiling self-regulation behaviors in STEM learning of engineering design. *Computers & Education*, *143*, Article 103669. https://doi.org/10.1016/j.compedu.2019.103669

Zhidkikh, D., Saarela, M., & Kärkkäinen, T. (2023). Measuring self-regulated learning in a junior high school mathematics classroom: Combining aptitude and event measures in digital learning materials. *Journal of Computer Assisted Learning*, *39*(6), 1834–1851. https://doi.org/10.1111/jcal.12842