Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Affective Student Engagement Detection using Machine Learning: A Systematic Survey on Methods, Modalities and Learning Environments

Shruthi H K.[1], Jayapriya J.[2], Vinay M.[3], Raju Ramakrishna Gondkar[4]

[1,2,3]Department of Computer Science, CHRIST University, Bengaluru, India.
[4]Department of Statistics, CHRIST University, Bengaluru, India.

E-mail: [1]shruthi.hk@res.christuniversity.in, [2]jayapriya.j@christuniversity.in, [3]vinay.m@christuniversity.in, [4]raju.ramakrishna@christuniversity.in

Orcid ID: [1]0009-0005-8089-2268, [2]0000-0001-6672-6865, [3]0000-0003-0297-3597, [4]0000-0001-6753-5067

Abstract

Student engagement is one of the essential factors an academic success, but tracking the internal affective conditions of learners is a complicated problem in the modern educational environment. The results show that there has been a major shift from using simple behavioral monitoring to more complex affective computing, though visual modalities (63%) are the most used because of their scalability and non-invasive nature. Nevertheless, the results highlight an increased use of multimodal fusion and deep learning systems, particularly, hybrid models (e.g. CNN-LSTM), to effectively learn fine-grained academic emotions (e.g. frustration, boredom, and enjoyment) in online, traditional, and hybrid classes. Furthermore, the review shows a methodological shift towards "in-the-wild" study designs with an emphasis on ecological validity in authentic and uncontrolled learning settings. Although there is significant technical advancement, major gaps remain in establishing standardized benchmark datasets, explaining AI models, and deploying ethical, privacy-protecting frameworks. Overall, this survey highlights that future developments should focus on promoting more responsive, inclusive, and empathetic learning through the development of real-time, pedagogically validated intervention systems.

Keywords: Student Engagement Detection, Emotion Recognition, Affective Computing, Machine Learning, Multimodal, Online and Hybrid Learning.

1. Introduction

Student engagement has been identified as one of the major factors in determining the effectiveness of learning, academic success, and learner satisfaction [1]. Engagement is usually a multidimensional construct in educational research and consists of behavioral, emotional, and cognitive aspects. Behavioral engagement is the observable behavior of students in terms of their involvement in learning processes, concentration, and involvement with learning materials. Cognitive engagement is a measure of the amount of mental effort that students

DOI: https://doi.org/10.36548/jtcsst.2026.3.005
Received: 29.04.2026, received in revised form: 01.06.2026, accepted: 23.06.2026, published: 04.07.2026

apply to the acquisition of tricky ideas and the resolution of problems. Emotional engagement, also known as affective engagement, entails students' feelings towards learning activities, instructors, and the learning environment, such as interest, boredom, frustration, or enjoyment [2]. Among these dimensions, emotional engagement is a very important aspect that determines learning outcomes, as emotional states largely affect attention, motivation, and memory processes. Curiosity and enjoyment are positive emotions that may support learning motivation and more in-depth thinking, and negative emotions like confusion or frustration may impair understanding and limit participation unless they are managed in time [3]. As a result, student's affective state has become a significant topic for learning analytics and educational technology research.

In earlier educational contexts, educators identified the emotional and behavioral engagement of students in the classroom by directly observing and interpreting student cues such as facial expressions, eye contact, posture, and participation patterns. With the growing use of online learning services, hybrid learning settings, and massive digital learning platforms, the traditional method of monitoring engagement has become much harder. The lack of physical contact and the increasing number of students on online platforms make it difficult to measure the affective states of learners and implement timely interventions [2]. To overcome these constraints, researchers have been paying more attention to automated student engagement detection systems based on machine learning and artificial intelligence methods. The goal of these systems is to interpret behavioral and affective cues observed by other sensing modalities, such as visual, audio, and multimodal signals. Among these, visual information such as facial expressions, eye gaze, and head posture has been well-studied due to its usefulness in providing educators valuable insight into the emotions of learners when they are learning. With recent advances in computer vision and deep learning, it has become possible to automatically extract such features from an image or video feed to infer the level of engagement [4].

Advances in recent years indicate that the area of engagement detection has progressed significantly via both machine learning and deep learning. Whereas traditional machine learning techniques rely on handcrafted features from behavioural indications, deep learning models (particularly CNNs and hybrid architectures) are able to automatically learn complex representations from large datasets. Researchers have also begun examining multimodal approaches using visual, auditory, and contextual cues because they provide a richer analysis of the students' affective states. Publicly available datasets that have annotated emotions or engagement provide researchers with the ability to effectively train and validate their machine learning models. These datasets were collected in a variety of different learning environments including controlled laboratory environments, and real-world educational environments, often referred to as "in-the-wild" conditions. The heterogeneity of the datasets, modalities, and experimental designs demonstrates the dynamic nature of this area of research, requiring a systematic approach to investigate the existing methodologies.

Numerous studies have investigated student engagement detection from different perspectives, but considerable limitations remain. Most previous works have been conducted using conventional machine learning approaches, facial expression analysis, or unimodal approaches, but they have largely neglected multimodal fusion, behavioral analysis, and structured student engagement modeling. In addition, although recent studies on deep learning methods and datasets are quite abundant, there still lacks an elaborate comparative analysis of the multimodal models as well as the structured classification of multimodal frameworks [5-9]. To fill these research gaps, this survey proposes a structured taxonomy by modalities, machine learning approaches, deep learning frameworks, datasets, learning environments, and practical problems as discussed in section 4.

However, despite these disadvantages, all the above-reviewed papers have made a substantial contribution to the scientific knowledge on engagement detection in education; nevertheless, there is still an opportunity to delve into the problem of student engagement detection from an emotional perspective. This survey aims at developing a more detailed review of emotion-oriented engagement detection, focusing on techniques, datasets, fusion methods, evaluations, and challenges. The taxonomy of approaches for affective student engagement detection is given in Figure 1.

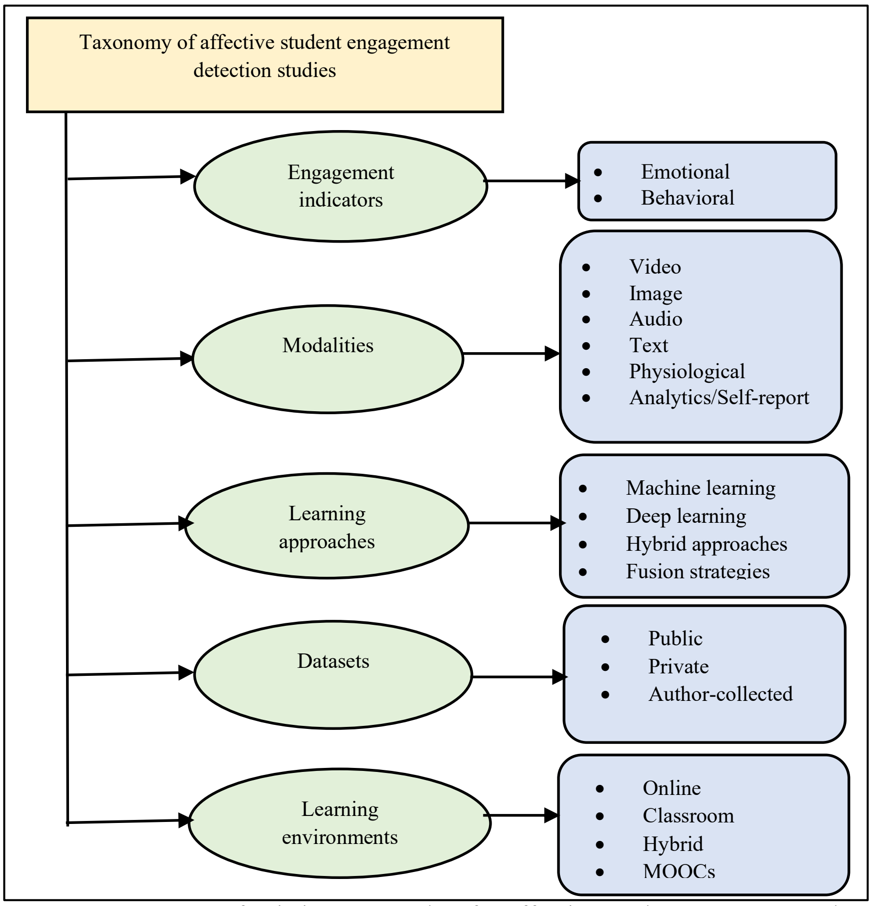

Figure 1. A taxonomy of existing approaches for affective student engagement detection

This survey is aims to advance the field of engagement detection systems based on emotions by synthesizing current knowledge on this topic. The research questions formulated for this systematic survey are as follows:

- RQ1: What are the emotional and behavioral indicators used to detect engagement in existing machine learning literature?
- RQ2: Which datasets, learning environments and learning methods are used in this field?
- RQ3: How do existing studies on affective student engagement detection compare?

The rest of the paper is organized as follows: Section 2 provides the related work. Section 3 describes the process of literature search and selection. Section 4 summarizes the results in accordance with the research questions. Section 5 discusses the review findings. Lastly, Section 6 presents the conclusion and future research opportunities.

2. Related Work

Student engagement is the basis of the learning process and has a tremendous effect on academic performance, retention, and satisfaction. It is not just physical attendance, but active engagement, and the cognitive and affective states of the student. All educators agree that emotion plays a crucial role in the learning process. Whatever the affective state of the learner might be be it curiosity, frustration, or enjoyment—it will affect their cognition, memory development, and academic performance. Traditionally, these patterns of behavior and affect were intuitively recognized by experienced teachers, who then made changes in the way they conducted their classes to create better learning experiences based on situational factors. However, the transition into the educational environment brings serious challenges to such an intuitive approach. The ideal learning environment ensures constant engagement and response to various emotional states of learners, and the provision of interventions when they are most required. However, because of the large size of contemporary learning environments and their asynchronous nature, individual and manual observation is nearly impossible. This often results in unnoticed student disengagement, which leads to poor academic performance, low retention rates, and inadequate comprehension of the curriculum. Lack of timely, personalized feedback will also contribute to feelings of detachment from the learning process on both sides [10].

To address these challenges, several recent reviews have provided useful insights into the field of student engagement detection, analyzing both machine learning and computer vision methods in the context of the educational setting. However, the depth of methodological coverage and the scope of these reviews differ substantially. Study [5] presented a literature review of automatic engagement estimation in smart learning settings and classified research according to broad computational strategies and engagement dimensions. This research is valuable as it systematizes engagement into several dimensions and identifies frequently used data sources to analyze engagement. Their results show that engagement estimation is becoming progressively employed with the use of machine learning methods. Similarly, study [7] concentrated mainly on facial expression recognition (FER) as part of education research. This research is significant because it provides a systematic summary of the types of classifiers and datasets and shows the increasing significance of FER in understanding student emotions. The results indicate that facial expression-based methods are the most prevalent in engagement-related studies, with deep learning models gaining popularity.

Similarly, the study [6], explored the academic emotion classification in online learning settings. According to their results, deep learning models are more dominant in emotion classification tasks. This research enhances the literature through the identification of popular datasets such as FER2013 and DAiSEE and the presentation of performance indicators that include the high rates of classification in different studies. Lastly, the recent research [9] conducted a meta-oriented analysis of machine learning models for predicting student engagement and performance in online higher education. The contribution of this research lies in its review of 96 primary studies and identification of popular machine learning models, particularly those related to classification methods. The findings showed that classification methods were prevalent while clustering methods were not, and engagement prediction was less explored than performance prediction. Likewise, the research [8] conducted an extensive

SLR study of student engagement detection using computer vision methods, and it explores visual cues such as facial expression, gaze, and posture. The results indicate that multimodal computer vision systems are highly accurate, with an average of 85% to 99% accuracy, and are increasingly being applied in online learning platforms.

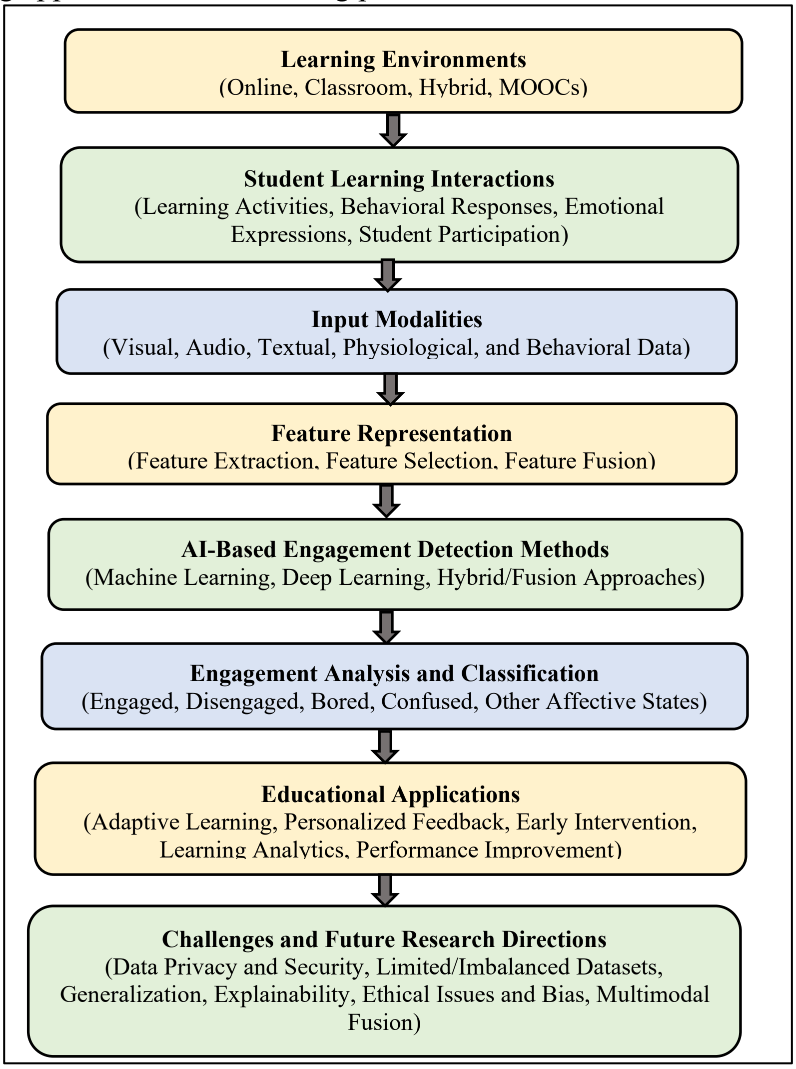

Figure 2. Proposed conceptual framework for the systematic survey on affective student engagement detection

Therefore, considering the rapid evolution of machine learning methods and the increasing access to multimodal educational data, it is important to have a comprehensive understanding of existing engagement detection approaches. The background given in this section sets the proposed conceptual framework of student engagement and affective computing in education, which underlies the further analysis of the literature and review of machine learning-based engagement-detecting methods, as illustrated in Figure 2.

3. Search Strategy and Selection Process

This section provides the methodology, based on PRISMA guidelines, applied to identify, select, and summarize relevant research articles in the sphere of automated student engagement detection. This includes the studies that utilize machine learning and deep learning algorithms in different modalities and learning contexts, as shown in Figure 3.

3.1 Search Strategy

The literature search was conducted solely on the Scopus database, chosen for its comprehensive search of high-impact, peer-reviewed journals and conference proceedings available across a wide variety of interdisciplinary topics related to this research. The advanced search features and standardized indexing of the database increase the accuracy and reproducibility of the literature search. To identify studies, a search was conducted for journal articles and conference papers published in English between 2014 and January 2026. The initial literature search in Scopus yielded 162 documents on the topic of affective student engagement detection. During the identification phase, manual verification of titles, authors, and publication information was conducted to identify possible duplicate records, but no significant duplicates were found. The retrieved studies were further screened based on inclusion and exclusion criteria through title, abstract, and full-text review. The screening and data extraction processes were conducted by multiple reviewers. In the first screening phase, 41 records were excluded for being unrelated to educational engagement, lacking an emotion recognition component, or not contributing to engagement detection. As a result, 121 records were retained for further evaluation. Subsequently, another 41 studies were eliminated due to lack of data, ambiguous methodology, or not being final research papers. The remaining 80 studies were assessed for eligibility through detailed examination. Finally, 6 studies were excluded because they did not fit within the scope of this review, resulting in 74 studies included in the qualitative analysis, as shown in Figure 3.

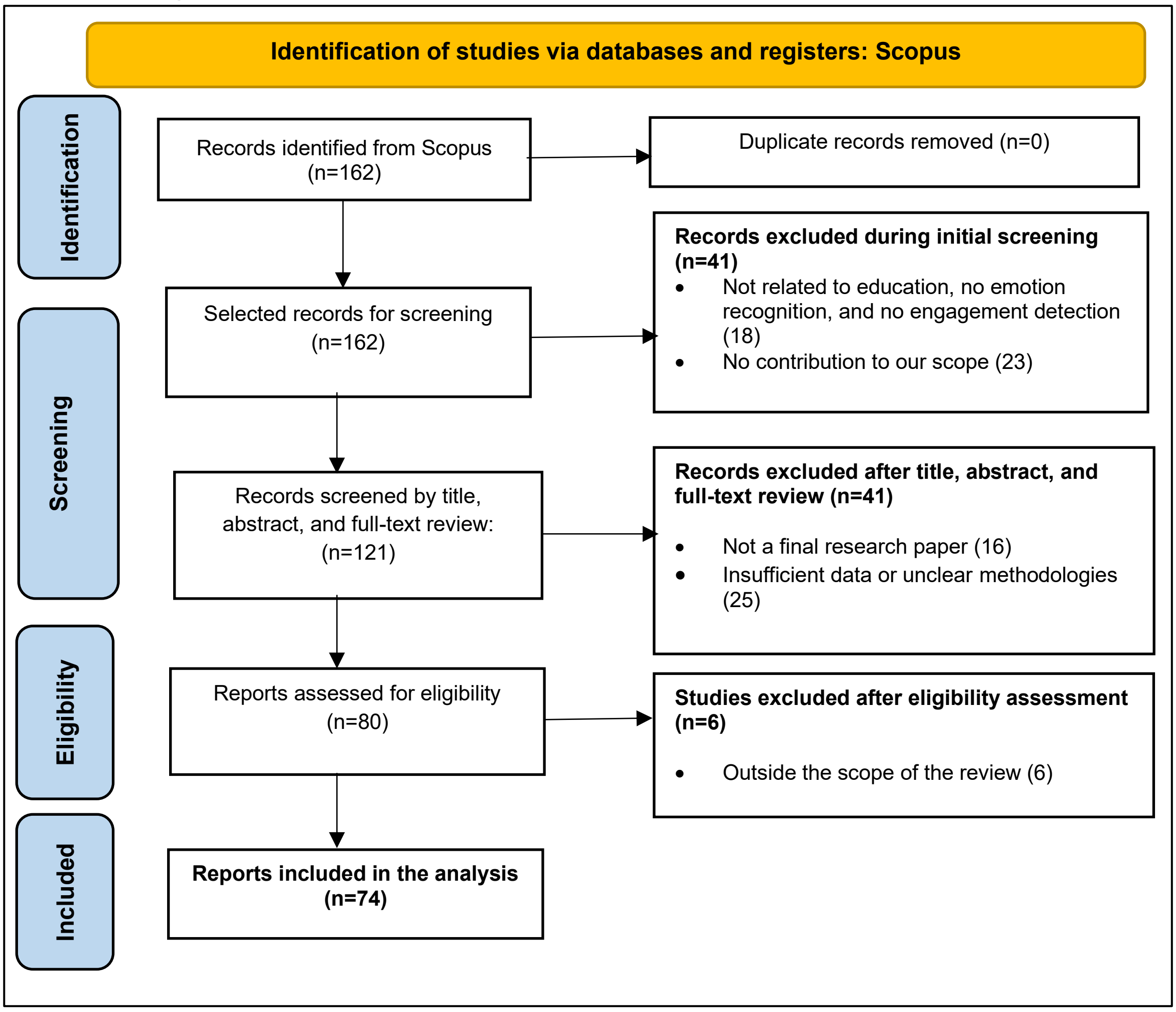

Figure 3. The PRISMA guidelines followed in the systematic survey process

The search query utilized a combination of keywords and Boolean operators to retrieve studies that covered different aspects of engagement detection based on emotions. The final query was formulated to include keywords related to "student engagement," "emotion detection," "machine learning," and "deep learning" to ensure wide coverage of the relevant literature in the field.

The final search query was applied in Scopus in the following way:

TITLE-ABS-KEY (("student engagement" OR "classroom engagement" OR "learner engagement") AND ("emotion recognition" OR "facial expression recognition" OR "affective computing" OR "emotion detection") AND ("machine learning" OR "deep learning" OR "convolutional neural network" OR CNN OR "multimodal learning")) AND (PUBYEAR > 2013 AND PUBYEAR < 2027)

3.2 Inclusion and Exclusion Criteria

To ensure that the selected studies had direct links to the objectives of the survey, the inclusion and exclusion criteria were adhered to strictly. This helped reduce the initial list of 162 articles to those that are relevant and of high quality and form part of the review's findings, as shown in Table 1.

Table 1. Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Topic focus	Studies focus on student engagement detection based on emotions using AI, machine learning, deep learning, or affective computing techniques in educational settings.	Research not focused on emotion/engagement detection, or not in an educational context. General AI/ML applications without a clear link to student emotion/engagement detection.
Methodology	Empirical research that includes information on experimental design, methods and findings.	Theoretical articles, opinion articles or editorials, or research that lacks empirical evidence or adequate information regarding experimental designs.
Technology use	Studies utilizing AI, machine learning, deep learning, or affective computing in emotion recognition.	Studies reporting pedagogical interventions that do not involve AI-driven emotion assessment, papers that do not explicitly use advanced computational models for emotion detection.
Level of detail	Publications that contained adequate information regarding the design, implementation, and assessment of models or systems.	Publications with inadequately detailed descriptions of methodology, model structure, data collection or evaluation metrics.
Publication type	Peer-reviewed journal articles, conference papers, and systematic reviews.	Books, book chapters, dissertations (unless published as papers), pre-prints not yet peer reviewed.
Language	Publications that are only in English.	Publications that are not in English.
Time frame	Studies published between 2014 and 2026 (January)	Studies published before 2014.

3.3 Data Analysis

This group consists of 74 articles covering student engagement detection based on emotions from 2014 to 2026. This 2014-2026 period was chosen for two reasons: (1) a ten-year period is commonly used in systematic surveys for maintaining relevance and convenience, and (2) it is sufficient for identifying emerging trends and existing research gaps within this specific area.

From a historical perspective, the development of educational applications in engagement monitoring began in 2014 when researchers were exploring facial expression recognition, head pose estimation, and gaze tracking in education. Thus, 2014 was chosen as the start of the review since it covers the recent wave of research inspired by deep neural networks, which is the state-of-the-art practice in this field. January 2026 was chosen as a deadline to include the newest publications available at the time of conducting the review. Among the 74 articles, 20 are conference proceedings, while 54 are peer-reviewed journal articles. Figure 4 depicts the distribution of publications per year.

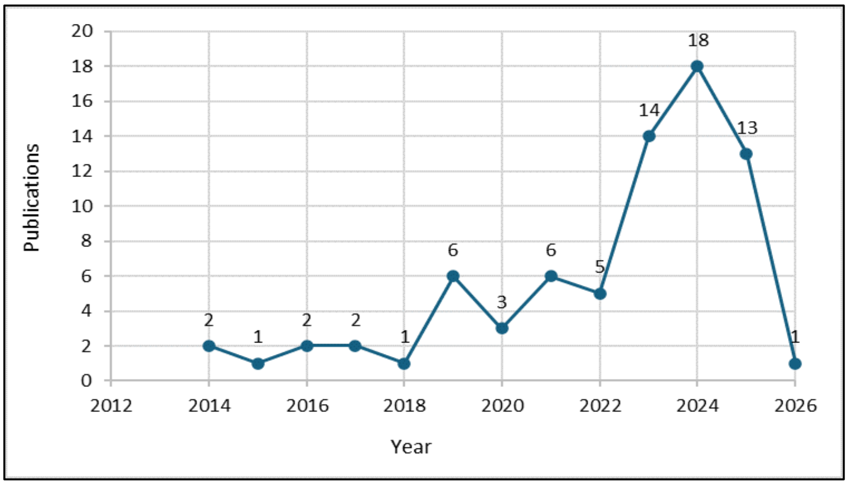

Figure 4. Number of publications included in the review per year (2014-2026)

To analyze the final set of selected studies, a structured data extraction process using Excel was utilized. The texts of the chosen studies were read manually, and the main information was carefully copied into predefined columns in Excel spreadsheets. These included publication year, keywords, types of engagement discussed, input modalities, machine learning techniques utilized, dataset characteristics, learning contexts (e.g., classroom, online, e-learning, virtual learning, MOOC learning, Virtual Reality, and Augmented Reality), and reported evaluation metrics. This systematic tabulation has helped in the thematic and quantitative arrangement of the studies, the identification of common patterns, and the examination of trends across the literature. The tabulated data were systematically derived into the thematic findings delineated in Section 4.

In data analysis, studies involving multiple modalities or hybrid analytical approaches were counted within each of the relevant categories to ensure the data reflect the diversity of ways in which emotion-based student engagement has been detected. For example, studies that used both audio and video modalities were counted under both modality categories, and hybrid architectures like CNN and LSTM or CNN and BERT were counted under each of the methodological categories. Therefore, the total number of reported frequencies may exceed the total number of reviewed studies.

4. Results

In this section, the significant findings of the survey are summarized in terms of major themes, datasets, methodologies, results, and applications identified in the reviewed literature in relation to the previously mentioned research questions.

The indicators extracted during coding were grouped into emotional and behavioral categories according to the dominant engagement cue reported in each study. Emotional indicators included affective states derived from facial expressions, speech emotion, physiological responses, or sentiment-related features. Behavioral indicators included gaze direction, head movement, posture, interaction behavior, and learning activity patterns. The studies performed with multimodal methods were categorized based on the main indicator reported by the authors.

In addition, the review of the included studies indicates a research inclination towards automated, non-intrusive monitoring. According to a wide variety of modalities, as shown in Figure 5, one of the most significant concentrations of modalities is on visual data (63% overall) to record both overt behavioral actions and latent emotional states. Image and video-based modalities are used as the primary sources of engagement detection.

In particular, the most common method of recognizing "learning-centered" emotions, such as boredom, confusion, and frustration, is Facial Expression Recognition, which usually employs deep learning models such as CNNs and MobileNetV3. In addition to general affect, 13% of the studies are dedicated to fine-grained ocular and positional indicators such as eye gaze, blink rate, and head pose, to identify attention changes and mind-wandering [8]. These spatiotemporal features are often obtained using technical frameworks such as Multi-Task Cascaded Convolutional Neural Network (MTCNN). Although most literature relies on visual data, about 37% of the literature includes non-visual or physiological indicators. This incorporates text and speech-based sentiment analysis to measure internal affective states, which cannot be observed using facial expressions. Also, physiological measures such as EEG offer high-resolution valence and arousal signals, whereas learning analytics, such as clickstream logs and mouse movements, relate system interactions to behavior participation [1]. These indicators cluster engagement into two dimensions, as follows: emotional engagement, regarding internal motivation and academic affect, and behavioral engagement, regarding outward participation through task-related physical actions. The combination of these internal and external indicators is meant to produce a multidimensional portrait of student involvement in the reviewed works.

The distribution of various modalities, as shown in Figure 5, is skewed toward visual data, with the most common indicators being facial expressions, followed by general images and head pose/movements. The least common data sources are contextual and internal modalities such as textual sentiment, speech, physiological signals and learning analytics and these are generally used in controlled experimental situations where a particular sensor or platform log is available. It must be mentioned that the total distributions and frequency of modalities presented in Figure 5 are greater than the total number of studies selected (n = 74), because some studies used more than one modality in a single engagement detection system.

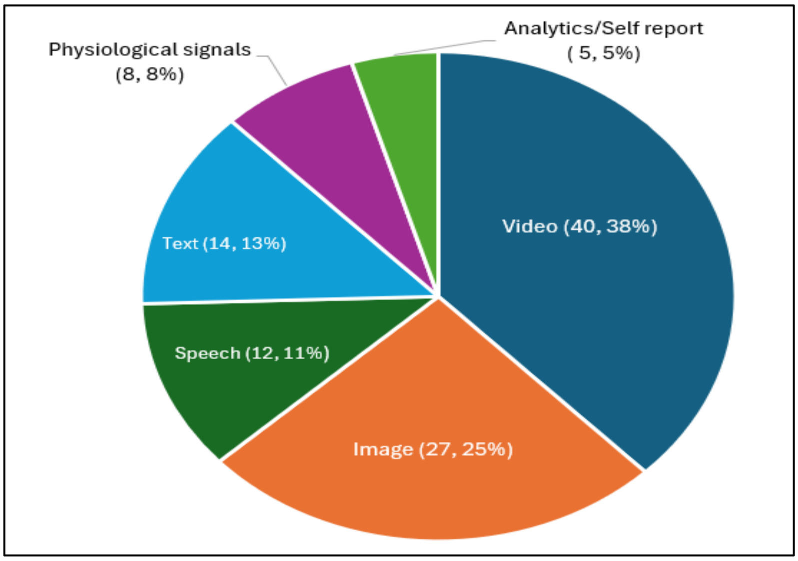

Figure 5. Distribution of modalities used for emotional and behavioral indicator extraction in the selected studies

The review of the chosen studies indicates a highly fragmented data landscape, where many researchers frequently balance established public datasets with context-specific, author-collected data. As shown in Figure 6, most research depends on primary datasets for model development and evaluation. These studies draw on author-collected data to measure engagement in natural educational context, e.g. MOOCs or virtual reality classrooms, which are not necessarily reflected in general-purpose datasets [7]. Some studies relied on publicly available secondary datasets, which are more commonly used due to their availability and standardized annotations. Additionally, a few studies followed a hybrid approach by combining both primary and secondary data. Among the publicly available standardized datasets, FER-2013 and CK+ are the most used in training facial expression recognition models, while education-specific datasets such as DAiSEE, AffectNet, and RAF-DB offer more in-the-wild examples of learning-focused affective states.

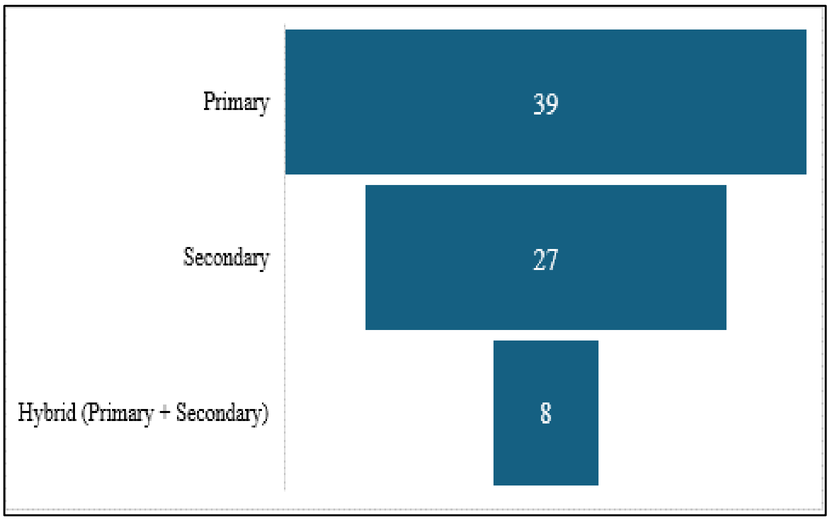

Figure 6. Distribution of dataset types in the included studies

There are two types of environments used for data sources, as indicated in Table 3: in-the-wild environments (such as Engagement and DAiSEE) and controlled laboratory environments (such as CK+). The use of in-the-wild datasets is favored by researchers for developing useful models that can be applied in actual online learning settings, considering variations in lighting and background noise. However, highly controlled laboratory settings provide quality-labeled data but have been criticized for not reflecting natural behaviors. The distribution of various learning contexts considered in this survey is depicted in Table 2. The results show that most research on student engagement detection based on emotions was conducted on online learning platforms.

Table 2. Distribution of learning environments considered in the reviewed studies

Learning Context	Number of Studies
Online learning	42
E-learning	10
Classroom learning	7
Virtual learning	5
Lab Settings	3
MOOC learning	3
Hybrid learning	2
Virtual reality	2

In the analysis of the chosen literature, a transition to more advanced computational models is noted, with deep learning noticeably more dominant in processing multidimensional affective data than traditional machine learning. As shown in Figure 7, most of the studies only use DL, a few are based on traditional ML, 20% are based on a hybrid ML-DL methodology, and only 7% use non-ML systems. These models are further integrated by data fusion strategies that are outlined in Figure 8. Although 75% of studies have not specified the fusion strategy, a few of these studies were based on a single modality. The remaining 25% of multimodal studies use Early Fusion (concatenating features at an early stage), Late/Decision-Level Fusion (independent signal processing), and Hybrid Fusion strategies. It is important to note that decision-level fusion has performed better than feature-level methods in MOOC settings and has achieved accuracies of 81.90% [10]. Such strategies enable a stronger examination of learning behaviors and sentiment, which leads to better emotional feedback and participation tracking.

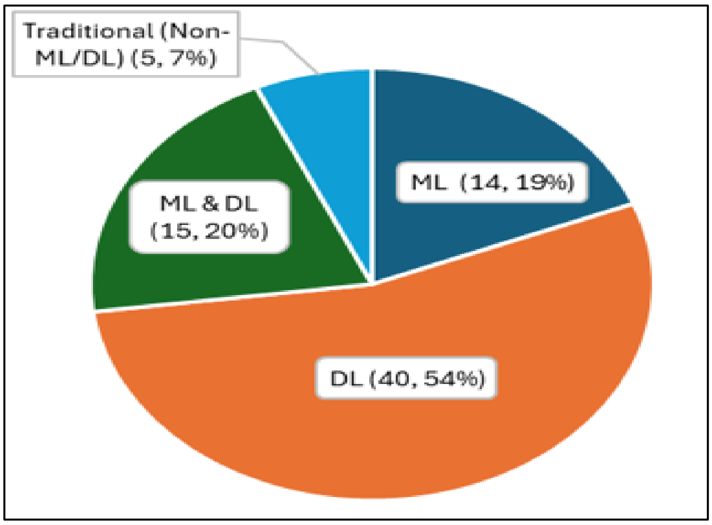

Figure 7. Distribution of ML/DL approaches included in the studies

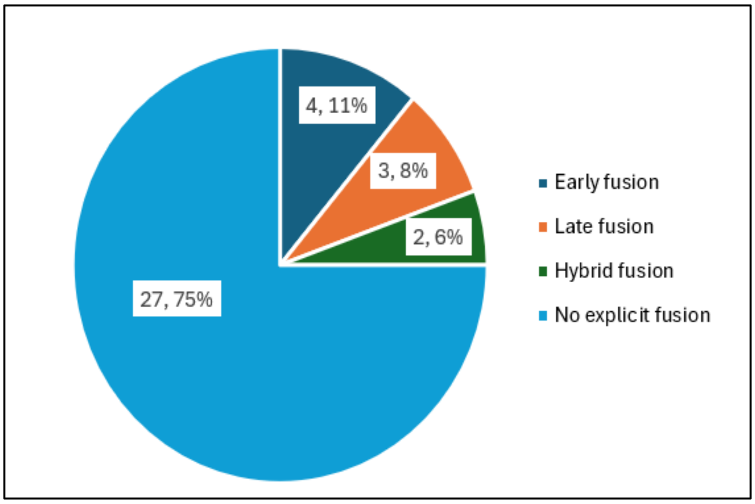

Figure 8. Fusion strategy distribution

From the comparison of the articles studied, there is a clear trade-off between the level of understanding emotions and the scalability of the systems used in practice. The types of modalities are video, images, speech, text, and physiological signals, each analyzed using various methods. Video, images, and speech signals are usually analyzed using deep learning models such as CNN, ResNet, and LSTM, while text signals are analyzed using BERT, Transformer, and traditional classifiers such as Naive Bayes and SVM. Physiological signals, on the other hand, are analyzed using traditional machine learning algorithms such as SVM, Random Forest, KNN, and boosting algorithms [6].

In addition, the comparative analysis demonstrates that deep learning and hybrid architectures are the most extensively used techniques for detecting student engagement, particularly in visual and multimodal environments. CNN, Transformer, and hybrid temporal models proved successful with large-scale databases such as AffectNet and FER-2013. The values of performance criteria reported in the papers under review are different and include accuracy, Pearson correlation coefficient (PCC), and mean squared error (MSE). One needs to treat the results cautiously since the high accuracy achieved in controlled physiological and MOOC-based datasets is likely to be biased owing to the lack of diverse participants, uneven distribution of the number of participants among the datasets, and more restrictive experimental settings. Moreover, sophisticated multimodal models are likely to be harder to apply in real-world settings due to their increased complexity. Table 3 provides an overview of selected studies. Evaluation of the performance of different studies is difficult owing to the differences in datasets used, sample sizes, criteria used for evaluation, and experimental setup. Therefore, the robustness and complexity of the dataset have been qualitatively evaluated in terms of size, diversity, quality of annotations, modality richness, and features of the model architecture.

Table 3. Benchmark comparison of selected machine learning-based student engagement detection studies

Study	Dataset (Type & Access)	Sample Size	Modalities	Learning Environment	Dataset Robustness	Computational Complexity	Model	Performance Metric
Singh et al. [11]	Author-collected experimental multi-faceted dataset: EngageNet, Public & Available at, (https://github.com/engagenet/engagenet_baselines)	127 subjects (83 males and 44 females), Age range of 18 to 37 years	Behavioral data, Cognitive data, Personality data	Real-world online learning	High	High	LSTM, CNN-LSTM, Temporal Convolutional Network (TCN), Transformer	Accuracy = 67.61%
Sassi et al. [12]	FER-2013 dataset, Public & Open access, available on Kaggle	35,887 grayscale images	Facial expressions	Online learning	Moderate-High	Moderate	CNN-RF hybrid model	Accuracy = 71.86%
Dey et al. [13]	Student Online Action Image dataset (SOAId), Public & Available at, (https://sites.google.com/view/student-actions/)	2029 student-centric images	Images	Online interactive learning	Moderate	High	Attention-Based AdaptSepCX Network	Accuracy = 96.60%
Su et al. [4]	DAiSEE and EmotiW-EP datasets, Public & DAiSEE: Restricted Access, EmotiW-EP: Registration required	DAiSEE dataset: 9,068 video clips, Collected from 112 students, EmotiW-EP dataset: 78 subjects (25 female, 53 male) aged between 19 and 27 years, contains 262 video instances	Video	Online learning	High	High	PANet (Part Attention Network), STformer (Spatial–Temporal Transformer)	DAiSEE dataset, Accuracy = 64.0%, EmotiW-EP dataset, MSE = 0.0729
Shomoye & Zhao [14]	AffectNet dataset, Public & Restricted Access, license agreement required	Containing over one million facial images annotated for eleven different expressions	Image: Facial expressions	Virtual Reality (VR)-based classroom	High	High	Custom-trained ResNet50-based CNN model	Accuracy = 91.11%
Gupta et al. [15]	WIDER Face dataset, CK+ dataset, FER-2013 dataset, Public & Restricted Access: WIDER Face, CK+, Open Access: FER-2013	WIDER Face contains 32,203 images, CK+ contains 593 video sequences, FER-2013 contains 35,887 facial images	WIDER Face, FER-2013: Image, CK+: Video	Real-world online learning	High	High	VGG19, ResNet50	Accuracy = 92.58%
Moise et al. [16]	DEAP dataset, Public & Restricted Access	32 participants (1,280 trials)	Physiological data: Reduced set of 5 Electroencephalogram channels (FP1, AF3, F7, T7, FP2)	Controlled classroom/laboratory	Moderate	Low-Moderate	1D-CNN	Accuracy = 99.91% for excitement

Banerjee et al. [17]	RAF-DB, AffectNet and FER-2013, Public & RAF-DB: Restricted Access, AffectNet: Restricted Access, FER-2013: Open Access	RAF-DB: 15,339 single-class photos tagged, seven main emotion classes, AffectNet: Over one million facial images, FER-2013: 35,887 facial images	Image	Real-world facial emotion recognition environment	High	Very High	CNN-based multi-scale feature extractors, ResNet50, Mixture of Experts (MoE) framework	AffectNet7:Accuracy = 74.40%, AffectNet8:Accuracy = 71.98%, RAF-DB: Accuracy = 83.41%, FER-2013:Accuracy = 67.05%
Kaur et al. [18]	Author-collected experimental dataset: Engagement in the Wild, Public & Available at, (http://iitrpr.ac.in/lasii)	264 videos collected from 91 participants (27 females, 64 males), Ages ranging from 19 to 27 years	Video (Facial expressions, Head movements, and Eye gaze)	Real-world online learning	Moderate-High	Moderate-High	Random Forest, LSTM-based Deep Sequence Network, Deep Multi-Instance Learning (DMIL) Network	Random Forest regressor, MSE = 0.09; PCC = 0.17
Tao et al. [19]	Author-collected experimental dataset, Private & Availability: Granted upon request and approval from University of Adelaide's Online Programs Team	4553 students in 2019 and 4258 students in 2020	Textual data: Discussion forum posts by MOOC students	MOOCs (Massive Open Online Courses) environment	High	Moderate	Naïve Bayes Classifier, Random Forest, Deep Learning Artificial Neural Network (ANN)	Accuracy = 99.35% for the 2020 audit student cohort

Table 4 provides a comparative analysis of recent selected studies based on their applications, strengths, and limitations. There have been major advancements in real-time tracking and resource optimization. An example of such applications is video-based models using CNN-Random Forest, enabling real-time visualization of affective states projected into psychological frameworks such as the Plutchik wheel, which allows for real-time intervention by instructors. Moreover, physiological applications based on 1D-CNNs EEG can achieve a superior accuracy of 99.90% on the DEAP dataset with just five channels, minimizing hardware limitations. The DEAP dataset was gathered from 32 subjects in a controlled experimental setting using music video stimuli and includes both physiological and emotional responses [16]. The implementation of a hybrid deep learning architecture, including a mixture of expert systems, has additionally enhanced the capacity of such architectures to generalize in various environmental conditions, including lighting and occlusion.

Despite these technical innovations, there are several common constraints to the general implementation of engagement detection systems. Most high-performing vision-based systems, including those that use XGBoost, are dependent on explicit behavioral indicators, such as hand tracking and head pose, and do not focus on more latent emotional assessments. Another persistent issue is data imbalance: the models are more precise when detecting shared emotions such as happiness and less precise when detecting less frequent states such as fear or disgust. Additionally, existing systems generally struggle with nuanced facial expressions and are not strictly validated by professional psychologists, which raises the question of pedagogical accuracy [7].

In terms of applications, the discipline is divided into real-time classroom intervention and post-hoc curriculum optimization. Real-time student participation tracking is facilitated mostly through vision and speech-based systems, and text-based mining based on BERT is used to review MOOCs to find areas of improvement in the course [10]. To obtain robust detection, as summarized in Table 4, more diverse sources of data and deeper psychological grounding will be needed.

Table 4. Overview of comparative analysis of selected studies: applications, strengths, and limitations

Study	Applications	Strengths	Limitations
Hossen & Uddin [20]	Student real-time attention monitoring in online classes	Integrate face recognition, hand tracking, and pose estimation for behavior tracking, Considers privacy and ethical factors	Reliance on observable behaviors, System lacks a deeper emotional analysis
Sassi et al. [12]	Online learning real-time monitoring and visualization of students' emotional states	Plutchik wheel–based mapping with real-time visualization	Limited detection of subtle facial expressions, Emotion state mapping lacks psychological validation
Xu et al. [3]	Automatic recognition of student academic emotions in images or videos	Spontaneous expressions in uncontrolled settings improve model robustness	Limited sample size, Limited gesture images in the dataset, Lacks multimodal signals such as speech, posture, and physiological data
Moise et al. [16]	Integration of automated emotion recognition in educational contexts	High accuracy using only five EEG channels, Reduced resource consumption and faster training	Class imbalance in the dataset affects robust model training, Data not collected in real learning environments
Qin et al. [21]	Analyze the emotions of college students in online education	Uses emotion data to enhance multimodal sentiment detection accuracy	Reliance on emoji labels limits model generalizability and robustness
Pei et al. [22]	Improve the overall quality of online courses	Identifies course attributes, learner satisfaction, attention, and improvement costs	Limited to two programming courses, reducing generalizability to other MOOCs
Wang et al. [23]	Intelligent multimodal learning path recommendation in digital education	Integrates learning behavior and sentiment analysis to improve detection accuracy	Single data source limits the generalizability of findings
Banerjee et al. [17]	Robust Facial Emotion Recognition (FER) system for detecting complex emotional expressions	Hybrid architecture improves generalization across datasets and real-world conditions	Class imbalance leads to biased emotion detection performance

The data modality-wise benchmarking analysis of the reviewed studies is presented in Table 5. Visual modalities achieved the best results on benchmark emotion recognition datasets like FER2013 and RAF-DB, while engagement-specific video datasets like DAiSEE reported lower results, indicating the higher complexity of authentic student engagement recognition. The physiological methods showed the highest accuracy, but these results were achieved in a controlled experimental context and may not be directly transferable to a real educational context. Multimodal and text-based methods also showed good results, highlighting the importance of contextual, behavioral and social interaction data for student analytics.

In addition, formal performance normalization, statistical aggregation, or meta-analysis was not performed because the reviewed studies utilized different datasets, modalities, sample sizes, evaluation metrics (e.g., accuracy, PCC, and MSE), and experimental protocols. These differences restrict the direct quantitative comparison of studies. Therefore, a modality-wise benchmarking approach is used to make a structured comparison between representative datasets, architectures and reported performance that also aims to highlight the impact of modality characteristics and the complexity of the datasets on engagement detection outcomes.

Table 5. Comparative benchmarking of representative studies based on data modality, representative datasets, best-performing architectures, and reported performance

Data Modalities	Representative Datasets	Best Architecture	Highest Reported Performance by Modality, (Study)
Visual (Image/Video)	FER2013, RAF-DB, RAFD, SCAUOL	MobileViT-Local-S with VOLO-D2 Knowledge Distillation	Accuracy = 94.96%, (Wang et al. [23])
Video (Engagement-specific)	DAiSEE	PANet and STformer	Accuracy = 64.0% , (Su et al. [4])
Text	Author-collected animated online education dataset	Information Block Bidirectional Long-Short term Memory (IB-BiLSTM)	Accuracy = 93.92%, (Jun et al. [24])
Physiological	DEAP	1D-CNN	Accuracy = 99.91%, (Moise et al. [16])
Multimodal (Text + Behavioral + Social Network)	Author-collected Multisource Dataset	BERT-GCN	Accuracy = 92.5%, (Wu [25])

5. Discussion

The findings of this survey indicate that student engagement detection has greatly evolved, with more complex forms of behavioral monitoring replaced by advanced, AI-based affective computing in online, traditional, and hybrid environments [2]. Furthermore, Figure 9 illustrates the primary research challenges, existing solution strategies, and future research requirements identified in the reviewed literature. This figure serves as a guide for the ongoing advancement of student engagement detection systems.

5.1 Integration of Affective and Behavioral Indicators

One of the main findings of this survey is the increasing focus on recognizing a student's emotional and cognitive state, rather than simply monitoring learning activities [4]. Visual modalities are the most prevalent methods, as they are non-invasive and have been aided by progress in deep learning computer vision technologies. Their effectiveness can be influenced by the environment, facial coverings, and the range of emotional expression [8]. Physiological and multimodal approaches, on the other hand, can deliver more comprehensive engagement information and sometimes better performance, but will require extra sensors, increased computational resources, and may be more privacy-sensitive. The results indicate that future engagement detection systems should incorporate a combination of emotional and behavioral data to more accurately and effectively measure engagement in real-world learning settings.

5.2 The Role of Deep Learning in Managing Complex Environments

The rise in the use of CNN-based and hybrid approaches highlights the importance of automating feature extraction and analyzing engagement in educational environments. While such methods are effective, they require quality annotations and training on different types of data gathered under controlled settings. Issues such as varying lighting conditions, occlusions, poor labeling, and cultural differences may influence the performance of such methods when applied in actual classroom settings. Engagement detection is a complex process that relies on combining various kinds of data, something made more difficult due to high computational demands [2]. The above discussion reveals an urgent need to develop engagement detection methods that provide robust, understandable, and human-centered analysis to educators.

5.3 Ecological Validity and the "In-The-Wild" Shift

The domination of author-collected data and "in-the-wild" research designs stands out as a transformative step towards ecological validity. A series of recent studies demonstrates a tendency towards greater emphasis on real-life learning conditions, including field recordings conducted in classrooms and wearables as means to reduce intrusion and increase ecological validity. The aim of this trend is to increase the quality of engagement detection models in terms of their ability to work in real-life situations and under conditions like various lighting, occlusions, and diversity of students, although laboratory experiments currently prevail in validating the methodology [15]. This method also introduces an additional level of environmental noise that face alignment cannot easily cope with, thus creating a need for the development of standardized evaluation schemes specific to the needs of education [7]. The results of the survey indicate that most student engagement detection research utilizes public visual datasets, namely FER-2013, CK+, DAiSEE, RAF-DB, and AffectNet, due to their availability and appropriateness for deep learning facial analysis tasks. The research focus is mostly on the online learning environment, reflecting the growing need for automated engagement monitoring in e-learning systems. While some research used multimodal and hybrid datasets to enhance robustness, the literature remains limited in the use of physiological, textual, and real-world classroom datasets, reflecting problems with data diversity, scalability, and application to a wide range of educational settings.

5.4 Balancing Precision, Privacy, and Practicality

The most prominent trade-off between the compared current technologies is that the ideal engagement detection system is both unobtrusive and morally acceptable. Whereas vision-based systems are extremely scalable for a daily use model, it is also found that these systems fail to record subtleties or internal emotional states, which can be detected by physiological indicators [4]. Thus, future technological advancements should consider the development of a multimodal learning analytics interface integrating pose dynamics and facial expressions within one inexpensive platform [26]. Also, problems regarding morality, such as privacy at home in the context of virtual learning and surveillance in classroom settings, should be resolved to ensure the sustained utilization of these technologies in K-12 and higher education sectors [14]. In addition, the survey results reveal the effectiveness of visual modalities, since the deep learning architectures of CNN, ResNet, and LSTM demonstrated high effectiveness in recording the spatial and temporal attributes in engagement detection tasks. Despite the high robustness and recognition accuracy of the multimodal and physiological approaches, many studies still face problems such as class imbalance, data diversity limitations, lack of real-world testing, and single-modal conditions. These challenges necessitate a call for the development of more general, ethically aware, and context-adaptive multimodal engagement detection systems.

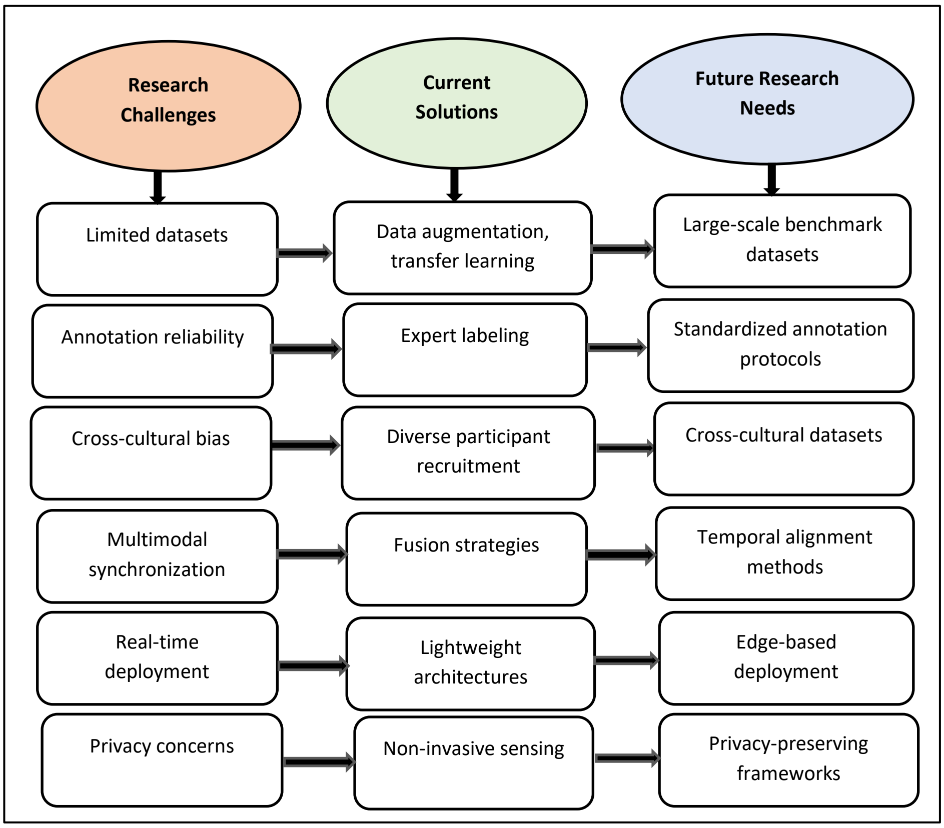

Figure 9. Challenge–solution matrix highlighting key challenges, existing solutions, and future research needs in affective student engagement detection

6. Conclusion

By summarizing existing literature, it is evident that the use of affective computing, which recognizes academic emotions like frustration, boredom, and happiness, offers a much more detailed idea of student engagement than behavioral tracking. The results show that although facial expression is still the most typical form of representation of these emotional states, the domain is rapidly shifting to multimodal deep learning models to address the nuance of internal affective changes, especially Convolutional Neural Networks (CNNs), being the dominant detection approach. Finally, changing the pedagogical emphasis from "what the student is doing" to "how the student is feeling" enables the development of more responsive and empathetic human-in-the-loop systems. These developments can potentially make both the digital and the physical classroom inclusive environments that actively support the affective process and long-term academic success. Despite its valuable findings, it has certain methodological limitations. Although this survey contains relevant literature, which is included from the Scopus database and published in the English language, there are still certain relevant studies in other databases. In addition, the reviewed studies varied in their datasets, experimental designs, and evaluation procedures, making direct comparison across approaches difficult. As a result, a formal statistical aggregation or meta-analysis was not feasible. The primary emphasis of this survey was on technological and methodological points, and less emphasis was placed on the large socio-economic issues that can affect student engagement. These problems could be solved in future reviews by increasing the search coverage to multiple language sources and by using multiple database snowballing methods to better capture the breadth of evidence of findings. Future research should be aimed at combining multimodal models with explainable AI to enhance the analysis of affective state, transparency, and pedagogical trust. Moreover, standardized real-world datasets, real-time privacy-preserving systems and longitudinal research are necessary to provide fairness, generalizability, and evaluation of the impact over time.

References

[1] Altuwairqi, Khawlah, Salma Kammoun Jarraya, Arwa Allinjawi, and Mohamed Hammami. "Student Behavior Analysis to Measure Engagement Levels in Online Learning Environments." Signal, image and video processing 15, no. 7 (2021): 1387-1395.

[2] Vanneste, Pieter, José Oramas, Thomas Verelst, Tinne Tuytelaars, Annelies Raes, Fien Depaepe, and Wim Van den Noortgate. "Computer Vision and Human Behaviour, Emotion and Cognition Detection: A Use Case on Student Engagement." Mathematics 9, no. 3 (2021): 287.

[3] Xu, Yaping, Yanyan Li, Yunshan Chen, Haogang Bao, and Yaqian Zheng. "Spontaneous Visual Database for Detecting Learning-Centered Emotions During Online Learning." Image and Vision Computing 136 (2023): 104739.

[4] Su, Rui, Lang He, and Mengnan Luo. "Leveraging Part-And-Sensitive Attention Network and Transformer for Learner Engagement Detection." Alexandria Engineering Journal 107 (2024): 198-204.

[5] Karimah, Shofiyati Nur, and Shinobu Hasegawa. "Automatic Engagement Estimation in Smart Education/Learning Settings: A Systematic Review of Engagement Definitions, Datasets, and Methods." Smart Learning Environments 9, no. 1 (2022): 31.

[6] Lek, Jeniffer Xin-Ying, and Jason Teo. "Academic Emotion Classification Using FER: A Systematic Review." Human Behavior and Emerging Technologies 2023, no. 1 (2023): 9790005.

[7] Fang, Bei, Xian Li, Guangxin Han, and Juhou He. "Facial Expression Recognition in Educational Research from the Perspective of Machine Learning: A Systematic Review." IEEE Access 11 (2023): 112060-112074.

[8] Qarbal, Ikram, Nawal Sael, and Sara Ouahabi. "Student's Engagement Detection Based on Computer Vision: A Systematic Literature Review." IEEE Access (2025). 140519–140545.

[9] Chong, Ke Ting, Noraini Ibrahim, Sharin Hazlin Huspi, Wan Mohd Nasir Wan Kadir, and Mohd Adham Isa. "A Systematic Review of Machine Learning Techniques for Predicting Student Engagement in Higher Education Online Learning." Journal of Information Technology Education: Research 24 (2025): 005.

[10] Bao, Jindi, Xiaomei Tao, and Yinghui Zhou. "An Emotion Recognition Method Based on Eye Movement and Audiovisual Features in MOOC Learning Environment." IEEE Transactions on Computational Social Systems 11, no. 1 (2022): 171-183.

[11] Singh, Monisha, Ximi Hoque, Donghuo Zeng, Yanan Wang, Kazushi Ikeda, and Abhinav Dhall. "Do I Have Your Attention: A Large Scale Engagement Prediction Dataset and Baselines." In Proceedings of the 25th International Conference on Multimodal Interaction, 2023, 174-182.

[12] Sassi, Ayoub, Wael Jaafar, Safa Cherif, Jihene Ben Abderrazak, and Halim Yanikomeroglu. "Video Traffic Analysis for Real-Time Emotion Recognition and Visualization in Online Learning." IEEE Access 11 (2023): 99376-99386.

[13] Dey, Arnab, Anubhav Anand, Subhajit Samanta, Bijay Kumar Sah, and Samit Biswas. "Attention-based AdaptSepCX Network for Effective Student Action Recognition in Online Learning." Procedia Computer Science 233 (2024): 164-174.

[14] Shomoye, Michael, and Richard Zhao. "Automated Emotion Recognition of Students in Virtual Reality Classrooms." Computers & Education: X Reality 5 (2024): 100082.

[15] Gupta, Swadha, Parteek Kumar, and Rajkumar Tekchandani. "A Multimodal Facial Cues Based Engagement Detection System in E-Learning Context Using Deep Learning Approach." Multimedia Tools and Applications 82, no. 18 (2023): 28589-28615.

[16] Moise, Gabriela, Elia Georgiana Dragomir, Daniela Șchiopu, and Lidia Angelica Iancu. "Towards Integrating Automatic Emotion Recognition in Education: A Deep Learning Model Based on 5 EEG Channels." International Journal of Computational Intelligence Systems 17, no. 1 (2024): 230

[17] Banerjee, Deeptimaan, Prateek Gothwal, and Ashis Kumer Biswas. "ExpressNet-MoE: A Hybrid Deep Neural Network for Emotion Recognition." Machine Learning with Applications (2026): 100830.

[18] Kaur, Amanjot, Aamir Mustafa, Love Mehta, and Abhinav Dhall. "Prediction and Localization of Student Engagement in the Wild." In 2018 Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2018, 1-8.

[19] Tao, Xiaohui, Aaron Shannon-Honson, Patrick Delaney, Christopher Dann, Haoran Xie, Yan Li, and Shirley O'Neill. "Towards an Understanding of the Engagement and Emotional Behaviour of MOOC Students Using Sentiment and Semantic Features." Computers and Education: Artificial Intelligence 4 (2023): 100116.

[20] Hossen, Muhammad Kamal, and Mohammad Shorif Uddin. "Attention Monitoring of Students During Online Classes Using XGBoost Classifier." Computers and Education: Artificial Intelligence 5 (2023): 100191.

[21] Qin, Xingguo, Ya Zhou, and Jun Li. "Multi-Modal Emotion Recognition for Online Education Using Emoji Prompts." Applied Sciences 14, no. 12 (2024): 5146.

[22] Yang, Pei, Ying Liu, Yuyan Luo, Zhong Wang, and Xiaoli Cai. "Text Mining and Multi-Attribute Decision-Making-Based Course Improvement in Massive Open Online Courses." Applied Sciences 14, no. 9 (2024): 3654.

[23] Wang, Jinfeng, Xiaomei Chen, and Zicong Zhang. "A Lightweight Teaching Assessment Framework Using Facial Expression Recognition for Online Courses." Applied Sciences 15, no. 23 (2025): 12461.

[24] Mao, Jun, Zhe Qian, and Terry Lucas. "Sentiment Analysis of Animated Online Education Texts Using Long Short-Term Memory Networks in the Context of the Internet of Things." IEEE Access 11 (2023): 109121-109130.

[25] Wu, Yan. "Research on Prediction Algorithm of College Students' Academic Performance Based on Bert-GCN Multi-Modal Data Fusion." Systems and Soft Computing 7 (2025): 200327.

[26] Shingjergji, Krist, Deniz İren, Corrie Urlings, and Roland Klemke. "Affective Computing in Online Higher Education: A Systematic Literature Review." Computers and Education: Artificial Intelligence 10 (2025): 100499.

***** Articles referred in primary study and supportive study enclosed in annexure.

Appendix A: Articles used in primary study

1. Pang, Jordan Min Han, Tee Connie, and Michael Kah Ong Goh. "Recognition of Academic Emotions in Online Classes." In 2021 International Conference on Computer and Information Technology (ICOICT), 445–450, 2021.

2. Bahri, Nisserine El, Zakaria Itahriouan, and Mohammed Ouazzani Jamil. "Emotion-Aware Education Through Affective Computing and Learning Analytics: Insights from a Moroccan University Case Study." Digital 5, no. 3 (2025): 45.

3. Thomas, Chinchu, and Dinesh Babu Jayagopi. "Predicting Student Engagement in Classrooms Using Facial Behavioral Cues." In Proceedings of the 19th ACM International Conference on Multimodal Interaction, 33–40, 2017.

4. Grafsgaard, Joseph F. "Multimodal Analysis and Modeling of Nonverbal Behaviors During Tutoring." In Proceedings of the International Conference on Multimodal Interaction, 404–408, 2014.

5. Liu, Shengxi, Xiaomei Tao, and Qiong Gui. "Research on Emotional State in Online Learning by Eye Tracking Technology." In Proceedings of the 2019 International Conference on Computer Science and Education Technology, 471–477, 2019.

6. Aly, Mohammed. "Revolutionizing Online Education: Advanced Facial Expression Recognition for Real-Time Student Progress Tracking Via Deep Learning Model." Multimedia Tools and Applications 84, no. 13 (2024): 12575–12614.

7. Wang, Shunye. "Online Learning Behavior Analysis Based on Image Emotion Recognition." Traitement du Signal 38, no. 3 (2021): 865–873.

8. Çimtay, Yücel, Erhan Ekmekcioğlu, and Seyma Caglar-Ozhan. "Cross-Subject Multimodal Emotion Recognition Based on Hybrid Fusion." IEEE Access 8 (2020): 168865–168878.

9. Sharma, Archana, and Vibhakar Mansotra. "Multimodal Decision-Level Group Sentiment Prediction of Students in Classrooms." International Journal of Innovative Technology and Exploring Engineering 8, no. 12 (2019): 4902–4909.

10. Hasnine, Mohammad Nehal, Huyen T. T. Bui, Thuy Thi Thu Tran, Ho Tran Nguyen, Gökhan Akçapınar, and Hiroshi Ueda. "Students' Emotion Extraction and Visualization for Engagement Detection in Online Learning." Procedia Computer Science 192 (2021): 3423–3431.

11. Bosch, Nigel, Sidney K. D'Mello, Ryan S. Baker, Jaclyn Ocumpaugh, Valerie J. Shute, Matthew Ventura, Lubin Wang, and Weinan Zhao. "Detecting Student Emotions in Computer-Enabled Classrooms." In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI), 4125–4129, 2016.

12. Zaharuddin, Yu Chen, and Gang Yao. "Enhancing Student Engagement with AI-driven Personalized Learning Systems." International Transactions on Education Technology (ITEE) 3, no. 1 (2024): 1–8.

13. Du, Chenguang, and Wei Wang. "Leveraging Cognitive Computing for Advanced Behavioral and Emotional Data Insights." International Journal of Cognitive Computing in Engineering 6 (2024): 183–190.

14. Saneiro, Mar, Olga C. Santos, Sergio Salmerón-Majadas, and Jesús G. Boticario. "Towards Emotion Detection in Educational Scenarios from Facial Expressions and Body Movements Through Multimodal Approaches." The Scientific World Journal 2014 (2014): 1–14.

15. Jagadeesh, M., and B. Baranidharan. "Facial Expression Recognition of Online Learners from Real-Time Videos Using a Novel Deep Learning Model." Multimedia Systems 28, no. 6 (2022): 2285–2305.

16. Hu, Xuhui, and Jian Gao. "Facial Expression Recognition Reveals Students' Engagement in Online Class: Correlations with Six Engagement Measurements." PLoS ONE 20, no. 10 (2025).

17. Abdellaoui, Benyoussef, Ahmed Remaida, Zineb Sabri, Mohammed Abdellaoui, Abderrahim El Hafidy, Younès El Bouzekri El Idrissi, and Aniss Moumen. "Analyzing Emotions in Online Classes: Unveiling Insights Through Topic Modeling, Statistical Analysis, and Random Walk Techniques." International Journal of Cognitive Computing in Engineering 5 (2024): 221–236.

18. Lou, Y. F., and Fan Li. "Design of an Online Education Student Learning Status Evaluation Model Based on Dual-Improved Neural Networks." Intelligent Systems with Applications 22 (2024): 200370.

19. Fwa, Hua Leong. "Fine-Grained Detection of Academic Emotions with Spatial Temporal Graph Attention Networks Using Facial Landmarks." In Proceedings of the International Conference on Computer Vision Theory and Applications, 27–34, 2022.

20. Li, Mingyong, Mingyuan Ge, Honggang Zhao, and Ziye An. "Modeling and Analysis of Learners' Emotions and Behaviors Based on Online Forum Texts." Computational Intelligence and Neuroscience 2022 (2022): 1–14.

21. Aly, Mohammed, and Nouf Saeed Alotaibi. "A Comprehensive Deep Learning Framework for Real Time Emotion Detection in Online Learning Using Hybrid Models." Scientific Reports 15, no. 1 (2025): 42012.

22. Sukumaran, Ajitha, Arun Manoharan, and Joewono Widjaja. "Student Engagement Recognition: Comprehensive Analysis Through EEG and Verification by Image Traits Using Deep Learning Techniques." IEEE Access (2025).

23. Sailer, Michael, Manuel Ninaus, Stefan E. Huber, Elisabeth Bauer, and Samuel Greiff. "The End is the Beginning is the End: The Closed-Loop Learning Analytics Framework." Computers in Human Behavior 158 (2024): 108305.

24. Chiaro, Diletta, Daniela Annuziata, Stefano Izzo, and Francesco Piccialli. "Unveiling Engagement in Virtual Classrooms: A Multimodal Analysis." In 2023 IEEE International Conference on Big Data (BigData), 4761–4769, 2023.

25. Ashwin, T. S., and Ram Mohana Reddy Guddeti. "Unobtrusive Behavioral Analysis of Students in Classroom Environment Using Non-Verbal Cues." IEEE Access 7 (2019): 150693–150709.

26. Sabuncuoğlu, Alpay, and T. Metin Sezgin. "Developing a Multimodal Classroom Engagement Analysis Dashboard for Higher-Education Students." In Proceedings of the ACM Conference on Learning at Scale, 2023.

27. Sharma, Kshitij, Zacharoula Papamitsiou, Jennifer K. Olsen, and Michail N. Giannakos. "Predicting Learners' Effortful Behaviour in Adaptive Assessment Using Multimodal Data." In Proceedings of the Tenth International Conference on Learning Analytics and Knowledge, 480–489, 2020.

28. Sümer, Ömer, Patricia Goldberg, Sidney K. D'Mello, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. "Multimodal Engagement Analysis from Facial Videos in the Classroom." IEEE Transactions on Affective Computing 14, no. 2 (2021): 1012–1027.

29. Vayadande, Kuldeep, Yogesh Bodhe, Amol A. Bhosle, Gitanjali Bhimrao Yadav, Ajit R. Patil, Jyoti Jayesh Chavhan, and Preeti Bailke. "Enhancing Facial Expression Recognition in Education with Hybrid Attention-Driven Feature Clustering." Journal of Engineering Education Transformations 39, no. 2 (2025): 200–213.

30. Salloum, Said A., Khaled Mohammad Alomari, Aseel M. Alfaisal, Rose A. Aljanada, and Azza Basiouni. "Emotion Recognition for Enhanced Learning: Using AI to Detect Students' Emotions and Adjust Teaching Methods." Smart Learning Environments 12, no. 1 (2025).

31. Pordoy, Jamie, Haleem Farman, Nevena Dicheva, Aamir Anwar, Moustafa M. Nasralla, Nasrullah Khilji, and Ikram Ur Rehman. "Multi-Frame Transfer Learning Framework for Facial Emotion Recognition in E-Learning Contexts." IEEE Access 12 (2024): 151360–151381.

32. Maddu, Rama Bhadra Rao, and S. Murugappan. "Online Learners' Engagement Detection via Facial Emotion Recognition in Online Learning Context Using Hybrid Classification Model." Social Network Analysis and Mining 14, no. 1 (2024).

33. Selim, Tasneem, Islam Elkabani, and Mohamed A. Abdou. "Students Engagement Level Detection in Online E-Learning Using Hybrid EfficientNetB7 Together with TCN, LSTM, and Bi-LSTM." IEEE Access 10 (2022): 99573–99583.

34. Giraldi, Luca, Marta Giovannetti, and Elena Cedrola. "Unveiling Emotional Reaction and Satisfaction in E-Learning with Face Tracking." Educational Research and Evaluation 30 (2025): 577–609.

35. Cárdenas-López, Héctor Manuel, Ramón Zataraín Cabada, María Lucía Barrón Estrada, and Hugo Mitre-Hernandez. "Semantic Fusion of Facial Expressions and Textual Opinions from Different Datasets for Learning-Centered Emotion Recognition." Soft Computing 27, no. 22 (2023): 17357–17367.

36. Moon, Jewoong, Fengfeng Ke, Zlatko Sokolikj, and Shayok Chakraborty. "Applying Multimodal Data Fusion to Track Autistic Adolescents' Representational Flexibility

Development During Virtual Reality-Based Training." Computers & Education X Reality 4 (2024): 100063.

37. Bahreini, Kiavash, Rob Nadolski, and Wim Westera. "Data Fusion for Real-Time Multimodal Emotion Recognition Through Webcams and Microphones In E-Learning." International Journal of Human-Computer Interaction 32, no. 5 (2016): 415–430.

38. Tran, Tien-Dzung. "TTNet: A Novel Machine Learning Model for Facial Emotion Detection in Online Learning Systems." SoftwareX 27 (2024): 101787.

39. Nadaud, Edouard, Antoun Yaacoub, Bénédicte Le Grand, and Lionel Prévost. "SEPT: Uncovering Student Difficulties Through Emotional and Physical Trajectories During Online Assessments." Procedia Computer Science 270 (2025): 2888–2897.

40. Chen, Mengnan, Lun Xie, Chiqin Li, and Zhiliang Wang. "Research on Emotion Recognition for Online Learning in a Novel Computing Model." Applied Sciences 12, no. 9 (2022): 4236.

41. Alyüz, Neşe, Eda Okur, Utku Genc, Sinem Aslan, Cagri Tanriover, and Asli Arslan Esme. "An Unobtrusive and Multimodal Approach for Behavioral Engagement Detection of Students." In Proceedings of the International Conference on Multimodal Interaction, 26–32, 2017.

42. Evangeline, D., and A. Parkavi. "Facial Emotion Recognition of Online Learners Using a Hybrid Deep Learning Model." International Journal of Intelligent Engineering & Systems 17, no. 6 (2024): 735–751.

43. Boulanger, David, M. Ali Akber Dewan, Vive Kumar, and Fuhua Lin. "Lightweight and Interpretable Detection of Affective Engagement for Online Learners." In 2021 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 176–184, 2021.

44. Harb, A. H., Abdalla Gad, Maha Yaghi, Marah Alhalabi, Huma Zia, Jawad Yousaf, Adel Khelifi, Kilani Ghoudi, and Mohammed Ghazal. "Diverse Distant-Students Deep Emotion Recognition and Visualization." Computers & Electrical Engineering 111 (2023): 108963.

45. Bahreini, Kiavash, Rob Nadolski, and Wim Westera. "Towards Multimodal Emotion Recognition in E-Learning Environments." Interactive Learning Environments 24, no. 3 (2014): 590–605.

46. Jiang, Peng, Y. Q. Wang, Chu Wang, and Satori Hachisuka. "Voice-Driven Emotion Recognition: Integrating Speaker Diarization for Enhanced Analysis." In 2024 International Conference on Information and Communication Engineering Technology (ICIET), 348–352, 2024.

47. Caballé, Santi. "Towards a Multi-Modal Emotion-Awareness E-Learning System." In 2015 International Conference on Intelligent Networking and Collaborative Systems (INCoS), 280–287, 2015.

48. Chango, Wilson, Miguel Sánchez-Santillán, Rebeca Cerezo, and Cristóbal Romero. "Predicting Students' Performance Using Emotion Detection from Face-Recording Video When Interacting with an ITS." In Proceedings of the 13th International Conference on Educational Data Mining, 2020.

49. Mahendar, Maragoni, Arun Malik, and Isha Batra. "Facial Micro-Expression Modelling-Based Student Learning Rate Evaluation Using VGG–CNN Transfer Learning Model." SN Computer Science 5, no. 2 (2024).

50. Bian, Cunling, Ya Zhang, Fei Yang, Wei Bi, and Weigang Lu. "Spontaneous Facial Expression Database for Academic Emotion Inference in Online Learning." IET Computer Vision 13, no. 3 (2018): 329–337.

51. Yang, Dedong, Abeer Alsadoon, P. W. C. Prasad, Ashutosh Kumar Singh, and Amr Elchouemi. "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment." Procedia Computer Science 125 (2018): 2–10.

52. Giannakos, Michail N., Kshitij Sharma, Ilias O. Pappas, Vassilis Kostakos, and Eduardo Velloso. "Multimodal Data as a Means to Understand the Learning Experience." International Journal of Information Management 48 (2019): 108–119.

53. Kabir, Md. Rayhan, M. Ali Akber Dewan, and Fuhua Lin. "Lightweight Model for Emotion Detection from Facial Expression in Online Learning." In 2023 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 174–179, 2023.

54. Mahmood, N., S. M. Bhatti, H. Dawood, M. R. Pradhan, and H. Ahmad. "Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models." Algorithms 17, no. 10 (2024): 458.

Appendix B. Articles used in supportive study

1. Corza-Vargas, Victor Manuel, Roberto Martínez-Maldonado, Boris Escalante-Ramírez, and Jimena Olveres. "Students' Ethical, Privacy, Design, and Cultural Perspectives on Visualizing Cognitive-Affective States in Online Learning." Journal of Learning Analytics (2024): 1–17.

2. Possaghi, Isabella, Boban Vesin, Feiran Zhang, Kshitij Sharma, Cecilie Knudsen, Håkon Bjørkum, and Sofia Papavlasopoulou. "Integrating Multi-Modal Learning Analytics Dashboard in K-12 Education: Insights for Enhancing Orchestration and Teacher Decision-Making." Smart Learning Environments 12, no. 1 (2025).