Ulusoy et al. *Research and Practice in Technology Enhanced Learning* (2027) 22:12
https://doi.org/10.58459/rptel.2027.22012

Research and Practice in
Technology Enhanced Learning

RESEARCH **Free and Open Access**

Learning analytics with multimodal data through the lens of AI in education: A systematic literature review

Onuralp Ulusoy *, Aysu Ismayilova and Johan Jeuring

*Correspondence:
o.ulusoy@uu.nl
Department of Information and Computing Sciences, Utrecht University; National Education Lab Artificial Intelligence (NOLAI), Radboud University, the Netherlands
Full list of author information is available at the end of the article

Abstract

Learning analytics (LA) collects, processes, analyzes and displays various types of learner data through various means, to give insights that support learning. Combining different data sources with possibly different modalities can further enhance these insights. This field of research is called multimodal learning analytics (MMLA). The availability of substantial amounts of data opens up possibilities for AI to further analyze MMLA data and use these analyses to help learners. In this paper, we systematically collect and analyze MMLA studies that make use of AI technologies to tackle research questions related to improving learning processes. Our systematic literature review gives an overview of the current state of the art of MMLA, provides results about the types of research questions, data types and how they are integrated, their target audiences, and the AI algorithms and methods used. We finish with providing insights about the current challenges in the field and possible future directions.

Keywords: learning analytics, literature review, multimodal, artificial intelligence

Introduction

Learning analytics (LA) is the field where various types of learner data are measured, collected, processed, analyzed, and displayed to understand the learning process and optimize it for efficient learning (Siemens, 2011). It is a very active research area with a dedicated conference and journal, and related publications in many other venues in technology-enhanced learning and AI in education. Learning analytics can be employed to model students based on their activities (Su et al., 2018), predict student success (Gardner & Brooks, 2018), measure emotions or cognitive capabilities of students (Ninaus et al., 2019), offer guidance to students in their learning (Sun et al., 2018), support teachers in observing the learning progress of students and intervene if needed (Xhakaj et al., 2017) and many other ways to create a suitable learning path for a student.

Using digital systems in education makes it possible to collect data from several sources, such as answers to online exams and interactions in online courses. This data can be used to improve many aspects of education. For example, teachers can use the data to adapt their courses (Supovitz & Klein, 2003), or obtain information about a student's learning progress (Mazza & Dimitrova, 2004). For students, the collected data can be used to present or recommend suitable activities to optimize their learning, paving the way for potentially improved self-regulated learning (Viberg et al., 2020).

Using various data sources with different modalities might give better insights into student learning progress and enable better support for learning goals of students (Blikstein, 2013). This brings out the notion of multimodality, where multiple modes contribute to an understanding of the meaning of data in some context. According to Kress (2009), a mode is "something that can be socially or culturally shaped to give something meaning". Speech, visuals, pieces of text, and sensory information of humans are all examples of different modes, and multimodality assumes the use of more than a single mode. Various modes can be captured with different devices in an educational setting. For example, a video recording can give visual cues about a student or a teacher, while an audio recording can capture discussions between students. Measuring facial expressions or detecting gaze of a student can give insights about the engagement level or emotional state of the student. The interactions in a digital learning environment can give information about a variety of aspects of a student's learning progress, such as their success rate in solving exercises, the time they spend on an activity and so on. All these examples show that multimodality can play an important part in learning analytics. To play this part, some challenges still need to be overcome. For example, Sharma and Giannakos show that there are challenges in collecting, synchronizing, and analyzing multimodal data (Sharma and Giannakos, 2020). First, aligning data streams with different sampling rates requires careful resampling and synchronization (Ochoa et al., 2018; Worsley and Martinez-Maldonado, 2018). Each modality contributes distinct feature types (e.g., facial emotion, gaze-based attention, EEG-derived workload), complicating feature engineering and fusion. Even after cleaning and synchronizing these streams, translating multimodal features into actionable, learning-analytics–specific guidelines remains challenging (Bakharia et al., 2016; Giannakos et al., 2019).

Artificial Intelligence (AI) has made big leaps in the last decades with contributions in the field of machine learning, language generation, and more. It is used in many daily life applications, such as recommender systems (Afsar et al., 2022; Ko et al., 2022; Wu et al., 2022), self-driving cars (Gupta et al., 2021; Ni et al., 2020), and text generation using generative AI systems (Baidoo-Anu & Ansah, 2023; Bandi et al., 2023; Brynjolfsson et al., 2023). The technical advancements also enable supporting many educational activities. AI supports education at various levels, including Massive open online courses (MOOCs) (Yu et al., 2017), higher education (Crompton & Burke, 2023), but also the early stages of education for younger students (Zafari et al., 2022). The use of digital data and AI in education raises questions about ethical values, which should be carefully considered and monitored (Zhang & Aslan, 2021). Ethical questions also open new grounds for AI in education, which should always consider values such as privacy, accountability, fairness, and transparency.

The availability of sometimes substantial amounts data opens up possibilities for AI to further analyze MMLA data and use these analyses to help learners. The goal of this paper is to systematically collect and analyze research on the use of multimodal data in learning analytics, through the lens of AI in education. We select papers that employ some AI technology to study the use of multimodal data for learning analytics. Using multimodal data for educational purposes has not been as widely investigated as many other applications of AI in education. There are few real-life cases of using multimodal data for learning analytics, and there are still many open research questions in the field. With this review we want to help researchers by showing what progress has already been made, and point out future directions to investigate the use of multimodality in learning analytics. We survey the state of the art for multimodal data learning analytics for all levels of education, including pre-schools, primary and secondary education (K-12), special education, as well as higher education and online learning environments. The main contribution of this paper is an investigation of the intersection between MMLA and AI techniques for all educational levels and a description of the current progress, limitations and future challenges. To our knowledge this has not been the main focus yet of a systematic review. The main research questions we address are:

- What is the state of the art in the education domain of learning analytics based on multiple data sources with different modalities, employing AI techniques?
- What type of AI Algorithms and statistical methods are applied in the use cases and methods to answer the research questions?
- What are the challenges and future directions for the field and how can these be used as guidance for future research on employing multimodality in learning analytics?

The first research question investigates the main research goals of the surveyed work and categorizes these to determine the areas in which multimodal data in learning analytics is used. Furthermore, we investigate which types of data are used, the purpose of multimodal data usage, and the target audience of the reviewed research. The second question delves into what AI algorithms or statistical methods applications of multimodal data use. To answer this research question we categorize the AI algorithms and the statistical methods used, and describe for what purpose they are used, such as classification of data or calculating prediction metrics. The third and final research question aims to offer some pointers to future research directions, reporting the future work and the limitations mentioned in the reviewed publications. Moreover, we also discuss our own observations about the challenges of the field and our views on how they could be overcome.

The rest of this paper is organized as follows. *Methodology* section introduces the method we use for our systematic literature review. *Related Work* section provides information about previous research surveying learning analytics methods using multimodal data, and discusses similarities and differences with our work. *Results* section gives the results of analyzing the surveyed work. *Discussion* session summarizes our findings and discusses how to interpret our results and lessons learned. *Conclusion* section concludes the paper with a summary and gives remarks and recommendations about future work on learning analytics and learner modelling in the context of multimodal data.

Methodology

This section outlines the method we use for our systematic literature review. We focus on the use of multimodal data in learning analytics and learner models through the lens of AI in education. This study aims to answer the research questions mentioned in the introduction section to analyze the latest research on learning analytics and learner models in education based on multiple data sources.

Systematic Literature Review Process

We follow the guidelines from Kitchenham and Charters (2007), which consist of several steps for conducting systematic literature reviews.

The five-step systematic literature review process starts with formulating the research questions, which we have done in the introduction section. Next, we select databases to search for relevant papers, and draw up queries to select papers from these databases. Subsequently, we define the criteria for including or excluding papers in our systematic review. We screen the papers we obtain through our queries to ensure they meet the inclusion criteria. Finally, we analyze and synthesize the findings from the selected papers to answer the research questions.

Literature Search Implementation and Results

We select search terms based on our research questions. We use the following search terms: “Education” AND (“learner models” OR “learning analytics”) AND (“multimodal data” OR “multimodal learning”) AND (“AI” OR “Artificial Intelligence”). Table 1 gives the rationale for selecting these terms.

Table 1

Keywords and rationale

Keyword	Rationale for Selection
Education	Ensures that all studies focus on education.
Learner Models	Captures studies focusing on computational representations of individual learners' knowledge, skills, and characteristics. This enables us to examine research on how these models are developed and applied to create adaptive and personalized learning experiences.
Multimodal Data	Identifies research that utilizes data-driven methods to analyze and optimize educational processes. This allows us to explore how educational data is being used to gain insights into learning behaviors and outcomes.
Multimodal Learning	Relates to learning that involves multiple modalities, which is increasingly common in digital education and addresses diverse learning needs.
AI, Artificial Intelligence	Explores the integration of artificial intelligence technologies and methodologies in educational contexts. This allows us to examine how AI is influencing and potentially transforming various aspects of education.

We conducted the search of academic databases up to December 31, 2024. The research papers were collected from Scopus, Web of Science, ACM, IEEEXplore and Springer. Scopus and Web of Science are large article databases that include content from many disciplines, which guarantees breadth in our search. To add depth, we include the ACM, IEEE Xplore, and Springer collections, which are the leading platforms for computer science and engineering. For example, the proceedings of all leading conferences on learning analytics (LAK, EDM, AIED, ECTEL) appear on these platforms.

This search resulted in a total of 1969 articles, after removing duplicates. We found 1111 journal articles, 745 conference papers, and 113 book chapters. We experimented with several alternative keywords to determine their effect on the results. Changing "learner models" to "student models" resulted in four additional papers, while modifying "learner models" to "learner modeling" yielded only one new paper. By including a broader set of keywords—specifically "student modeling," "learner modeling," and "student models" to our initial set of keywords—we identified an additional 41 papers on top of the 1969 papers obtained from the previous keywords. Therefore, our findings indicate that these changes do not significantly impact the overall results, suggesting that variations in terminology have a limited effect on the number of relevant studies identified. We selected papers without specifying an initial date. We did not set a lower-bound year for the search in order to ensure a comprehensive overview of the field. In practice, no relevant studies were found prior to 2012. This is consistent with the development of the field of learning analytics, which only began to be established as a distinct research domain around 2011 with the emergence of dedicated conferences and formal definitions (Guan et al., 2020). The studies retrieved from the databases were uploaded to Rayyan QCRI, an AI-based online platform that helps to screen abstracts and titles by using a process of semi-automation including machine learning for screening, natural language processing for text analysis, and support vector machines for classification (Ouzzani et al., 2016).

Inclusion and Exclusion Criteria

To include or exclude studies, we use the inclusion and exclusion criteria in Table 2 and Table 3, respectively. We only check the exclusion criteria for papers that satisfy the inclusion criteria.

Table 2

Inclusion Criteria

Criteria	Description
Publication Domain	The study focuses on educational contexts such as K-12, higher education, or massive open online courses (MOOCs).
Publication Type	The study is peer-reviewed and published in a journal or a conference proceedings.
Data Source	The study uses multimodal data.
Implementation Details	The study includes details about systems or methodologies implemented using/combining multiple data sources.

Table 3

Exclusion Criteria

Criteria	Description
Short Papers	Papers shorter than five pages are excluded to ensure sufficient detailed information and depth.
Literature Reviews	Literature reviews and surveys are excluded.
Language	The study is not written in English.

Screening Process

Two of the authors selected the papers. They independently reviewed the relevance of all of the 1969 studies. The paper selection method involved two phases. First, we evaluated titles, abstracts, and keywords of each publication to determine its relevance. We excluded publications falling outside the predefined inclusion criteria. For the majority of the papers, we were able to decide to include or exclude them based on the information in the abstract. If the abstract did not have enough information to make this decision, we assigned the paper a provisional "Maybe" status, indicating the need for a more comprehensive examination of the full text.

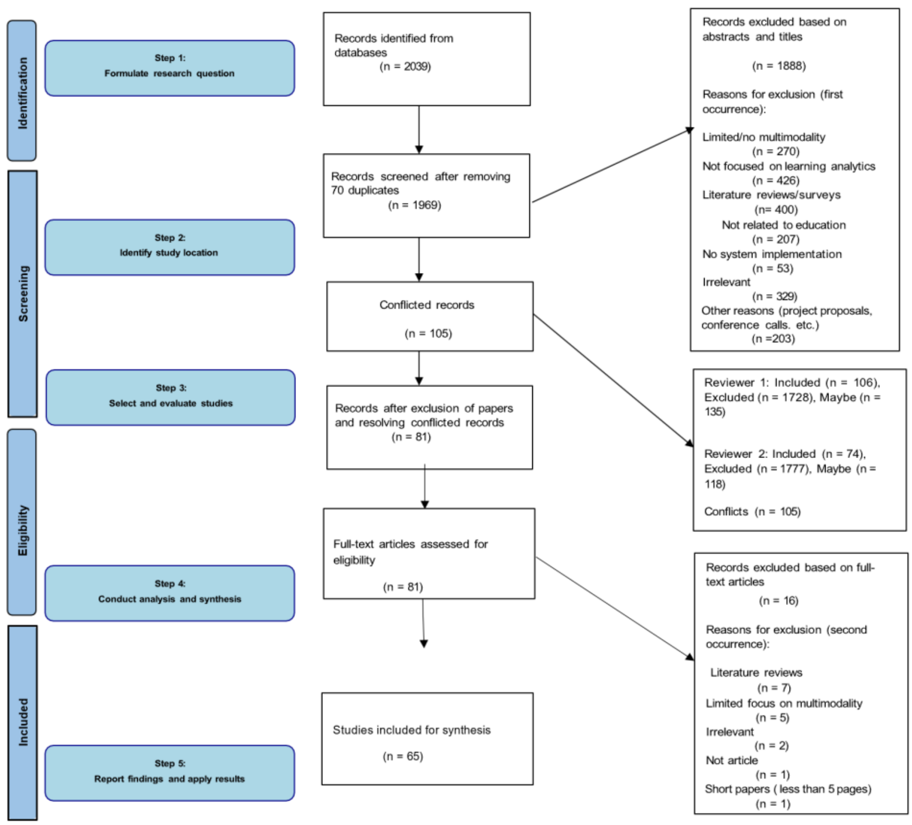
In the second stage, the same researchers independently reviewed the full texts of the publications that passed the first screening. The same inclusion and exclusion criteria were applied to ensure each study met the requirements. During this process, we employed specific keywords to categorize papers, facilitating discussions regarding the reasons behind the exclusion of certain papers. We used Cohen's Kappa to measure the agreement between the researchers (Kitchenham & Charters, 2007) and obtained a value of 0.9, which indicates an almost perfect agreement according to Landis and Koch (1977).

Results of Screening

We excluded 1888 papers in the initial screening phase. From these excluded studies, 270 did not use multimodality or used it in a very limited fashion. 426 studies did not have learner modelling as a primary focus. Additionally, we identified 400 studies as literature reviews or surveys, and 207 unrelated to education. 53 studies lacked any description of a system implementation, and 329 papers were classified as irrelevant due to their lack of both AI implementation and multimodality. 203 of the studies were excluded for other reasons, such as being project proposals or conference calls, having an extended follow-up study that is already screened etc. We disagreed on whether or not to include 105 papers. We defined these disagreements as "conflicts", which occurred when one reviewer marked a paper for inclusion with a "Yes", while another reviewer marked it as "No" or "Maybe". We discussed all papers labelled with a conflict. As a result, we reached consensus on including in total 81 papers. We conducted a full text review with respect to the inclusion and exclusion criteria for these 81 papers. This led to discarding another 16 papers. We ended up with 65 papers for the final review, see Figure 1.

Figure 1

PRISMA flow diagram and five steps of the review process (Page et al., 2021)



Data Coding Process

Data coding is the process of labeling different pieces of information from the articles with a word or short phrase that summarizes its content based on predefined criteria or categories.

In this systematic literature review, we describe each paper using a structured set of metadata: a unique identifier, the title, the authors, and a label indicating the publication type. To address our research questions, we code the following aspects of the papers we review:

- **Year of Publication:** To analyze temporal trends and the evolution of research in the field.
- **Main Research Question:** To understand the core focus of research of each study.
- **Data Types Used:** Classifies the empirical data utilized in the research.
- **Target Group:** Identifies the population or user group that is the subject of the research.
- **Multimodal Data Sources:** To indicate whether and how the study integrates different data types from a single source or from multiple sources.

- **Type of AI Algorithms Employed:** To identify the AI algorithms or statistical methods applied in the study.
- **Use Cases of AI Algorithms:** To identify the specific purposes for which the AI algorithms are employed within the studies.
- **Challenges and Limitations:** Summarizes the main obstacles or constraints reported in the studies.
- **Future Research Directions:** Captures the research gaps and opportunities for further work identified by the authors.

For categories such as data types, multimodal data sources, and type of AI algorithms employed, we utilize deductive coding based on the coding introduced by Samuelsen et al. (2019). For research questions, target groups, limitations, and future research directions, we employed inductive coding conducted by two researchers. When screening the retrieved papers, two authors created a coding for the different kinds of research questions. Then 15 out of 65 selected papers were coded separately by two researchers, and the inter-coder reliability was again measured with Cohen's Kappa, which was 0.9012. Because of the satisfactory inter-coder reliability, the rest of the publications were divided between two researchers and coded by a single researcher.

Related Work

With new technological advancements enabling easier data collection using various devices or applications, the use of multimodal data in education and learning analytics is becoming a topic of interest for researchers. In the last few years, there have been some studies that investigate and review several aspects of learning analytics with multimodal data, each focusing on different perspectives from multiple disciplines. In this section, we briefly discuss the findings of some of these papers, while comparing the similarities and differences with what we aim to investigate in this paper.

Samuelsen et al. (2019) has similar research goals as our work. It systematically reviews the studies that focus on integrating multiple data sources for learning analytics in the context of higher education. The study investigates the data types, data integration approaches, the research questions and the methods used in the reviewed studies. The authors review 20 studies, and conclude that most of the studies lack details about technical solutions, do not involve stakeholders and miss educational data specifications. They mostly combine similar data formats, which is an easier task than combining data with different modalities. The main difference between our work and Samuelsen et al. (2019) is that we review studies from all educational levels instead of just higher education, and we include online learning environments and special education. Our focus is on the use of AI methods for learning analytics and learner models with multimodal data, instead of the data integration, which Samuelsen et al. (2019) investigate.

Lee et al. (2023) is a scoping review that investigates the role of multimodal learning systems in technology enhanced learning. Their research goals align with our review for investigating collected data types and their purposes. However, they focus specifically on using the collected data to give feedback to learners and how multimodality can be integrated into the design of learning systems. While the data typology and their research purposes are thoroughly reviewed, the authors do not focus on learning analytics and the application of AI in multimodal studies, which is what we focus on in this paper.

Cukurova et al. (2020) describe the potential of learning analytics with multimodal data, and formulate challenges for future research. The authors claim that multimodality can lead to interdisciplinary collaborations that combine fields such as learning sciences, affective computing and human-computer interaction to reach better prediction, analysis and learner support. They state the main challenges about practicality (e.g., the participation of children, parental involvement, potential surveillance issues) and the need to rely on grounded methodology (e.g., ensuring high data quality, ecologically valid studies). The authors explain that these challenges are rooted from the research, development, and deployment phases taking place in a moral vacuum. The notion of moral vacuum is explained in the paper as very limited groundwork, no clear guidelines and no regulations during all the mentioned phases. The authors also investigate the major ethical concerns around data collection, analysis and processing. Worsley and Martinez-Maldonado (2018) also study multimodal learning analytics, discussing the past, present and future of MMLA by reviewing the literature. The authors note increasing accessibility, employing deep learning and simplifying data collection as the main challenges for future research. In follow up work, Worsley et al. (2021) articulate twelve commitments that researchers should adhere to while doing research on MMLA. These commitments address challenges in data collection, data analysis and inference, meaningful feedback and dissemination of the resulting data.

Some researchers focus on specific aspects of MMLA and learner models. Chango et al. (2022) review data fusion studies that use multimodal data for learning analytics and combine data of different modalities for a goal. They categorize the studies according to their fusion point, techniques and algorithms used and their learning analytics objectives. The fusion point is defined by the authors as "the moment in which the fusion is done" and categorized as early, late and hybrid, which is a mix of both early and late fusion. Mu et al. (2020) also provide a systematic literature review for data fusion for MMLA. The authors provide a categorization of multimodal data, in relation with learning indicators. They also show what types of data are integrated in what manner, such as many-to-one, many-to-many and mutual verification (i.e., the verification of the same results by different types of learning data) between separate multimodal data streams. Sharma and Giannakos (2020) conduct a systematic literature review to investigate capabilities of multimodal data for human learning, and identify six main objectives for MMLA research: behavioral trajectories, learning outcomes, learning-task performance, teacher support, engagement and student feedback. They provide a data type categorization, as well as the number of modalities observed in each surveyed work. The authors also associate research objectives with the modalities used in the papers. Noroozi et al. (2020) provide a systematic literature review that categorizes papers that use multimodal data, and they investigate the relation between different modalities and their abilities to capture emotional, cognitive and motivational learning processes.

Since learning analytics with multimodal data requires collecting data from multiple sources related to learning activities, further to be processed, analyzed and employed for some research goal, it sometimes raises ethical questions. Crescenzi-Lanna (2020) conducts a systematic review of the use of MMLA with young children. The review investigates the purpose of use for various data types in the studies that collect multimodal data from young children. The authors investigate ethical questions when discussing what kind of approaches to and what types of data in learning analytics are more suitable for young children. Yan et al. (2022) systematically review studies on the methodological, practical, and ethical challenges for the scalability and sustainability of MMLA. The challenges discussed include lack of replicability, using small samples, low technological readiness, limited user interfaces, privacy, bias and equality issues in the surveyed work. Alwahaby et al. (2022) also systematically review ethical considerations for MMLA, categorising ethical issues. The authors conclude that the majority of the studies address very few ethical challenges such as privacy or bias.

We found several review and overview papers that investigate various aspects related to MMLA. Monsalves et al. (2023) study how Artificial Intelligence as a Service (AIaaS) could be integrated into MMLA applications and research. The authors define eight points where AIaaS can support and enhance MMLA applications, such as analysis of social and collaborative interactions, identification of success factors and prediction of outcomes and analysis of behavior and engagement. Their findings align with what we found in this review for the research goals within the field of MMLA. Martinez-Maldonado et al. (2023) provide a longitudinal study that employed MMLA and give an overview of the challenges faced within the scope of logistics, privacy and ethics. Some of the stated shortcomings are in line with what we report as limitations of the reviewed studies in this paper, such as the difficulties in obtaining a variety of modalities, and collecting trustworthy data continuously and sustainably. Chejara et al. (2023c) describes another similar longitudinal study and reaches similar conclusions about the limitations of MMLA in real life applications. Basystiuk et al. (2023) presents a brief overview of how a data collection methodology should be designed to use MMLA to provide support for teachers and learners. They define a layered architecture that could be adopted by future studies which aim to have the MMLA research to end up in classroom use. Ouhaichi et al. (2023) and Pei et al. (2023) are two mapping studies that investigate the field of MMLA through keyword analysis of the publications to find out the trends in MMLA research. Ouhaichi et al. (2023) identify four research themes covering all research topics, namely learning context, learning process, systems and modalities and technologies. The authors also state that AI is the most exploited technology in MMLA research, which indicates our focus on AI in this paper is addressing an essential theme for MMLA. Pei et al. (2023) investigate keyword frequencies in MMLA research and point out common occurrences of keywords such as "Performance Prediction", “Collaborative Learning” and “Machine Learning Related Approaches”, which align with our findings after thoroughly reviewing the selected studies in the scope of this paper.

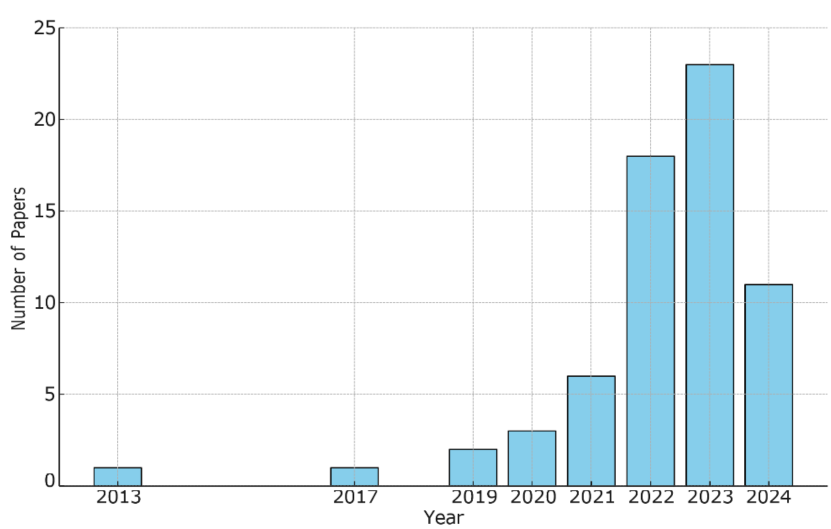
All these reviews tackle important aspects that are related to MMLA. With the research goals of our systematic literature review, we aim to study the educational research with a focus on data typology, the relations between different kinds of data within applications, the research questions, the target audiences and the use of AI related algorithms and methods. To our knowledge, these topics have not been the main focus of a systematic literature review at all education levels before. Furthermore, we investigate the stated challenges and future directions in the studies, and provide guidance on future MMLA related research focusing on data.

Results

Based on our review of publication trends up to December 2024, the field is clearly growing, see Figure 2.

Figure 2

Number of Papers Included in the Review Published by Year



As part of our systematic coding process for this literature review, we present the key findings from the complete coding (see Availability of data and materials in Appendix). These tables summarize important aspects of the reviewed studies, including the data types analyzed, research objectives, education levels targeted, and AI algorithms employed. For a comprehensive explanation of the abbreviations used in the coding scheme tables, see Abbreviations in Appendix.

Research Questions of the Surveyed Studies

Researchers investigate a variety of educational research questions based on learning analytics using multimodal data. We categorize the topics of the main research questions of the 65 studies we have reviewed, to investigate which research topics are tackled most. The categories are obtained using an inductive coding approach, in which we label each paper with the research topics mentioned. Research questions that appear infrequently (encountered in two studies or less) and share similar research goals and methods with another category that is more frequently encountered, are categorized under the latter category. For example, only one study investigates Socially Shared Regulation of Learning, but the study shares research goals with Collaborative Learning, and is included in the latter category. Table 4 shows the categories and the number of papers in these categories. Some studies answer research questions about more than a single topic, so the sum of the frequencies is higher than the total number of papers.

Table 4

Main research questions of the surveyed studies and their frequencies.

Research Question	Frequency
Collaborative Learning	27
Emotion/Engagement Detection/Analysis	16
Performance Prediction	13
Student Feedback/Support	7
Student Modeling	7
Self-regulated Learning	6
Human Computer/Robot Interaction (HCI/HRI)	6
Game-based/Problem-based Learning	4
Data Visualization/Dashboards	4
Collaborative Problem Solving	3
Intelligent Tutoring Systems	1

The results in Table 4 show that the most frequently encountered topic on which research questions are answered is Collaborative Learning. Multiple studies conduct user experiments with groups of learners collaborating to solve learning tasks, collecting data from multiple sources with various modalities. Another common research question deals with performance prediction, where multimodal data collected from learner activities are used for predicting success or failure in future endeavours. Detection of emotion or engagement of learners is another theme in learning analytics that makes use of multimodal data, while some studies also further analyze the outcomes to find correlations between learning progress and detected emotions and/or engagement levels of the learners. Student feedback and support also is a point of interest, where information obtained with multimodal data is used to help learners in current or future learning tasks. Although there are some studies that aim to enhance the learning process of individual learners using multimodal data, fewer studies focus on research questions like self-regulated learning, human-computer interaction or human-robot interaction. We encountered data visualization and dashboards as end products to provide fine-tuned analysis and information to students or teachers in a limited number of studies.

Data Types and Data Integration

Since learning analytics research with multimodal data requires data of different modalities to be collected, the reviewed studies contain a variety of data types that are employed for the research goals addressed. We start with the categories of data types used by Samuelsen et al. (2019) and Molenaar et al. (2023), and introduce a new category whenever we encounter a category that is not present in their work. Table 5 shows the frequencies of the data types encountered in the reviewed studies. Similar to the frequencies of research questions in Table 4, the sum of the frequencies is higher than the number of surveyed studies, since each study employs more than one data type as they are all about multimodal data.

Table 5

Data types used in the surveyed studies and their frequencies.

Research Question	Frequency
Audio Data	25
Video Data	24
Eye Tracking	24
Log Data	22
Electrodermal Activity (EDA) Data	16
Face Tracking	15
Questionnaire	13
Motion Data	12
Text Interaction Data	11
Learning Management System (LMS) Data	10
Screen Recordings	8
Electroencephalogram (EEG) Data	5
Controller Data	3
Self Reports	1
Environmental Data	1

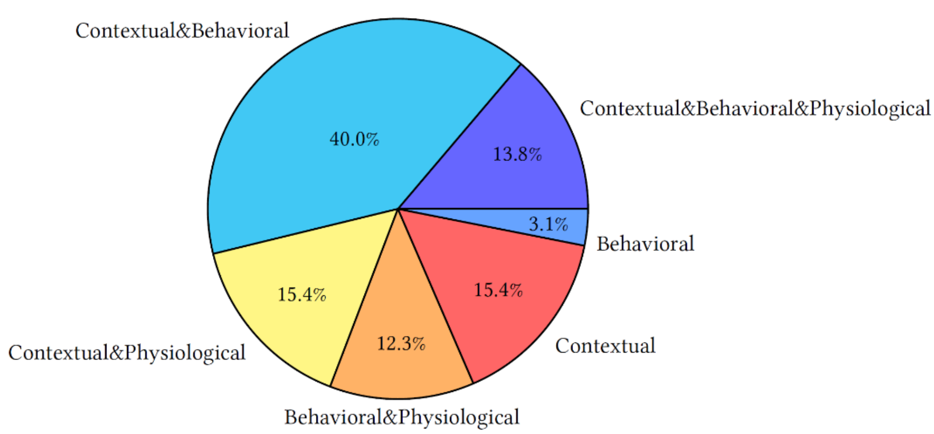
To further analyze the data types, we cluster them into three main categories, namely contextual, behavioral and physiological. This categorization follows Molenaar et al. (2023), in which the authors categorize data types used in self-regulated learning processes using multimodal data. Contextual data types are data types that are directly related to the context of the study. The information contained in data from these types comes from learner or teacher actions that require some deliberation by the learners or teachers, and knowledge of the context that is essential for taking these actions. An example of contextual data is video data in which the recorded subjects perform actions while working on a learning task. Another example is a questionnaire, where subjects answer questions to provide information for a particular educational or research goal. Behavioral data types contain data consisting of actions by people that are not explicitly tied to the context of the study. A prime example for behavioral data is eye tracking, with which eye movements, focal points or pupil dilation of a participant in an experiment is recorded. The data types in this category are related to the physical actions of learners or teachers that do not directly affect the context of the study. The third type of data is physiological data, which includes data such as Electrodermal Activity (EDA) Data and Electroencephalogram (EEG) Data, which often record some bodily reaction with a wearable device of participants in an experiment. We also include environmental data in this category, which consists of data that indirectly has an effect on the physiology of the participants in an experiment, such as the room temperature or the brightness of the location. The various kinds of data categorized under these main categories are given below, followed by an example case that describes these categories in a real-life setting.

- **Contextual data:** Video Data, Audio Data, LMS Data, Screen Recordings, Text Interaction Data, Questionnaire, Self Reports
- **Behavioral data:** Log Data, Eye Tracking, Face Tracking, Motion Data, Controller Data
- **Physiological data:** EDA, EEG, Environmental Data

Example 4.1. Tang et al. (2022) use multimodal data to investigate learner attention states when collaboratively solving problems in an online learning environment. In an experiment with higher-education students in a computer science course they collect data of various modalities. They collect students' prior knowledge using questionnaires, and their knowledge after the course using a post test. During the course, the authors collect screen recordings to investigate the actions of the students. All these data types describe contextual data, since they are directly related to the context of the study. The authors also use video recordings of the students to track their motions during the experiment, which is behavioral data. Moreover, electro-encephalogram (EEG) data of the students are collected with a headset, which measures brainwave activity to capture attention levels. Since this data type is not related to the context of the course, nor to any intentional behavior of the students, it is physiological data.

Figure 3

Data Type Combinations in the Reviewed Studies



Using the categorization from Molenaar et al. (2023), we investigate which types of data are combined in each study, and show the frequencies of these combinations in Figure 3. The results show that the most frequent combination of data types are from contextual and behavioral data (40.0%). This is followed up by the combinations of contextual and physiological data types (15.4%) and only contextual data types (15.4%). We consider the combination of only contextual data types in these results still multimodal data, since although all data in such studies is contextual, there are still different modes collected, such as a combination of questionnaires and text interaction data. The combination of contextual, behavioral and physiological data in the same study is the fourth most frequent in the reviewed studies (13.8%), with 9 out of the 65 reviewed studies. The remaining data type combinations occur in comparatively few studies: behavioral and physiological (12.3%) and only behavioral (3.1%). We did not encounter combinations of different types of just physiological data.

We have also investigated the multimodality aspect of the reviewed studies, according to the amount of sources the data is collected from. Because multimodal data requires data with different modalities, all studies have multiple data types. However, it is possible to collect different types of data from a single source, such as a dataset containing multiple types of data, or a learning management system collecting and storing more than one type of data. Therefore, we categorized the studies we review as single source and multiple source studies. This categorization aims to inspect how many of the studies collect and combine data from multiple sources, which might require merging data for further use. Only 8 out of 65 studies collect multimodal data from a single source, while the rest uses multiple sources. Although these studies make use of multiple data sources, most of them utilize collected data in the learning process and do not merge data from different sources to store for further use. In a majority of the studies, separate data sources are processed for specific parts, such as defining components of a student model, or setting up parameters for a machine learning algorithm. We encountered one study developing an architecture to collect, process and store data in a learning record store to infer additional information from the merged data (Di Mitri et al., 2017). The results show that methodologies for merging data for further use are still not widely applied in MMLA studies. This might be an interesting direction for future research.

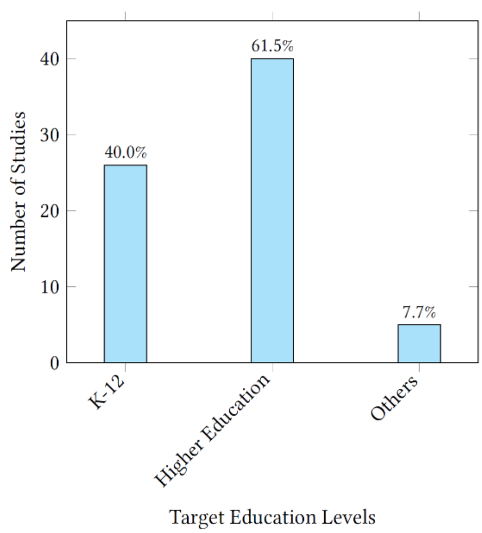
Target Audiences

Research Questions of the Surveyed Studies subsection shows that learning analytics using multimodal data in education can be employed to tackle a variety of research questions. The studies target a variety of audiences, learners at different levels of education, or different groups with varying roles in an educational system. This section reports on which education level and which group(s) the 65 studies we review target. We divide the education levels into three main categories: K-12, Higher Education and Others. These categories cover all the education levels from kindergarten up to higher education. K-12 has the following subcategories: primary education, secondary education and vocational schools. The others category contains special education, pre-school education and online learning that does not target a specific education level, while if specified online learning studies are included in their related categories.

Figure 4 shows the frequencies for the target education levels of the reviewed studies. Higher education is studied most, in 40 of the 65 studies (61.5%). K-12 is the target education level for 26 out of 65 of the reviewed studies (40.0%). When we break down K-12 into subcategories, 6 of the studies focus on primary education, 9 on secondary education (high schools or pre-high & post-primary education) and 2 on vocational schools, while 9 studies either target both primary and secondary education or they don't explicitly state a subcategory of K-12. Since some studies target more than one of these categories, the sum of the percentages in Figure 4 does not add up to 100%.

Figure 4

Target Education Levels of the Reviewed Studies

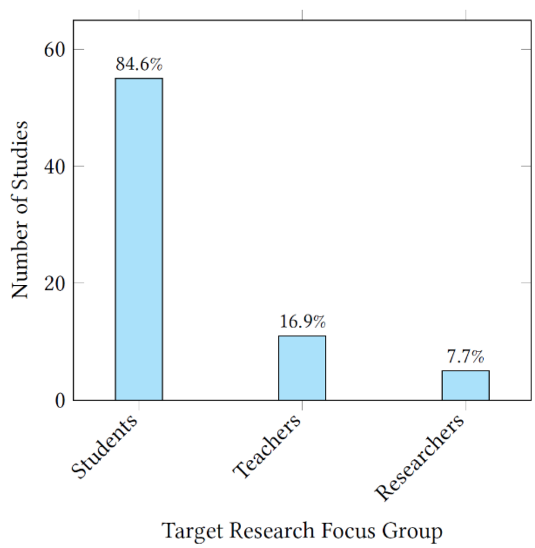


The second categorization of the target audiences is obtained by looking at the group(s) to which the research participants belong. We use three group categories: students, teachers and researchers. The student category is used for studies that have an output that can be used by a student to improve an aspect in their learning process. Examples of such outputs are support for self-regulated learning, or feedback for students. The teacher category is used to label studies that have outputs that target teachers. Examples of such outputs are performance prediction and data visualization, where teachers can track the learning progress of their pupils or predict future success or failure. The third category is researchers. This category is used to label studies that do not specify outputs to be used in the classroom or other learning environments, but rather showcase some results for future research. The studies in this category provide some algorithmic results or findings of user experiments that could be beneficial for other researchers, such as providing an optimal parameter setup for a machine learning task, or identifying the most suitable data types to employ for a specific research topic.

Figure 5 depicts the number of studies categorized according to their group targets. Similar to Fig. 4, the sum of the percentages does not add up to 100% since some studies have multiple groups as their research target. With 55 studies out of 65 (84.6%), the majority of the studies focus on student related research goals. Only 11 studies focus on teachers (16.9%), and we can imagine that future MMLA related research would look more at supporting teachers. 5 studies target researchers (7.7%), where the algorithmic results and the outcomes can be used by other researchers in their future work.

Figure 5

Target Research Focus Group of the Reviewed Studies



AI algorithms/methods used in the studies

Out of the 65 papers reviewed, 59 mention the AI algorithms they use. The remaining six studies use AI systems such as computer vision for position estimation, and electrodermal activity, without mentioning any specific algorithm.

We categorise AI algorithms into two groups: Machine Learning (ML) algorithms and Statistical methods. Among the reviewed studies, the most commonly used Machine Learning algorithms are Random Forest (RF), Decision Trees, Support Vector Machines (SVM), Long Short-Term Memory networks (LSTM), Convolutional Neural Networks (CNN), and a hybrid method combining LSTM and CNN (LSTM + CNN). For statistical methods, the most commonly used techniques include the ANOVA test, Epistemic Network Analysis (ENA), Hidden Markov Models (HMMs), and K-means clustering.

Some studies use multiple algorithms in their research. See Table 6 for a summary of the usage percentages of various machine learning algorithms and statistical methods across the reviewed studies. The percentages are calculated based on the total number of studies that employ AI algorithms (n = 59). Since some studies apply both methods, the percentages do not sum up to 100%.

For instance, Chejara et al. (2023a) use RF classifiers to estimate collaboration quality and its dimensions (e.g., argumentation, structuring problem-solving, time management, knowledge exchange, and collaboration flow) using three audio-log multimodal datasets. Giannakos et al. (2019) use the RF algorithm to predict learning performance in learner–computer interaction. Some authors compare various machine learning algorithms to determine which algorithm performs best for a specific research purpose. For instance, Yusuf et al. (2024) compare four supervised machine learning algorithms, namely multiple linear regression, RF, Naïve Bayes, and SVM to predict students' academic achievement based on their learning behavior. Similarly, Sharma et al. (2020) predict students' effort for the next task by using behavioral patterns identified with HMMs and the Viterbi algorithm. The authors compare these predictions with results from machine learning algorithms such as SVM and RF. Tato and Nkambou (2023) introduce Us-DMN, a new user-sensitive deep multi-modal network model for adaptive platforms like Intelligent Tutoring Systems and serious games. They evaluate its performance against traditional methods for opinion polarity detection (i.e., Positive, Negative, Neutral), including a basic LSTM model, a basic CNN model, a CNN model incorporating additional user data (CNN+User_data), and a hybrid model combining LSTM and CNN (CNN+LSTM). The authors found that the Us-DMN model performs better than other models, showing higher accuracy and greater stability. It also generalizes well with limited data and converges quickly, although it takes longer to train due to the complexity.

The LSTM algorithm has been predominantly used for prediction tasks. For instance, Järvelä et al. (2023) use LSTM neural networks to automatically predict regulatory patterns, for socially shared regulation, in order to optimize collaborative learning. Similarly, Olsen et al. (2020) employ LSTM to predict test scores and normalized learning gains. CNNs are mostly used for classification tasks. For example, Nguyen et al. (2023) use the CNN model to classify sequences of regulatory and physiological activities. It is used to differentiate between successful and less successful collaborative learning sessions.

As statistical methods, HMMs are predominantly used to uncover and analyze hidden states and transitions in collaborative learning processes. For instance, Nasir et al. (2022) use HMMs to differentiate learners with varying learning gains by analyzing patterns in their collaborative activities. Ouyang et al. (2023) employ HMMs to find patterns in collaborative activities, showing the sequences that lead to effective collaboration. Similarly, Tang et al. (2022) use HMMs to identify and analyze the underlying states of attention during collaborative activities. Another commonly used statistical method is ANOVA, which is frequently employed to evaluate whether there are significant differences between the means of three or more groups. Dubovi (2022) employs a one-way ANOVA to investigate cognitive engagement among nursing students during a Virtual Reality (VR) simulation for medication administration. The analysis examines eye-tracking metrics across three Areas of Interest (AOIs): correct medication, expired medication, and incorrect medication, revealing significant differences in cognitive engagement between the AOIs. This shows that VR can create different cognitive responses depending on the task and content. Sharma et al. (2022) use repeated-measures ANOVA to analyze students' cognitive and physiological responses, such as eye-tracking data and physiological signals, during various phases of a motion-based educational game. This analysis highlights how ANOVA can reveal significant variations in engagement and performance, which are used to design adaptive feedback mechanisms in AI-driven learning environments. Similarly, in another paper, Sharma and Giannakos (2023) use ANOVA to examine differences in cognitive and physiological responses across various scaffolding conditions during the learning process. This analysis identifies significant variations in how students respond to different feedback strategies, offering insights into the most effective approaches for enhancing learning outcomes. We categorize the use cases of machine learning algorithms in education into three groups: engagement detection and analysis (n=12, 24.0%), predictive modeling of learning outcomes (n=6, 12.0%), and collaboration quality assessment (n=7, 14.0%). For the use cases of statistical methods we identify four categories: analysis of collaborative learning processes (n=7, 30.4%), prediction and assessment of learner behavior (n=4, 17.4%), evaluation of learners' engagement (n=2, 8.7%), and developing AI-driven learning support (n=1, 4.3%). The percentages are calculated relative to the total number of machine learning algorithms (n=50) and statistical methods (n=23). Since some studies applied both machine learning and statistical methods, the percentages within each category do not sum to 100%. An overview of these use cases, the corresponding algorithms, and their percentage distributions is provided in Tables 6 and 7.

Table 6

Overview of AI algorithms and their applications in the reviewed studies

Category	Item	No. of Studies (n)	Percentage
Machine Learning Algorithms	Random Forest (RF)	15	25.4%
	Support Vector Machines (SVM)	13	22.0%
	Long Short-Term Memory (LSTM)	8	13.5%
	Convolutional Neural Networks (CNN)	6	10.1%
	Decision Trees	5	8.4%
	Hybrid (LSTM + CNN)	3	5.0%
Statistical Methods	ANOVA Test	8	13.5%
	Epistemic Network Analysis (ENA)	6	10.1%
	Hidden Markov Models (HMMs)	6	10.1%
	K-means Clustering	3	5.0%

Based on our investigation, we find that researchers employ a diverse range of algorithms tailored to specific research topics, see Table 4. Our analysis shows that RF algorithms are often used to analyze student interactions in Collaborative Learning, identifying patterns that help predict successful collaboration strategies. For Performance Prediction, SVMs frequently classify and forecast student outcomes using historical data.

Table 7

Use Cases of Machine Learning Algorithms and Statistical Methods in the Reviewed Studies

Category	Item	No. of Studies (n)	Percentage
The Use Cases of Machine Learning Algorithms	Engagement detection and analysis	12	24.0%
	Collaboration quality assessment	7	14.0%
	Predictive modeling of learning outcomes	6	12.0%
The Use Cases of Statistical Methods	Analysis of collaborative learning processes	7	30.4%
	Prediction and assessment of learner behavior	4	17.4%
	Evaluation of learners' engagement	2	8.7%
	Developing AI-driven Learning Support	1	4.3%

In studies on Self-regulated Learning, LSTM networks are commonly employed to track and predict students' regulatory behaviors over time. Additionally, we find that authors predominantly use CNNs in research on Student Feedback/Support.

Challenges and Future Work

The reviewed studies use multimodal data to tackle a multitude of research questions about education, employing different methods and techniques. The authors of these studies also state some challenges and limitations, and give directions for future work. This subsection analyzes and describes challenges and limitations that are often mentioned, and the suggestions for future directions.

Challenges and Limitations

(1) Limited number of participants: One of the main challenges with the majority of the studies is related to the amount of participants for the experiments or data collection, which is mentioned in 28 of the reviewed studies. Since most of the academic lab experiments have limited time or work resources, the studies are conducted with a limited number of participants. It would be good to verify that we obtain similar results if we perform an experiment again with a much larger group of participants. Ouyang et al. (2023) is one of the studies that tackle collaborative problem solving, where the authors created a dataset from groups of students doing collaborative activities. The study only consists of 4 student groups doing 24 collaborative activities. The authors mention the small sample size being "a reason for causing the insignificant differences in the final concept map products". They also add that "future empirical research needs to expand the sample size and experiment with different courses to test, validate, or modify the implications".

(2) Limited variety of participants: Similar to the challenges around the limited number of participants, studies generally collect data of participants from a limited variety of demographics, such as only participants from higher education, or participants of a single classroom. This raises questions about the generalizability of the proposed methods and findings. 22 of the reviewed studies mention this as a limitation. Li X. et al. (2021) aims to understand student engagement in agent-based environments with multimodal data. They conduct a user study with 24 undergraduate students, where the participants were divided into groups of 3 to make 8 groups. The groups worked on an English literature task with the help of either a virtual human agent or a text-based agent. Although the authors find that virtual human agents result in much higher student engagement, they also mention the small sample size, and they question the extensibility of their results due to not reaching statistically significant outcomes.

(3) Limited multimodality: Even though studies make use of multimodal data to some extent, 18 of the reviewed studies mention that having a larger variety of modality could provide better insights. Data from some types is more challenging to collect due to the need of additional equipment, such as EDA, motion, or eye tracking data. Understandably, studies often collect feasible data, and cannot investigate the effects of data from other modalities. Fahid et al. (2023) investigates behavioral engagement in collaborative game-based learning, where they conduct a user study with 28 middle school students to understand the effects of different modalities in behavioral engagement detection. The authors use 10-second-segments of video data from the game-based learning system. Although they give insights about the modalities, they also mention integration of further modalities such as game interaction logs and in-game collaborative processes could improve the quality of engagement detection, which is currently limited to the modalities that are employed in their study.

(4) Confined Experiment Settings: Another limitation commonly stated (12 of the reviewed studies) is that experiments are conducted in confined settings, where participants are recruited and the experiment is performed under special conditions. Although a good setup can give valuable insights, it does not necessarily replicate in real-life situations. The participants are aware that they are in an experiment where their data is collected, and might therefore act differently than they would in real-life. Thus, results in real-life settings might differ from those obtained in confined experiments. Yusuf et al. (2024) use MMLA to model students' learning behavior for animated programming. They conduct an experiment in a confined classroom setting with 35 participants. The authors studied the same research question before, getting different results, and mention their outcomes might be specific to animated programming setting in the experiment setup they created.

(5) **Lack of Algorithm/Parameter Variety:** Studies applying machine learning algorithms on the data collected usually use a few generic methods, such as random forest for classification or LSTM for deep learning. Moreover, in many cases the parameters used in these algorithms are static, and a range of parameters are not explored. Better settings or more efficient algorithms might not always be explored in the studies, and fine-tuning might not always lead to the best results. 11 of the reviewed studies mention these issues as a limitation.

(6) **Partial/Missing Data:** In a few cases, studies that use previously collected datasets mention discrepancies in the data due to human or technical errors, such as missing or incorrect information (Ciordas-Hertel et al., 2022; Moon et al., 2022; Sabuncuoglu & Sezgin, 2023). These issues make analyzing data challenging for researchers, and might result in erroneous implications depending on the severity of the issues.

(7) **Hardware Issues:** The studies that collect physiological data require wearable devices for collecting EDA or EEG data. Face and eye tracking data also depends on specific cameras and tools. Studies that collect these types of data often mention challenges around calibrating devices and the quality of the equipment, and the collected data might not always completely reflect the real-life information (Di Mitri et al., 2017; Hasnine et al., 2021; Li X. et al., 2023; Sabuncuoglu & Sezgin, 2023; Tang et al., 2022).

Future Work

(1) **Experiments on a larger scale:** A common goal in the suggestions for future work in the reviewed studies is to expand the experiments by collecting data from a larger target group. Authors often mention this as an approach to develop more reliable and generalizable methods or systems. These improvements could help resolving drawbacks from Limitations 1 and 2.

(2) **Longitudinal Studies:** The research presented in the reviewed studies is often evaluated by means of data collected in brief experiments in confined settings. Some studies mention that conducting a longitudinal study would significantly improve the validity of the work, which can give more realistic results than the confined experiment results mentioned in Limitation 4.

(3) **Increasing Modality:** Lack of multimodality affects the validity and the reliability of the results, as stated in Limitation 3. Therefore, in many of the reviewed studies the authors suggest to expand their work by adding more modalities by employing additional devices or applications. This could also implicitly help with Limitations 6 and 7, since a larger variety of modalities can help overcome data errors and hardware issues by providing more types of data. In addition, we expect that reliability and validity of the collected data increases with more data points collected through various means.

(4) Exploration of Methods and Algorithms: Many studies mention that they want to expand the experiments by using different methods or algorithms. Machine learning and especially deep learning is often employed in the studies, and authors mention that using a wider variety of methods and algorithms might give better results or more precise insights. This would be a solution for Limitation 5, which discusses the issues emerging from the lack of algorithm and parameter variety.

Discussion

The results of our review show that researchers who employ multimodal data in education focus on a variety of research questions. The outcomes of these studies provide beneficial insights through the experiments conducted and learning applications developed. But there are still many open questions, and some research topics remain to be thoroughly investigated. This section discusses the results of our systematic review, answers the research questions formulated in the introduction section, and presents the key findings that might provide guidance for future research.

R.Q.1 - learning analytics with multimodal data: state of the art

To analyze this research question, we have investigated the main research topics of the reviewed studies, the data types employed, data integration levels, the amount of data sources and the target audiences of the studies.

Result 1: Research focus is mainly on collaborative learning, emotion/engagement detection and performance prediction. The results in Table 4 show that more than one third of the studies focus on collaborative learning, tackling issues in topics such as collaborative programming, group dynamics and collaborative game-based learning. Capturing contextual, behavioral and physiological aspects of group interactions in collaborative learning settings would be difficult to obtain with a single source of data and no multimodality. This leads to collaborative learning being a point of interest for multimodal learning analytics studies. Chandra (2015) investigates the concept of collaborative learning in education, and finds that interaction and "doing" are of primary importance when learning in groups. Analyzing actions and interactions between peers in groups requires data from several modes, including audial and visual. This study supports our finding that multimodality is an essential part of collaborative learning. Since collaboration results in more measurable data than individual learning, collaborative learning is an obvious source for multimodal data research.

Detection and analysis of learner emotion and engagement are also common research topics in the reviewed studies. This is as expected, since physiological data measurements on top of other modalities naturally plays an essential role in emotion and engagement detection. Garcia-Garcia et al. (2017) conduct a technological review of emotion detection and categorize various approaches through detection from speech, text, body gestures and movements, facial expressions, and physiological states. Combining emotion detection with learning data results in multimodal data, which is in line with our findings. Dewan et al. (2019) provides a review of engagement detection in online learning. They divide engagement detection approaches into three categories, namely manual, semi-automatic and automatic. The automatic approaches use log data, sensor data and computer vision data, consisting amongst others of facial expressions, gaze, gesture and posture data. Their categorization also shows the need of multimodality for learning analytics for automatic engagement detection.

The third main research topic is performance prediction, studied in more than one fifth of the reviewed articles. In many cases, prediction algorithms benefit from various data sources with multimodal features. Each modality can provide valuable information to capture data patterns to calculate accurate predictions. Thus, studies employing multimodal data for performance prediction offer solutions in topics such as success/failure prediction of students for their learning goals and effort prediction for an upcoming learning tasks. Pelima et al. (2024) systematically review performance prediction approaches using machine learning for higher education. The authors describe the number of papers per data source on which performance prediction is based. Although there are many applications that use only contextual academic data or only behavioral data, they also show that a combination of both, which requires multimodal data, has the second highest frequency, below only contextual data applications and above only behavioral data applications. Their results show that multimodality is already used in many performance prediction related studies, which aligns with our findings.

Other research topics such as data visualization and dashboards, human-computer or human-robot interaction and intelligent tutoring systems are discussed in only a few of the reviewed studies. We expect that the mentioned research topics could benefit from employing and performing experiments with multimodal data.

Result 2: Video data, eye tracking, log data, audio data and EDA data are the most common data types used in the studies. As expected from previous research on multimodal data (Molenaar et al., 2023), log data is one of the most commonly used data types in the reviewed studies. This can be explained with all the learning and teaching applications containing important pieces of information in their log data, such as data of learners working on tasks, time spent on learning tasks and clickstream data. With multimodality being a focal point of the reviewed studies, the research also often uses video, audio and eye-tracking data, which are often collected with video cameras and microphones. Since the devices required for these data types are often easy to obtain and use, they are frequently used by researchers to collect both contextual and behavioral data. EDA data, the most common type of physiological data, is typically obtained via wristbands worn by the participants in experiments. EDA data is commonly used in studies focusing on emotion and engagement detection.

Result 3: Contextual data types are most common in the studies, and the combination of contextual data with behavioral data is widely applied. Contextual information about learning activities is often essential for research topics such as performance prediction, self-regulated learning or data visualization. Without the actual context of the learning process, it would not be possible to measure the success of learners on their learning goals. Because contextual data is so important, almost all of the reviewed studies use data types that contain contextual information. The most common data type combination is contextual and behavioral data, used in almost 40% of the studies. Research topics such as collaborative learning and emotion/engagement analysis usually benefit from this combination, since both contextual and behavioral aspects of learning are important for these topics. Physiological data is encountered less than both contextual and behavioral data in the reviewed studies. We argue that the main reason for this is the need of wearable devices for physiological data collection. In larger classroom settings, each learner wearing a device to collect data can be expensive and cumbersome. Studies that use physiological data are therefore not common. The breakdown of data types for various educational research in the reviews of Pelima et al. (2024) and Garcia-Garcia et al. (2017) both align with our findings, where the modalities we consider as contextual are included in most of the reviewed studies, while combination of contextual and behavioral data is also common.

Result 4: The majority of the studies focus on higher education, while very few studies tackle issues in special education. Over 60% of the studies focus on higher education, by either making use of data collected in higher education institutes, or tackling research questions that address issues that occur in higher education. We expected this result because the majority of the studies are conducted by researchers working in universities, which makes it easier to recruit participants from higher education for data collection. Moreover, higher education students are usually of an age that doesn't require parental consent, which makes the recruitment process, the ethical approval of the experiment, and the data collection easier than studies with younger students. Despite these challenges, slightly over 40% of the reviewed studies focus on K-12 education. However, in most of the studies the application of the proposed methods and the developed systems in real-life settings are mentioned as future work. We have found few studies performing research experiments in special education. We expect that conducting user studies and collecting longitudinal data is often more challenging in special education than in regular schools, because students are more frequently absent, or some types of modal data might be hard to collect in a similar manner from all the students. However, a variety of different modalities, such as audial and visual cues, motion detection, facial data analysis can be helpful for learners in special education (Chan et al., 2023; Ting, 2014). Therefore, the benefits of applying learning analytics with multimodal data can become essential for improving the learning and teaching quality in this domain. Zhai et al. (2021) reviews studies on AI in education, and provides a breakdown of education levels for the studies they surveyed. Their breakdown resembles ours, with a majority of higher-education applications. They also state the lack of studies for special education, reviewing just a single study about a text-to-diagram application to help visually impaired students.

Result 5: Research is mainly aimed at student-related research questions; not many studies focus on research questions targeting teachers. A significant majority of the studies offer answers to research questions targeting students. Since learning activities are performed by students, we expected this. Although student activities produce various kinds of data in their learning trajectory, such as log data in a learning environment when solving exercises, teacher activities might not be easy to capture. For example, assessing students for a task might be related to teacher's intuition, therefore the reasoning behind would be difficult to collect as data. However, there are also research questions related to teachers, such as overview dashboards for the learning progress of pupils in a classroom, feedback for a teacher, and teacher professionalization tools. Although some of the reviewed studies tackle such questions, the number of learning analytics studies employing multimodal data that focus on teacher-related research questions is significantly lower than the number of studies focusing on students. Alfredo et al. (2024) systematically review human-centric learning analytics applications that employ AI in education. Their analysis of stakeholders in the reviewed studies shows that 71% of all studies target students, either as participants or supporting a design. Only 20% of the studies target teachers, although 39% of the studies include teachers in the design process, even if the solution targets students. Zhai et al. (2021) also report a relatively low number of studies investigating teacher-related questions in their review on AI in education.

R.Q.2 - AI algorithms and statistical methods employed

The discussion on AI algorithms and the statistical methods used refers to the outcomes of *AI algorithms/methods used in the studies* subsection, where we investigated which studies applied which methods for which purposes.

Result 6: Machine learning is the most used AI approach, where deep learning methods are commonly employed. The majority of the studies use machine learning methods for developing AI solutions for their research goals. Since many of the state of the art machine learning solutions involve deep learning methods, they are often employed in the reviewed studies, mostly for prediction or classification tasks using multimodal data as input. The typical use of machine learning algorithms involve processing various collected data types as input, and running the algorithm to produce output for testing their hypotheses. Pre-processing data and labeling or classification of data points by human experts is common practice, while merging data from different sources before giving them as input to machine learning algorithms is rarely encountered. Martin et al. (2024) conduct a systematic review on the use of AI in K-12 education. They classify machine learning methods used in AIED applications as reinforcement, unsupervised, or supervised learning, where they combine the latter category with deep learning. Their results show that almost all included studies employ supervised and deep learning. This is in line with our results.

Result 7: Statistical methods are frequently applied in the studies. Most studies use statistical methods like ANOVA (Analysis of Variance) and Hidden Markov Models (HMMs) in collaborative learning research. HMMs uncover patterns in collaboration, such as effective group dynamics and attention shifts, while ANOVA identifies differences in engagement and performance across groups or conditions.

R.Q.3 - learning analytics with multimodal data: state of the art

Based on the results in *Challenges and Future Work* subsection, we discuss a summary of the challenges and possible directions to take for future research.

Result 8: Most experiments are performed in a confined setting with a limited number of participants; few studies apply MMLA in real-life situations. We and the authors of most reviewed studies observe that one of the main challenges of MMLA is that experiments are often conducted in a confined environment, during a limited period, and with a limited number of participants with little variety in demographics. Therefore, generalizability of the results or the developed systems stays an open question for the majority of the studies. Moreover, data collection for experiments uses several devices for audio and video recording, eye-tracking, motion-tracking and EDA or EEG measurements. Although it is possible to collect data with these devices in the scope of the study experiments, using these devices in real-life situations, satisfying security and privacy requirements and collecting the data for long periods with optimal calibration is challenging. In most of the studies these limitations are acknowledged, but solutions for obtaining more generalizable systems are usually not explored and stated as goals for future research. Eden et al. (2024) provide an overview of the challenges for integrating AI in education. The authors mention the difficulties of ensuring equitable access to solutions due to socio-economical differences and geographical conditions. They also raise concern for bias in solutions, possibly stemming from research being conducted by a relatively small group of researchers with limited data. Borenstein and Howard (2021) focus on ethical challenges of using AI in education, and mention the potential bias and fairness related issues that should be addressed by researchers considering a large target audience.

Conclusion

This systematic literature review investigates studies in the scientific literature that use multimodal data in educational research and apply learning analytics with AI methods. We selected 65 studies published before December 31, 2024, coded them based on the research questions stated in the introduction section, and analyzed the results to gain insights about the current developments in the domain. In our first research question, we analyze the research areas of the reviewed studies, using a categorization of the research topics that the studies tackle. We also provide a classification of data types used in these studies, and show which types are most often combined with each other. We examine the target audiences in the reviewed studies distinguishing various education levels and the stakeholders for learning. Moreover, we explore the types of AI algorithms and statistical methods employed and show their purposes. Finally, we describe the challenges and possible future work stated in the reviewed studies, and point out possible future directions that researchers can take.

The yearly increasing number of publications shows that using multimodal data for learning analytics in education is rapidly becoming more common. With our systematic keyword search, only a handful of studies were retrieved from 2012 until 2020. The field did not have peer-reviewed work before 2012, because learning analytics became its own field only in the last decade and a half. The last three years saw a significant increase in the number of publications, and we expect a further increase in the coming years, with many open research questions still to tackle. Moreover, the recent introduction of large language models enables the use of AI chatbots for educational purposes. With the use of multimodal data, large language models can offer a variety of solutions for student support, self-regulated learning and many other research areas. Therefore, they are starting to gain a significant amount of attention by the learning analytics researchers. We have found that many of the efforts end after an experimentation phase and do not end up in the daily classroom. We expect this to change in the near future, when multimodal learning analytics research provides efficient solutions that are much harder to achieve with prior applications.

Our review has a limitation that will be addressed in future work. Although we found over a thousand publications using our keyword string, it is possible that some related publications were not in the list of retrieved publications, for example because they do not use the terms that we are searching for such as learning analytics or multimodal. A snowballing approach from the references of the reviewed studies and other related systematic literature reviews, and the publications citing these studies could help us retrieve more related papers that would help answering our research questions.

Appendix

Abbreviations

ANOVA: Analysis of Variance; CNN: Convolutional Neural Network; DNN: Deep Neural Network; DT: Decision Tree; HMM: Hidden Markov Model; kNN:k-Nearest Neighbor; MEM: Linear Mixed Effects Models; LMS: Learning Management System; LSTM: Long Short-Term Memory; MHMM: Multilevel Hidden Markov Model; MLR: Multiple Linear Regression; RF: Random Forest; RNN: Recurrent Neural Network; SVM: Support Vector Machine

Author's contributions

Onuralp Ulusoy conducted the systematic literature review, contributed to the selection and coding of the studies, wrote Introduction, Related Work, Results, Discussion and Conclusion sections.

Aysu Ismayilova conducted the systematic literature review, contributed to the selection and coding of the studies, wrote Methodology and Results (AI algorithms/methods used in the studies) sections.

Johan Jeuring contributed to the keyword selection and coding design of the systematic literature review, overviewed the structure of the study, edited all sections to improve the text.

Author's information

Onuralp Ulusoy is a postdoctoral researcher in the Department of Information and Computing Sciences at Utrecht University and National Education Lab Artificial Intelligence (NOLAI) at Radboud University

Aysu Ismayilova is a PhD candidate in the Department of Information and Computing Sciences at Utrecht University and National Education Lab Artificial Intelligence (NOLAI) at Radboud University

Johan Jeuring is a Professor in the Department of Information and Computing Sciences at Utrecht University and National Education Lab Artificial Intelligence (NOLAI) at Radboud University

Funding

Funding is provided by the Dutch National Growth Fund through NOLAI, the National Education Lab Artificial Intelligence

Availability of data and materials

Link to full coded tables: [Learning Analytics with Multimodal Data Through the Lens of AI in Education a Systematic Literature Review · GitLab]

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

All authors are affiliated with the Department of Information and Computing Sciences at Utrecht University, the Netherlands, and National Education Lab Artificial Intelligence (NOLAI), Radboud University, Nederland

Received: 9 June 2025 Accepted: 30 December 2025

Published online: 1 January 2027 (Online First: 15 April 2026)

References

Afsar, M. M., Crump, T., & Far, B. (2022). Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7), 1–38.

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215.

Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022). The evidence of impact and ethical considerations of multimodal learning analytics: A systematic literature review. *The multimodal learning analytics handbook*, 289–325.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of Chatgpt in promoting teaching and learning. *Journal of AI*, 7 (1), 52–62.

Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. *Proceedings of the sixth international conference on learning analytics & knowledge*, 329–338.

Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative AI: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, 15(8), 260.

Basystiuk, O., Melnykova, N., & Rybchak, Z. (2023). Multimodal learning analytics: An overview of the data collection methodology. *2023 IEEE 18th International Conference on Computer Science and Information Technologies*, 1–4.

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the third international conference on learning analytics and knowledge*, 102–106.

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1(1), 61–65.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work (tech. rep.)*. National Bureau of Economic Research.

Chan, R. Y.-Y., Wong, C. M. V., & Yum, Y. N. (2023a). Predicting behavior change in students with special education needs using multimodal learning analytics. *IEEE Access*, 11, 63238–63251.

Chandra, R. (2015). Collaborative learning for educational achievement. *IOSR Journal of Research & Method in Education*, 5(3), 2320–7388.

Chango, W., Lara, J. A., Cerezo, R., & Romero, C. (2022). A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(4), e1458.

Chejara, P., Kasepalu, R., Prieto, L. P., Rodríguez-Triana, M. J., Ruiz-Calleja, A., & Shankar, S. K. (2023a). Multimodal learning analytics research in the wild: Challenges and their potential solutions. *LAK23: 13th International Learning Analytics and Knowledge Conference, CrossMMLA@LAK Workshop*, 36–42.

Chejara, P., Prieto, L. P., Dimitriadis, Y., Rodríguez-Triana, M. J., Ruiz-Calleja, A., Kasepalu, R., & Shankar, S. K. (2024). The impact of attribute noise on the automated estimation of collaboration quality using multimodal learning analytics in authentic classrooms. *Journal of Learning Analytics*, 11(2), 73–90.

Chejara, P., Prieto, L. P., Rodriguez-Triana, M. J., Kasepalu, R., Ruiz-Calleja, A., & Shankar, S. K. (2023b). How to build more generalizable models for collaboration quality? Lessons learned from exploring multi-context audio-log datasets using multimodal learning analytics. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 111–121.

Chejara, P., Prieto, L. P., Rodriguez-Triana, M. J., Ruiz-Calleja, A., & Khalil, M. (2023c). Impact of window size on the generalizability of collaboration quality estimation models developed using multimodal learning analytics. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 559–565.

Ciordas-Hertel, G.-P., Biedermann, D., Winter, M., Mordel, J., & Drachsler, H. (2022). How can interaction data be contextualized with mobile sensing to enhance learning engagement assessment in distance learning? *Companion Publication of the 2022 International Conference on Multimodal Interaction*, 105–112.

Crescenzi-Lanna, L. (2020). Multimodal learning analytics research with young children: A systematic review. *British Journal of Educational Technology*, 51(5), 1485–1504.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, 51(5), 1441–1449.

Dewan, M., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, 6(1), 1–20.

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. *Proceedings of the seventh international learning analytics & knowledge conference*, 188–197.

Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education*, 183, 104495.

Eden, C. A., Chisom, O. N., Adeniyi, I. S., et al. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, 10(2), 006–013.

Fahid, F. M., Lee, S., Mott, B., Vandenberg, J., Acosta, H., Brush, T., Glazewski, K., Hmelo-Silver, C., & Lester, J. (2023). Effects of modalities in detecting behavioral engagement in collaborative game-based learning. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 208–218.

Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017). Emotion detection: A technology review. *Proceedings of the XVIII international conference on human computer interaction*, 1–8.

Gardner, J., & Brooks, C. (2018). Student success prediction in moocs. *User Modeling and User-Adapted Interaction*, 28, 127–203.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48, 108–119.

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134–147.

Gupta, A., Anpalagan, A., Guan, L., & Khwaja, A. S. (2021). Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array*, 10, 100057.

Hasnine, M. N., Bui, H. T., Tran, T. T. T., Nguyen, H. T., Akçapınar, G., & Ueda, H. (2021). Students' emotion extraction and visualization for engagement detection in online learning. *Procedia Computer Science*, 192, 3423–3431.

Heinemann, B., Görzen, S., & Schroeder, U. (2023). Teaching the basics of computer graphics in virtual reality. *Computers & Graphics*, 112, 1–12.

Järvelä, S., Nguyen, A., Vuorenmaa, E., Malmberg, J., & Järvenoja, H. (2023). Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior*, 144, 107737.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report EBSE-2007-01*.

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics*, 11(1), 141.

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 159-174.

Lee, Y., Limbu, B., Rusak, Z., & Specht, M. (2023). Role of multimodal learning systems in technology-enhanced learning (tel): A scoping review. *European Conference on Technology Enhanced Learning*, 164–182.

Lee, Y., & Specht, M. (2023). Can we empower attentive e-reading with a social robot? An introductory study with a novel multimodal dataset and deep learning approaches. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 520–530.

Li, X., Bao, H., & Li, Y. (2021). Understanding student engagement based on multimodal data in different types of agent-based collaborative learning contexts. *2021 Tenth International Conference of Educational Innovation through Technology*, 239–244.

Li, X., Yan, L., Zhao, L., Martinez-Maldonado, R., & Gasevic, D. (2023). Cvpe: A computer vision approach for scalable and privacy-preserving socio-spatial, multimodal learning analytics. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 175–185.

Martin, F., Zhuang, M., & Schaefer, D. (2024). Systematic review of research on artificial intelligence in K-12 education (2017–2022). *Computers and Education: Artificial Intelligence*, 6, 100195.

Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., Li, X., Dix, S., Jaggard, H., Wotherspoon, R., et al. (2023). Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction*, 31(1), 1–41.

Mazza, R., & Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, 154–161.

Moissa, B., Bonnin, G., & Boyer, A. (2021). Measuring and predicting students' effort: A study on the feasibility of cognitive load measures to real-life scenarios. *Technology-Enhanced Learning for a Free, Safe, and Sustainable World:* 16th *European Conference on Technology Enhanced Learning, ECTEL 2021*, Proceedings 16, 363–367.

Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, 139, 107540.

Monsalves, D., Riquelme, F., & Cornide-Reyes, H. (2023). Enhancing education through multimodal learning analytics and AI-as-a-service. *2023 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies*, 1–6.

Moon, J., Ke, F., Sokolikj, Z., & Dahlstrom-Hakki, I. (2022). Multimodal data fusion to track students' distress during educational gameplay. *Journal of Learning Analytics*, 9(3), 75–87.

Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors*, 20(23), 6856.

Nasir, J., Abderrahim, M., Kothiyal, A., & Dillenbourg, P. (2022). Temporal pathways to learning: How learning emerges in an open-ended collaborative activity. *Computers and Education: Artificial Intelligence*, 3, 100093.

Nguyen, A., Järvelä, S., Rosé, C., Järvenoja, H., & Malmberg, J. (2023). Examining socially shared regulation and shared physiological arousal events with multimodal learning analytics. *British Journal of Educational Technology*, 54(1), 293–312.

Ni, J., Chen, Y., Chen, Y., Zhu, J., Ali, D., & Cao, W. (2020). A survey on theories and applications for self-driving cars based on deep learning methods. *Applied Sciences*, 10(8), 2749.

Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O., & Moeller, K. (2019). Increased emotional engagement in game-based learning–A machine learning approach on facial emotion detection data. *Computers & Education*, 142, 103641.

Noroozi, O., Pijeira-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies*, 25, 5499–5547.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). The rap system: Automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. *Proceedings of the 8th international conference on learning analytics and knowledge*, 360–364.

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, 51(5), 1527–1547.

Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, 4, 100136.

Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(1), 39–66.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, 5, 1–10.

Oviatt, S., & Cohen, A. (2013). Written and multimodal representations as predictors of expertise and problem-solving success in mathematics. *Proceedings of the 15th ACM on International conference on multimodal interaction*, 599–606.

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., et al. (2021). *The prisma 2020 statement: An updated guideline for reporting systematic reviews*. Bmj, 372.

Pei, B., Xing, W., & Wang, M. (2023). Academic development of multimodal learning analytics: A bibliometric analysis. *Interactive Learning Environments*, 31(6), 3543–3561.

Pelima, L. R., Sukmana, Y., & Rosmansyah, Y. (2024). Predicting university student graduation using academic performance and machine learning: A systematic literature review. *IEEE Access*, 12, 23451–23465.

Sabuncuoglu, A., & Sezgin, T. M. (2023). Developing a multimodal classroom engagement analysis dashboard for higher-education. *Proceedings of the ACM on Human-Computer Interaction*, 7 (EICS), 1–23.

Samuelsen, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics—Review of literature. *Research and Practice in Technology Enhanced Learning*, 14(1), 11.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, 51(5), 1450–1484.

Sharma, K., & Giannakos, M. (2023). Carry-forward effect: Early scaffolding learning processes. *Proceedings of the 2023 Symposium on Learning, Design and Technology*, 43–52.

Sharma, K., Lee-Cultura, S., & Giannakos, M. (2022). Keep calm and do not carry-forward: Toward sensor-data driven AI agent to enhance human learning. *Frontiers in Artificial Intelligence*, 4, 713176.

Sharma, K., Papamitsiou, Z., Olsen, J. K., & Giannakos, M. (2020). Predicting learners' effortful behaviour in adaptive assessment using multimodal data. *Proceedings of the tenth international conference on learning analytics & knowledge*, 480–489.

Siemens, G. (2011). Call for papers of the 1st international conference on learning analytics & knowledge (Lak 2011).

Su, Y., Liu, Q., Liu, Q., Huang, Z., Yin, Y., Chen, E., Ding, C., Wei, S., & Hu, G. (2018). Exercise-enhanced sequential modeling for student performance prediction. *Proceedings of the AAAI conference on artificial intelligence*, 32(1).

Sun, Z., Xie, K., & Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *The internet and higher education*, 36, 41–53.

Supovitz, J. A., & Klein, V. (2003). Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement. *Consortium for Policy Research in Education Philadelphia*, PA.

Tang, H., Dai, M., Yang, S., Du, X., Hung, J.-L., & Li, H. (2022). Using multimodal analytics to systemically investigate online collaborative problem-solving. *Distance Education*, 43(2), 290–317.

Tato, A., & Nkambou, R. (2023). Towards a multi-modal deep learning architecture for user modeling. *The International FLAIRS Conference Proceedings*, 36.

Ting, K.-y. (2014). Multimodal resources to facilitate language learning for students with special needs. *International Education Studies*, 7 (8), 85–93.

Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. *Proceedings of the tenth international conference on learning analytics & knowledge*, 524–533.

Worsley, M., & Martinez-Maldonado, R. (2018). Multimodal learning analytics' past, present, and potential futures. *LAK18: 8th International Learning Analytics and Knowledge Conference, CrossMMLA@LAK Workshop*, 2.

Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A new era in multimodal learning analytics: Twelve core commitments to ground and grow MMLA. *Journal of Learning Analytics*, 8(3), 10–27.

Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: A survey. *ACM Computing Surveys*, 55(5), 1–37.

Xhakaj, F., Aleven, V., & McLaren, B. M. (2017). Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, ECTEL 2017*, Proceedings 12, 315–329.

Yan, L., Martinez-Maldonado, R., Swiecki, Z., Zhao, L., Li, X., & Gašević, D. (2024). Dissecting the temporal dynamics of embodied collaborative learning using multimodal learning analytics. *Journal of Educational Psychology,* 117(1), 106.

Yan, L., Zhao, L., Gasevic, D., & Martinez-Maldonado, R. (2022). Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th international learning analytics and knowledge conference*, 13–23.

Yu, H., Miao, C., Leung, C., & White, T. J. (2017). Towards AI-powered personalization in MOOC learning. *NPJ Science of Learning*, 2(1), 15.

Yusuf, A., Noor, N. M., & Bello, S. (2024). Using multimodal learning analytics to model students' learning behavior in animated programming classroom. *Education and Information Technologies*, 29(6), 6947–6990.

Zafari, M., Bazargani, J. S., Sadeghi-Niaraki, A., & Choi, S.-M. (2022). Artificial intelligence applications in K-12 education: A systematic literature review. *Leee Access*, 10, 61905–61921.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021(1), 8812542.

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. https://doi.org/https://doi.org/10.1016/j.caeai.2021.100025

Publisher's Note