applied sciences

MDPI

Review

A Comprehensive Review of Multimodal Analysis in Education

Jared D. T. Guerrero-Sosa [1], Francisco P. Romero [1,*], Víctor H. Menéndez-Domínguez [2], Jesus Serrano-Guerrero [1], Andres Montoro-Montarroso [1] and Jose A. Olivas [1]

[1] Department of Information Technologies and Systems, University of Castilla La Mancha, Paseo de la Universidad, 4, 13071 Ciudad Real, Spain; jareddavid.guerrero@uclm.es (J.D.T.G.-S.); jesus.serrano@uclm.es (J.S.-G.); andres.montoro@uclm.es (A.M.-M.); joseangel.olivas@uclm.es (J.A.O.)

[2] Mathematics School, Autonomous University of Yucatan, Anillo Periférico Norte, Tablaje Cat. 13615, Merida 97119, Mexico; mdoming@correo.uady.mx

* Correspondence: franciscop.romero@uclm.es

Abstract: Multimodal learning analytics (MMLA) has become a prominent approach for capturing the complexity of learning by integrating diverse data sources such as video, audio, physiological signals, and digital interactions. This comprehensive review synthesises findings from 177 peer-reviewed studies to examine the foundations, methodologies, tools, and applications of MMLA in education. It provides a detailed analysis of data collection modalities, feature extraction pipelines, modelling techniques—including machine learning, deep learning, and fusion strategies—and software frameworks used across various educational settings. Applications are categorised by pedagogical goals, including engagement monitoring, collaborative learning, simulation-based environments, and inclusive education. The review identifies key challenges, such as data synchronisation, model interpretability, ethical concerns, and scalability barriers. It concludes by outlining future research directions, with emphasis on real-world deployment, longitudinal studies, explainable artificial intelligence, emerging modalities, and cross-cultural validation. This work aims to consolidate current knowledge, address gaps in practice, and offer practical guidance for researchers and practitioners advancing multimodal approaches in education.

Keywords: multimodal learning analytics; educational data mining; learning analytics; multimodal feature extraction

check for updates

Academic Editor: Juan Francisco De Paz Santana

Received: 24 April 2025
Revised: 19 May 2025
Accepted: 20 May 2025
Published: 23 May 2025

Citation: Guerrero-Sosa, J.D.T.; Romero, F.P.; Menéndez-Domínguez, V.H.; Serrano-Guerrero, J.; Montoro-Montarroso, A.; Olivas, J.A. A Comprehensive Review of Multimodal Analysis in Education. *Appl. Sci.* **2025**, *15*, 5896. https://doi.org/10.3390/app15115896

1. Introduction

The evolution from unimodal to multimodal learning analytics (MMLA) represents a significant methodological shift in educational research. Unlike traditional approaches that rely on single data sources, MMLA integrates and triangulates a range of data types—such as speech, eye gaze, posture, and physiological signals—to offer a comprehensive representation of learning processes [1]. This integration facilitates the modelling of complex learning environments and enhances the accuracy of predictive models when compared to unimodal approaches [2].

MMLA contributes to a more holistic understanding of both observable behaviours and internal cognitive-affective states. By leveraging trace data from multiple channels, this approach captures nuanced learner activity, engagement, and emotion that would otherwise go undetected [3]. It supports diverse learning modalities and acknowledges the variety of ways in which students interact with content, thereby enabling more personalised and adaptive learning environments [4]. Moreover, the use of artificial intelligence (AI), machine learning (ML), and advanced sensing technologies has enabled the real-time generation of feedback and dynamic instructional adaptation, further enhancing educational responsiveness [5,6].

The evolution from unimodal to multimodal approaches offers clear advantages in understanding both observable behaviours and internal cognitive-affective states. Unlike unimodal systems, which often capture only a single dimension of learner activity, MMLA can provide a richer, more comprehensive view by integrating diverse data sources. For instance, integrating visual, auditory, and physiological signals can reveal deeper insights into engagement, cognitive load, and collaboration, supporting more accurate assessment and personalised interventions [7,8]. MMLA also facilitates the modelling of complex learning behaviours, such as group dynamics and emotional responses, which are critical for effective collaborative learning and feedback systems [9,10]. This approach enables the detection of context-specific learning patterns that would be challenging to identify with unimodal methods alone, enhancing the overall interpretability and impact of educational analytics [11–13].

The application of multimodal analysis holds particular value in studying learner engagement. By analysing multimodal discourse—including both verbal and non-verbal cues—researchers can access deeper layers of meaning, such as intention and emotion, in educational interactions [14]. This insight supports the development of adaptive systems that align instructional strategies with learners' needs and preferences [15]. Importantly, multimodal analysis also offers the potential to inform inclusive learning design by accommodating multiple intelligences and individual learning styles.

Despite these advancements, several gaps persist in the field. Research has tended to prioritise verbal communication, with limited emphasis on non-verbal behaviours such as facial expressions and gestures, which play a crucial role in classroom interaction and learner engagement [16]. The integration of heterogeneous data sources remains methodologically challenging, particularly when aligning modalities that differ in granularity and temporal resolution [17]. Furthermore, MMLA has been underutilised in primary education, where its potential to support early learning remains largely unexplored [1].

There is also a lack of comparative research evaluating the performance of predictive models that incorporate low-intrusiveness modalities, such as wearable sensors. Exploring such combinations could lead to more naturalistic applications of MMLA in real-world settings [18]. Finally, the application of multimodal approaches in the context of teacher professional development has received little attention, despite its potential to enhance understanding of instructional practices and engagement in such settings [16].

The conceptualisation of MMLA can be traced back to the need for more comprehensive frameworks that capture the complexity of human learning. Early foundational work aimed to bridge the gap left by unimodal approaches, integrating data from speech, gaze, gesture, and physiological signals to create a more holistic representation of learner behaviour. Over time, this approach has evolved to incorporate a diverse range of data types, including spatial positioning, digital interaction logs, and context-aware data, reflecting a broader shift within the learning sciences towards capturing the full spectrum of educational interactions. This evolution has enabled the development of sophisticated data fusion frameworks that allow researchers to map low-level sensor data to higher-order cognitive constructs, providing deeper insights into the cognitive, affective, and social dimensions of learning.

In light of these opportunities and challenges, this review aims to synthesise recent developments in multimodal analysis in education. Drawing upon 177 peer-reviewed articles, the review addresses the following research questions:

- What are the foundational theories underpinning multimodal analysis in education?

- Which methodologies and tools are most commonly employed, and how are they implemented?
- How is multimodal analysis used to support learning and assessment across domains?
- What are the main challenges, limitations, and ethical concerns reported?
- What promising directions are emerging for future research in MMLA?

By addressing these questions, the review seeks to consolidate the state of knowledge in MMLA, clarify conceptual boundaries, and identify practical and methodological advances for future inquiry.

2. Background and Theoretical Foundations

The emergence of MMLA can be attributed to the limitations of traditional learning analytics in capturing the complexity of human learning. Instead of depending on a solitary data source, MMLA employs a multifaceted integration of heterogeneous data types, encompassing audio, video, eye-tracking, physiological signals and user logs. This comprehensive approach facilitates the acquisition of a more profound and contextualised comprehension of learning processes. This evolution is indicative of a more extensive trend within the domain of the learning sciences, namely the acknowledgement that meaningful learning occurs across a variety of channels and modalities. It follows that analytics must take these dynamics into account in order to support robust interpretation and feedback.

2.1. Origins and Development of MMLA

The conceptualisation of MMLA was initiated with the objective of establishing a unified framework for the integration of multimodal data into the domain of educational research. Early foundational work in multimodal learning can be traced back to studies in science education and semiotics, which explored the interplay between different modes of communication, including speech, gesture, and visual representations. These studies highlighted how multimodal resources allow for richer and more complex meaning-making processes that go beyond purely linguistic descriptions, integrating visual and verbal elements in scientific texts to form cohesive multimodal narratives [19]. Other research has examined how gestures serve as a bridge between physical experiences in science laboratories and the abstract language of scientific discourse, facilitating the transition from concrete actions to symbolic representation [20]. These foundational studies underscored the importance of integrating multiple data streams to capture the full spectrum of human learning, recognising that meaning is constructed through a complex interplay of language, symbols, physical actions, and social interactions.

The foundational work in this area, including that of Di Mitri et al. [21], established the theoretical underpinnings for the subsequent integration of low-level sensor data into higher-order learning constructs. These conceptual models were found to be of paramount importance in terms of facilitating the harmonisation of terminology and providing a framework for collaborative research initiatives across various academic disciplines.

Subsequent studies expanded the scope of MMLA by exploring the potential of real-time analysis and adaptive feedback, particularly in digital learning environments. Researchers such as Chango et al. [17] have emphasised the necessity for effective data fusion methods to synthesise diverse data sources, which range from video and electrodermal activity (EDA) to eye-tracking and interaction logs. Cohn et al. [22] made a further contribution to the field by proposing a taxonomy of multimodal domains that categorises data into five broad groups. The following terms are to be considered: natural language, video, sensors, human-centred design and environment logs.

As time has passed, the application of MMLA has become increasingly diverse, encompassing a variety of educational contexts. These include higher education, early childhood education [23], and music instruction [24]. This expansion has also prompted greater scrutiny of the ethical implications of collecting and analysing multimodal data, particularly regarding student privacy [25].

2.2. Theoretical Foundations

The theoretical underpinnings of MMLA are drawn from a multidisciplinary body of work, combining insights from educational psychology, cognitive science, computer science, and AI. However, it is important to recognise that multimodal learning is not purely a cognitive phenomenon. The epistemological foundations of MMLA also draw from fields like semiotics, sociocultural theory, and embodied cognition, which emphasise that meaning is constructed through a complex interplay of language, symbols, physical actions, and social interactions. For example, gestures have been shown to play a critical role in the formation of scientific language, serving as an intermediary between physical actions and abstract discourse [20]. Other studies have demonstrated how students construct scientific concepts like work and energy through the integration of multiple modalities, including language, mathematical symbolism, and physical actions, highlighting the importance of coordinating these modalities for coherent scientific understanding [26]. This integration is essential for the development of a coherent understanding of scientific concepts, as it allows students to bridge the gap between concrete experiences and abstract reasoning.

To provide a comprehensive framework for understanding the diverse approaches to MMLA, this section draws on well-established theoretical perspectives that have been widely applied in the field. These six core frameworks were identified through a combination of literature analysis and thematic review, focusing on perspectives that capture the cognitive, social, and computational dimensions of multimodal learning. While a wide range of theoretical approaches were considered, these six emerged as particularly relevant for interpreting the complex interactions captured through multimodal data, reflecting a stable core of foundational theories within the current MMLA landscape.

The most frequently applied frameworks include:

- Self-Regulated Learning (SRL): focuses on metacognitive processes, motivation, and strategic behaviour. MMLA enables the modelling of SRL by capturing behavioural and physiological traces during learning activities [27].
- Embodied cognition: emphasises the interaction between mind and body, suggesting that learning is grounded in sensorimotor experiences. This perspective has guided studies in mathematics learning and psychomotor skills training [28].
- Collaborative learning: investigates the co-construction of knowledge in group settings. MMLA facilitates the analysis of social interaction and synchronisation among learners [29,30].
- Data fusion frameworks: provide methodological scaffolding for combining multimodal data streams. These include many-to-one and many-to-many fusion strategies for managing diverse inputs and validating integrated models [17].
- Cognitive load theory: centers on how learners manage mental effort and working memory. Physiological indicators such as eye-tracking and heart rate are often used to assess cognitive load during tasks [31].
- AI and ML: support the development of predictive models that identify patterns in multimodal data to forecast learning outcomes and tailor interventions [32,33].

By acknowledging the epistemological aspects of learning, MMLA can provide a more holistic view of educational processes, recognising that learning is not merely the accumulation of knowledge but also the development of meaningful understanding through multimodal interactions. This broader perspective is essential for capturing the complexities of real-world educational contexts.

Each of these frameworks offers a lens through which multimodal learning data can be interpreted, thereby enabling researchers to link raw sensory inputs with higher-level educational constructs. Whilst the majority of research has historically concentrated on the implementation of a single framework, recent studies have indicated a discernible shift towards the integration of multiple theoretical frameworks in order to address the multifaceted nature of learning.

As MMLA continues to evolve as a research domain, there is a growing need to synthesise empirical evidence on how these theoretical frameworks are operationalised across diverse educational settings. To this end, the present review systematically maps the landscape of multimodal analysis in education through a structured literature search and selection process, as described in the following subsection.

2.3. Search and Selection Process

To ensure a comprehensive and representative review of multimodal analysis in educational settings, a structured search was conducted across four major bibliographic databases: Scopus, Web of Science, IEEE Xplore, and the ACM Digital Library. These databases were selected because they provide extensive coverage of high-quality, peer-reviewed research in the fields of computer science, education, and AI, which are core to the multidisciplinary nature of MMLA. Unlike other platforms, these databases offer more rigorous indexing, metadata quality, and citation analysis tools, ensuring a more reliable and replicable review process. Additionally, they allow for advanced filtering by document type, publication year, and subject area, which is crucial for reviews.

The search strategy was designed to retrieve documents that addressed multimodal analysis or multimodal learning in educational contexts, including terms such as "multimodal analysis", "multimodal learning", "multimodal data", and "multimodal learning analytics", combined with educational terms such as "education", "educational technology", "learning environment", "classroom", and "intelligent tutoring system".

Only articles in English were considered, as this is the dominant language in the academic publications that form the basis of the scientific discourse in this field. This choice reduces the potential for misinterpretation and inconsistency in the translation of technical terminology, ensuring that the findings can be more widely understood and integrated into the global research community.

The following Boolean logic was used as a base structure for the queries (adjusted to each database's syntax and field indexing):

```
(‘‘multimodal analysis’’ OR ‘‘multimodal learning’’ OR ‘‘multimodal data’’
OR ‘‘multimodal learning analytics’’)
AND (education OR ‘‘educational technology’’ OR ‘‘learning environment’’
OR classroom OR ‘‘intelligent tutoring system’’)
```

To ensure a manageable yet comprehensive corpus, the search strategy focused on explicit multimodal terminology, while also leveraging indexed keywords and subject categories where available. This approach recognises that multimodal research can encompass a wide range of techniques—such as eye-tracking, electroencephalogram (EEG), gesture recognition, and physiological monitoring—that may not always be labelled explicitly as "multimodal" in titles or abstracts, but are nonetheless categorised as such by the databases. This decision reflects a balance between inclusivity and specificity, prioritising studies that clearly align with the theoretical and analytical frameworks of MMLA, while avoiding an unmanageable volume of unrelated work.

Queries were applied to document titles, abstracts, keywords, or indexed terms, and results were restricted to journal articles, conference proceedings, and book chapters in the fields of computer science and education. This process retrieved a total of 1667 documents:

603 from Scopus, 730 from Web of Science, 198 from IEEE Xplore, and 136 from the ACM Digital Library. After deduplication, 1195 unique entries were obtained.

Regarding the publication period, the search did not impose a specific time range. However, the final set of 177 peer-reviewed studies included in this review spans from 2008 to 2025. This range reflects the distribution of available publications that met the inclusion criteria, rather than an intentional cut-off. The starting point of 2008 corresponds to the earliest relevant study identified during the screening phase, as the inclusion criteria focused on studies that explicitly addressed multimodal data collection, processing, or analysis in educational settings. Although early theoretical work on multimodality dates back to the 1990s, the practical integration of multimodal data in educational research gained momentum in the late 2000s, driven by advances in data capture technologies and machine learning. Each entry was carefully screened for relevance, prioritising studies that addressed core aspects of MMLA rather than peripheral topics. This process resulted in a final corpus of 177 peer-reviewed studies, which form the empirical basis of this comprehensive review.

3. Methodologies in Multimodal Analysis

The methodological foundation of multimodal analysis in education is anchored in the integration of diverse data streams, the application of advanced analytical techniques, and the systematic transformation of raw signals into meaningful educational indicators. This section categorises and describes the principal methods and tools employed in MMLA, structured across data collection modalities, feature extraction and preprocessing, modelling techniques and tools and frameworks.

To bridge the theoretical foundations established in Section 2 with the methodological approaches detailed in this section, it is important to outline the process by which the four key components of MMLA were identified. The 177 studies selected for this review underwent a structured qualitative analysis to capture the diversity of methodologies employed in multimodal research. This process included multiple stages: initial coding, thematic clustering, and iterative refinement. Initially, each study was coded based on its primary methodological focus, including data collection strategies, feature extraction methods, modelling approaches, and the computational frameworks used to process multimodal data. During this phase, studies were grouped according to the types of data they utilised (e.g., visual, auditory, physiological), the preprocessing techniques applied, the modelling methods employed, and the specific tools or platforms used.

Subsequently, an iterative thematic analysis was conducted to identify common patterns and refine these initial groupings. This involved several rounds of manual review to ensure consistency and comprehensiveness, allowing for the identification of emerging trends and methodological gaps. The final structure, comprising four core components, emerged as a natural grouping of the most frequently cited methods and tools: (1) data collection modalities, (2) feature extraction and preprocessing, (3) modelling techniques, and (4) tools and frameworks. This hierarchical organisation captures the critical steps required to collect, process, and interpret multimodal data in educational contexts, reflecting both the technical and conceptual frameworks that guide the interpretation of multimodal signals. Figure 1 provides a visual summary of this structure, illustrating the relationships between the various components.

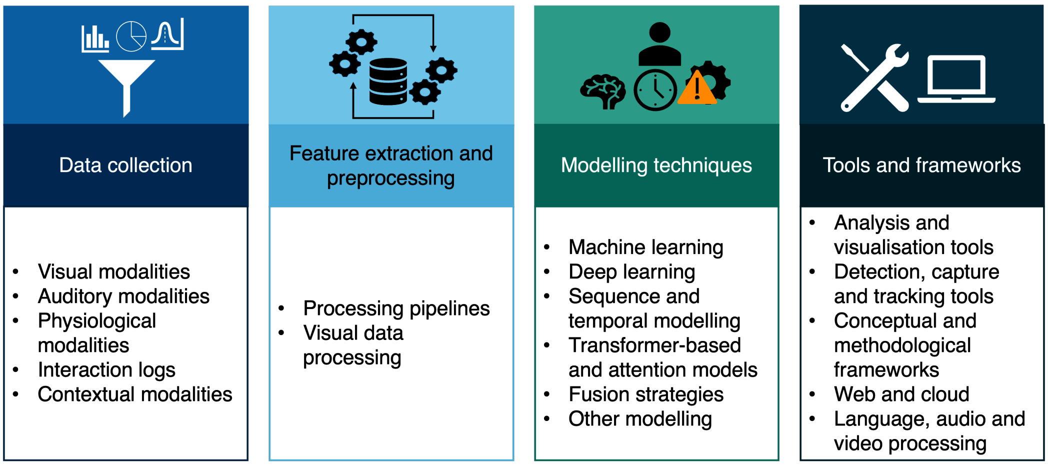


Figure 1. Key methodological components of multimodal learning analytics (MMLA), from data collection to tools and frameworks.

3.1. Data Collection Modalities

MMLA leverages a diverse array of data collection modalities to capture the cognitive, affective, and behavioural dimensions of learning. The primary modalities include visual, auditory, physiological, interaction logs, and contextual data.

The distribution of studies within this section was guided by the primary data type analysed in each work. Each paper was categorised based on its core focus, whether it involved visual data (e.g., video and eye-tracking), auditory signals (e.g., speech and paralinguistic features), physiological measures (e.g., EDA and heart rate), interaction logs (e.g., mouse clicks and keystrokes), or contextual information (e.g., spatial positioning and ambient conditions). This approach aimed to accurately reflect the diversity of data sources used in MMLA, ensuring comprehensive coverage across modalities while minimising overlap.

3.1.1. Visual Modalities

Visual data is the most widely used modality, commonly collected through video recordings, facial expression tracking, eye gaze analysis, and posture recognition. Studies frequently employ cameras to capture student behaviours in classrooms [34], analyse facial cues during online sessions [35], or use computer vision systems such as Kinect or OpenFace to extract body posture and expressions [3,36]. Eye-tracking is also prevalent in identifying attention patterns and engagement levels [37–39]. For example, some studies used dual eye-tracking (DUET) setups to understand collaboration dynamics [40], while others employed gaze analysis for cognitive assessment [16]. Visual modalities also include skeletal data capture using depth sensors or motion capture systems [41,42].

3.1.2. Auditory Modalities

Auditory data provides insights into discourse patterns, collaboration, and emotional tone through voice and paralinguistic signals. It is typically gathered using microphones in classroom or online settings [43–45]. Audio features such as Mel Frequency Cepstral Coefficients (MFCCs) and Voice Activity Detection (VAD) are used to analyse speech quality and emotional engagement [46,47]. In some cases, speech is synchronized with video to investigate participation and discussion quality [9,48]. Mobile setups like AirPods or lapel mics have also been applied to record in situ conversations [49].

3.1.3. Physiological Modalities

Physiological sensors enrich MMLA by capturing internal states such as stress, attention, or cognitive load. EDA, heart rate variability (HRV), and EEG are the most commonly used signals [50,51]. Wearables like the Empatica E4 wristband have been used to measure arousal and engagement [41,50,52]. Some platforms even infer physiological signals from webcams (e.g., heart rate via skin tone changes) [53]. These data provide a window into affective states and are often triangulated with behavioural data to enrich the interpretation [4,54,55].

3.1.4. Interaction Logs

Interaction logs are critical for tracking learners' engagement with digital platforms. These include mouse clicks, keystrokes, scrolls, touch events, or system events recorded during interaction with educational software [16,56,57]. Logs enable the detection of problem-solving strategies, learning paths, or system navigation patterns. Some studies employed custom platforms (e.g., CoTrack or Knowledge Forum) to record and analyse collaborative writing or problem-solving [58–60]. Logs have also been combined with other modalities, like gaze and voice, to assess learning effectiveness [40].

3.1.5. Contextual Modalities

Contextual data provides environmental or spatial information that influences learning, such as seating arrangements, physical proximity, location, and ambient conditions. These are often collected through positioning sensors, manual observations, or system metadata [52,61,62]. For example, spatial positioning with Pozyx tags or motion capture enables analysis of embodied interaction in classrooms [52,61]. Observational notes and metadata (e.g., timestamps, roles) further enrich interpretation by situating learner actions in real-world or virtual environments [57,63].

3.2. Feature Extraction and Preprocessing

3.2.1. Overview of Processing Pipelines

Feature extraction and preprocessing are foundational steps in MMLA, serving to transform raw data into structured, analysable inputs suitable for modelling. These processes vary according to the type and source of data but often involve a combination of cleaning, synchronisation, segmentation, and dimensionality reduction.

Several studies implement structured pipelines that begin with noise reduction and data transformation. For example, raw physiological signals such as EDA or EEG are frequently filtered and normalised to improve data quality prior to analysis [39,64]. This step is often followed by resampling or interpolation to handle missing values and align data streams temporally [65]. The integration of heterogeneous data sources, such as video, audio, and interaction logs, necessitates careful synchronisation and time alignment across modalities [66].

Many frameworks employ segmentation methods to divide continuous data into analysis-ready units. These units may be defined by fixed time windows [47], event-based triggers [45], or learning phases such as planning and reflection [67]. Some approaches leverage automated tools to streamline this process, such as CoTrack for multimodal annotation and segmentation [68].

Dimensionality reduction techniques like principal component analysis (PCA) and factor analysis are often employed to simplify high-dimensional feature spaces while retaining relevant variance [69]. Feature selection algorithms, such as RELIEF-F, have also been applied to optimise model inputs [49]. In certain cases, attention mechanisms or hierarchical representations are used to refine feature representations and reduce the need for manual engineering [8,69–71].

Finally, various frameworks stress the ethical and privacy implications of preprocessing, particularly when sensitive learner data are involved. Strategies include anonymisation, localised processing, or limiting raw data storage [58,66]. These pipelines not only enhance model performance but also ensure data integrity and user trust throughout the analytic process.

3.2.2. Visual Data Processing

Processing of visual data in MMLA involves a range of techniques designed to extract meaningful features from videos, images, and spatial representations of learners' behaviours. Several studies utilised deep neural networks (DNNs) and convolutional models to extract high-level features from facial expressions, head pose, and posture information. In some cases, feature extraction was performed through VGG16 and ResNet-101 for facial and pose estimation, respectively, while Mediapipe provided landmark detection [35]. To ensure clarity and robustness, images can be cropped, resized, and normalised to reduce background interference and enhance classifier accuracy [72].

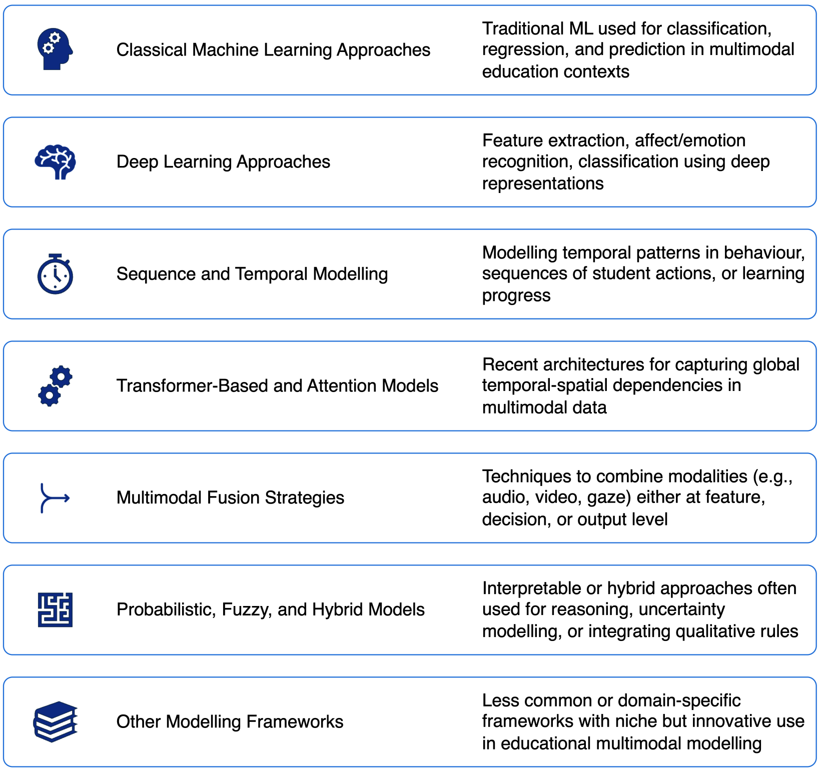
Some approaches focused on classroom video analysis using pyramid-based architectures and context-sensitive prediction modules [73]. Low-level features were extracted using Inception-ResNet-V2 to detect students' engagement with high precision. In other frameworks, facial features were processed frame-by-frame using OpenFace and OpenPose, yielding detailed facial action unit intensities and 3D keypoints to represent body pose in collaborative scenarios [74]. Standardisation of visual features within cross-validation splits helped mitigate bias across training and testing data [1].

Visual features were also analysed through blend shape coefficients from iPhone depth cameras to measure movement in key facial areas, such as eyelids and mouth, enabling proxies for cognitive or affective states [49]. Motion capture systems employing infrared cameras and Smart Tracker software enabled precise detection of movement features by translating marker coordinates into behavioural indicators [75]. To support visual emotion detection, video frames were preprocessed using tools such as DeepFace, with additional transformation to gray-scale and resizing before analysis [76].

In some cases, spatial data extracted from videos were synchronised with interaction metrics and normalised to preserve behavioural fidelity across different sessions [65]. Posture analysis was facilitated through skeletal heat-maps and salient video frames fused in a common embedding space, providing robust feature representations for model training [15]. Additionally, optical flow algorithms such as TV-L1 were employed to analyse motion in physical education contexts, particularly when evaluating embodied learning [42]. These diverse visual data processing strategies contribute significantly to capturing the complexity of learner behaviours in multimodal educational environments.

3.3. Modelling Techniques

A wide variety of modelling techniques have been employed, each addressing different analytical needs and characteristics of multimodal data. These techniques range from traditional ML classifiers to more recent deep and sequential models, as well as hybrid and probabilistic approaches. Their selection often depends on the nature of the task (e.g., prediction, classification, temporal analysis), the structure of the data, and the level of interpretability required for educational applications. Figure 2 presents an overview of the main categories identified in the reviewed literature, along with the primary purpose or application of each modelling approach within educational settings.



Classical Machine Learning Approaches	Traditional ML used for classification, regression, and prediction in multimodal education contexts
Deep Learning Approaches	Feature extraction, affect/emotion recognition, classification using deep representations
Sequence and Temporal Modelling	Modelling temporal patterns in behaviour, sequences of student actions, or learning progress
Transformer-Based and Attention Models	Recent architectures for capturing global temporal-spatial dependencies in multimodal data
Multimodal Fusion Strategies	Techniques to combine modalities (e.g., audio, video, gaze) either at feature, decision, or output level
Probabilistic, Fuzzy, and Hybrid Models	Interpretable or hybrid approaches often used for reasoning, uncertainty modelling, or integrating qualitative rules
Other Modelling Frameworks	Less common or domain-specific frameworks with niche but innovative use in educational multimodal modelling

Figure 2. Classification of modelling techniques used in MMLA, along with their primary analytical purposes in educational contexts.

3.3.1. Classical Machine Learning Approaches

Support Vector Machine (SVM): SVMs have been widely applied in MMLA due to their robustness in handling high-dimensional data and their effectiveness in classification tasks. In the context of engagement detection and behavioural analysis, SVMs have been used to classify eye states from visual input [76], detect patterns in socio-spatial behaviour [77], and model facial expressions using Gabor features [73]. Several studies have integrated SVMs with ensemble approaches to improve predictive accuracy across educational tasks. For instance, SVMs were employed alongside decision trees and neural networks to estimate collaboration quality and analyse behavioural data [13,78]. In simulation-based environments, they have contributed to identifying relationships between spatial-procedural behaviours and team performance [65]. SVMs have also been used in conjunction with random forest (RF) and gradient boosting in informal learning contexts, such as modelling visitor engagement in museums using multimodal sensor data [1].

Moreover, the adaptability of SVM kernels has been leveraged in predictive modelling tasks, with linear, radial, and polynomial kernels employed to enhance stability and accuracy [41]. In content analysis and student artefact evaluation, SVMs have been integrated into ML pipelines using open-source tools like scikit-learn [79]. The technique has also supported music-related learning applications, such as genre classification and intelligent labelling, contributing to the development of adaptive educational tools [80]. Overall, the use of SVMs spans diverse educational domains and serves various analytical goals, from behavioural prediction to artefact quality estimation, illustrating their ongoing relevance in multimodal modelling frameworks.

Decision trees: decision tree-based models are frequently employed in multimodal educational research due to their interpretability, low computational cost, and capacity to handle diverse data types. Several studies have utilised standard decision trees or their ensemble variants to support modelling efforts in learning analytics. For instance, decision trees were applied to classify students based on their cluster frequencies and learning gains in engineering design tasks, highlighting their utility in tracing conceptual progression [81]. Similarly, they have been employed for rule generation and behavioural pattern prediction in student engagement studies, where their transparency has proven beneficial for educational practitioners [82]. In the context of SRL, decision trees have also served as part of quantitative learning analytics approaches using mobile-generated data to identify student types [83].

More complex implementations, such as Gradient Boosted Decision Trees (GBDTs), have demonstrated strong predictive performance in applications like dropout detection in K-12 online environments, particularly when combined with data augmentation techniques to mitigate class imbalance [84]. Decision trees have also been adopted for multimodal modelling of affective and cognitive engagement, integrating features derived from brain waves, eye movement, and emotion metrics [54]. In collaborative learning contexts, they are frequently included in comparative analyses with other ML methods to assess modelling robustness in noisy environments [13,48,78].

Random Forest (RF): RF algorithms are frequently adopted in MMLA for their robustness, scalability, and ability to handle heterogeneous data. They have been employed for various educational tasks such as modelling engagement [85], predicting behavioural changes [86], assessing artefact quality [79], and evaluating collaboration quality [48]. Studies have also leveraged RFs to predict student progression from socio-spatial behaviours [77], estimate lesson performance [3], and model simulation-based team behaviours [65]. In more specialised applications, hierarchical RFs were used to estimate learning interest from fused multimodal features [87], while ensemble comparisons showed RFs as strong performers in tasks like public speaking assessment [88], artefact classification [1], and collaboration modelling with noisy data [59,78]. Their integration into multimodal pipelines has been shown to improve classification reliability in educational data mining [13], often in combination with data augmentation and feature normalisation techniques [79]. Finally, RFs have also supported exploratory modelling of engagement and learning outcomes through mobile and classroom-based multimodal recordings [65].

K-Nearest Neighbour (KNN): KNN has been applied in multimodal educational modelling for classification tasks due to its simplicity and effectiveness in handling small to medium-sized datasets. It has been used to analyse students' spatial-procedural behaviours in simulation-based learning, alongside other common classifiers such as logistic regression and RFs [65]. In broader MMLA frameworks, KNN has supported predictions of behavioural changes in students with special educational needs, contributing to model ensembles evaluated through standard validation protocols [13]. Additionally, KNN has proven valuable in identifying low-progress learners based on socio-spatial behaviours, offering a non-parametric alternative to linear classifiers in student performance prediction [77].

Logistic/linear regression: logistic and linear regression models offer interpretable solutions for predicting binary and continuous outcomes in MMLA contexts. Logistic regression has been used to model students' spatial-procedural behaviours in simulation-based learning [65], estimate collaboration quality from multimodal indicators [78], and identify low-progress students based on socio-spatial patterns [77]. In parallel, linear regression approaches, including multiple and ridge regression, have been employed to analyse collaborative problem-solving scores [67], assess learner performance and personalise support in instructional environments [89], and predict lesson activity outcomes by linking discourse patterns with action unit metrics [3].

Naive Bayes: naive Bayes classifiers serve as efficient probabilistic models for estimating collaboration quality and predicting learner behaviours in MMLA. These probabilistic models have been included in comparative evaluations alongside other ML algorithms to model constructs in classroom collaboration scenarios [78], support predictive modelling in special education contexts [13], and analyse multimodal signals—such as audio, video, and physiological data—to assess collaborative behaviours in educational settings [48].

Gradient Boosting (XGBoost, AdaBoost): these gradient boosting techniques have been applied in MMLA to model collaboration quality and improve classification accuracy. These approaches have been evaluated in regression-based models to assess learning sub-dimensions [59], as part of diverse classifier ensembles in simulation-based learning settings [13], and to estimate collaboration outcomes using robust model development pipelines [78]. Additionally, AdaBoost has been explored alongside other ML algorithms in the modelling of multimodal collaboration scenarios [48].

3.3.2. Deep Learning Approaches

Convolutional Neural Network (CNN): several studies employ CNNs as core components of multimodal systems, integrating them with complementary techniques such as Long Short-Term Memory (LSTM) and Transformer networks to capture both spatial and temporal features [72,73]. CNNs have been used for emotion recognition tasks in classroom contexts, often achieving high accuracy by leveraging multimodal inputs like speech and facial expressions [8,90]. In music education, CNNs were employed to extract features from audio signals, forming part of complex architectures that also incorporate recurrent layers and reinforcement learning modules [91]. Additionally, adaptive deep CNNs have been applied to facial expression recognition using entropy-based optimisation and transfer learning (TL) strategies [71]. Some studies have used CNNs in combination with fuzzy logic to model students' emotional states in language learning [90], while others embedded them within pipelines for analysing learner artefacts in learning analytics platforms [79]. Overall, these approaches demonstrate the versatility of CNNs in handling various data modalities and modelling needs across educational research.

Multi-Layer Perceptron (MLP): for instance, MLPs have been used to classify student behaviours in special education contexts, alongside other classifiers within standard ML pipelines, including balancing, cross-validation, and hyperparameter tuning [13]. In another study, MLPs were employed at the decision-fusion level, integrating facial and EEG features to infer cognitive and emotional states, showcasing their effectiveness in multimodal engagement detection frameworks [35].

Deep Neural Network (DNN): DNNs have been utilised in MMLA to address both classification and regression tasks. In one study, DNNs were implemented as part of an ML pipeline to model behaviour changes in students with special education needs, demonstrating the capacity of deep models to capture complex patterns in multimodal data [13]. Another investigation employed a two-stage approach where DNNs were applied to predict artefact quality in project-based learning, outperforming traditional models and confirming their potential for evaluating educational outcomes based on diverse data sources [92].

Autoencoders: autoencoders function as unsupervised feature reduction mechanisms in predictive student modelling, especially when processing multimodal data from game-based learning environments. In one study, autoencoders were used alongside TL and multi-task learning (MTL) to enhance the accuracy of models predicting post-test performance and student interest [93]. The approach integrated diverse modalities such as gameplay logs, eye gaze, facial expressions, and reflective writing, enabling a comprehensive representation of student learning behaviours.

3.3.3. Sequence and Temporal Modelling

Recurrent Neural Network (RNN): RNNs are useful for capturing temporal dependencies in sequential data, particularly in the context of educational applications involving music. One study integrated RNNs within a hierarchical note-level language model as part of the MusicARLtrans Net architecture, which combines convolutional layers, recurrent layers, and reinforcement learning to model musical elements efficiently and adaptively [91].

Long Short-Term Memory (LSTM): One study employed LSTMs to classify dialogue acts in game-based learning by processing facial action units and game trace logs, demonstrating superior performance compared to CRFs [70]. In classroom activity detection, LSTMs were integrated into a Siamese neural architecture alongside GRUs and transformers, facilitating deep representation learning of audio signals [94]. An enhanced variant, the Information Block Bidirectional LSTM (IB-BiLSTM), was proposed for sentiment analysis in animated online educational content, improving sequence segmentation and convergence performance [95]. Other works utilised LSTMs to analyse learner-generated artefacts [79], detect engagement via multimodal behavioural signals [73], and compare them to alternatives like Temporal Convolutional Network (TCN) for long-term sequential encoding [72].

Gated Recurrent Unit (GRU): GRUs have been applied in educational modelling tasks that involve sequential data streams, offering a computationally efficient alternative to traditional recurrent architectures. One study proposed a multimodal, multi-task stealth assessment framework that integrates GRUs to analyse game trace logs and student reflections, employing multiple fusion strategies—early, late, and hybrid—and achieving effective predictions of post-test scores and reflection ratings [96]. In a separate study focused on classroom activity detection, GRUs were incorporated within a Siamese neural framework alongside LSTMs and transformers to capture temporal dependencies in audio representations, demonstrating the flexibility of GRUs in deep sequence modelling contexts [94].

Temporal Convolutional Network (TCN): TCNs have emerged as a viable alternative to recurrent models for sequential data modelling in educational settings. A study highlighted the ability of TCNs to extract features from sequential inputs efficiently while acknowledging challenges in multimodal recognition. The research integrated TCNs with other architectures, such as transformers and CNNs, to enhance the detection of student engagement, thereby showcasing the effectiveness of hybrid deep learning strategies in modelling complex educational behaviours [72].

Markov Model (MM): they have been employed to analyse sequential data by capturing the probabilistic transitions between discrete states, enabling the identification of patterns and system evolution over time. In the context of learning analytics, they offer interpretable transition probabilities that can illuminate behavioural dynamics and support data-driven decisions. For instance, ref. [50] used MMs to model transitions between sub-codes in educational sequences, revealing significant differences in transition patterns and demonstrating the model's effectiveness in uncovering learning behaviours.

Hidden Markov Model (HMM): these models enable the identification of unobservable learning states through observable patterns, making them particularly suitable for analysing dynamic learner behaviours. For instance, ref. [5] applied HMMs to explore productivity and interaction challenges in tabletop environments, while ref. [97] utilised them to classify time-variant behaviours and behavioural states in learning-by-teaching scenarios. Additionally, ref. [36] used HMMs in conjunction with motion vector analysis to uncover latent motion states, combining them with optimal matching algorithms to model temporal hand movement data. These studies underscore the value of HMMs in revealing underlying behavioural patterns and enhancing MMLA.

Conditional Random Fields (CRFs): In [70], CRFs were used alongside LSTMs to analyse multimodal data streams comprising facial action units, physiological skin responses and game trace logs. Although LSTMs outperformed CRFs in terms of predictive accuracy, the study demonstrated the complementary role of CRFs in sequence labelling tasks and highlighted their potential for modelling structured interactions in game-based learning environments.

3.3.4. Transformer-Based and Attention Models

Transformers: Transformer-based architectures have gained prominence in multimodal educational modelling due to their capacity to capture long-range dependencies across modalities and time. In [94], a Siamese neural framework incorporates Transformers alongside LSTM and GRU models for classroom activity detection, enhancing representation learning. Similarly, ref. [2] leverages Bidirectional Encoder Representations from Transformers (BERTs) to predict collaborative problem-solving skills based on student communications. In the context of engagement detection, Transformers are integrated with TCN and CNN in a hybrid approach to improve performance in multimodal environments [72]. For physical education contexts, ref. [42] introduces a Transformer-based structure (TEMS) to extract temporal features from skeletal data, while ref. [15] proposes HRformer, a dual-stream Transformer framework that unifies skeletal and video inputs through self-attention and cross-attention mechanisms for effective behaviour recognition.

Attention mechanisms: these models have been increasingly employed to enhance multimodal modelling by enabling models to prioritise salient features and manage noisy inputs effectively. In [69], an attention-optimised DNN is proposed for predicting teaching quality, demonstrating improved accuracy through the integration of complex network theory and multimodal data sources such as academic performance, behavioural cues, and psychological attributes. Similarly, ref. [15] utilises both self-attention and cross-attention within the HRformer framework to align skeletal and video modalities, facilitating robust behaviour recognition in physical education contexts through enhanced multimodal fusion strategies.

3.3.5. Multimodal Fusion Strategies

Multimodal fusion strategies are widely employed to integrate diverse data sources for enhanced modelling. Studies have adopted early, late, and hybrid fusion techniques to combine modalities like EEG, facial expressions, gaze, and body posture [51,96,98]. Decision- and feature-level fusion methods have also been applied to improve recognition of engagement and mental states [35,46,49,99].

Some approaches integrate deep learning and attention mechanisms to align multimodal inputs, as seen in Transformer-based models and dual-stream frameworks for behaviour recognition [15,42,100]. Techniques such as PCA-based, entropy-optimised, and ensemble fusion further enhance model robustness and accuracy [59,71,101]. Software tools are used for feature extraction in hybrid models combining audio and visual cues [8,102].

Several studies have proposed innovative methods for integrating multimodal data in educational settings. For instance, the use of fast-slow neural networks has been explored for detecting student engagement, effectively combining asynchronous data streams like EEG and pose information for accurate on- and off-task state prediction [98].

Hierarchical fusion approaches, such as hierarchical random forest for attention estimation and CRF for affective learning, have also been utilised to model student interest, achieving notable accuracy improvements [87].

Additionally, complex network models incorporating attention mechanisms have been employed to capture teaching dynamics and student interactions, demonstrating superior predictive accuracy in classroom settings [69].

For the analysis of group interactions, some approaches rely on GRUs to integrate written reflections and game trace logs, implementing both early and late fusion strategies for enhanced student assessment [96].

Hybrid approaches that combine physical and digital modalities, including facial expressions, head posture, eye gaze, and EEG signals, have also been explored to create comprehensive engagement models for educational applications [51].

Moreover, deep learning models like the state-aware deep item response theory integrate spatial features extracted from facial videos to assess student cognitive states, effectively bridging the gap between traditional assessment and real-time multimodal analytics [101].

Other studies have implemented ensemble models that combine facial and body tracking for engagement detection, integrating sentiment analysis from audio signals for a more comprehensive understanding of student interactions [76].

Finally, the use of novel algorithms like adaptive entropy minimisation and hybrid deep restricted Boltzmann machines has been proposed to optimise feature extraction and fusion, enhancing the generalisation ability of multimodal learning models [71].

3.3.6. Probabilistic, Fuzzy, and Hybrid Models

Bayesian networks: Bayesian networks have been applied to enhance the modelling in complex multimodal contexts. One study incorporated a hybrid deep restricted Boltzmann machine with Bayesian network methodologies to improve facial expression recognition and multimodal teaching behaviour analysis, showcasing the potential of probabilistic reasoning in data fusion tasks [71]. Another study applied Bayesian Ridge Regression (BRR) alongside other regression models to analyse collaborative problem-solving scores, highlighting its role in capturing dependencies among multiple variables in learning scenarios [67].

Fuzzy logic: fuzzy logic has been incorporated to address the inherent uncertainty and imprecision in human behaviour and emotion recognition. One study employed fuzzy mathematics in the evaluation framework for English translation teaching, enhancing the interpretability of emotion recognition models in classroom contexts [90]. Similarly, fuzzy conceptions were used for classifying musical intelligence in a study that combined AI and multimodal techniques to support music education, demonstrating fuzzy logic's role in handling vague or complex categorisations [80].

Latent Profile Analysis (LPA): LPA can serve as a statistical approach to uncover distinct behavioural patterns among students over time. In one study, LPA was used alongside HMMs to analyse students' time-variant behaviours based on multimodal data, facilitating the identification of behavioural profiles and their association with learning outcomes [97].

Hybrid models: hybrid modelling techniques have been increasingly adopted to address the complexity of multimodal educational data. One study introduced a behavioural disengagement detection framework that integrated deep language models with facial expression and gaze features, achieving higher accuracy than unimodal approaches [100]. Another work employed a hybrid deep restricted Boltzmann machine, combining deep learning with Bayesian networks and entropy minimisation to improve facial expression recognition in teaching scenarios [71]. Similarly, a hybrid architecture combining TCNs, Transformers, and CNNs was proposed to overcome limitations of individual models in student engagement detection [72].

3.3.7. Other Modelling Frameworks

Epistemic Network Analysis (ENA): ENA has been applied to examine the co-occurrence and sequencing of coded features in learning processes, integrating temporality and structure into network-based interpretations. This method uses techniques such as Means Rotation and Singular Value Decomposition to simplify the dimensionality of behavioural connections. In one study, ENA was combined with a BERT-based predictive model to analyse student communications and forecast collaborative problem-solving skills, further enriched by MMLA that included verbal, spatial, and physiological data [2].

Inverse Reinforcement Learning (IRL): IRL has been employed to model SRL by inferring the reward functions underlying students' observed behaviours, enabling the assessment of skill mastery and the delivery of personalised interventions. One study integrated IRL into a teacher dashboard system, MetaDash, to support instructional decision-making through real-time visualisations of learning trajectories [103].

Multi-Task Learning (MTL): MTL plays a significant role in enhancing predictive accuracy through the shared representation of related tasks. For example, one study introduced a multimodal, multi-task stealth assessment framework that combined game trace logs and written reflections. Within this setup, MTL was leveraged to concurrently predict post-test scores and reflection ratings, allowing the model to generalise more effectively across distinct yet interconnected learning outcomes [96].

Transfer Learning (TL): TL has emerged as a valuable approach in multimodal educational modelling, enabling systems to leverage prior knowledge from external datasets. In one study, TL was incorporated into a hybrid deep learning framework for facial expression recognition, contributing to improved generalisation and performance in multimodal teaching behaviour analysis [71]. Similarly, TL was employed to enhance predictive models in a game-based learning context by reusing learned representations across different tasks and modalities, including gameplay data, facial expressions, and reflective text [93].

Adaptive factorisation machines: the study by [57] presents an individualised Adaptive Factorisation Machine (iAFM) model, extending traditional AFM approaches by employing linear mixed models to estimate student-specific learning rates in problem-solving tasks. Through binomial regression, the model predicts whether students can solve steps without tutor assistance, accounting for initial difficulty and the rate of learning per knowledge component. This personalised approach enables the differentiation of student learning trajectories, supporting the evaluation of teacher practices in AI-supported educational environments.

Rule-based and expert systems: the study by [82] explores the use of rule-based systems and decision trees for classifying student engagement and behavioural patterns, emphasising their interpretability and suitability for educational contexts. The approach highlights the efficiency and clarity of rule generation from decision tree models, offering educators an accessible means to understand and apply ML outcomes in practice.

3.4. Tools and Frameworks

3.4.1. Multimodal Analysis and Visualisation Tools

A wide range of tools and frameworks have been developed to support the visualisation, annotation, and interpretation of multimodal data in educational research. These instruments include systems for capturing group interactions (e.g., CoTrack, Moodoo), platforms for dashboard-based decision support (e.g., MetaDash, KNIGHT), toolkits for data collection and synchronisation (e.g., EZ-MMLA, IMotions Lab), and software for behavioural coding and annotation (e.g., HELP, LEDLAB). Others focus on collaborative learning environments (e.g., Group Scribbles, MTClassroom), spatial tracking (e.g., Moodoo), or advanced data alignment and stream integration (e.g., STREAMS, colaborative learning mechanisms (CLM)). Table 1 summarises these tools along with their purposes and representative works.

Table 1. Tools for multimodal analysis and visualisation, their purposes, and references.

Tool	Purpose	Works
Collaborative learning mechanism (CLM)	Framework analysing collaborative discussion and coordination in group learning	[40]
CoTrack	Collaboration visualisation and analysis	[58,68,78,104]
CognitOS	Student attention and emotion monitoring	[87]
EDA Explorer	EDA signal analysis	[105]
EZ-MMLA	Web-based multimodal data collection	[53]
FACT	Multimodal data capture in classrooms	[106]
Group Scribbles	Classroom collaboration support	[106]
HELP coding framework	Video annotation for disengagement detection	[100]
IMotions Lab	Eye-tracking and physiology synchronisation	[58]
KNIGHT framework	Personalised learning analytics	[107]
LEDLAB	EDA signal decomposition	[64]
MetaDash	Dashboard for teacher decision support	[103]
Multimodal Data Value Chain (M-DVC)	Conceptual model to process and transform multimodal learning data	[47,66,108]
MLeAM	Machine learning (ML)-enhanced learning analytics model	[47,108]
Moodoo	Spatial behaviour tracking in classrooms	[109]
MTClassroom	Table-top collaborative environment	[106]
SNAPP framework	Online engagement tracking	[107]
STREAMS	Temporal alignment of multimodal streams	[16,110]
TeachLivE	Virtual reality classroom simulation	[111]
Web-based Programming Grading Assistant (WPGA)	Augmented feedback system	[106]
Writer(s)-Within-Community Model	Writing analytics in social contexts	[112]
Wits Intelligent Teaching System	Real-time feedback via emotion/engagement	[113]

3.4.2. Detection, Capture and Tracking Tools

To support MMLA, a wide array of tools has been employed for detecting, capturing, and tracking learners' behaviours and physiological signals. These include integrated platforms like iMotions for synchronising affective, visual, and physiological measures; motion capture devices such as Kinect and LELIKËLEN for embodied learning and gesture analysis; and specialised toolkits like MOVES-NL and PyramidApp for collecting skeletal, electrodermal, and behavioural data. Computer vision-based approaches leverage libraries such as OpenCV, OpenFace, and OpenPose, often combined with deep learning architectures like EfficientNetB2 or ResNet-101, to enhance the accuracy of facial expression, posture, and emotion recognition. Additional tools such as SMART Notebooks and VAD enable the monitoring of content interaction and verbal engagement. Table 2 summarises these tools, their purposes, and representative works.

Table 2. Detection, capture and tracking tools used in MMLA studies.

Tool	Purpose	Works
EfficientNetB2	Deep learning architecture used for emotion recognition by analysing video and audio inputs	[76]
iMotions	Integrated platform for synchronised recording and analysis of multimodal data such as eye-tracking, physiological signals, and facial expressions	[37,53,58]
Kinect	Motion sensing and body-tracking tool used for gesture recognition, movement capture, and embodied learning analysis in classroom and game environments	[12,36,41,59,88,114–116]
LELIKËLEN system	Tool for organising and managing multimodal data captured with Kinect during student presentations	[115]
MOVES-NL	Interactive tool that collects physiological and skeletal movement data using devices such as wristbands and depth cameras for embodied learning feedback	[50]
NISPI framework	Framework for analysing video-recorded classroom interactions by coding nonverbal physical participation to assess engagement	[75]
OpenCV	Computer vision library used for extracting real-time facial landmarks, posture and gesture features	[35]
OpenFace	Toolkit for facial behaviour analysis, including action units, gaze, and head pose estimation	[36,74,100]
OpenPose	Framework for detecting human body pose and movements based on deep learning techniques	[74]
PyramidApp	Web-based platform used for deploying computer-supported collaborative learning (CSCL) activities while collecting electrodermal, behavioural, and observational data	[10]
ResNet-101	Deep convolutional neural network (CNN) used for estimating head pose and recognising facial expressions in student videos	[35]
SMART Notebooks	Educational tool supporting multimodal teaching; used to track visual and content-based engagement during lessons	[117]
Voice Activity Detection (VAD)	Method for separating speech and non-speech segments in audio data, applied for detecting turn-taking and verbal engagement	[50]

3.4.3. Conceptual and Methodological Frameworks

Numerous conceptual and methodological frameworks have been introduced in MMLA research to guide the design, interpretation, and implementation of multimodal systems. These frameworks support diverse objectives, from modelling collaborative learning and personalising feedback to structuring the collection and integration of heterogeneous data. Examples include domain-specific models like the KNIGHT framework for personalised learning analytics and FESSA for sentiment-based student feedback, as well as more generalisable structures like CLM and DUET for analysing collaborative dynamics. Some frameworks are tailored for mobile or embodied contexts, such as Mobile Multimodal Learning Analytics (MOLAM) and MaTHiSiS, while others like transmodal analysis (TMA) and transmodal ordered network analysis model (T/ONA) offer theoretical underpinnings for analysing transmodal and ontological dimensions. Table 3 summarises the main frameworks, their intended purposes, and associated studies.

Table 3. Conceptual and methodological frameworks in MMLA.

Tool	Purpose	Works
AI-enhanced learning assistant (AIELA)	Framework for enhancing adaptive intelligent environments through multimodal feedback mechanisms	[118]
Align Before Fuse (ALBEF)	Vision-language model adapted to support alignment and fusion in multimodal educational settings	[91]
AmbiLearn	Ambient intelligence-based framework for interpreting learner engagement and performance	[119]
ASSURE	Instructional design framework applied to multimodal learning environments to structure content delivery and evaluation	[120]
CLM	Conceptual framework for analysing collaborative learning processes using synchronised multimodal data	[40]
Dual eye-tracking (DUET)	Methodological framework for disentangling user engagement and task performance in collaborative contexts	[40]
FESSA	Sentiment-aware feedback framework based on AI and multimodal sentiment analysis	[99]
KNIGHT framework	Personalised learning analytics model using multimodal indicators to adapt feedback and instruction	[107]
MaTHiSiS	Multimodal learning framework for adaptive education using robotics and affective computing	[116]
Mobile Multimodal Learning Analytics (MOLAM)	Conceptual model for collecting and analysing multimodal data via mobile devices in learning scenarios	[83]
Transmodal Analysis (TMA)	Theoretical framework for identifying transitions and synchrony across different modalities of learning	[57]
Transmodal ordered network analysis model (T/ONA)	Ontological model for representing multimodal learning events and relationships among modalities	[57]
Writer(s)-Within-Community Model	Framework for analysing writing processes in collaborative settings using multimodal evidence	[112]

3.4.4. Web and Cloud Technologies

Web-based and cloud technologies have become increasingly central to the development and deployment of MMLA systems, facilitating scalable data processing, real-time interaction, and integration of diverse services. Cloud APIs, such as GPT-4-turbo and Google Cloud Speech-to-Text, enable intelligent analysis and transcription of multimodal content at scale. Web platforms like CoTrack, MetaDash, and EZ-MMLA provide synchronous data visualisation, dashboard analytics, and multimodal data collection through browser-based interfaces. PyramidApp extends this functionality to support collaborative learning scenarios enriched with behavioural and physiological monitoring. These technologies lower the barriers for implementing multimodal analytics in diverse educational contexts and are summarised in Table 4.

Table 4. Web and cloud technologies for MMLA.

Tool	Purpose	Works
API GPT-4	Cloud-based generative language model used for intelligent feedback and dialogue generation in adaptive learning environments	[118]
CoTrack	Web platform for real-time visualisation and coordination of multimodal data in collaborative learning settings	[58,68,78,104]
EZ-MMLA	Browser-based system for synchronised collection and monitoring of multimodal learning data across distributed environments	[53]
Google Cloud Speech-to-Text	Cloud service for real-time audio transcription and speech recognition in educational video and classroom recordings	[76]
MetaDash	Online dashboard for supporting teacher decision-making by integrating multimodal process data	[103]
PyramidApp	Web-based tool for deploying CSCL activities while collecting behavioural and physiological signals from participants	[10]

3.4.5. Language, Audio and Video Processing

A range of tools have been applied in MMLA to extract and analyse information from text, audio, and video sources. For natural language processing, models such as embeddings from language model (ELMo) generate deep contextual embeddings that help identify emotional and semantic patterns in educational discourse. Audio-related tools include FFmpeg, which facilitates audio extraction and preprocessing from multimedia files, and Google Cloud Speech-to-Text, which supports automatic speech recognition for analysing classroom interactions. EmotionCLIP enables joint analysis of visual and textual data to infer learners' affective states. On the visual side, OpenCV is widely used for video processing and feature detection, while OpenFace supports facial landmark and gaze analysis. Additionally, OpenPose provides robust body pose estimation, allowing researchers to capture posture and gesture dynamics in educational environments. An overview of these tools and their respective applications is provided in Table 5.

Table 5. Language, audio and video processing tools in MMLA.

Tool	Purpose	Works
Embeddings from language model (ELMo)	Deep contextualised word embeddings used for emotion and discourse analysis in educational texts	[96,100]
FFmpeg	Multimedia framework used for audio extraction and format conversion from video sources	[76]
Google Cloud Speech-to-Text	Cloud service for automatic speech recognition and transcription in classroom recordings	[76]
EmotionCLIP	Vision-language model for recognising emotions from video and text data	[121]
OpenCV	Computer vision library for real-time video processing and feature extraction	[35]
OpenFace	Facial analysis toolkit for detecting gaze, head pose, and action units	[36,74,100]
OpenPose	Pose estimation framework for tracking body movements in video data	[74]

4. Applications in Education

The reviewed literature reveals a rich and diverse spectrum of application domains, reflecting the versatility of multimodal approaches in addressing pedagogical challenges, enhancing teaching practices, and supporting learners. Based on a thematic classification of the selected works, seven major categories of application were identified. These categories reflect both the breadth and specificity of multimodal learning applications in contemporary educational research.

The identification of the seven application categories presented in this section was based on a structured thematic analysis of the selected studies. This process began with an initial open coding phase, where each study was manually reviewed to capture its primary educational focus, intended learning outcomes, and core application domain. During this phase, studies were grouped based on their main objectives, including enhancing engagement, supporting collaboration, personalisation, emotional monitoring, and inclusive education.

Following this initial coding, the studies were iteratively organised into broader thematic clusters, allowing for the identification of common patterns and emerging trends. This approach ensured that the resulting categories accurately reflected the diversity and specificity of multimodal learning applications in the reviewed literature, while minimising subjective bias. The final set of categories was refined through multiple rounds of manual review to ensure consistency and comprehensive coverage, resulting in the seven major categories presented in this section. A complete overview of the citations included in each category is presented in Table 6.

Table 6. Categorisation of reviewed papers by application domain.

Application Category	Works
Enhancing Engagement and Learning Performance	[1,5,14,34–36,42,43,46,47,50,54,55,57,62–64,71,72,74,77,79–83,85,91,94,95,99,103,106,108,109,111,112,116–118,120,122–162]
Supporting Collaborative Learning and CSCL	[2,9,10,34,40,44,45,47,48,52,58,59,61,67,68,75,78,99,100,104,119,131,163–179]
Educational Games and Simulation-Based Learning	[11,12,36–38,40,41,65,70,93,114,180]
Multimodal Assessment and Personalisation	[3,4,6,7,15,23,53,56,62,66,69,73,84–89,92,96–98,102,107,110,115,181]
Monitoring Emotional and Cognitive States	[8,10,49,60,76,90,99,101,105,113,116,145,182]
Special Needs and Inclusive Education	[13,18,39,51,116,183,184]
Ethical and Participatory Multimodal Learning Analytics (MMLA)	[7,25,77,108,185,186]

4.1. *Enhancing Engagement and Learning Performance*

This category encompasses studies that leverage multimodal data to increase students' motivation, participation, and academic outcomes. Applications include systems that provide real-time feedback, adaptive learning environments, and tools to monitor classroom dynamics. These works commonly integrate visual, auditory, and behavioural modalities to assess and stimulate learner activity, with the goal of improving knowledge acquisition, skill development, and retention.

For instance, Criswell et al. [143] conducted a study in a high-school chemistry classroom, demonstrating how multimodal analysis can track student positioning and sensemaking through gestures, verbal cues, and physical interactions. Their findings highlighted the importance of recognising diverse forms of student expression, including gestures, to support sensemaking and participation in scientific practices. Similarly, Alabdeli et al. [125] proposed a virtual reality (VR)-based idea mapping approach for second language learners, using multimodal data (e.g., behavioural logs, gaze tracking, and mouse movement) to identify learning patterns and improve vocabulary acquisition. Their model achieved high engagement rates and demonstrated the benefits of combining multiple data streams for real-time learning feedback.

4.2. Supporting Collaborative Learning and Computer-Supported Collaborative Learning (CSCL)

Articles within this category focus on the use of MMLA to understand, support, and assess collaborative learning processes, often in the context of CSCL. The emphasis is placed on group dynamics, communication patterns, coordination, and socially shared regulation of learning. These studies utilise modalities such as speech, gesture, spatial positioning, and interaction logs to uncover collaborative strategies, identify group roles, and support teacher interventions aimed at improving teamwork effectiveness.

For example, Echeverria et al. [9] developed the TeamSlides dashboard, which integrates spatial and audio data from high-fidelity healthcare simulations to support teacher-guided reflection in physical learning environments. Their study demonstrated the value of combining multimodal data streams to capture complex group dynamics, providing actionable insights for teachers to guide post-scenario debriefs and improve team coordination. Similarly, Hakami et al. [10] investigated the orchestration load experienced by teachers in scripted CSCL settings using a multimodal approach that included electrodermal activity sensors, log data, and self-reported questionnaires. Their findings highlight the importance of capturing both physiological and observational data to understand the cognitive demands placed on teachers when managing collaborative learning activities.

4.3. Educational Games and Simulation-Based Learning

This category includes research that explores the integration of MMLA within interactive learning environments, such as digital games or simulations. The main aim is to create interesting learning experiences that get students to think more deeply and improve their performance. Multimodal data from gameplay, physiological sensors, or motion tracking are commonly used to assess learning progress, emotional reactions, and behavioural patterns during simulation-based activities.

For example, Ding et al. [11] implemented a scaffolded game-based learning science unit with immersive VR to enhance the acquisition of scientific knowledge in middle school students. Their study found that students using the immersive VR version demonstrated greater improvement in targeted science knowledge compared to those using a non-immersive desktop version, highlighting the benefits of whole-body movements and multimodal meaning-making for conceptual understanding. Similarly, Na and Sung [12] designed an embodied augmented reality (AR) learning game for fourth-grade geometry education, integrating computer vision-based gesture recognition to support embodied learning of geometric concepts. Their results revealed significant cognitive gains in students' understanding of angles and shapes, demonstrating the potential of embodied interactions for enhancing spatial reasoning and engagement in mathematics.

4.4. Multimodal Assessment and Personalisation

Studies in this category aim to personalise educational experiences and improve assessment practices through multimodal evidence. By combining diverse data sources—such as eye tracking, audio, video, written reflections, and physiological signals—these approaches seek to construct holistic learner models that support adaptive feedback, early detection of learning risks, and tailored interventions. This line of work highlights the potential of MMLA to support formative and stealth assessment strategies in a range of disciplines.

For instance, Yang et al. [7] investigated the use of multimodal classroom data to support teacher reflection, integrating physiological signals, speech, and classroom interaction logs. Their study identified critical factors that influence teachers' ability to reflect on their practices, including the need for timely, context-rich feedback that aligns with teachers' privacy concerns and preferences for data sharing. Similarly, Li et al. [69] developed a multimodal teaching quality prediction model that combines complex network analysis and attention mechanisms to predict teaching outcomes. This approach achieved a high level of accuracy by integrating multimodal data such as grades, classroom behaviour, and psychological characteristics, demonstrating the potential of complex network structures for personalised education.

4.5. Monitoring Emotional and Cognitive States

This group of papers focuses on the detection and interpretation of learners' emotional and cognitive states using multimodal signals. Applications typically involve the recognition of engagement, confusion, frustration, or cognitive overload, using facial expressions, body posture, speech, or physiological signals such as EDA. The insights derived from these data streams are used to inform adaptive interventions, improve instructional strategies, or encourage emotional awareness in educational settings.

For example, Lin [8] developed a recognition model for classroom state monitoring in music education, combining depth-separable convolution and LeNet-5 networks to identify students' engagement levels based on facial expressions and behavioural data. The model achieved a recognition accuracy of 94.12%, with a significant reduction in computational complexity compared to traditional convolutional approaches, making it particularly suitable for real-time classroom analysis. Similarly, Zhang [90] proposed a multimodal emotion recognition framework for English translation teaching, integrating speech and expression recognition through convolutional neural networks. This approach demonstrated a high recognition accuracy of 86.4%, supporting the design of more responsive and personalised teaching interventions based on real-time emotional feedback.

4.6. Special Needs and Inclusive Education

This category refers to applications that aim to support learners with diverse needs, including disabilities, language delays, or age-related limitations. These studies apply multimodal technologies to create more accessible, differentiated, and supportive learning environments. Examples include inclusive learning platforms, mobile systems designed for learners with cognitive or physical challenges, or approaches that value gestural and non-verbal communication as valid learning expressions.

For instance, Chan et al. [13] developed an IoT-based system for Applied Behaviour Analysis (ABA) therapies targeting students with special education needs (SENs). Their system integrates physiological sensors, environmental data, and motion capture to enhance the predictive accuracy of behaviour change in SEN students, achieving an impressive 98% accuracy and 97% precision in behaviour change prediction. Similarly, Gunnars [18] explored the use of smartbands in primary education settings to monitor student stress levels and guide behavioural interventions. Their approach leverages HRV and EDA to provide real-time insights into student well-being, supporting more precise and contextually informed interventions for students with SEN.

4.7. Ethical and Participatory MMLA

Finally, a subset of the literature addresses ethical and participatory dimensions of multimodal analytics in education. These contributions advocate for learner-centred designs, transparent data practices, and enhanced informed consent mechanisms. They emphasise the importance of fairness, agency, and responsible use of analytics in learning environments, especially when dealing with sensitive multimodal data.

For example, Alwahaby and Cukurova [186] developed an ethical framework for MMLA, based on 60 in-depth interviews with educational stakeholders, including students, educators, and researchers. Their study identified critical ethical themes, such as privacy, transparency, and algorithmic bias, which must be addressed to encourage trust and support the ethical use of multimodal technologies in educational contexts. Similarly, Ouhaichi et al. [108] conducted a qualitative study to identify key design considerations for MMLA systems, including the integration of user-centric design principles, data protection, and the importance of contextual awareness in ethical decision-making. Their findings highlight the need for more comprehensive guidelines to ensure that MMLA systems are both technically robust and ethically sound.

5. Discussion

5.1. Summary of Key Findings

The comprehensive review of 177 studies revealed two primary themes: the diverse methodologies used to capture and process multimodal data, and the wide range of educational applications.

- Methodologies and tools: MMLA research employs a variety of data collection modalities, including visual (e.g., facial expressions, gestures), auditory (e.g., speech, tone), physiological (e.g., heart rate, electrodermal activity), interaction logs, and contextual data. These data streams are processed using feature extraction methods like computer vision, speech analysis, and sensor fusion, supported by machine learning algorithms and multimodal frameworks such as OpenPose, CoTrack, and iMotions. The studies reviewed highlight the importance of accurate synchronisation and alignment of heterogeneous data sources to model complex learning behaviours effectively.
- Applications in education: MMLA has been applied across diverse educational contexts, demonstrating its potential to enhance student engagement, support collaborative learning, personalise assessments, monitor emotional and cognitive states, and address special needs. For example, multimodal systems have been used to track student sensemaking in classrooms, support reflection in healthcare training, and monitor cognitive load in collaborative learning. Other studies have explored immersive learning through VR and AR, personalised assessments based on physiological and behavioural signals, and ethical considerations in data-driven education.

Together, these findings highlight the potential of MMLA to capture the complexity of learning processes, providing richer insights into learner behaviour and cognitive states. However, successful implementation often depends on the quality of the collected data, the robustness of feature extraction methods, and the interpretability of the resulting models.

5.2. Challenges and Methodological Limitations

Rapid advances in MMLA have led to increasingly sophisticated frameworks for capturing and interpreting complex learning behaviours. However, the integration of these systems into educational practice remains fraught with unresolved issues spanning technical, methodological, ethical and pedagogical dimensions. These challenges not only hinder current implementation but also serve as important points of reflection for future research and development.

A critical observation from this review is the relatively sparse consideration of multimodal approaches in many of the selected studies. This gap reflects a broader challenge in educational research, where the adoption of MMLA is often hindered by technical, financial, and institutional barriers. The high cost of multimodal sensors, the complexity of data synchronisation, and the need for specialised expertise limit the scalability of MMLA in real-world educational settings. Furthermore, many institutions lack the infrastructure required to support these data-intensive approaches, and there is often a reluctance to adopt unfamiliar methodologies without clear evidence of their pedagogical impact. This highlights the need for broader empirical validation and theoretical grounding to ensure that future MMLA models are not only technically sound but also contextually relevant, ethically transparent, and pedagogically meaningful. Without addressing these foundational gaps, the full potential of MMLA to enhance educational outcomes may remain unrealised.

Technically, multimodal integration remains a primary bottleneck. The heterogeneity of data sources from video and audio to physiological and log data makes synchronisation and alignment difficult. While promising frameworks have emerged, the latency and high dimensionality of features often hamper real-time applications, especially when using deep learning models that are both resource-intensive and opaque [91,107,150,174]. The labour-intensive nature of data pre-processing exacerbates these problems, requiring manual annotation or domain-specific configurations that are difficult to scale [97,134].

From a modelling perspective, the tension between accuracy and interpretability remains unresolved. Although DNNs often outperform traditional models in classification or prediction tasks [69], their lack of transparency hinders both user confidence and pedagogical applicability. Educators and stakeholders are often reluctant to rely on systems whose inner workings are not understandable or actionable [71]. In addition, the generalisability of findings is undermined by small or institution-specific data sets, calling into question the ecological validity of many experimental designs [129,140].

Another persistent concern is that of scalability. The robustness of MMLA tools in naturalistic classrooms is frequently compromised by environmental variability, such as lighting, occlusion and background noise, which affects the quality of video and audio inputs [8,78]. A significant number of educational institutions also lack the necessary infrastructure to support the use of data-intensive tools, which hinders their adoption in institutions with less advanced technological capabilities [18,85].

Ethical issues have become increasingly salient, particularly in light of concerns around surveillance and privacy. While some studies attempt to address consent and data security, there is often insufficient transparency regarding how data are processed, stored, and interpreted [7,58,99]. Teachers and students express ambivalence or resistance to persistent video and audio monitoring, especially when such practices lack clear pedagogical benefits or explanations [169].

A further layer of complexity arises in translating MMLA outputs into pedagogically meaningful actions. Despite the existence of systems capable of generating real-time feedback or visualisations, numerous teachers have reported difficulties in interpreting these outputs. These difficulties can be attributed to a number of factors, including time constraints, a lack of training, or a misalignment with curricular objectives [3,57]. Furthermore, the temporal dynamics of learning are frequently under-represented in current models, which tend to focus on static snapshots rather than continuous trajectories [48].

Methodological limitations remain a significant barrier to reproducibility and rigour. Studies often rely on self-reported measures or small sample sizes, introducing potential biases and reducing statistical power [55,162]. Furthermore, a lack of control groups or comparative benchmarks limits the interpretability of observed effects, making it difficult to distinguish between true intervention impacts and context-specific noise [12,180].

Finally, the field is still in an exploratory stage, and many promising avenues—such as MMLA in programming education, informal learning contexts, or collaborative problem-solving—remain underexplored [120]. This underscores the need for broader empirical validation and theoretical grounding to ensure that future models not only perform well but are also meaningful, equitable, and contextually relevant.

6. Future Directions

The continued advancement of MMLA presents promising opportunities for both theoretical and practical contributions in educational research. The future directions outlined in this section reflect a synthesis of the insights gained from the reviewed literature, combined with the authors' critical assessment of current challenges and emerging opportunities in the field. These directions are intended to guide future research efforts and highlight areas where significant impact can be achieved.

Real-World Deployment and Scalability: A significant proportion of current MMLA research remains confined to controlled or small-scale settings. Future work should prioritise the deployment of MMLA systems in real-world classrooms across diverse educational contexts. This includes primary and secondary education, non-formal learning environments, and under-represented geographical regions. Addressing issues related to technical scalability, such as latency reduction, robustness to environmental noise, and cost-effective hardware integration, will be essential to ensure broader adoption.

Enhancing Model Interpretability and Transparency: While deep learning models have demonstrated remarkable performance, their opaque nature hinders practical implementation by educators and stakeholders. It is recommended that future research concentrate on the development of interpretable models that can provide teachers and learners with actionable feedback. The integration of explainable AI (XAI) techniques with multimodal pipelines has the potential to engender greater levels of trust and facilitate meaningful pedagogical decision-making processes.

Ethical Frameworks and Participatory Design: The ethical implications of MMLA remain a pressing concern, especially regarding privacy, consent, and data ownership. Future research should explore participatory approaches that involve students, teachers, and caregivers in the design and governance of MMLA systems. Emphasising transparency, fairness, and inclusivity can help mitigate concerns about surveillance and data misuse while supporting agency and equity in learning analytics.

Integration of Emerging Modalities and Technologies: Although current MMLA studies predominantly utilise visual, auditory, and physiological signals, the integration of emerging modalities such as tactile data, brain–computer interfaces, and augmented reality remains under-explored. Moreover, the incorporation of large language models (LLMs) and generative AI for contextual interpretation of multimodal inputs could enhance feedback generation and learner modelling, particularly in naturalistic learning environments.

Longitudinal and Cross-Cultural Studies: The majority of MMLA studies are short-term and often context-specific. Longitudinal studies are necessary to evaluate the sustained impact of multimodal interventions on learning trajectories and educational outcomes. Additionally, cross-cultural investigations are crucial to assess the generalisability and cultural sensitivity of multimodal models and frameworks, especially in multilingual and socio-economically diverse settings.

Professional Development and Teacher Support: MMLA has the potential to significantly enhance teacher professional development by providing insights into classroom dynamics, instructional effectiveness, and student engagement. Future research should explore how multimodal analytics can be integrated into teacher training programmes and continuous professional development, with an emphasis on usability and alignment with teachers' pedagogical practices.

Standardisation and Benchmarking: There is a pressing need for standardised protocols and shared datasets to support replication, benchmarking, and cross-study comparisons. Establishing open-access repositories, benchmark tasks, and evaluation frameworks will strengthen methodological rigour and accelerate progress in the field.

Fusion of Theoretical and Computational Approaches: Finally, future research should strive to align computational methods with well-established learning theories. Bridging this gap can ensure that the development of MMLA systems is not only technologically advanced but also pedagogically meaningful. The design of multimodal systems grounded in constructivist, socio-cultural, or embodied learning frameworks can yield richer interpretations and more effective interventions.

7. Conclusions

This review provides a comprehensive synthesis of recent advances in MMLA in education, drawing on peer-reviewed studies to examine theoretical foundations, data modalities, modelling techniques, tools, and educational applications. The findings demonstrate that MMLA offers a robust framework for capturing the complexity of learning processes by integrating diverse data sources—visual, auditory, physiological, contextual, and interaction-based.

A key strength of MMLA lies in its capacity to model both observable behaviours and internal cognitive-affective states, which supports more personalised and adaptive learning environments. Methodologically, the field benefits from the integration of advanced ML and deep learning models, with growing interest in fusion strategies, interpretable AI, and real-time feedback mechanisms. The application domains explored range from engagement monitoring and collaborative learning to inclusive education and simulation-based training.

Nevertheless, the implementation of MMLA in real-world contexts remains challenged by technical constraints, data synchronisation complexities, model opacity, and ethical considerations surrounding privacy and consent. Furthermore, there is a need to improve the scalability and reproducibility validity of multimodal systems, especially in under-represented educational settings.

Looking ahead, future research should prioritise not only technical and methodological advancements but also a deeper consideration of the epistemological and pedagogical dimensions of learning. MMLA technologies must be designed to capture the nuanced ways in which learners engage with, interpret, and produce knowledge within diverse educational contexts. This involves recognising that learning is not merely a cognitive process but also a social and cultural one, shaped by the negotiation and transformation of knowledge. To ensure that MMLA systems yield meaningful, equitable, and contextually relevant insights, researchers and practitioners should emphasise interpretability, fairness, and pedagogical alignment, alongside longitudinal studies, cross-cultural validation, and participatory design approaches that actively involve educators and learners in the co-development of these systems.

This review aims to consolidate existing knowledge, identify ongoing gaps, and offer guidance for researchers and practitioners seeking to design, implement, and evaluate MMLA in education. By doing so, it contributes to shaping a future where multimodal evidence meaningfully informs teaching, learning, and assessment across diverse learning environments.

Author Contributions: Conceptualization, J.D.T.G.-S., F.P.R. and V.H.M.-D.; investigation, J.D.T.G.-S. and F.P.R.; writing—original draft preparation, J.D.T.G.-S., F.P.R., V.H.M.-D., J.S.-G. and A.M.-M.; writing—review and editing, J.D.T.G.-S., F.P.R., V.H.M.-D., J.S.-G. and J.A.O.; visualization, J.D.T.G.-S. and F.P.R.; supervision, F.P.R. and V.H.M.-D.; funding acquisition, J.S.-G. and J.A.O. All authors have read and agreed to the published version of the manuscript.

Funding: The Spanish Government has partially supported this work under the grant SAFER: PID2019-104735RB-C42 (ERA/ERDF, EU), and project PLEC2021-007681 funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest. The authors have no relevant financial or non-financial interests to disclose.

Abbreviations

The following abbreviations are used in this manuscript:

MMLA	Multimodal learning analytics
ML	Machine learning
AI	Artificial intelligence
SRL	Self-Regulated Learning
MFCCs	Mel Frequency Cepstral Coefficients
VAD	Voice Activity Detection
EDA	Electrodermal activity
HRV	Heart rate variability
EEG	Electroencephalogram
PCA	Principal component analysis
SVM	Support Vector Machine
GBDT	Gradient Boosted Decision Trees
RF	Random Forest
KNN	K-Nearest Neighbours
CNN	Convolutional Neural Networks
MLP	Multi-Layer Perceptrons
DNN	Deep Neural Networks
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
IB-BiLSTM	Information Block Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Units
TCN	Temporal Convolutional Networks
MM	Markov Model
HMM	Hidden Markov Models
CRF	Conditional Random Fields
BERT	Bidirectional Encoder Representations from Transformers
TEMS	Transformer-based structure
BRR	Bayesian Ridge Regression
LPA	Latent Profile Analysis
ENA	Epistemic Network Analysis
IRL	Inverse Reinforcement Learning
MTL	Multi-Task Learning
TL	Transfer learning
iAFM	individualized Adaptive Factorisation Machine
MOLAM	Mobile Multimodal Learning Analytics
CSCL	Computer-supported collaborative learning
XAI	Explainable artificial intelligence
LLM	Large language model
CLM	Collaborative learning mechanism
M-DVC	Multimodal Data Value Chain
AIELA	Artificial Intelligence-enhanced learning assistant
DUET	Dual eye-tracking
TMA	Transmodal analysis
T/ONA	Transmodal ordered network analysis model
ELMo	Embeddings from language model
AR	Augmented reality
VR	Virtual reality
SEN	Special education needs

References

1. Emerson, A.; Henderson, N.; Rowe, J.; Min, W.; Lee, S.; Minogue, J.; Lester, J. Early Prediction of Visitor Engagement in Science Museums with Multimodal Learning Analytics. In Proceedings of the 2020 International Conference on Multimodal Interaction, Utrecht, The Netherlands, 25–29 October 2020 ; ICMI '20; pp. 107–116. [CrossRef]
2. Yan, L.; Martinez-Maldonado, R.; Swiecki, Z.; Zhao, L.; Li, X.; Gasevic, D. Dissecting the Temporal Dynamics of Embodied Collaborative Learning Using Multimodal Learning Analytics. *J. Educ. Psychol.* **2025**, *117*, 106–133. [CrossRef]
3. Moon, J.; Yeo, S.; Banihashem, S.; Noroozi, O. Using multimodal learning analytics as a formative assessment tool: Exploring collaborative dynamics in mathematics teacher education. *J. Comput. Assist. Learn.* **2024**, *40*, 2753–2771. [CrossRef]
4. Mangaroska, K.; Martinez-Maldonado, R.; Vesin, B.; Gašević, D. Challenges and opportunities of multimodal data in human learning: The computer science students' perspective. *J. Comput. Assist. Learn.* **2021**, *37*, 1030–1047. [CrossRef]
5. Worsley, M.; Abrahamson, D.; Blikstein, P.; Grover, S.; Schneider, B.; Tissenbaum, M. Situating multimodal learning analytics. In Proceedings of the International Conference of the Learning Sciences, ICLS, Singapore, 20–24 June 2016; Volume 2, pp. 1346–1349.
6. Perveen, A. Facilitating Multiple Intelligences Through Multimodal Learning Analytics. *Turk. Online J. Distance Educ.* **2018**, *19*, 18–30. [CrossRef]
7. Yang, K.; Borchers, C.; Falhs, A.C.; Echeverria, V.; Karumbaiah, S.; Rummel, N.; Aleven, V. Leveraging Multimodal Classroom Data for Teacher Reflection: Teachers' Preferences, Practices, and Privacy Considerations. In Proceedings of the Lecture Notes in Computer Science, Krems, Austria, 16–20 September 2024; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 15159 , pp. 498–511. [CrossRef]
8. Lin, H. Online Professional-Creative Fusion Music Major Students' Classroom State Recognition Based on the Integration of DSC and LeNet-5 Models. *J. Artif. Intell. Technol.* **2024**, *4*, 216–226. [CrossRef]
9. Echeverria, V.; Yan, L.; Zhao, L.; Abel, S.; Alfredo, R.; Dix, S.; Jaggard, H.; Wotherspoon, R.; Osborne, A.; Buckingham Shum, S.; et al. TeamSlides: A Multimodal Teamwork Analytics Dashboard for Teacher-guided Reflection in a Physical Learning Space. In Proceedings of the 14th Learning Analytics and Knowledge Conference, Kyoto, Japan, 18–22 March 2024; LAK '24; pp. 112–122. [CrossRef]
10. Hakami, L.; Hernandez-Leo, D.; Amarasinghe, I.; Sayis, B. Investigating teacher orchestration load in scripted CSCL: A multimodal data analysis perspective. *Br. J. Educ. Technol.* **2024**, *55*, 1926–1949. [CrossRef]
11. Ding, A.C.E.; Huang, K.T.T.; DuBois, J.; Fu, H. Integrating immersive virtual reality technology in scaffolded game-based learning to enhance low motivation students' multimodal science learning. *Educ. Technol. Res. Dev.* **2024**, *72*, 2083–2102. [CrossRef]
12. Na, H.; Sung, H. Learn math through motion: A technology-enhanced embodied approach with augmented reality for geometry learning in K-12 classrooms. *Interact. Learn. Environ.* 2025, *in press*. [CrossRef]
13. Chan, R.Y.-Y.; Wong, C.M.V.; Yum, Y.N. Predicting Behavior Change in Students with Special Education Needs Using Multimodal Learning Analytics. *IEEE Access* **2023**, *11*, 63238–63251. [CrossRef]
14. Vivante, I.; Vedder-Weiss, D. Examining science teachers' engagement in professional development: A multimodal situated perspective. *J. Res. Sci. Teach.* **2023**, *60*, 1401–1430. [CrossRef]
15. Zhao, Y.; Yu, B. Construction of a Classification Model for Teacher and Student Behavior in Physical Education Classrooms-Based on Multimodal Data. *Appl. Math. Nonlinear Sci.* **2024**, *9*. [CrossRef]
16. Basystiuk, O.; Melnykova, N.; Rybchak, Z. Multimodal Learning Analytics: An Overview of the Data Collection Methodology. In Proceedings of the 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), Lviv, Ukraine, 19–21 October 2023; pp. 1–4. [CrossRef]
17. Chango, W.; Lara, J.; Cerezo, R.; Romero, C. A review on data fusion in multimodal learning analytics and educational data mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*. [CrossRef]
18. Gunnars, F. Smartbands and Behavioural Interventions in the Classroom: Multimodal Learning Analytics Stress-Level Visualisations for Primary Education Teachers. *Int. J. Disabil. Dev. Educ.* 2024, *in press*. [CrossRef]
19. Lemke, J. Multiplying meaning: Visual and verbal semiotics in scientific text. In *Reading Science: Critical and Functional Perspectives on Discourses of Science*; Routledge: London, UK, 1998; pp. 87–113. [CrossRef]
20. Roth, W.M.; Welzel, M. From activity to gestures and scientific language. *J. Res. Sci. Teach.* **2001**, *38*, 103–136. [CrossRef]
21. Di Mitri, D.; Schneider, J.; Specht, M.; Drachsler, H. From signals to knowledge: A conceptual model for multimodal learning analytics. *J. Comput. Assist. Learn.* **2018**, *34*, 338–349. [CrossRef]
22. Cohn, C.; Davalos, E.; Vatral, C.; Fonteles, J.H.; Wang, H.D.; Ma, M.; Biswas, G. Multimodal Methods for Analyzing Learning and Training Environments: A Systematic Literature Review. *arXiv* **2024**, arXiv:2408.14491.
23. Emerson, A.; Cloude, E.B.; Azevedo, R.; Lester, J. Multimodal learning analytics for game-based learning. *Br. J. Educ. Technol.* **2020**, *51*, 1505–1526. [CrossRef]
24. Volta, E.; Volpe, G. Exploiting multimodal integration in adaptive interactive systems and game-based learning interfaces. In Proceedings of the 5th International Conference on Movement and Computing, MOCO '18, Genoa, Italy, 28–30 June 2018; [CrossRef]

25. Prinsloo, P.; Slade, S.; Khalil, M. Multimodal learning analytics—In-between student privacy and encroachment: A systematic review. *Br. J. Educ. Technol.* **2023**, *54*, 1566–1586. [CrossRef]
26. Tang, K.; Tan, S.C.; .; Yeo, J. Students' Multimodal Construction of the Work–Energy Concept. *Int. J. Sci. Educ.* **2011**, *33*, 1775–1804. [CrossRef]
27. Cloude, E.B.; Azevedo, R.; Winne, P.H.; Biswas, G.; Jang, E.E. System design for using multimodal trace data in modeling self-regulated learning. *Front. Educ.* **2022**, *7*, 928632. [CrossRef]
28. Di Mitri, D. The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences. Ph.D. Thesis, Open Universiteit, Heerlen, The Netherlands, 2020.
29. Khan, S.M. Multimodal Behavioral Analytics in Intelligent Learning and Assessment Systems. In *Innovative Assessment of Collaboration*; von Davier, A.A., Zhu, M., Kyllonen, P.C., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 173–184. [CrossRef]
30. Aoyama Lawrence, L.; Weinberger, A. Being in-sync: A multimodal framework on the emotional and cognitive synchronization of collaborative learners. *Front. Educ.* **2022**, *7*, 867186. [CrossRef]
31. Mangaroska, K.; Sharma, K.; Gasevic, D.; Giannakos, M. Multimodal Learning Analytics to Inform Learning Design: Lessons Learned from Computing Education. *J. Learn. Anal.* **2020**, *7*, 79–97. [CrossRef]
32. Sharma, K.; Papamitsiou, Z.; Giannakos, M. Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *Br. J. Educ. Technol.* **2019**, *50*, 3004–3031. [CrossRef]
33. Yan, L.; Gasevic, D.; Echeverria, V.; Jin, Y.; Zhao, L.; Martinez-Maldonado, R. From Complexity to Parsimony: Integrating Latent Class Analysis to Uncover Multimodal Learning Patterns in Collaborative Learning. In Proceedings of the 15th International Learning Analytics and Knowledge Conference, Dublin, Ireland, 3–7 March 2025; pp. 70–81. [CrossRef]
34. Matsumoto, Y. Material Moments: Teacher and Student Use of Materials in Multilingual Writing Classroom Interactions. *Mod. Lang. J.* **2019**, *103*, 179–204. [CrossRef]
35. Xie, N.; Liu, Z.; Li, Z.; Pang, W.; Lu, B. Student engagement detection in online environment using computer vision and multi-dimensional feature fusion. *Multimed. Syst.* **2023**, *29*, 3559–3577. [CrossRef]
36. Andrade, A. Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, LAK '17, Vancouver, BC, Canada, 13–17 March 2017; pp. 70–79. [CrossRef]
37. Obade, C.; Kim, H.W.; Cook-Chennault, K. WIP: Using Multimodal Approaches to Understand the Attention and Focus of Students Engaging in Intuition-Based Online Engineering Learning Games. In Proceedings of the 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, 18–21 October 2023; pp. 1–5. [CrossRef]
38. Gomes, J.; Yassine, M.; Worsley, M.; Blikstein, P. Analysing engineering expertise of high school students using eye tracking and multimodal learning analytics. In Proceedings of the 6th International Conference on Educational Data Mining, EDM 2013, Memphis, TN, USA, 6–9 July 2013.
39. Tamura, K.; Lu, M.; Konomi, S.; Hatano, K.; Inaba, M.; Oi, M.; Okamoto, T.; Okubo, F.; Shimada, A.; Wang, J.; et al. Integrating Multimodal Learning Analytics and Inclusive Learning Support Systems for People of All Ages. In Proceedings of the Lecture Notes in Computer Science, Orlando, FL, USA, 26–31 July 2019; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 11577, pp. 469–481. [CrossRef]
40. Sharma, K.; Leftheriotis, I.; Giannakos, M. Utilizing Interactive Surfaces to Enhance Learning, Collaboration and Engagement: Insights from Learners' Gaze and Speech. *Sensors* **2020**, *20*, 1964. [CrossRef] [PubMed]
41. Sharma, K.; Lee-Cultura, S.; Giannakos, M. Keep Calm and Do Not Carry-Forward: Toward Sensor-Data Driven AI Agent to Enhance Human Learning. *Front. Artif. Intell.* **2022**, *4*, 713176. [CrossRef] [PubMed]
42. Liu, G. Multimodal Analysis and Optimisation Strategy of Teaching Behaviour in Physical Education Classroom. *Appl. Math. Nonlinear Sci.* **2024**, *9*, 20241506. [CrossRef]
43. Hellmich, E.; Castek, J.; Smith, B.E.; Floyd, R.; Wen, W. Student perspectives on multimodal composing in the L2 classroom: tensions with audience, media, learning and sharing. *Engl. Teach.-Pract. Crit.* **2021**, *20*, 210–226. [CrossRef]
44. Noel, R.; Riquelme, F.; Lean, R.; Merino, E.; Cechinel, C.; Barcelos, T.; Villarroel, R.; Munoz, R. Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course. *IEEE Access* **2018**, *6*, 67783–67798. [CrossRef]
45. Chejara, P.; Prieto, L.P.; Rodriguez-Triana, M.J.; Kasepalu, R.; Ruiz-Calleja, A.; Shankar, S.K. How to Build More Generalizable Models for Collaboration Quality? Lessons Learned from Exploring Multi-Context Audio-Log Datasets using Multimodal Learning Analytics. In Proceedings of the LAK23: 13th International Learning Analytics and Knowledge Conference, LAK2023, Arlington, TX ,USA, 13–17 March 2023; pp. 111–121. [CrossRef]
46. Ma, F. Construction and Evaluation of College English Translation Teaching Model Based on Multimodal Integration. *Appl. Math. Nonlinear Sci.* **2024**, *9*, 20241774. [CrossRef]

47. Chejara, P.; Prieto, L.; Ruiz-Calleja, A.; Rodríguez-Triana, M.; Shankar, S.; Kasepalu, R. EFAR-MMLA: An Evaluation Framework to Assess and Report Generalizability of Machine Learning Models in MMLA. *Sensors* **2021**, *21*, 2863. [CrossRef]
48. Chejara, P.; Kasepalu, R.; Prieto, L.P.; Rodriguez-Triana, M.J.; Ruiz Calleja, A.; Schneider, B. How well do collaboration quality estimation models generalize across authentic school contexts? *Br. J. Educ. Technol.* **2024**, *55*, 1602–1624. [CrossRef]
49. Peng, S.; Nagao, K. Recognition of Students' Mental States in Discussion Based on Multimodal Data and its Application to Educational Support. *IEEE Access* **2021**, *9*, 18235–18250. [CrossRef]
50. Cosentino, G.; Anton, J.; Sharma, K.; Gelsomini, M.; Giannakos, M.; Abrahamson, D. Hybrid teaching intelligence: Lessons learned from an embodied mathematics learning experience. *Br. J. Educ. Technol.* **2024**, *56*, 621–649. [CrossRef]
51. Boulton, H.; Brown, D.; Taheri, M.; Van Isacker, K.; Burton, A.; Shopland, N. Mobile Developments to Support Learners in Mainstream Education. In Proceedings of the EDULEARN19: 11th International Conference on Education and New Learning Technologies, Palma, Spain, 1–3 July 2019; Chova, L., Martinez, A., Torres, I., Eds.; EDULEARN Proceedings; pp. 3921–3927.
52. Yan, L.; Echeverria, V.; Jin, Y.; Fernandez-Nieto, G.; Zhao, L.; Li, X.; Alfredo, R.; Swiecki, Z.; Gasevic, D.; Martinez-Maldonado, R. Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning. *Br. J. Educ. Technol.* **2024**, *55*, 1900–1925. [CrossRef]
53. Hassan, J.; Leong, J.; Schneider, B. Multimodal data collection made easy: The EZ-MMLA toolkit: A data collection website that provides educators and researchers with easy access to multimodal data streams. In Proceedings of the LAK21: 11th International Learning Analytics and Knowledge Conference, LAK21, Irvine, CA, USA, 12–16 April 2021; pp. 579–585. [CrossRef]
54. Xiao, J.; Jiang, Z.; Wang, L.; Yu, T. What can multimodal data tell us about online synchronous training: Learning outcomes and engagement of in-service teachers. *Front. Psychol.* **2023**, *13*, 1092848. [CrossRef]
55. Du, X.; Zhang, L.; Hung, J.L.; Li, H.; Tang, H.; Dai, M. Analyzing the effects of instructional strategies on students' on-task status from aspects of their learning behaviors and cognitive factors. *J. Comput. High. Educ.* **2024**, *36*, 29–56. [CrossRef]
56. Eradze, M.; Rodríguez-Triana, M.; Milikic, N.; Laanpere, M.; Tammets, K. Contextualising Learning Analytics with Classroom Observations: A Case Study. *Interact. Des. Archit.* **2020**, *44*, 71–95. [CrossRef]
57. Borchers, C.; Wang, Y.; Karumbaiah, S.; Ashiq, M.; Shaffer, D.W.; Aleven, V. Revealing Networks: Understanding Effective Teacher Practices in AI-Supported Classrooms using Transmodal Ordered Network Analysis. In Proceedings of the 14th Learning Analytics and Knowledge Conference, LAK '24, Kyoto, Japan, 18–22 March 2024; pp. 371–381. [CrossRef]
58. Chejara, P.; Kasepalu, R.; Prieto, L.; Rodrıǵuez-Triana, M.J.; Ruiz-Calleja, A. Bringing Collaborative Analytics using Multimodal Data to Masses: Evaluation and Design Guidelines for Developing a MMLA System for Research and Teaching Practices in CSCL. In Proceedings of the 14th Learning Analytics and Knowledge Conference, LAK '24, Kyoto, Japan, 18–22 March 2024; pp. 800–806. [CrossRef]
59. Chejara, P.; Prieto, L.; Ruiz-Calleja, A.; Rodríguez-Triana, M.; Shankar, S.; Kasepalu, R. Quantifying collaboration quality in face-to-face classroom settings using MMLA. In Proceedings of the Lecture Notes in Computer Science, Tartu, Estonia, 8–11 September, 2020; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 12324, pp. 159–166. [CrossRef]
60. Zhu, G.; Xing, W.; Costa, S.; Scardamalia, M.; Pei, B. Exploring emotional and cognitive dynamics of Knowledge Building in grades 1 and 2. *User Model. User-Adapt. Interact.* **2019**, *29*, 789–820. [CrossRef]
61. Zhao, L.; Gasevic, D.; Swiecki, Z.; Li, Y.; Lin, J.; Sha, L.; Yan, L.; Alfredo, R.; Li, X.; Martinez-Maldonado, R. Towards automated transcribing and coding of embodied teamwork communication through multimodal learning analytics. *Br. J. Educ. Technol.* **2024**, *55*, 1673–1702. [CrossRef]
62. Shankar, S.K.; Rodríguez-Triana, M.J.; Ruiz-Calleja, A.; Prieto, L.P. .; Chejara, P.; Martínez-Monés, A. Multimodal Data Value Chain (M-DVC): A Conceptual Tool to Support the Development of Multimodal Learning Analytics Solutions. *IEEE Rev. Iberoam. Tecnol. Aprendiz.* **2020**, *15*, 113–122. [CrossRef]
63. Levinsen, K.; Sørensen, B. Digital literacy and subject matter learning. In Proceedings of the European Conference on e-Learning, ECEL, Hatfield, UK, 29–30 October 2015; pp. 305–312.
64. Zhang, J.; Wang, K.; Zhang, Y. Physiological Characterization of Student Engagement in the Naturalistic Classroom: A Mixed-Methods Approach. *Mind Brain Educ.* **2021**, *15*, 322–343. [CrossRef]
65. Yan, L.; Martinez-Maldonado, R.; Zhao, L.; Dix, S.; Jaggard, H.; Wotherspoon, R.; Li, X.; Gasevic, D. The role of indoor positioning analytics in assessment of simulation-based learning. *Br. J. Educ. Technol.* **2023**, *54*, 267–292. [CrossRef]
66. Ramírez, J.A.R.; Glasserman-Morales, L.D. Use of Multimodal Data Value Chain as a Contribution to the Management of the Teaching-Learning Process in Higher Education Institutions. In Proceedings of the 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, Virtual Event, 15–17 December 2021; pp. 1–6. [CrossRef]
67. Spikol, D.; Ruffaldi, E.; Cukurova, M. Using multimodal learning analytics to identify aspects of collaboration in project-based learning. In Proceedings of the 12th International Conference on Computer Supported Collaborative Learning—Making a Difference: Prioritizing Equity and Access in CSCL, CSCL 2017, Philadelphia, PA, USA, 18–22 June 2017; Volume 1, pp. 263–270.

68. Chejara, P.; Kasepalu, R.; Prieto, L.; Rodríguez-Triana, M.; Ruiz-Calleja, A.; Shankar, S. Multimodal Learning Analytics Research in the Wild: Challenges and their Potential Solutions. In Proceedings of the CEUR Workshop Proceedings, Arlington, TX, USA, 13–17 March 2023; Volume 3439, pp. 36–42.
69. Li, C.; Liu, C.; Ju, W.; Zhong, Y.; Li, Y. Prediction of teaching quality in the context of smart education: application of multimodal data fusion and complex network topology structure. *Discov. Artif. Intell.* **2025**, *5*, 19. [CrossRef]
70. Min, W.; Vail, A.; Frankosky, M.; Wiggins, J.; Boyer, K.; Wiebe, E.; Pezzullo, L.; Mott, B.; Lester, J. Predicting dialogue acts for intelligent virtual agents with multimodal student interaction data. In Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, NC, USA, 29 June–2 July 2016; pp. 454–459.
71. Hou, P.; Yang, M.; Zhang, T.; Na, T. Analysis of English classroom teaching behavior and strategies under adaptive deep learning under cognitive psychology. *Curr. Psychol.* **2024**, *43*, 35974–35988. [CrossRef]
72. Xie, N.; Li, Z.; Lu, H.; Pang, W.; Song, J.; Lu, B. MSC-Trans: A Multi-Feature-Fusion Network with Encoding Structure for Student Engagement Detecting. *IEEE Trans. Learn. Technol.* **2025**, *18*, 243–255. [CrossRef]
73. Ashwin, T.S.; Guddeti, R.M.R. Unobtrusive Behavioral Analysis of Students in Classroom Environment Using Non-Verbal Cues. *IEEE Access* **2019**, *7*, 150693–150709. [CrossRef]
74. Sabuncuoglu, A.; Sezgin, T.M. Developing a Multimodal Classroom Engagement Analysis Dashboard for Higher-Education. *Proc. ACM Hum.-Comput. Interact.* **2023**, *7*, 1–23. [CrossRef]
75. Vujovic, M.; Hernandez-Leo, D.; Tassani, S.; Spikol, D. Round or rectangular tables for collaborative problem solving? A multimodal learning analytics study. *Br. J. Educ. Technol.* **2020**, *51*, 1597–1614. [CrossRef]
76. D. Chiaro.; D. Annuziata.; S. Izzo.; F. Piccialli. Unveiling engagement in virtual classrooms: A multimodal analysis. In Proceedings of the 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 15–18 December 2023; pp. 4761–4769. [CrossRef]
77. Yan, L.; Martinez-Maldonado, R.; Cordoba, B.G.; Deppeler, J.; Corrigan, D.; Gasevic, D. Mapping from proximity traces to socio-spatial behaviours and student progression at the school. *Br. J. Educ. Technol.* **2022**, *53*, 1645–1664. [CrossRef]
78. Chejara, P.; Prieto, L.; Dimitriadis, Y.; Rodríguez-Triana, M.; Ruiz-Calleja, A.; Kasepalu, R.; Shankar, S. The Impact of Attribute Noise on the Automated Estimation of Collaboration Quality Using Multimodal Learning Analytics in Authentic Classrooms. *J. Learn. Anal.* **2024**, *11*, 73–90. [CrossRef]
79. Liao, C.H.; Wu, J.Y. Deploying multimodal learning analysis models to explore the impact of digital distraction and peer learning on student performance. *Comput. Educ.* **2022**, *190*, 104599. [CrossRef]
80. Yang, S.; Huang, Y.; Wang, Y.; Li, W. The Classroom Education Model of Artificial Intelligence and Information Technology Supported by Wireless Networks. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 8797317. [CrossRef]
81. Worsley, M. (Dis)engagement matters: Identifying efficacious learning practices with multimodal learning analytics. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, LAK '18, Sydney, Australia, 7–9 March 2018; pp. 365–369. [CrossRef]
82. Camacho, V. L.; de la Guía, E.; Olivares, T.; Flores, M. J.; Orozco-Barbosa, L. Data Capture and Multimodal Learning Analytics Focused on Engagement with a New Wearable IoT Approach. *IEEE Trans. Learn. Technol.* **2020**, *13*, 704–717. [CrossRef]
83. Khalil, M. Mobile Multimodal Learning Analytics Conceptual Framework to Support Student Self-Regulated Learning (MOLAM). In *SpringerBriefs in Open and Distance Education*; Springer: Singapore, 2022; pp. 63–75. [CrossRef]
84. Li, H.; Ding, W.; Liu, Z. Identifying At-Risk K-12 Students in Multimodal Online Environments: A Machine Learning Approach. In Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Virtual Event, 10–13 July 2020; pp. 137–147.
85. Bouktif, S.; Nimmi, K. Real-Time Detection of Student Attention in Online Learning: A Scalable Solution for Crisis-Driven Remote Education. In Proceedings of the 2024 IEEE International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 20–22 December 2024; pp. 75–79. [CrossRef]
86. Wang, C. Course evaluation of preschoolhygiene under the multimodal learning model. *Appl. Math. Nonlinear Sci.* **2024**, *9*, 20243194. [CrossRef]
87. Luo, Z.; Jingying, C.; Guangshuai, W.; Mengyi, L. A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology. *Interact. Learn. Environ.* **2022**, *30*, 1117–1130. [CrossRef]
88. Shirol, A.; Gadad, J.; M, V. Unleashing Potential: Transforming Oral Presentations Through Multimodal Learning Analytics. In Proceedings of the 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Bengaluru, India, 9–12 December 2024; pp. 1–7. [CrossRef]
89. Qushem, U.B.; Christopoulos, A.; Laakso, M.-J. The Value Proposition of An Integrated Multimodal Learning Analytics Framework. In Proceedings of the 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 23–27 May 2022; pp. 666–671. [CrossRef]
90. Zhang, C. Research on the Path of English Translation Teaching and External Communication Based on the Multimodal Analysis Method. *Appl. Math. Nonlinear Sci.* **2024**, *9*, 20241143. [CrossRef]

91. Chang, J.; Wang, Z.; Yan, C. MusicARLtrans Net: A multimodal agent interactive music education system driven via reinforcement learning. *Front. Neurorobotics* **2024**, *18*, 1479694. [CrossRef] [PubMed]
92. Spikol, D.; Ruffaldi, E.; Dabisias, G.; Cukurova, M. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *J. Comput. Assist. Learn.* **2018**, *34*, 366–377. [CrossRef]
93. Emerson, A.; Min, W.; Rowe, J.; Azevedo, R.; Lester, J. Multimodal Predictive Student Modeling with Multi-Task Transfer Learning. In Proceedings of the LAK23: 13th International Learning Analytics and Knowledge Conference, LAK2023, Arlington, TX, USA, 13–17 March 2023; pp. 333–344. [CrossRef]
94. Li, H.; Wang, Z.; Tang, J.; Ding, W.; Liu, Z. Siamese Neural Networks for Class Activity Detection. In Proceedings of the Lecture Notes in Computer Science, Ifrane, Morocco, 6–10 July 2020; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 12164, pp. 162–167. [CrossRef]
95. Mao, J.; Qian, Z.; Lucas, T. Sentiment Analysis of Animated Online Education Texts Using Long Short-Term Memory Networks in the Context of the Internet of Things. *IEEE Access* **2023**, *11*, 109121–109130. [CrossRef]
96. Gupta, A.; Carpenter, D.; Min, W.; Rowe, J.; Azevedo, R.; Lester, J. Multimodal Multi-Task Stealth Assessment for Reflection-Enriched Game-Based Learning. In Proceedings of the CEUR Workshop Proceedings, Utrecht, The Netherlands, 10–14 June 2021; Volume 2902, pp. 93–102.
97. Yusuf, A.; Noor, N.M.; Bello, S. Using multimodal learning analytics to model students' learning behavior in animated programming classroom. *Educ. Inf. Technol.* **2024**, *29*, 6947–6990. [CrossRef]
98. Zhang, L.; Hung, J.L.; Du, X.; Li, H.; Hu, Z. Multimodal Fast–Slow Neural Network for learning engagement evaluation. *Data Technol. Appl.* **2023**, *57*, 418–435. [CrossRef]
99. Tantry, R.; Shenoy, S.U.; Acharya, S.; Prathibha, K.N. Artificial Intelligence Assisted Student States monitoring based on Enhancing Cognitive and Emotional Feedback Mechanism using Collaborative Learning Environments. In Proceedings of the 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 9–10 May 2024; pp. 1–5. [CrossRef]
100. Fahid, F.; Acosta, H.; Lee, S.; Carpenter, D.; Mott, B.; Bae, H.; Saleh, A.; Brush, T.; Glazewski, K.; Hmelo-Silver, C.; et al. Multimodal Behavioral Disengagement Detection for Collaborative Game-Based Learning. In Proceedings of the Lecture Notes in Computer Science, Durham, UK, 27–31 July 2022; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 13356, pp. 218–221. [CrossRef]
101. Zhou, Y.; Suzuki, K.; Kumano, S. State-Aware Deep Item Response Theory using student facial features. *Front. Artif. Intell.* **2023**, *6*, 1324279. [CrossRef] [PubMed]
102. Lin, C.J.; Wang, W.S.; Lee, H.Y.; Huang, Y.M.; Wu, T.T. Recognitions of image and speech to improve learning diagnosis on STEM collaborative activity for precision education. *Educ. Inf. Technol.* **2024**, *29*, 13859–13884. [CrossRef]
103. Wiedbusch, M.D.; Kite, V.; Yang, X.; Park, S.; Chi, M.; Taub, M.; Azevedo, R. A Theoretical and Evidence-Based Conceptual Design of MetaDash: An Intelligent Teacher Dashboard to Support Teachers' Decision Making and Students' Self-Regulated Learning. *Front. Educ.* **2021**, *6*, 570229. [CrossRef]
104. Chejara, P.; Kasepalu, R.; Ruiz-Calleja, A.; Rodriguez-Triana, M.; Prieto, L. Co-Designing a Multimodal Dashboard for Collaborative Analytics. In Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning-CSCL 2022, Hiroshima, Japan, 6–10 June 2022; Volume 2022, pp. 577–578.
105. Ronda-Carracao, M.; Santos, O.; Fernandez-Nieto, G.; Martinez-Maldonado, R. Towards Exploring Stress Reactions in Teamwork using Multimodal Physiological Data. In Proceedings of the CEUR Workshop Proceedings, Virtual Event, 14 June 2021; Volume 2902, pp. 49–60.
106. Paredes, Y.; Huang, P.K.; Hsiao, I.H. Utilising behavioural analytics in a blended programming learning environment. *New Rev. Hypermedia Multimed.* **2019**, *25*, 89–111. [CrossRef]
107. Sharif, M.; Uckelmann, D. Multi-Modal LA in Personalized Education Using Deep Reinforcement Learning Based Approach. *IEEE Access* **2024**, *12*, 54049–54065. [CrossRef]
108. Ouhaichi, H.; Bahtijar, V.; Spikol, D. Exploring design considerations for multimodal learning analytics systems: An interview study. *Front. Educ.* **2024**, *9*, 1356537. [CrossRef]
109. Martinez-Maldonado, R.; Echeverria, V.; Mangaroska, K.; Shibani, A.; Fernandez-Nieto, G.; Schulte, J.; Buckingham Shum, S. Moodoo the Tracker: Spatial Classroom Analytics for Characterising Teachers' Pedagogical Approaches. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 1025–1051. [CrossRef]
110. Liu, R.; Stamper, J.; Davenport, J.; Crossley, S.; McNamara, D.; Nzinga, K.; Sherin, B. Learning linkages: Integrating data streams of multiple modalities and timescales. *J. Comput. Assist. Learn.* **2019**, *35*, 99–109. [CrossRef]
111. Barmaki, R.; Hughes, C.E. Providing Real-time Feedback for Student Teachers in a Virtual Rehearsal Environment. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI '15, Seattle, WA, USA, 9–13 November 2015; pp. 531–537. [CrossRef]

112. Taub, M.; Banzon, A.M.; Outerbridge, S.; Walker, L.R.; Olivera, L.; Salas, M.; Schneier, J. Towards scaffolding self-regulated writing: Implications for developing writing interventions in first-year writing. *Metacognition Learn.* **2023**, *18*, 749–782. [CrossRef]
113. Hasnine, M.; Bui, H.; Tran, T.; Nguyen, H.; Akçapõnar, G.; Ueda, H. Students' emotion extraction and visualization for engagement detection in online learning. In Proceedings of the Procedia Computer Science, Szczecin, Poland, 8–10 September 2021; Volume 192, pp. 3423–3431. [CrossRef]
114. Lee-Cultura, S.; Sharma, K.; Giannakos, M. Using multimodal learning analytics to explore how children experience educational motion-based touchless games. In Proceedings of the CEUR Workshop Proceedings, Virtual Event, 24 March 2020; Volume 2610, pp. 34–39.
115. Vieira Roque, F.; Cechinel, C.; Merino, E.; Villarroel, R.; Lemos, R.; Munoz, R. Using multimodal data to find patterns in student presentations. In Proceedings of the 13th Latin American Conference on Learning Technologies, LACLO 2018, São Paulo, Brazil, 1–5 October 2018; pp. 256–263. [CrossRef]
116. Boulton, H.; Brown, D.; Standen, P.; Belmonte, M.; Kwiatkowska, G.; Hughes-Roberts, T.; Taheri, M. Multi-Modalities in Classroom Learning Environments. In Proceedings of the 12th International Technology, Education and Development Conference (INTED), Valencia, Spain, 5–7 March 2018; Chova, L., Martinez, A., Torres, I., Eds.; INTED Proceedings; pp. 1542–1547.
117. Yu, F.; Chang, L.; Kim, M. An Exploration of a Novice Kindergarten Teacher's Enactment of Multiliteracies Pedagogy during the Pandemic: A Case Study of a Virtual Kindergarten Classroom. In Proceedings of the 30th International Conference on Computers in Education, ICCE 2022—Proceedings, Kuala Lumpur, Malaysia, 28 November–2 December 2022; Volume 1, pp. 527–536.
118. Ramasamy, V.; Kulpinski, E.; Beaupre, T.; Antreassian, A.; Jeong, Y.; Clarke, P.J.; Aiello, A.; Ray, C. Enhancing CS Education with LAs Using AI-Empowered AIELA Program. In Proceedings of the 2024 IEEE Frontiers in Education Conference (FIE), Washington, DC, USA, 13–16 October 2024; pp. 1–9. [CrossRef]
119. Hyndman, J.; Lunney, T.; Mc Kevitt, P. AmbiLearn: Multimodal assisted learning. *Int. J. Ambient Comput. Intell.* **2011**, *3*, 53–59. [CrossRef]
120. Wulansari, R.; Sakti, R.; Saputra, H.; Samala, A.; Novalia, R.; Tun, H. Multimodal Analysis of Augmented Reality in Basic Programming Course: Innovation Learning Modern Classes. *J. Appl. Eng. Technol. Sci.* **2024**, *6*, 115–137.
121. Teotia, J.; Zhang, X.; Mao, R.; Cambria, E. Evaluating Vision Language Models in Detecting Learning Engagement. In Proceedings of the 2024 IEEE International Conference on Data Mining Workshops (ICDMW), Abu Dhabi, United Arab Emirates, 9 December 2024; pp. 496–502. [CrossRef]
122. Bullen, T.R.; Brown, K.; Ogle, K.; Liu, Y.T.; Jurjus, R.A. Using ultrasound to teach living anatomy to non-medical graduate students. *Surg. Radiol. Anat.* **2020**, *42*, 1383–1392. [CrossRef]
123. Wang, L.; Xiao, J.; Qi, Y.; Yu, T. Research on the Live Teaching Effect Based on Multimodal Data Fusion. In Proceedings of the 2021 2nd International Conference on Information Science and Education (ICISE-IE), Chongqing, China, 26–28 November 2021; pp. 935–939. [CrossRef]
124. Suparmi. Engaging Students through Multimodal Learning Environments: An Indonesian Context. In Proceedings of the 4th International Conference on Language, Society and Culture in Asian Contexts (LSCAC), Malang, Indonesia, 24–25 May 2016; Widiati, U., Ed.; KnE Social Sciences; pp. 202–209. [CrossRef]
125. Alabdeli, H.; Obaid, M.K.; Krebat, M.K.; Al-Ani, M.; Muhsin, A. Large Scale Online Vocabulary Classes for prediction of College Student Learning Patterns. In Proceedings of the 2024 International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 12–14 December 2024; pp. 1–6. [CrossRef]
126. Unsal, Z.; Jakobson, B.; Wickman, P.O.; Molander, B.O. Gesticulating science: Emergent bilingual students' use of gestures. *J. Res. Sci. Teach.* **2018**, *55*, 121–144. [CrossRef]
127. Liu, Y.; Sathishkumar, V.; Manickam, A. Augmented reality technology based on school physical education training. *Comput. Electr. Eng.* **2022**, *99*, 107807. [CrossRef]
128. Ahmad, I.; Khusro, S.; Alam, I.; Khan, I.; Niazi, B. Towards a Low-Cost Teacher Orchestration Using Ubiquitous Computing Devices for Detecting Student's Engagement. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 7979766. [CrossRef]
129. Ober, T.; Cheng, Y.; Carter, M.; Liu, C. Leveraging performance and feedback-seeking indicators from a digital learning platform for early prediction of students' learning outcomes. *J. Comput. Assist. Learn.* **2024**, *40*, 219–240. [CrossRef]
130. Martinez-Maldonado, R.; Echeverria, V.; Prieto, L.; Rodriguez-Triana, M.; Spikol, D.; Curukova, M.; Mavrikis, M.; Ochoa, X.; Worsley, M. 2nd Crossmmla: Multimodal learning analytics across physical and digital spaces. In Proceedings of the CEUR Workshop Proceedings, Arlington, TX, USA, 13–17 March 2018; Volume 2163.
131. Olivier, J. Short Instructional Videos as Multimodal Open Educational Resources in a Language Classroom. *J. Educ. Multimed. Hypermedia* **2019**, *28*, 381–409.
132. Magana, A.; Serrano, M.; Rebello, N. A sequenced multimodal learning approach to support students' development of conceptual learning. *J. Comput. Assist. Learn.* **2019**, *35*, 516–528. [CrossRef]
133. Alkhasawneh, I.M.; Mrayyan, M.T.; Docherty, C.; Alashram, S.; Yousef, H.Y. Problem-based learning (PBL): Assessing students' learning preferences using vark. *Nurse Educ. Today* **2008**, *28*, 572–579. [CrossRef]

134. Ouyang, Y.; Gao, W.; Wang, H.; Chen, L.; Wang, J.; Zeng, Y. MEDSQ: Towards personalized medical education via multi-form interaction guidance. *Expert Syst. Appl.* **2025**, *267*, 126138. [CrossRef]
135. Tytler, R.; Prain, V.; Kirk, M.; Mulligan, J.; Nielsen, C.; Speldewinde, C.; White, P.; Xu, L. Characterising a Representation Construction Pedagogy for Integrating Science and Mathematics in the Primary School. *Int. J. Sci. Math. Educ.* **2023**, *21*, 1153–1175. [CrossRef]
136. Tan, S.; Wiebrands, M.; O'Halloran, K.; Wignell, P. Analysing student engagement with 360-degree videos through multimodal data analytics and user annotations. *Technol. Pedagog. Educ.* **2020**, *29*, 593–612. [CrossRef]
137. Shoukry, L.; Bedair, Y.; Elgedawy, I. ClasScorer: Towards a Gamified Smart Classroom. In Proceedings of the 2022 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Alamein New City, Egypt, 18–21 December 2022; pp. 194–201. [CrossRef]
138. Sankey, M.; Birch, D.; Gardiner, M. Engaging students through multimodal learning environments: The journey continues. In Proceedings of the 27th Annual ASCILITE Conference, Sydney, Australia, 5–8 December 2010; pp. 852–863.
139. Samarakoon, L.; Fernando, T.; Rodrigo, C.; Rajapakse, S. Learning styles and approaches to learning among medical undergraduates and postgraduates. *BMC Med. Educ.* **2013**, *13*, 42. [CrossRef]
140. Chinnapun, D.; Narkkul, U. Enhancing Learning in Medical Biochemistry by Teaching Based on VARK Learning Style for Medical Students. *Adv. Med. Educ. Pract.* **2024**, *15*, 895–902. [CrossRef]
141. Zhou, C.; Cai, W.; Shao, F.; Li, M. Intelligent Analysis of Teacher Classroom Management Features Based on Video Stream Data. In Proceedings of the 2024 4th International Conference on Educational Technology (ICET), Wuhan, China, 13–15 September 2024; pp. 516–520. [CrossRef]
142. Arufe Giraldez, V.; Sanmiguel-Rodriguez, A.; Ramos Alvarez, O.; Navarro-Paton, R. Can Gamification Influence the Academic Performance of Students? *Sustainability* **2022**, *14*, 5115. [CrossRef]
143. Criswell, B.; Demir, K.; Zoss, M. A Sequence of Sensemaking in a High School Chemistry Classroom: Tracking Student Thinking and Positioning. *Sci. Educ.* **2025**, *109*, 650–672. [CrossRef]
144. Gaddis, M.L. Faculty and Student Technology Use to Enhance Student Learning. *Int. Rev. Res. Open Distrib. Learn.* **2020**, *21*, 39–60. [CrossRef]
145. Luckin, R.; Cukurova, M. Designing educational technologies in the age of AI: A learning sciences-driven approach. *Br. J. Educ. Technol.* **2019**, *50*, 2824–2838. [CrossRef]
146. Aldosari, M.A.; Aljabaa, A.H.; Al-Sehaibany, F.S.; Albarakati, S.F. Learning style preferences of dental students at a single institution in Riyadh, Saudi Arabia, evaluated using the VARK questionnaire. *Adv. Med. Educ. Pract.* **2018**, *9*, 179–186. [CrossRef]
147. Ally, F.; Pillay, J.D.; Govender, N. Teaching and learning considerations during the COVID-19 pandemic: Supporting multimodal student learning preferences. *Afr. J. Health Prof. Educ.* **2022**, *14*, 13–16. [CrossRef]
148. Moreno, A.C.R.; Taschner, N.P.; Piantola, M.A.F.; Armellini, B.R.C.; Lellis-Santos, C.; Ferreira, R.d.C.C. Real-Lab-Day: Undergraduate scientific hands-on activity as an authentic learning opportunity in microbiology education. *FEMS Microbiol. Lett.* **2023**, *370*, fnad062. [CrossRef]
149. Uchinokura, S. Primary and lower secondary students' perceptions of representational practices in science learning: Focus on drawing and writing. *Int. J. Sci. Educ.* **2020**, *42*, 3003–3025. [CrossRef]
150. Baksh, F.; Zorec, M.; Kruusamäe, K. Open-Source Robotic Study Companion with Multimodal Human–Robot Interaction to Improve the Learning Experience of University Students. *Appl. Sci.* **2024**, *14*, 5644. [CrossRef]
151. Tomlinson, M.M. The model of the "Space of Music Dialogue": Three instances of practice in Australian homes and classrooms. *Music Educ. Res.* **2018**, *20*, 83–101. [CrossRef]
152. Laszcz, M.; Dalvi, T. Studying the affordances of a technology-based nanoscience module to promote student engagement in learning novel nanoscience and nanotechnology concepts at the middle school level. *Res. Sci. Technol. Educ.* **2023**, *41*, 700–716. [CrossRef]
153. Shenoy, N.; Shenoy, A.K.; Ratnakar, U.P. The Perceptual Preferences in Learning Among Dental Students in Clinical Subjects. *J. Clin. Diagn. Res.* **2013**, *7*, 1683–1685. [CrossRef]
154. Azcona, D.; Hsiao, I.-H.; Smeaton, A.F. Personalizing Computer Science Education by Leveraging Multimodal Learning Analytics. In Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 3–6 October 2018; pp. 1–9. [CrossRef]
155. Ren, Z. Optimization of Innovative Education Resource Allocation in Colleges and Universities Based on Cloud Computing and User Privacy Security. *Wirel. Pers. Commun.* 2023, *in press*. [CrossRef]
156. Chaitanya, P.; Sorokina, S.; Basov, O. Multimodal Analysis of Online Webinars Conducted in Zoom. In Proceedings of the International Conference on Research in Education and Science, Antalya, Turkey, 1–4 April 2021; Volume 7, pp. 162–171.
157. Breckler, J.; Joun, D.; Ngo, H. Learning styles of physiology students interested in the health professions. *Adv. Physiol. Educ.* **2009**, *33*, 30–36. [CrossRef]
158. Jalali, A.; Jeong, D.; Sutherland, S. Implementing a Competency-Based Approach to Anatomy Teaching: Beginning with the End in Mind. *J. Med. Educ. Curric. Dev.* **2020**, *7*, 1–5 . [CrossRef]

159. Yaylaci, S.; Ulman, Y.I.; Vatansever, K.; Senyurek, G.; Turkmen, S.; Aldinc, H.; Gun, C. Integrating patient management, reflective practice, and ethical decision-making in an emergency medicine intern boot camp. *BMC Med. Educ.* **2021**, *21*, 536. [CrossRef]
160. Gachago, D.; Cronje, F.; Ivala, E.; Condy, J.; Chigona, A. Stories of resistance: Digital counterstories among South African pre-service student educators. In Proceedings of the International Conference on e-Learning, ICEL, Cape Town, South Africa, 27–28 June 2013; pp. 149–156.
161. Nazir, M.A.; Al-Ansari, A.; Farooqi, F.A. Influence of Gender, Class Year, Academic Performance and Paternal Socioeconomic Status on Learning Style Preferences among Dental Students. *J. Clin. Diagn. Res.* **2018**, *12*, ZC04–ZC08. [CrossRef]
162. Ma, Y.; Zuo, M.; Gao, R.; Yan, Y.; Luo, H. Interrelationships among College Students' Perceptions of Smart Classroom Environments, Perceived Usefulness of Mobile Technology, Achievement Emotions, and Cognitive Engagement. *Behav. Sci.* **2024**, *14*, 565. [CrossRef] [PubMed]
163. Riquelme, F.; Noel, R.; Cornide-Reyes, H.; Geldes, G.; Cechinel, C.; Miranda, D.; Villarroel, R.; Munoz, R. Where are You? Exploring Micro-Location in Indoor Learning Environments. *IEEE Access* **2020**, *8*, 125776–125785. [CrossRef]
164. Cukurova, M.; Luckin, R.; Millán, E.; Mavrikis, M. The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Comput. Educ.* **2018**, *116*, 93–109. [CrossRef]
165. Dooly, M.; Sadler, R. Filling in the gaps: Linking theory and practice through telecollaboration in teacher education. *ReCALL* **2013**, *25*, 4–29. [CrossRef]
166. Jarvela, S.; Nguyen, A.; Hadwin, A. Human and artificial intelligence collaboration for socially shared regulation in learning. *Br. J. Educ. Technol.* **2023**, *54*, 1057–1076. [CrossRef]
167. Ioannou, A.; Vasiliou, C.; Zaphiris, P.; Arh, T.; Klobučar, T.; Pipan, M. Creative Multimodal Learning Environments and Blended Interaction for Problem-Based Activity in HCI Education. *TechTrends* **2015**, *59*, 47–56. [CrossRef]
168. Vasiliou, C.; Ioannou, A.; Zaphiris, P. Measuring students' flow experience in a multimodal learning environment: A case study. In Proceedings of the Lecture Notes in Computer Science, Heraklion, Greece, 22–27 June 2014; Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Volume 8523, pp. 346–357. [CrossRef]
169. Ouhaichi, H.; Spikol, D.; Vogel, B. Guiding the Integration of Multimodal Learning Analytics in the Glocal Classroom: A Case Study Applying MAMDA. In Proceedings of the International Conference on Computer Supported Education, CSEDU-Proceedings, Angers, France, 2–4 May 2024; Volume 1, pp. 478–485. [CrossRef]
170. Kasepalu, R. Overcoming the Difficulties for Teachers in Collaborative Learning Using Multimodal Learning Analytics. In Proceedings of the 2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT), Tartu, Estonia, 6–9 July 2020; pp. 393–395. [CrossRef]
171. Li, X.; Bao, H.; Li, Y. Understanding student engagement based on multimodal data in different types of agent-based collaborative learning contexts. In Proceedings of the 2021 Tenth International Conference of Educational Innovation Through Technology (EITT), Chongqing, China, 16-20 December 2021; pp. 239–244. [CrossRef]
172. Spanjaard, D.; Garlin, F.; Mohammed, H. Tell Me a Story! Blending Digital Storytelling Into Marketing Higher Education for Student Engagement. *J. Mark. Educ.* **2023**, *45*, 167–182. [CrossRef]
173. Ouyang, F.; Dai, X.; Chen, S. Applying multimodal learning analytics to examine the immediate and delayed effects of instructor scaffoldings on small groups' collaborative programming. *Int. J. STEM Educ.* **2022**, *9*, 45. [CrossRef]
174. Cohn, C.; Snyder, C.; Fonteles, J.H.; Ashwin, T.S.; Montenegro, J.; Biswas, G. A multimodal approach to support teacher, researcher and AI collaboration in STEM plus C learning environments. *Br. J. Educ. Technol.* **2024**, *56*, 595–620. [CrossRef]
175. Fahid, F.M.; Lee, S.; Mott, B.; Vandenberg, J.; Acosta, H.; Brush, T.; Glazewski, K.; Hmelo-Silver, C.; Lester, J. Effects of Modalities in Detecting Behavioral Engagement in Collaborative Game-Based Learning. In Proceedings of the LAK23: 13th International Learning Analytics and Knowledge Conference, Arlington, TX, USA, 13–17 March 2023; LAK2023, pp. 208–218. [CrossRef]
176. Celdrán, A.; Ruipérez-Valiente, J.; Clemente, F.; Rodríguez-Triana, M.; Shankar, S.; Pérez, G. A scalable architecture for the dynamic deployment of multimodal learning analytics applications in smart classrooms. *Sensors* **2020**, *20*, 2923. [CrossRef] [PubMed]
177. López, M.; Strada, F.; Bottino, A.; Fabricatore, C. Using multimodal learning analytics to explore collaboration in a sustainability co-located tabletop game. In Proceedings of the European Conference on Games-based Learning, Brighton, UK, 23–24 September 2021; Volume 2021, pp. 482–489.
178. D'Angelo, C.M.; Rajarathinam, R.J. Speech analysis of teaching assistant interventions in small group collaborative problem solving with undergraduate engineering students. *Br. J. Educ. Technol.* **2024**, *55*, 1583–1601. [CrossRef]
179. Gachago, D.; Cronje, F.; Ivala, E.; Condy, J.; Chigona, A. Using digital counterstories as multimodal pedagogy among south African pre-service student educators to produce stories of resistance. *Electron. J. E-Learn.* **2014**, *12*, 29–42.
180. Aloizou, V.; Linardatou, S.; Boloudakis, M.; Retalis, S. Integrating a movement-based learning platform as core curriculum tool in kindergarten classrooms. *Br. J. Educ. Technol.* **2025**, *56*, 339–365. [CrossRef]
181. Whitehead, R.; Nguyen, A.; Järvelä, S. Generative Multimodal Analysis (GMA) for Learning Process Data Analytics. In Proceedings of the CEUR Workshop Proceedings, Kyoto, Japan, 18–22 March 2024; Volume 3667, pp. 214–218. [CrossRef]

182. Lajoie, S.P.; Zheng, J.; Li, S.; Jarrell, A.; Gube, M. Examining the interplay of affect and self regulation in the context of clinical reasoning. *Learn. Instr.* **2021**, *72*, 101219. [CrossRef]
183. Deeg, M.T.; Farrand, K.M.; Oakes, W.P. Creating space for interactive dialogue during preschool circle time using play-based pedagogies and dramatic inquiry. *J. Early Child. Res.* **2020**, *18*, 387–403. [CrossRef]
184. Omoush, M.H.A.; Mehigan, T. Leveraging Robotics to Enhance Accessibility and Engagement in Mathematics Education for Vision-Impaired Students. In Proceedings of the 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkiye, 8–10 June 2023; pp. 1–6. [CrossRef]
185. Beardsley, M.; Martinez Moreno, J.; Vujovic, M.; Santos, P.; Hernandez-Leo, D. Enhancing consent forms to support participant decision making in multimodal learning data research. *Br. J. Educ. Technol.* **2020**, *51*, 1631–1652. [CrossRef]
186. Alwahaby, H.; Cukurova, M. Navigating the ethical landscape of multimodal learning analytics: A guiding framework. In *Ethics in Online AI-Based Systems: Risks and Opportunities in Current Technological Trends*; Elsevier: Amsterdam, The Netherlands, 2024; pp. 25–53. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.