I.J. Modern Education and Computer Science, **2026, 3, 190-203**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2026.03.12

Multimodal Assessment of Student Engagement by Fusing EEG, Facial Expressions, and Body Posture in an Offline Classroom

Min Song*
Department of Educational Science, Ganesha University of Education (Undiksha), Bali, Indonesia
E-mail: Songmin@nbu.edu.cn
ORCID iD: https://orcid.org/0009-0006-4601-5760
*Corresponding author

I Gusti Putu Sudiarta
Department of Mathematics, Ganesha University of Education (Undiksha), Bali, Indonesia
E-mail: gussudiarta@undiksha.ac.id
ORCID iD: https://orcid.org/0000-0001-8181-6674

Putu Kerti Nitiasih
Department of English Education, Ganesha University of Education (Undiksha), Bali, Indonesia
Email: kertinitiasih@undiksha.ac.id
ORCID iD: https://orcid.org/0000-0003-4016-0757

Putu Nanci Riastini
Department of Educational Science Program Study, Ganesha University of Education (Undiksha), Bali, Indonesia
E-mail: putunanci.riastini@undiksha.ac.id
ORCID iD: https://orcid.org/0000-0002-6727-0959

Zhang Wang
Faculty of Information Engineering, College of Science and Technology Ningbo University, Ningbo, China
E-mail: wangzhang@nbu.edu.cn
ORCID iD: https://orcid.org/0009-0002-7840-9191

Junyi Chai
Faculty of Information Engineering, College of Science and Technology Ningbo University, Ningbo, China
E-mail: chaijunyiningbo@icloud.com
ORCID iD: https://orcid.org/0009-0007-7792-0904

Received: February 2, 2026; Revised: March 3, 2026; Accepted: March 28, 2026; Published: June 8, 2026

Abstract: An accurate and comprehensive assessment of student engagement in classrooms is crucial for enabling data-driven teaching and personalized education. Current approaches primarily rely on teacher observation or student self-reports, which are often subjective, delayed, and unable to capture cognitive engagement. To address these limitations, this study proposes a Multimodal Cognitive-Attention Fusion (MCA Fusion) framework, grounded in Fredricks' three-dimensional engagement model. The framework integrates electroencephalography (EEG), facial expressions, and body posture to simultaneously quantify cognitive, emotional, and behavioral engagement. Built on a Transformer architecture, it employs self-attention to extract temporal features within each modality and introduces a cognition-guided cross-attention mechanism to dynamically integrate multimodal signals. To validate the framework, experiments were conducted with 36 undergraduate students in real classroom settings. The results demonstrate that our framework significantly outperforms all single-modality baselines, achieving an accuracy of 92% and an F1-score of 94.87%. Compared with the best single-modality model (EEG), the F1-score improves by 34.58 percentage points. Ablation studies further confirm the critical role of the cognitive modality (EEG) and the MCA Fusion mechanism, the removal of which leads to F1-score reductions of 62.58 and 56.16 percentage points, respectively. The proposed approach not only provides a theoretically informed and technically evaluated framework for engagement recognition but also

provides a methodological foundation for future closed-loop “perception–assessment–feedback” systems in intelligent learning environments.

Keywords: Student Engagement, multimodal fusion, MCA Fusion, EEG, Facial Expression Recognition, Body Posture Analysis

1. Introduction

The digital transformation of education has intensified the demand for data-driven teaching and personalized interventions. At the heart of this shift lies a central challenge: the real-time, comprehensive, and objective assessment of student learning states [1,2]. Among these states, student engagement stands out as particularly crucial, given its well-established influence on learning outcomes and educational quality [3-5]. Traditionally, engagement has been assessed through teacher observation and student self-reports. These approaches, however, are highly subjective, lack timeliness, and are difficult to adapt to large-scale dynamic classroom environments, leading to growing concerns over their reliability [6]. Consequently, the adoption of automated and objective technologies has become an inevitable trend.

A powerful lens for analyzing this challenge is Fredricks' widely applied three-dimensional engagement framework from educational psychology, which conceptualizes student engagement across cognitive, emotional, and behavioral dimensions [7]. This theoretical perspective not only provides a structured taxonomy but also immediately reveals the inherent limitations of approaches relying on a single type of measurement. Consequently, existing automated studies have made progress from different, yet often isolated, angles. Computer vision-based methods analyze facial expressions and body posture to infer emotional and behavioral engagement, offering advantages of being non-invasive and easy to deploy [8-10]. However, such methods struggle to detect deeper cognitive states, such as “pseudo-engagement” [11]. To overcome this, some researchers have turned to physiological signals like electroencephalography (EEG), often regarded in prior laboratory-based research as a “gold standard” for assessing cognitive load and attention [12], particularly when implemented with multi-channel research-grade systems [13,14]. Nevertheless, EEG also has notable limitations, including its intrusiveness, complex signal interpretation, and lack of behavioral context [15].

Multimodal Learning Analytics (MMLA) has gradually emerged as a promising research direction to address these gaps [16-19]. Yet, existing studies still reveal significant shortcomings [20]. Some focus only on integrating external behavioral modalities without delving into the cognitive dimension [9,21,22], while others incorporate physiological signals but neglect emotional aspects [23-25]. Therefore, constructing a unified framework that effectively integrates cognitive physiological signals (EEG), emotional visual cues (facial expressions), and behavioral dynamics (body posture) remains an urgent and unmet research need.

In response, this study develops a novel Transformer-based multimodal fusion model and makes three key contributions:

- Providing a comprehensive assessment framework that integrates EEG, facial expressions, and body posture, directly mapping these modalities to the three dimensions of engagement.
- Proposing a novel cognition-guided cross-attention mechanism, which uses EEG features as queries to dynamically filter and weight emotional and behavioral signals, enabling a more informed fusion process.
- Offering empirical validation in real classroom settings, demonstrating superior performance over established unimodal and multimodal baselines, thereby enhancing the ecological relevance of the approach within authentic classroom settings.

2. Methodology

The methodology elaborates on the multimodal framework for student engagement assessment, which is grounded in Fredricks’ three-dimensional engagement theory. The framework integrates electroencephalography (EEG), facial expressions, and body posture data to construct an end-to-end deep learning system. In accordance with ethical principles, our study was approved by the institutional review board of our organization. Furthermore, all data were anonymized after collection to protect participant privacy.

2.1. Overall Framework Design

A preliminary experiment was conducted in this research. The overall architecture, as illustrated in Figure 1, comprises four main stages: (1) data acquisition and preprocessing, (2) single-modality feature representation, (3) multimodal data fusion, and (4) engagement classification. The central idea is to build a three-dimensional representation space that comprehensively reflects students’ classroom engagement by leveraging the collaborative

perception and dynamic fusion of EEG and visual behavioral data. The visual behavioral data are processed through two parallel streams: Facial Expression Recognition (FER) and Body Posture Recognition (BPR).

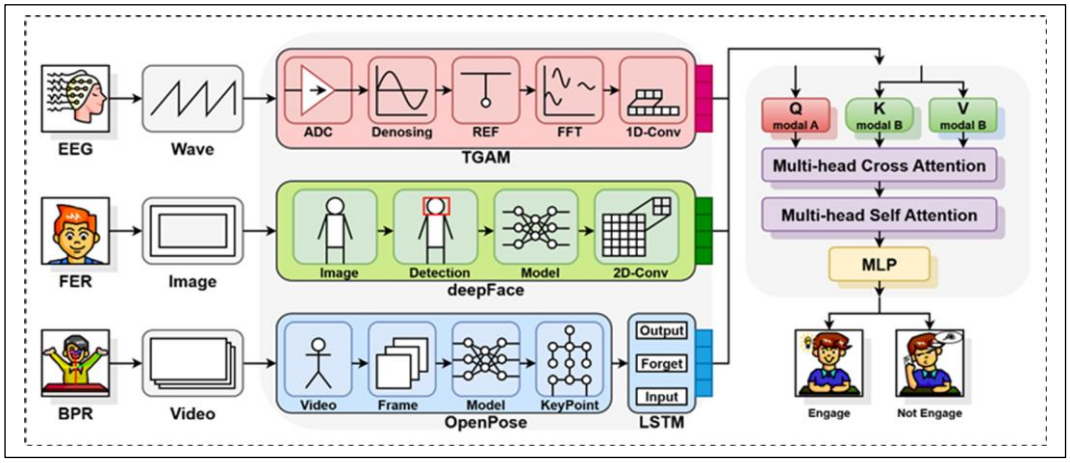

Fig. 1. Framework of the multimodal engagement assessment system.

2.2. Data Collection and Preprocessing

2.2.1. Experimental Design and Data Collection

This study was carried out within a smart classroom located at a Chinese university. Visual behavioral data were collected non-intrusively via a front-facing HD camera permanently installed above the blackboard, while cognitive data were obtained through a wearable EEG headset. All other intelligent functions were disabled to preserve ecological validity during the regular class sessions [26].

- Participants: Thirty-six first-year undergraduate students majoring in Software Engineering participated in the study (mean age = 18.9 ± 0.7 years; 19 males, 17 females). All participants reported normal or corrected vision, without any neurological or psychological conditions, and each gave written informed consent before participation.

  All participants were first-year undergraduate students majoring in Software Engineering from a single university. Although the data were collected in authentic classroom settings, the homogeneity of the sample may limit the generalizability of the findings across different disciplines and institutions. Future studies will extend validation to more diverse and multi-institutional populations.
- Experimental Task: Participants attended their regular compulsory Python programming course. Data were recorded during six standard instructional sessions (40 minutes each), randomly selected from the curriculum. The pedagogical design of these sessions—which included lectures, programming practice, and group discussions—was intentionally varied to naturally elicit a wide spectrum of student engagement states.
- Data Acquisition and Annotation Protocol: The specifications for multimodal data collection and the engagement labeling scheme are detailed in Table 1.

Table 1. Data Acquisition and Annotation Protocol.

Category / Component	Device/Platform	Key Parameters / Description
EEG	TGAM dry electrode module	Single channel (Fp1); 512 Hz sampling rate; 0.5–50 Hz bandwidth
Facial Expression	HD camera + DeepFace	1920×1080 resolution; 30 fps; real-time valence analysis
Body Posture	Same camera + OpenPose	2D skeletal keypoint detection (18 keypoints)
Annotation (Ground Truth)	Label Studio 1.18	Manual annotation; 10-second window; based on the three-dimensional engagement framework

The single-channel Fp1 dry-electrode configuration was adopted to minimize classroom disruption and ensure deployment feasibility in authentic instructional environments. Although this setup does not provide high-resolution spatial brain mapping comparable to multi-channel research-grade EEG systems, it is sufficient for capturing coarse-grained frontal attentional fluctuations relevant to engagement modeling. The design prioritizes ecological consistency and practical applicability over laboratory-level spatial precision.

2.2.2. Multimodal Data Preprocessing and Feature Engineering

Following Fredricks' three-dimensional framework, cognitive, emotional, and behavioral features were extracted. All features were computed within a 10-second sliding window (stride = 5 seconds) to ensure temporal alignment and cross-modal consistency [27]. The process included signal preprocessing, feature encoding, and vector representation for each modality.

i. Cognitive Engagement via EEG

Cognitive engagement, reflecting the level of attention and mental resource allocation, was assessed through EEG spectral analysis. In line with established methodologies, this study employed the ratio of beta (13–30 Hz) to theta (4–8 Hz) band power as an indicator of attention and cognitive load [28]. The detailed procedure is as follows:

a. Filtering:

The EEG signals were first detrended and mean-centered, followed by a 50 Hz notch filter to suppress power-line interference, and then subjected to an infinite impulse response (IIR) band-pass filter with a range of 0.5–50 Hz to eliminate low-frequency drift and high-frequency disturbances while retaining physiologically relevant components.

To further mitigate ocular and muscle artifacts commonly present in frontal single-channel recordings, discrete wavelet transform (DWT)-based denoising with adaptive soft-thresholding was performed on overlapping signal windows. In addition, a signal-quality mask was applied to exclude segments with abnormal amplitude or abrupt fluctuations before spectral feature computation.

b. Frequency Decomposition:

Perform a Fast Fourier Transform (FFT) on the preprocessed EEG signals to calculate the relative power spectral density (PSD) across four standard frequency bands: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), and β (13–30 Hz).

c. Convolutional Encoding:

The extracted band-power features were processed using a one-dimensional convolutional neural network (1D-CNN) to learn short-term temporal dependencies along the time sequence [29]. The operation for the *i*-th convolutional channel is given by:

$$y^{(i)}[t]=f\left(\sum_{k=0}^{K-1} w^{(i)}[k]\cdot x[t-k]+b^{(i)}\right) \tag{1}$$

where $x[t]$ is the input sequence, $w^{(i)}[k]$represents the convolution kernel weights, $b^{(i)}$ is the bias term, and $f(\cdot)$ denotes a nonlinear activation function.

d. Feature Representation:

All convolutional channel outputs were integrated through global average pooling (GAP) and concatenated to construct the final cognitive feature vector:

$$\mathbf{h}_{eeg}=Concat\left(GAP\left(\mathbf{y}^{(1)}\right),GAP\left(\mathbf{y}^{(2)}\right),\ldots,GAP\left(\mathbf{y}^{(n)}\right)\right) \tag{2}$$

The resulting $\mathbf{h}_{eeg}$ serves as a high-level representation of students' cognitive engagement for subsequent multimodal fusion.

ii. Emotional Engagement via Facial Expression

Emotional engagement, reflecting students' emotional states, was operationalized as continuous emotional valence. The detailed procedure is as follows:

a. Face Preprocessing:

Facial regions were detected and aligned using YOLOv11n and MTCNN.

b. Emotion Feature Extraction:

2D-CNN was utilized to extract salient features from the facial images for emotion recognition. This network learns hierarchical representations of expressions by sequentially applying multiple layers of convolution and pooling operations [30]. Let the k-th local receptive field of the input image be $x[k]$, the convolution kernel weights be $w^{(i)}[k]$, the bias term be $b^{(i)}$, and the nonlinear activation function be $f(\cdot)$. The activation value for the *i*-th filter at a specific location is given by:

$$z^{(i)} = f\left(\sum_{k=0}^{K-1} w^{(i)}[k] \cdot x[k] + b^{(i)}\right) \tag{3}$$

c. Feature Representation:

The CNN outputs were aggregated via GAP to obtain a fixed-dimensional emotional feature vector:

$$z^{emotion} = Concat\left(GAP\left(z^{(1)}\right), GAP\left(z^{(2)}\right), \ldots, GAP\left(z^{(n)}\right)\right) \tag{4}$$

The resulting feature vector was subsequently processed by a fully connected regression layer to predict a continuous valence score $v \in [-1,1]$. In this scale, higher scores correspond to positive emotional states (e.g., enjoyment, satisfaction), whereas lower scores represent negative emotional responses (e.g., confusion, frustration). This score was treated as the representation of students' emotional engagement in classroom contexts and subsequently fed into the multimodal fusion model for integrated analysis.

iii. Behavioral Engagement via Body Posture

Behavioral engagement reflects students' outward participation in classroom activities and is often manifested through their posture stability, attention direction, and movement amplitude. In this study, students' body movements were quantitatively analyzed using OpenPose to extract skeletal keypoints and construct a temporal representation of postural dynamics. The main procedure involved three stages:

a. Keypoint Extraction:

Each video frame was processed using OpenPose to identify the two-dimensional coordinates of major skeletal keypoints—such as the head, shoulders, and hands—denoted as $p_t \in R^d$. By arranging these coordinates over time, a temporal posture sequence $\{p_1, p_2, \ldots, p_T\}$was generated, providing a continuous description of students' movement trajectories during classroom learning.

b. Temporal Feature Modeling:

Inputting pose sequences into a Long Short-Term Memory (LSTM) network to capture dynamic pose patterns through its recurrent structure [31]. The recurrent process can be expressed as:

$$s_t, c_t = LSTM\left(p_t, s_{t-1}, c_{t-1}\right) \tag{5}$$

where s_t is the hidden state and c_t is the memory cell state at time t.

c. Feature Representation:

The sequence of hidden states was summarized using GAP to produce a compact behavioral feature vector:

$$\boldsymbol{h}_{pose} = GAP(s_1, s_2, \ldots, s_T) \tag{6}$$

This vector encodes the temporal and spatial characteristics of students' posture changes, serving as the representation of behavioral engagement. It was subsequently integrated with EEG and facial expression features in the multimodal fusion stage for comprehensive engagement assessment.

2.2.3. Engagement Label Definition and Annotation

This study employed a manual annotation methodology to establish robust engagement labels. The coding framework was operationalized based on Fredricks' theoretical model of engagement, which was adaptively refined to accurately capture student participation in the observed learning context, thereby ensuring both the construct validity and coding consistency of the labels.

i. Label Definition

The student engagement status for each data segment was categorized into three classes based on Fredricks' three-dimensional engagement framework. The definitions for each category are provided in Table 2. For detailed operational criteria and behavioral indicators, annotators referred to a comprehensive annotation manual to ensure consistent and objective labeling.

Engagement labeling was grounded in task-aligned behavioral consistency rather than isolated visual features. Annotators evaluated whether students' behaviors were continuously aligned with the ongoing instructional task, including sustained task focus, synchronization with instructional events, responsiveness to learning content, and absence of non-learning activities. Observable cues such as gaze direction, posture, or facial expression served only as evidence carriers to support holistic behavioral interpretation and were not treated as independent labeling rules.

Table 2. Student engagement label definitions.

Category	Definition
Engaged (Task-related)	The student remains focused on the current learning task and exhibits positive emotions or sustained classroom behaviors, such as attentive gaze, nodding responses, active gestures, or stable posture.
Not engaged (Task-unrelated)	The student demonstrates behaviors or emotional states unrelated to the learning task, such as daydreaming, chatting with peers, using a mobile phone, or showing obvious boredom or resistance.
Unidentifiable	Due to occlusion, missing frames, or highly ambiguous postures, annotators cannot reliably determine the learning state.

For experimental validity and model reproducibility, only Engaged and Not engaged samples were retained in the classification task. Segments labeled as Unidentifiable were excluded from both training and testing, as they lacked discernible features for reliable modeling. These segments were excluded to prevent label noise rather than to simplify the classification boundary.

ii. Annotation Procedure

To ensure the consistency and reliability of the ground-truth labels, a rigorous multi-step annotation procedure was implemented.

a. Preparation and Training:

A standardized annotation manual was developed, detailing label definitions, operational criteria, and visual examples. All annotators underwent unified training and were required to pass a consistency test before formal annotation.

b. Annotation Process:

During formal annotation, a 10-second time window with a 5-second stride was used. For each window, annotators made a comprehensive assessment by integrating cues from students' gaze, facial expressions, body posture, and the classroom context. Importantly, annotators did not have access to model-derived features or EEG signals during the labeling process.

c. Reliability Assessment:

To quantify inter-rater reliability, each data segment was independently annotated by two raters. The agreement was assessed using Cohen's Kappa statistic, yielding an average value of 0.88, which indicates almost perfect agreement [32].

d. Adjudication:

In cases of disagreement, a third expert rater provided a final adjudication. The resolved label was used as the definitive ground truth for model training and evaluation.

2.3. Model Architecture

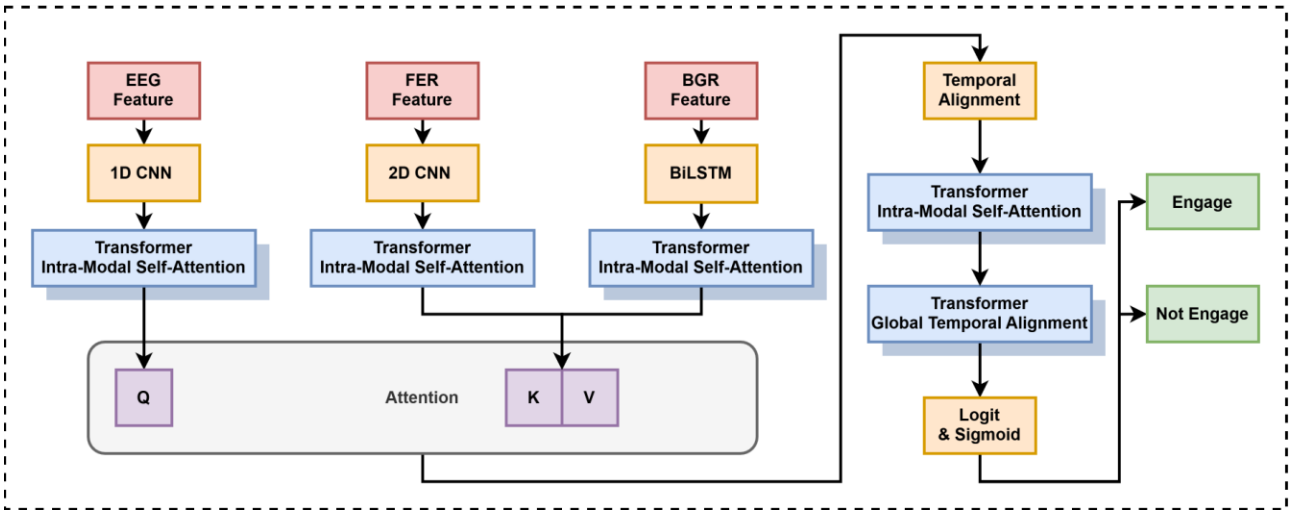

Fig. 2. Overall architecture of the Multimodal Cognitive-Attention Fusion (MCA Fusion) framework.

This study adopts a Multimodal Cognitive-Attention Fusion (MCA Fusion) framework. The framework follows a late fusion paradigm, where high-level features are first extracted from each modality and then integrated through a dedicated cognition-guided fusion module. This approach preserves the unique spatiotemporal characteristics of each modality while enabling dynamic interaction. The overall architecture, illustrated in Figure 2, consists of three core components: modality-specific encoders, the MCA Fusion module, and an engagement classification module.

2.3.1. Modality-Specific Encoders

Each modality input is processed by an independent sub-network to extract high-level semantic features:

i. EEG Branch

This branch models the preprocessed EEG signals using a 1D-CNN to capture local temporal patterns and spatial correlations across different frequency bands. The network outputs the cognitive feature vector $\mathbf{h}_{\text{eeg}}$.

ii. Facial Expression Branch

This branch takes aligned facial images as input. 2D-CNN is applied to extract deep, multi-level emotional representations. The network outputs the emotional feature vector $\mathbf{z}^{emotion}$ along with the corresponding continuous valence score v.

iii. Body Posture Branch

This branch analyzes the sequential dynamics of body posture by processing skeletal keypoints from successive video frames using a LSTM network. The network outputs the behavioral feature vector $\mathbf{h}_{\text{pose}}$. The detailed architecture of this branch is illustrated in Figure 3.

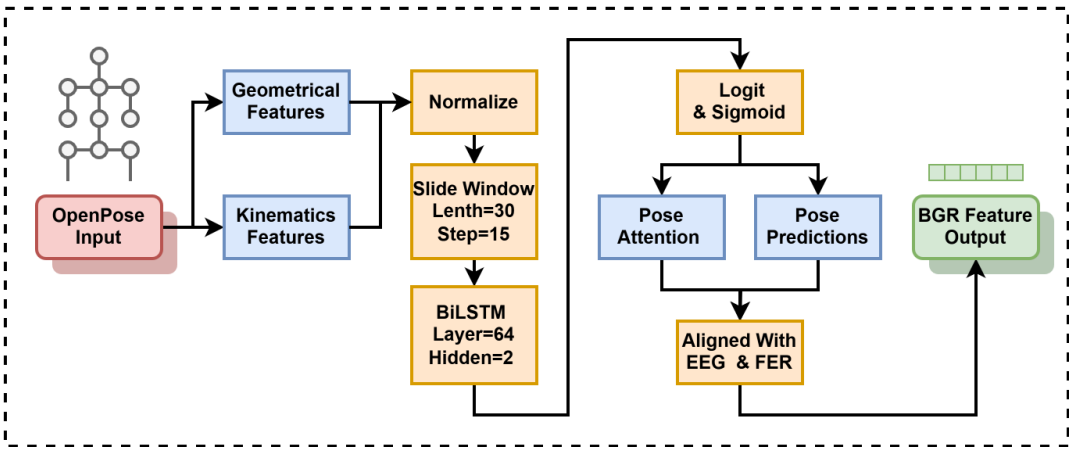

Fig. 3. Architecture of the Body Posture Branch.

2.3.2. Cross-Modal Fusion Module

This module is constructed using a Transformer-based architecture, employing self-attention and cross-attention mechanisms to enhance inter-modal interaction and collaborative representation learning [33-35]. Specifically, in this study, the EEG feature vector $\mathbf{h}_{\mathbf{eeg}}$ is designated as the query vector (Q), while the feature vectors derived from facial expressions and body posture are used as keys (K) and values (V), respectively.

This design is motivated by the central assumption that internal cognitive states are the core of engagement and should therefore serve as a guiding signal to filter and modulate the relative importance of external behavioral and emotional features. The corresponding attention mechanism is mathematically expressed as:

$$\text{Attention}(Q,K,V)=\text{Softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right) \tag{7}$$

where d_k denotes the dimensionality of the key vectors. The final output is a cross-modal fused representation modulated and enhanced by cognitive signals.

2.3.3. Engagement Classification Module (Classifier Design)

It should be clarified that the three-dimensional engagement framework serves as representational grounding for multimodal feature modeling rather than as the direct prediction label space. The cognitive, emotional, and behavioral dimensions are preserved in the internal representation learning stage, while the final binary classification functions as a decision-level objective to ensure stable model optimization under limited sample conditions.

The integrated multimodal representations were processed through a dense layer that employed a Softmax activation mechanism to generate the final engagement classification outcomes. The model was optimized using a Weighted Binary Cross-Entropy (BCE) Loss function to address the challenge of imbalanced class distributions, which is common in authentic classroom recordings [36]:

$$L=-\frac{1}{N}\sum_{i=1}^{N}\left(w_1 \cdot y_i \cdot \log \hat{p}_i + w_0 \cdot (1-y_i) \cdot \log(1-\hat{p}_i)\right) \quad (8)$$

Here, $y_i \in \{0,1\}$ denotes the ground-truth engagement state for the i-th sample(0= not engaged, 1= engaged), $\hat{p}_i$ represents the predicted probability of engagement, and w_1 and w_0 are the weights for positive and negative samples, respectively, to mitigate class imbalance.

While the MCA Fusion model introduces additional cross-attention layers, the overall architectural depth and feature dimensionality remain within a comparable scale to the baseline fusion models. This design ensures that performance differences primarily reflect differences in fusion strategy rather than arbitrary increases in model capacity.

A standard 70/15/15 split was applied to partition the dataset for training, validation, and testing. To prevent subject-level data leakage and ensure a realistic evaluation setting, the dataset was partitioned following a subject-independent protocol rather than a random window-level split. Specifically, all samples belonging to a single participant were assigned exclusively to only one subset. The training process utilized the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32 [37], coupled with a Cosine Annealing Scheduler for adaptive learning rate decay [38]. Additionally, early stopping based on the validation F1-score (patience = 15 epochs) was applied to reduce the risk of overfitting [39].

3. Results and Discussion

This section is organized into a systematic assessment of the MCA Fusion model's performance. The empirical assessment includes comparative analysis with baseline models and ablation experiments. These quantitative results offer a robust foundation for the subsequent discussion of the model's efficacy and behavioral implications.

3.1. Model Performance Comparison

An empirical validation was conducted to assess the efficacy of the MCA Fusion model by comparing it against several baseline approaches on an identical test set. The baseline models included three unimodal models (EEG, facial expression, and body posture), an early fusion model (feature concatenation + MLP), and a benchmark late fusion model (direct concatenation of modality-specific feature vectors followed by classification). The performance of all these models, as measured by Accuracy, Precision, Recall, and F1-Score, is comprehensively compared in Table 3.

To ensure the robustness of the experimental results, all models were trained and evaluated five times using different random seeds under the same subject-independent data partition protocol, and the results are reported as mean ± standard deviation.

For the early fusion model, features from EEG, facial expression, and body posture are concatenated into a single vector and fed into a three-layer MLP for classification. For the late fusion model, each modality is encoded separately, weighted via a gating network, and combined with pairwise cross-modal interaction features. A residual projection preserves modality-specific information before the final MLP produces the binary classification output.

Table 3. Classification Results of the MCA Fusion versus Baseline Models (Note: Bold values indicate the best performance in each column. Results are reported as mean ± standard deviation across five runs.).

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EEG (Unimodal)	46 ±0.84	87.23 ±1.12	46.07 ±0.95	60.29 ±0.88
Facial Expression (Unimodal)	66.63 ±0.76	21.7 ±0.91	30.54 ±1.03	25.37 ±0.94
Body Posture (Unimodal)	66.41 ±0.81	32.92 ±1.05	67.7 ±0.92	44.3 ±0.89
Early Fusion	86.7 ±0.65	85.9 ±0.72	85.0 ±0.68	85.4 ±0.71
Late Fusion	60 ±0.91	63.79 ±0.86	66.07 ±0.79	64.91 ±0.82
MCA Fusion	**92 ±0.48**	**93.67 ±0.53**	**96.1 ±0.44**	**94.87 ±0.79**

The results presented in Table 3 reveal several key observations:

- Performance of Multimodal vs. Unimodal Models: All multimodal fusion models (Early Fusion, Late Fusion, and MCA Fusion) all recorded higher F1-scores than the best-performing unimodal model (EEG, F1-score = 60.29%).
- Comparison of Fusion Strategies: The benchmark Late Fusion model (F1-score = 64.91%) yielded a lower F1-score than the Early Fusion model (F1-score = 85.40%).

- Performance of the MCA Fusion Model: The MCA Fusion model achieved the highest values across all four evaluation metrics. Its F1-score (94.87%) shows an improvement of 29.96 percentage points over the benchmark late fusion model and 34.58 percentage points over the best unimodal model.

To further analyze the fairness of the comparison, the parameter sizes of the fusion models were examined. The Early Fusion and Late Fusion models contain 13,441 and 11,664 parameters respectively, while the proposed MCA Fusion model contains 96,547 parameters. The increase in parameters mainly comes from the cross-modal attention module used in the MCA Fusion architecture.

Although the MCA Fusion model introduces additional parameters, the improvement in performance cannot be attributed solely to increased model capacity. The cross-modal attention mechanism enables the model to capture interactions among EEG, facial expression, and body posture features more effectively than simple concatenation-based fusion strategies. This allows the MCA Fusion model to better exploit complementary multimodal information.

3.2. Ablation Study

A comprehensive ablation study was conducted to deconstruct the contributions of core components within the MCA Fusion architecture. In this process, particular modules were selectively excluded or replaced with simplified alternatives, and the resulting variations in model performance were examined. All ablation experiments were repeated five times with different random seeds under the same subject-independent data partition protocol, and the mean performance is reported. The relatively small variation across runs (see F1 ± std) indicates that the observed performance gaps are not caused by random fluctuations or training instability. The corresponding results from this experimental procedure are documented in Table 4.

Table 4. Ablation Analysis on the Contributions of Model Components (Note: Δ F1 indicates the drop in F1-score compared with the full model. Results are reported as mean ± standard deviation across five runs.).

Ablation Setting	Accuracy (%)	F1-score (%)	Δ F1 (pp)
Full Model (MCA Fusion)	92 ±0.48	94.87 ±0.79	—
(a) Removing EEG modality	63.43 ±1.12	32.29 ±1.36	-62.58
(b) Removing cross-attention mechanism	81 ±0.95	38.71 ±1.18	-56.16
(c) Replacing with simple feature concatenation	64.68 ±1.07	43.14 ±1.25	-51.73

The ablation study led to the following findings:

- Effect of EEG Modality Removal: The exclusion of the EEG modality (setting a) resulted in the most severe performance drop, with the F1-score decreasing by 62.58 percentage points. This result highlights the dominant role of cognitive-state information in engagement modeling and confirms that the multimodal framework cannot rely solely on affective or behavioral cues.
- Effect of Fusion Mechanism Alteration: Both the removal of the cross-attention mechanism (setting b) and its replacement with simple feature concatenation (setting c) led to substantial performance degradation, with F1-scores falling by 56.16 and 51.73 percentage points, respectively. The cross-attention module explicitly models conditional dependencies between cognitive and affective modalities, enabling dynamic modality alignment and noise suppression. When this interaction modeling is removed, the network loses its ability to capture cross-modal semantic dependencies, leading to a non-linear degradation in discriminative representation quality. The relatively small standard deviations further confirm that the large performance gaps are structurally induced rather than caused by overfitting or instability.

3.3. Discussion

This section provides an in-depth interpretation of the experimental results, explores their theoretical and practical implications, compares them with existing research, and objectively discusses the study's limitations and future directions.

3.3.1. Interpretation of Results and Theoretical Implications

The experimental findings demonstrate that the MCA Fusion model, which integrates cognitive, emotional, and behavioral information, significantly outperforms all unimodal models and conventional multimodal baselines. This result is consistent with the multidimensional perspective of engagement proposed by Fredricks et al. [7], which posits that classroom engagement is a complex, multi-faceted construct that cannot be fully captured by a single modality. Among unimodal models, the behavioral (posture) and emotional (facial expressions) modalities performed relatively poorly, suggesting that relying solely on external observations is susceptible to "pseudo-engagement". In contrast, the

EEG unimodal model achieved the best baseline performance, underscoring the central role of the cognitive dimension in engagement assessment.

More importantly, the MCA Fusion yielded significant performance gains compared with simple feature concatenation and standard late fusion. This finding indicates that effective multimodal fusion is not a mere aggregation of information but rather a dynamic and selective decision-making process. By using EEG features as the query, the model actively retrieves emotional and behavioral signals relevant to the current cognitive state, thereby shifting from "static fusion" to "cognition-driven dynamic fusion." This introduces a new paradigm for learning analytics, in which internal physiological signals serve as anchors to interpret external behaviors.

3.3.2. Methodological Contributions and Practical Implications

The ablation study further reinforced these conclusions. Removing the EEG modality led to the largest performance drop, highlighting the indispensable role of cognitive signals in accurate assessment. Removing the cross-attention mechanism confirmed that the fusion strategy itself—rather than the mere increase in parameters—contributes to the performance improvement.

From a practical perspective, this study translates a psychological theory into a feasible technological pathway for supporting intelligent teaching systems. The proposed framework enables earlier and more accurate identification of students' disengagement, thereby offering objective data support for personalized instructional interventions. Furthermore, the collection and validation of data in real classroom environments significantly enhance ecological validity, demonstrating strong potential for real-world application.

3.3.3. Comparison with Existing Studies

This study shares a common goal with recent influential works [23,24,40,41], which aim to enhance the accuracy of student engagement assessment through multimodal fusion. However, it makes differentiated contributions in three key aspects:

- **Theoretical integrity:** Guided by Fredricks' model of student engagement, the study simultaneously collected cognitive, emotional, and behavioral data, avoiding the limitation of missing dimensions found in some prior research.
- **Architectural advancement:** By introducing a Transformer-based cross-attention mechanism, the model achieved asymmetric and dynamic fusion across modalities, surpassing commonly used symmetric or static fusion approaches.
- **Authenticity of validation:** Unlike studies relying on controlled laboratory tasks, the data in this study were collected from continuous real classroom teaching, making the conclusions more reflective of authentic classroom dynamics within the investigated context.

3.3.4. Limitations and Future Work

Although this study has yielded positive results, several limitations must be acknowledged, which also illuminate productive pathways for future research.

- **Sample size and generalizability:** All participants came from the same major and grade, with a limited sample size (N = 36). Future research should validate the model's generalizability across larger and more diverse populations.
- **Data quality and equipment trade-offs:** Although portable dry-electrode EEG systems improved ecological validity, their signal-to-noise ratio remained lower than that of laboratory-grade wet electrodes. Future studies could explore advanced signal processing methods (e.g., deep denoising) or high-precision devices.
- **Challenges in real-time application:** The current system relies on offline analysis. To enable real-time feedback, further research is required on model lightweighting and distributed computing to meet the low-latency requirements of classroom use.
- **Granularity of state analysis:** The present study focused on binary classification (engaged vs. disengaged). Future work could extend to multi-level engagement (high/medium/low) or finer-grained learning states (e.g., confusion, focus, boredom) to provide more precise instructional support.

4. Conclusion

This study comprehensively investigated intelligent student engagement analysis by developing and empirically evaluating a novel multimodal fusion framework informed by Fredricks et al.'s multidimensional engagement theory. The core of this work lay in operationalizing this theoretical structure into a computational model, where EEG signals, facial expressions, and body postures were strategically selected to represent cognitive, emotional, and behavioral dimensions, respectively. The proposed MCA fusion model leveraged a cognition-guided cross-modal attention mechanism, which was empirically demonstrated to be highly effective in assessing student classroom engagement.

Experimental results confirmed that the proposed approach not only outperformed both unimodal and traditional multimodal baselines but also validated the indispensability of the cognitive dimension and the advantages of dynamic, asymmetric fusion over static fusion strategies.

The findings of this study hold significant theoretical and practical implications. Theoretically, this work demonstrates how a well-established psychological framework can inform the design of a data-driven computational system, contributing an operational perspective to multidimensional engagement research. Practically, it offers a feasible and robust technical pathway for developing next-generation learning analytics tools capable of delivering granular, objective assessments of student engagement in real classroom settings.

Looking ahead, this work opens several promising research avenues. Future studies will focus on extending the validation of this paradigm to diverse educational contexts and larger populations. The natural evolution of this technology will involve progressing from binary classification toward modeling a continuous spectrum of engagement and more nuanced learning-centered states. Furthermore, efforts will be directed at advancing the system toward practical, real-time classroom applications, with the ultimate goal of empowering educators with precise, actionable insights to enhance teaching and learning outcomes.

All the Declarations and Statements

Author Contributions Statement

Min Song – Conceptualization, Methodology, and Writing – Original Draft: Proposed the research idea, designed the overall research framework, and prepared the initial manuscript.
I Gst. Putu Sudiarta – Supervision, Review and Editing: Provided academic supervision, reviewed the manuscript, and contributed to improving the clarity and quality of the paper.
Putu Kerti Nitiasih – Methodology and Academic Guidance: Provided methodological guidance and academic suggestions during the research process.
Putu Nanci Riastini – Validation and Educational Analysis: Contributed to the validation of the research results and provided educational analysis related to the study.
Zhang Wang – Software Implementation and Experimental Design: Assisted with software implementation and the design of experimental procedures.
Junyi Chai – Data Processing and Visualization: Responsible for data preprocessing and visualization of experimental results.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

None.

Data Availability Statement

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Ethical Declarations

This study was conducted in accordance with institutional ethical standards. All participants were informed about the purpose of the study and consent was obtained prior to participation.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used ChatGPT to assist with English language editing and grammar refinement. The authors carefully reviewed and revised the generated content and take full responsibility for the final manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

MCA Fusion-Multimodal Cognitive-Attention Fusion
EEG- Electroencephalography
FER-Facial Expression Recognition
BPR-Body Posture Recognition

Appendix

None.

References

[1] E. Mangina and G. Psyrra, "Review of Learning Analytics and Educational Data Mining Applications," EDULEARN21 Proc., vol. 1, no. August, pp. 949–954, 2021, doi: 10.21125/edulearn.2021.0250.

[2] V. T. Tran and N. H. Tran, "A Review of Smart Education and Lessons Learned for An Effective Application in Binh Duong Province, Vietnam," Pegem Egit. ve Ogr. Derg., vol. 13, no. 1, pp. 234–240, 2022, doi: 10.47750/pegegog.13.01.25.

[3] A. Radloff and H. Coates, "Doing More for Learning: Enhancing Engagement and Outcomes. Australasian Survey of Student Engagement (AUSSE) Report," 2010. [Online]. Available: https://www.acer.org/files/AUSSE_Australasian-Student-Engagement-Report-ASER-2009.pdf

[4] S. G. T. Ong and G. C. L. Quek, "Enhancing teacher–student interactions and student online engagement in an online learning environment," Learn. Environ. Res., vol. 26, no. 3, pp. 681–707, 2023, doi: 10.1007/s10984-022-09447-5.

[5] H. Ross, Y. Cen, and Z. Zhou, "Assessing student engagement in China: Responding to local and global discourse on raising educational quality," Curr. Issues Comp. Educ., vol. 14, no. 1, pp. 24–37, 2011.

[6] T. L. Hofkens and E. Ruzek, "Measuring student engagement to inform effective interventions in schools," in Handbook of student engagement interventions: Working with disengaged students, S. L. Fredricks, J. A., Reschly, A. L., & Christenson, Ed., Elsevier Academic Press, 2019, pp. 309–324. doi: 10.1016/B978-0-12-813413-9.00021-8.

[7] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," Rev. Educ. Res., vol. 74, no. 1, pp. 59–109, 2004, [Online]. Available: https://doi.org/10.3102/00346543074001059

[8] S. G. Khenkar, S. K. Jarraya, A. Allinjawi, S. Alkhuraiji, N. Abuzinadah, and F. A. Kateb, "Deep Analysis of Student Body Activities to Detect Engagement State in E-Learning Sessions," Appl. Sci., vol. 13, no. 4, 2023, doi: 10.3390/app13042591.

[9] I. Qarbal, N. Sael, and S. Ouahabi, "Student ' s Engagement Detection Based on Computer Vision : A Systematic Literature Review," IEEE Access, vol. 13, no. August, pp. 140519–140545, 2025, doi: 10.1109/ACCESS.2025.3596885.

[10] Q. Liu and X. Jiang, "Classroom Behavior Recognition Using Computer Vision : A Systematic Review," Sensors, vol. 25, no. 2, p. 373, 2025, doi: https://doi.org/10.3390/s25020373.

[11] S. Arefnejad, A. Khadivi, and F. Alipour, "Challenges and Applications of Artificial Intelligence in Education: A Systematic Review," J. Knowledge-Research Stud., vol. 3, no. 4, p. 2024, 2024, doi: 10.22034/jkrs.2024.63182.1106.

[12] C. Berka, D. J. Levendowski, M. N. Lumicao, and A. Yau, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," Aviat. Space. Environ. Med., vol. 78, no. 5, pp. B231–B244, 2007.

[13] K. Yin, H. Bin Shin, D. Li, and S. W. Lee, "EEG-based Multimodal Representation Learning for Emotion Recognition," Int. Winter Conf. Brain-Computer Interface, BCI, 2025, doi: 10.1109/BCI65088.2025.10931743.

[14] A. Sukumaran and A. Manoharan, "Student Engagement Recognition: Comprehensive Analysis Through EEG and Verification by Image Traits Using Deep Learning Techniques," IEEE Access, vol. 13, no. January, pp. 11639–11662, 2025, doi: 10.1109/ACCESS.2025.3526187.

[15] L. Wei, Y. Yu, Y. Qin, and S. Zhang, "A Survey of EEG-Based Approaches to Classroom Attention Assessment in Education," Information, vol. 16, no. 10, p. 860, 2025, doi: 10.3390/info16100860.

[16] S. K. D'Mello, E. Dieterle, and A. L. Duckworth, "Advanced, analytic, automated (AAA) Measurement of Engagement During Learning," Educ. Psychol., vol. 52, no. 2, pp. 104–123, 2017, doi: 10.1080/00461520.2017.1281747.

[17] J. D. T. Guerrero-sosa, F. P. Romero, V. H. Men ández-dom nguez, J. Serrano-guerrero, A. Montoro-montarroso, and J. A. Olivas, "A Comprehensive Review of Multimodal Analysis in Education," Appl. Sci., vol. 15, no. 11, p. 5896, 2025.

[18] O. R. Yürüm, "Technology-Enhanced Multimodal Learning Analytics in Higher Education : A Systematic Literature Review," vol. 13, no. May, pp. 92057–92073, 2025, doi: 10.1109/ACCESS.2025.3572467.

[19] H. Ouhaichi, D. Spikol, and B. Vogel, "Research trends in multimodal learning analytics: A systematic mapping study," Comput. Educ. Artif. Intell., vol. 4, p. 100136, 2023, doi: https://doi.org/10.1016/j.caeai.2023.100136.

[20] N. Bergdahl, M. Bond, J. Sjöberg, M. Dougherty, and E. Oxley, "Unpacking student engagement in higher education learning analytics : a systematic review," Int. J. Educ. Technol. High. Educ., 2024, doi: 10.1186/s41239-024-00493-y.

[21] K. Mangaroska, K. Sharma, D. Gašević, and M. Giannakos, "Multimodal learning analytics to inform learning design: Lessons learned from computing education," J. Learn. Anal., vol. 7, no. 3, pp. 79–97, 2020, doi: 10.18608/JLA.2020.73.7.

[22] A. Sabuncuoglu and T. M. Sezgin, "Developing a Multimodal Classroom Engagement Analysis Dashboard for Higher-Education," Proc. ACM Human-Computer Interact., vol. 7, no. EICS, 2023, doi: 10.1145/3593240.

[23] L. Zhang, J. L. Hung, X. Du, H. Li, and Z. Hu, "Multimodal Fast–Slow Neural Network for learning engagement evaluation," Data Technol. Appl., vol. 57, no. 3, pp. 418–435, 2023, doi: 10.1108/DTA-05-2022-0199.

[24] K. Mallibhat, "Student attention detection using multimodal data fusion," in 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), IEEE, 2024, pp. 295–297. doi: 10.1109/ICALT61570.2024.00092.

[25] M. Mohammadi, E. Tajik, R. Martinez-maldonado, S. Sadiq, W. Tomaszewski, and H. Khosravi, "Artificial intelligence in

multimodal learning analytics: A systematic literature review," Comput. Educ. Artif. Intell., vol. 8, no. May, p. 100426, 2025, doi: 10.1016/j.caeai.2025.100426.
[26] S. Dikker et al., "Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom," Curr. Biol., vol. 27, no. 9, pp. 1375–1380, 2017, doi: 10.1016/j.cub.2017.04.002.
[27] P. Chejara, L. P. Prieto, M. J. Rodriguez-Triana, A. Ruiz-Calleja, and M. Khalil, "Impact of window size on the generalizability of collaboration quality estimation models developed using Multimodal Learning Analytics," ACM Int. Conf. Proceeding Ser., vol. 1, no. 1, pp. 559–565, 2023, doi: 10.1145/3576050.3576143.
[28] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, "Using Electroencephalography to Measure Cognitive Load," Educ. Psychol. Rev., vol. 22, no. 4, pp. 425–438, 2010, doi: 10.1007/s10648-010-9130-y.
[29] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," J. Neural Eng., vol. 15, no. 5, pp. 1–30, 2018, doi: 10.1088/1741-2552/aace8c.
[30] A. Abedi and S. S. Khan, "Improving state-of-the-art in Detecting Student Engagement with ResNet and TCN Hybrid Network," in Proceedings - 2021 18th Conference on Robots and Vision, CRV 2021, IEEE, 2021, pp. 151–157. doi: 10.1109/CRV52889.2021.00028.
[31] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021, doi: 10.1109/TPAMI.2019.2929257.
[32] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics," Comput. Linguist., vol. 34, no. 4, pp. 555–596, 2008, doi: 10.1162/coli.07-034-R2.
[33] G. Y. Li, J. Chen, S. I. Jang, K. Gong, and Q. Li, "SwinCross: Cross-modal Swin transformer for head-and-neck tumor segmentation in PET/CT images," Med. Phys., vol. 51, no. 3, pp. 2096–2107, 2024, doi: 10.1002/mp.16703.
[34] X. Jiang, "Deep Learning-Based Multimodal Fusion Algorithm for Assessing Online Learning Engagement," in Proceedings of the 10th International Conference on Cyber Security and Information Engineering (ICCSIE 2025), Association for Computing Machinery (ACM), 2026, pp. 88–93. doi: 10.1145/3759179.3759192.
[35] D. Dresvyanskiy, A. Karpov, and W. Minker, "A Cross-Multi-modal Fusion Approach for Enhanced Engagement Recognition BT - Speech and Computer," in Speech and Computer: 26th International Conference, SPECOM 2024, A. Karpov and V. Delić, Eds., Cham: Springer Nature Switzerland, 2024, pp. 3–17. doi: https://doi.org/10.1007/978-3-031-78014-1_1.
[36] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.
[37] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15, 2015.
[38] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc., pp. 1–16, 2017.
[39] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT review, vol. 22, no. 4. 2016.
[40] E. Fan, M. Bower, and J. Siemon, "From heartbeats to actions : Multimodal learning analytics of cognitive and behavior engagement in real classrooms," Learn. Instr., vol. 103, no. January, p. 102325, 2026, doi: 10.1016/j.learninstruc.2026.102325.
[41] C. Li, X. Weng, Y. Li, and T. Zhang, "Multimodal Learning Engagement Assessment System: An Innovative Approach to Optimizing Learning Engagement," Int. J. Human–Computer Interact., vol. 41, no. 5, pp. 3474–3490, Mar. 2025, doi: 10.1080/10447318.2024.2338616.

Authors' Profiles

Min Song was born in Zhejiang Province, China, in 1988. She received the M.S. degree in Marine Biology in 2015. After completing her master's degree, she began working in higher education and has been engaged in teaching management. She is currently a Ph.D. candidate in Educational Science at Ganesha University of Education (Undiksha), Indonesia. She works in teaching management at the Faculty of Information Engineering, College of Science and Technology, Ningbo University, China. Her research interests include educational science, teaching management, classroom assessment, and multimodal approaches to analyzing students' learning engagement.

I Gst. Putu Sudiarta received the M.Si. degree in mathematics from Universitas Gadjah Mada, Indonesia, and the doctoral degree in mathematics from Charité Universität Medizin Berlin, Germany. His major field of study is mathematics and mathematics education. He is currently a Professor with Ganesha University of Education (Undiksha), Bali, Indonesia. He has published numerous articles in national and international journals related to mathematics education and educational technology. His research interests include mathematics education, mathematical problem solving, digital learning environments, and innovative approaches to mathematics learning. Prof. Sudiarta has been actively involved in research, academic publications, and community service programs related to mathematics education and educational development.

Putu Kerti Nitiasih received the B.A. degree in English language education, the M.A. degree in linguistics, and the doctoral degree in applied linguistics from Udayana University, Denpasar, Indonesia, in 1985, 2002, and 2007, respectively. Her major field of study is applied linguistics and English language education. She is currently a Professor with the Department of English Education, Ganesha University of Education (Undiksha), Bali, Indonesia. She has published research in national and international journals and authored the book Bilingualism and Bilingual Education. Her research interests include applied linguistics, sociolinguistics, bilingual education, and English language teaching. Prof. Nitiasih has been actively involved in research, teacher training, and community service programs related to English language education.

Putu Nanci Riastini received the bachelor's degree in chemistry education and the master's degree in elementary science education from Ganesha University of Education (Undiksha), Bali, Indonesia, and the doctoral degree in educational science from Yogyakarta State University, Indonesia. Her major field of study is science education. She is a Lecturer with the Department of Educational Science, Ganesha University of Education (Undiksha), Bali, Indonesia. She teaches in the primary teacher education program and the doctoral program in educational science. Her research interests include science education, innovative teaching methods, and teacher professional development. Dr. Riastini has participated in various academic activities, including research, community service, and teacher training programs.

Zhang Wang is a Lecturer with the Faculty of Information Engineering, College of Science and Technology, Ningbo University, Ningbo, China. He is currently a Ph.D. candidate in computing and informatics at the University Malaysia Sabah, Malaysia. His major field of study is computing and informatics. He conducts research in artificial intelligence and computer vision. His research interests include AI-generated content, computer vision, and medical image enhancement, with a focus on applying these technologies in healthcare-related fields. Mr. Wang is actively involved in research on intelligent algorithms and image processing techniques for improving the quality and accuracy of medical images.

Junyi Chai is currently a student with the Faculty of Information Engineering, College of Science and Technology, Ningbo University, Ningbo, China. His major field of study is computer science and information engineering. He is engaged in research related to intelligent computing and data-driven technologies. His research interests include multimodal fusion, computer vision, deep learning, and digital twins. Mr. Chai is currently involved in academic research projects in the field of artificial intelligence and intelligent systems.

How to cite this paper

Min Song, I Gusti Putu Sudiarta, Putu Kerti Nitiasih, Putu Nanci Riastini, Zhang Wang, Junyi Chai, "Multimodal Assessment of Student Engagement by Fusing EEG, Facial Expressions, and Body Posture in an Offline Classroom", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.3, pp. 190-203, 2026. DOI:10.5815/ijmecs.2026.03.12