Classroom Engagement Evaluation Using Decision-Level Fusion of Student Behavior and Emotion Recognition

Tengyuan Zhao[1], Bemnet Wondimagegnehu Mersha[2[0000−0002−1000−8984]], Yaping Dai[3[0000−0001−8795−5333]], Kaoru Hirota[4[0000−0001−5347−6182]], Wei Dai[5], and Yumin Lin[6]

[1] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: 1516238565@qq.com
[2] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: MershaBemnetW@bit.edu.cn
[3] Beijing Institute of Technology Zhuhai, Tangjiawan, Zhuhai City, 519088, Guangdong, China email: daiyaping@bit.edu.cn
[4] School of Automation, Beijing Institute of Technology, Beijing, 100081, China email: hirota@bit.edu.cn
[5] River Security Technology Co., LTD, Shanghai 200336, China email:daiwei@gmail.com
[6] River Security Technology Co., LTD, Shanghai 200336, China email:ymlin@riversecurity.com

Abstract. Student engagement evaluation is a critical component of smart education. The traditional method of evaluating student engagement relies on self-reports and external observations, which can be time-consuming and subjective. In order to solve this problem, a decision-level fusion approach is introduced for student engagement evaluation. The proposed method combines student behavior (action) recognition and facial emotion recognition. It also uses weighted averages with Fuzzy logic to better interpret engagement levels. A custom-made dataset was created in a simulated classroom, and students' engagement levels were assessed using questionnaires. The accuracy of the proposed method was compared with behavior-only and emotion-only student engagement evaluation methods. The behavior-only evaluation achieved an accuracy of 68.75%, while the emotion-only evaluation achieved 43.75%. Notably, the proposed combined approach significantly outperformed both, with an accuracy of 87.5%.

Keywords: Facial Emotion Recognition · Behavior recognition · Decision-level fusion · Fuzzy Logic.

1 Introduction

Education plays a crucial role in any society. It can shape the future generation, making them competent and well-equipped for their country's progress. One of

the most important components of strong education is the evaluation of students' classroom engagement. Evaluating students' classroom engagement is a valuable tool that can enhance the quality of educational curriculums[1].

Student engagement is a student's emotional, cognitive, and behavioral investment in learning[2]. Research has shown that low-achieving students tend to spend a considerable amount of time during class engaging in non-academic activities, as compared to high-achieving students[1]. This makes studying student behavior in classrooms crucial, as it has significant implications for enhancing student performance and promoting effective instructional strategies.

The traditional method of assessing student engagement involves using self-reports and external behavior observations. Self-reports are cost-effective, practical, and simple to administer to a large sample, making them useful for measuring engagement and other objectives[3]. However, self-reports have limitations, as they rely on participant compliance and diligence[4]. External behavior observations are a well-established method in education research and are valuable for assessing student participation. The traditional method for observing external behavior is typically carried out by expert observers in the classroom. However, this approach is often time-consuming and subjective.

Another alternative method is to use machine learning-based methods for student engagement evaluation. There has been much research that focuses on student engagement evaluation using machine learning methods[5, 6]. However, most of the research focuses on either the student's emotional engagement evaluation or the student's behaviour evaluation. There is a need for research that combines both the emotional and behavioral components of student engagement.

In our research, a method that integrates behavioral (action recognition) and facial emotion information using a decision-level fusion algorithm to establish a comprehensive criterion for determining students' classroom engagement states is proposed (see Figure 1). Machine learning models have been developed to recognize student actions and emotions in the classroom. These models' outputs are combined using weighted average decision-level fusion. Fuzzy logic is then used to determine the level of student engagement. The proposed method is compared with other methods that use either emotion-only recognition or action-only recognition for evaluating student engagement.

The paper is structured as follows: Section two introduces the custom dataset used in the research. Section three describes the proposed method for evaluating student engagement. Section four outlines the results obtained, and Section five concludes the paper.

2 Dataset

2.1 Custom-made dataset for action recognition

A custom-made dataset was created for behavior (action) recognition. A simulated classroom environment was set up to record six specific student behaviors (actions): using cell phones, sleeping, standing, raising hands, sitting, and writing. The cameras were placed at a height of 180 cm on tripods and positioned

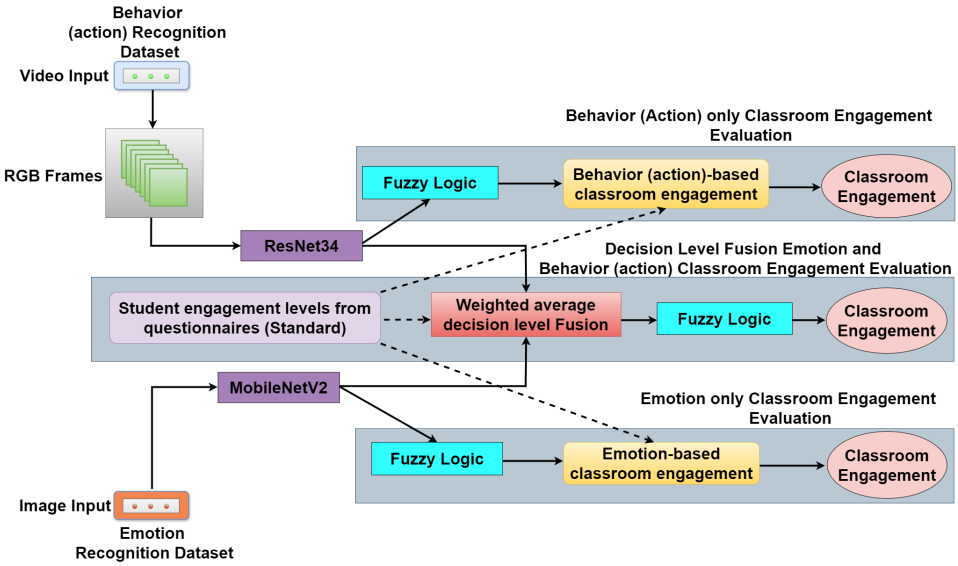

Fig. 1: Proposed weighted averages with Fuzzy logic decision level Fusion

in front and to the side of the students to ensure everyone was within the frame. One hundred videos were recorded. Each video segment was approximately 7 seconds long and recorded at a frame rate of 30 frames per second.

(a) Dataset with multiple behavior (b) Dataset of single behavior

Fig. 2: Custom Made Dataset

After recording classroom behavior using a camera, the video file was converted to RGB frames. OpenCV was used to convert the video into images. Following this, YOLOv4[7] was utilized to identify and extract individual person images from the group images. This method essentially transforms each multi-person behavioral image into six single-person images, as illustrated in Figure 2b. A total of 6,695 single-person behavioral images were processed. However, 287 images were removed due to recognition or cropping issues, resulting in 6,408 eligible images for further analysis.

2.2 Dataset for emotion recognition

Various datasets are currently available for facial emotion recognition. This study utilized the publicly available Emotion-Domestic (Asia) Expression Recognition dataset[8]. This dataset primarily comprises photos of East Asian faces gathered from the internet. It includes approximately 55,000 facial images categorized into seven emotions: anger, disgust, fear, happy, neutral, sad, and surprise.

3 Proposed Student engagement evaluation method

The proposed method fuses behavior recognition (action recognition) and facial emotion recognition using the weighted average fusion method. In this section, the student engagement state determination is based on behavioral(action) information, the student engagement state determination is based on facial emotion information, and the proposed fusion-based method is described in detail.

3.1 Student state determination based on behavioral information

The ResNet34 network[9] was used for behavior (action) recognition because of its effective feature extraction with residual learning. The training utilized a custom behavioral recognition dataset (see Section 2.1). The dataset was split into three categories: training (60%), validation (20%), and testing (20%). To prevent overfitting, early stopping was implemented in the training code. The batch size was set to ten, the learning rate to 0.001, the maximum number of iterations to 50, and the early stopping time to 5 epochs.

In order to develop the correspondence between classroom students' behavior (action) and their classroom engagement level, a fuzzy statistical experiment was conducted. Questionnaires were used to gather data about the relationship between classroom behavior (action) and engagement level. The questionaries are used to label the student classroom behavior (action) as "positive," "neutral," and "negative." Questionnaire surveys often involve respondents' subjective answers, which can be ambiguous. The Fuzzy statistical methods used are particularly effective in addressing this ambiguity.

According to the fundamental principles of fuzzy theory[10], the domain U_a ={sitting, standing, playing cell phone, sleeping, raising hand, writing}, and the fuzzy sets S_{a1}, S_{a2}, and S_{a3} are used to denote the three students' engagement states of "positive," "neutral," and "negative," respectively. The frequency of each behavior's (action) connection with the three fuzzy sets was recorded. Table 1 displays the results of calculating each behavior's affiliation to the three fuzzy sets: $S_{a1} = \{0.4, 0.97, 0, 0, 0.93, 0.7\}$, $S_{a2} = \{0.6, 0.03, 0, 0, 0, 0.07, 0.3\}$, and $S_{a3} = \{0, 0, 1, 1, 0, 0\}$. The principle of maximal affiliation was used to handle the outcomes of each behaviour's affiliation with the three fuzzy sets. The fuzzy set with the greatest association for each behavior (action) was chosen as the final determination state. Table 1 illustrates the principles for determining the condition of classroom engagement based on student behavior (action).

Table 1: Student behavioral affiliation calculations and Student state

State / Actions	Positive	Neutral	Negative	Selected State
sitting	0.4	0.6	0	Neutral
raising hand	0.97	0.03	0	Positive
playing phone	0	0	1	Negative
sleeping	0	0	1	Negative
standing	0.93	0.07	0	Positive
writing	0.7	0.3	0	Positive

3.2 Student state determination based on emotion

For classroom facial emotion recognition, the MobileNetV2[11] network was used because of its lightweight. The training utilized the Emotion-Domestic Expression Recognition dataset. To prevent overfitting, the training employed an early-stopping approach. The specific training parameters included a batch size of 64, a learning rate of 0.001, a maximum of 100 iterations, early stopping patience set at five epochs, and four parallel data loading tasks (num_workers = 4).

In order to determine the relationship between students' emotion and their classroom engagement state, fuzzy logic experiments are used. To do this, a questionnaire was used to classify student engagement based on facial emotion. A fuzzy set U_e = {Happy, Neutral, Disgusted} is created. The facial emotions are classified into three engagement states ("positive," "neutral," and "negative") using fuzzy sets $\{S_{e1}$, S_{e2}, and $S_{e3}\}$. The frequency of each emotion in the three fuzzy sets was used to calculate the fuzzy logic of the student's facial emotion. The degree of affiliation between each facial emotion and the three fuzzy sets are shown in Table 2. The fuzzy sets are defined as $S_{e1} = \{0.57, 0.37, 0\}$, $S_{e2} = \{0.37, 0.63, 0.13\}$, and $S_{e3} = \{0.06, 0, 0.87\}$. Based on the principle of maximum affiliation, the fuzzy set that exhibits the highest affiliation with each facial emotion is selected as the determined state. The parameters for evaluating the engagement state, which is based on students' facial emotion information, are detailed in Table 2.

Table 2: Student emotion affiliation calculations and Student state

State / Emotion	Positive	Neutral	Negative	Selected State
Happy	0.57	0.37	0.06	Positive
Neutral	0.37	0.63	0	Neutral
Disgust	0	0.13	0.87	Negative

3.3 The Proposed student engagement state determination combining behavioral (action) and facial emotion

In order to assess students' states based on their behavior and facial emotion, data fusion is used. The main purpose of data fusion is to improve the accuracy of identifying information and detection by combining data from one or

more sensory nodes across various platforms. There are three main types of data fusion: data layer fusion, feature layer fusion, and decision layer fusion. To determine students' engagement state, behavioral (action) and facial emotions were combined at the decision level using weighted average data fusion. The basic steps for creating the weighted data fusion algorithm are as follows:

(1) Two different data sources are used, where each data source classifier has N_i $(i = 1, 2)$ categories;
(2) The recall of the i-th data source classifier to recognize the j-th category $(j = 1, 2, \ldots, N_i)$ is r_{ij}, so the weight ω_i is assigned to the i-th data source classifier, whose matrix form is

$$\omega_i = \begin{pmatrix} r_{i1} & 0 & \cdots & 0 \\ 0 & r_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{iN} \end{pmatrix} \tag{1}$$

(3) Assuming that the probability that the i-th data source classifier recognizes the j-th classification result $(j = 1, 2, \ldots, N_i)$ is p_{ij}, the probability of each category p_i that the i-th data source classifier outputs at this time through the Softmax layer can be obtained, and the matrix form of which is

$$p_i = \begin{pmatrix} p_{i1} & p_{i2} & \cdots & p_{iN_i} \end{pmatrix}^T \tag{2}$$

(4) Fusing the ω_i and p_i obtained from the above two steps, the coefficients P_i of all the N_i class categories in the i-th data source classifier can be obtained, whose matrix form is

$$P_i = \omega_i p_i = \begin{pmatrix} p_{i1}r_{i1} & 0 & \cdots & 0 \\ 0 & p_{i2}r_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{iN}r_{iN} \end{pmatrix} \tag{3}$$

The weighted average fusion approach integrates the softmax layer of the facial emotion recognition and behavior (action) recognition machine learning models. The coefficients P_1 and P_2 indicate the level of confidence in the machine learning model's classification for each category. Using the weighted data fusion algorithm, we derive the coefficients as

$$P_1 = \begin{pmatrix} p_{11}r_{11} & p_{12}r_{12} & p_{13}r_{13} & p_{14}r_{14} & p_{15}r_{15} & p_{16}r_{16} \end{pmatrix}^T \tag{4}$$

for all categories in the behavioral (action) classifier, which includes six classifications, and

$$P_2 = \begin{pmatrix} p_{21}r_{21} & p_{22}r_{22} & p_{23}r_{23} \end{pmatrix}^T \tag{5}$$

for the categories in the facial emotion classifier.

After obtaining the coefficients P_i for all categories in the behavior (action) and facial emotion classifiers, the next step is establishing a relationship between each category in behavior (action) and facial emotion to the student's classroom engagement. Students' engagement is given a value within the range of 0 to 1. A classroom engagement value close to 1 indicates active engagement. A classroom engagement value close to 0 represents non-engagement. Table 3 shows how several engagement values are assigned to different types of student behaviors (actions).

Sitting and writing are common classroom actions linked with a neutral state, but standing and raising hands suggest active involvement and are classed as positive. In contrast, playing with cell phones and napping suggest disengagement and are categorized as negative behaviors. The specific values associated with each behavior are determined through manual assessment. Using an empirical method, the state assignment for student behaviors (actions) in the classroom is calculated, represented as $s_1 = (0, 0.9, 0.5, 0, 0.95, 0.8)$. Similarly, based on statistical results presented in Table 2, different student behaviors (actions) are assigned values corresponding to their levels of classroom engagement. Table 3 provides a summary of the results of this assignment.

Table 3: Student Behavioral and Facial Emotion Assignments

Category	**Assignments**
Behaviors	
Sitting	0.5
Standing	0.95
Playing phone/sleeping	0
Raising hand	0.9
Writing	0.8
Emotions	
Happy	0.5
Neutral	0.6
Disgust	0.1

In the classroom, the "neutral" emotion is the most popular and represents a neutral state. The "disgust" facial emotion shows negative participation and is considered a negative state. While a "happy" facial emotion may suggest complete participation in classroom activities, it may also reflect involvement in unrelated recreational activities; consequently, "happy" is considered a neutral mood. The values assigned to each facial emotion are based on expert knowledge evaluation. This empirical method assures that the classification of student facial emotions in the classroom is consistent with common sense. As a result, the facial emotion assignment matrix is defined as $s_2 = (0.1, 0.5, 0.6)$.

The weighted average fusion computations using behavioral and emotional data can be performed once the relationship between students' facial emotions and behaviors (action) has been established. Our experimental results indicate that students' behavior (action) in the classroom is more influential than their facial emotion in determining their classroom engagement status (see Section

4). For instance, when a student uses a cell phone, their classroom engagement status is poor, regardless of their facial emotion. Conversely, when a student stands and responds to a question, even with a disgusted facial emotion, they are still engaged in the classroom, so their engagement status is not negative. Therefore, the assignment of behavioral weights is set slightly higher than that of facial emotion weights, with behavioral (action) weights set to $\delta_1 = 0.55$ and facial emotion weights to $\delta_2 = 0.45$.

In order to obtain accurate state determination results, it is important to have an effective decision-making mechanism at the decision-making level. The weighted average fusion calculation formula, which combines the aforementioned parameters, can be given by

$$State - Value = \left(\frac{s_1 P_1}{\sum_{j=1}^{6} P_{1j}} \right) \cdot \delta_1 + \left(\frac{s_2 P_2}{\sum_{j=1}^{3} P_{2j}} \right) \cdot \delta_2. \tag{6}$$

It is necessary to understand the relationship between the State_Value and the three engagement states. Questionnaires were distributed, the level of classroom participation for each behavior (action) and facial emotion mentioned above was counted, and the following state determination rules were obtained:

(1) If state_Value < 0.35, it is a negative state;
(2) If $0.35 <$ state_Value < 0.7, it is a neutral state;
(3) If state_Value > 0.7, it is a positive state.

4 Experiments and Results analysis

4.1 The result for the behavior(action) recognition machine learning model model

The model for behavior recognition was trained on the custom-made classroom behavior (action) dataset. The training results indicate that early stopping was implemented during the 14th epoch, resulting in a training loss of 0.18 and a training accuracy of 93.86%. The validation loss was 0.03, with a validation accuracy of 96.84%. At this stage, both the training and validation set loss function values had stabilized, indicating the completion of model training. Loss and accuracy curves are depicted in Figure 3. Additionally, a confusion matrix was generated, and recall and precision for each classification were computed (see Figure 6 and Table 4. Since the custom-made dataset does not generalize well, the loss and accuracy curves on the validation set are flatter and do not follow the trend of the curves on the test set. In further study, we will continue to enhance the generalization of the dataset.

4.2 The result for the facial emotion recognition machine learning model model

The model for facial emotion was trained on the Emotion-Domestic Expression Recognition dataset, which consists of a test set and a training set. The results

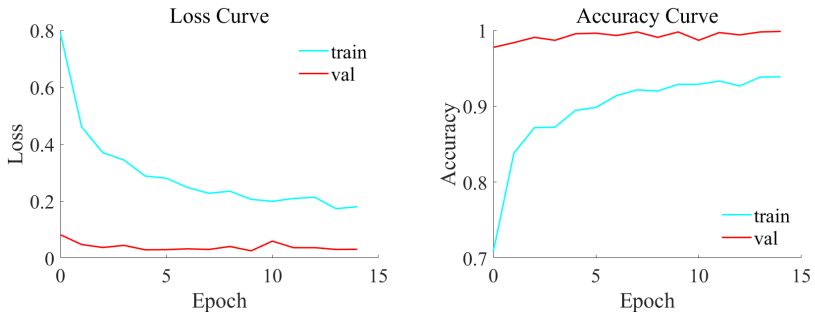

Fig. 3: The loss and accuracy for action recognition

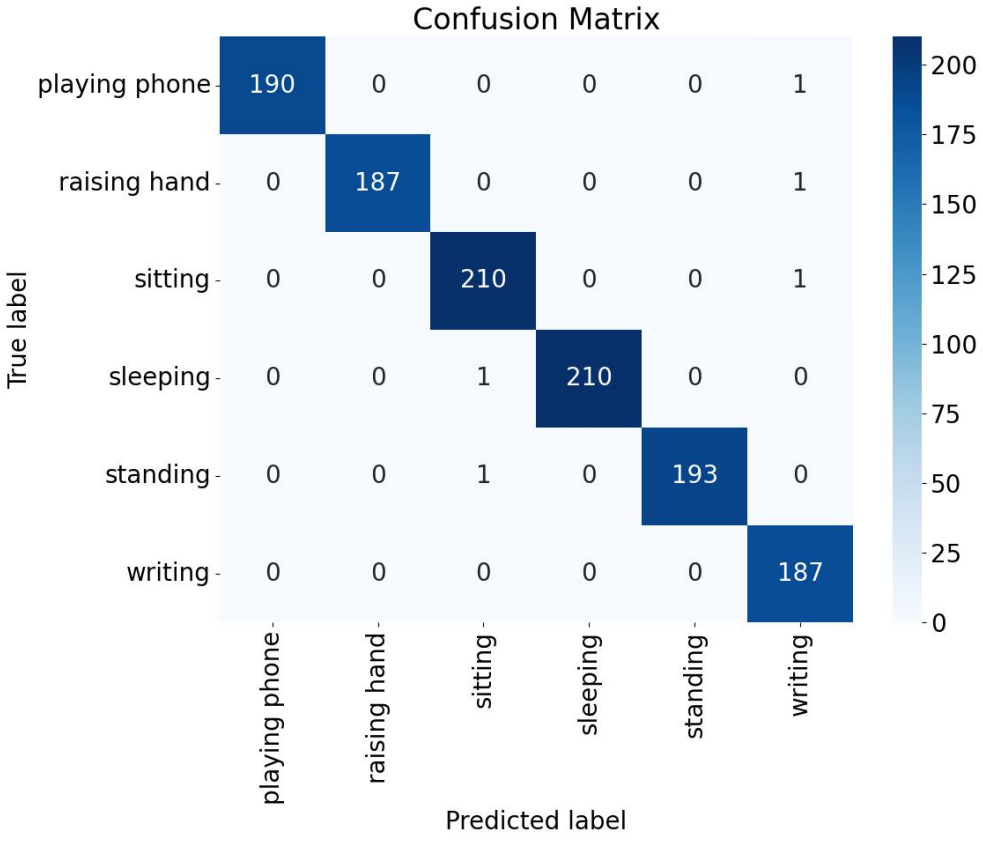

Fig. 4: Behavioral identification confusion matrix

Table 4: Recall and precision for six behaviors

Actions	**Recall**	**Precision**
playing phone	99.5%	100%
raising hand	99.5%	98.9%
sitting	99.1%	99.5%
sleeping	99.5%	100%
standing	99.5%	100%
writing	100%	98.4%

from the training show that early stopping was applied during the 27th iteration of training, resulting in a training loss of 0.18 and a training accuracy of 92.6%. The loss on the test set was 0.06, and the accuracy was 97.71% (see Figure 5). Furthermore, a confusion matrix was created, and the recall and precision for each category were assessed. The results are displayed in Figure 6 and Table 5.

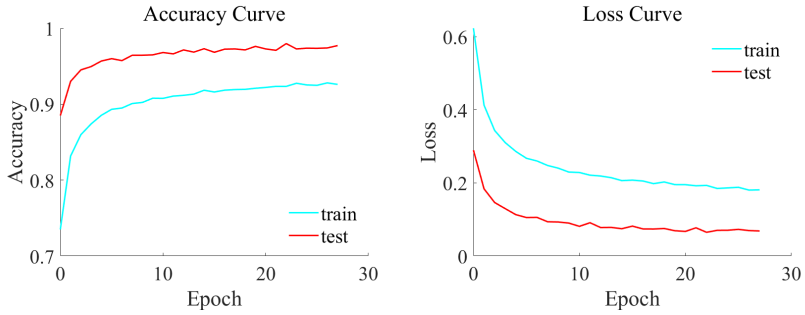

Fig. 5: The loss and accuracy for Emotion recognition

Table 5: Recall and Precision for Emotion

Emotion	**Recall**	**Precision**
Disgust	89.1%	92.1%
Happy	98.7%	99%
Neutral	96.7%	95.1%

4.3 The fusion based experimental result

The weighted average fusion approach was used to evaluate students' engagement levels. For instance, Figure 7 illustrates the outcomes of a test experiment in which the observed behavior involved raising hands, the facial emotion expressed is conveyed as disgust and the engagement state by the weighted average fusion is "Neutral." Additionally, Table 6 presents the parameters and results of the weighted average fusion method for the experiment shown in Figure 7.

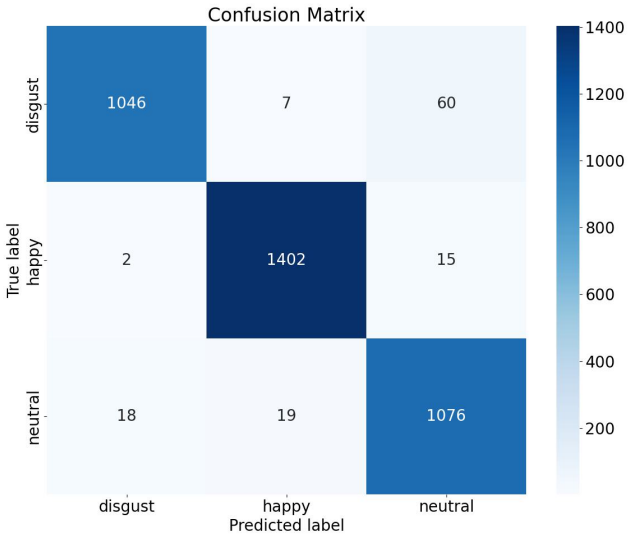

Fig. 6: Emotion identification confusion matrix

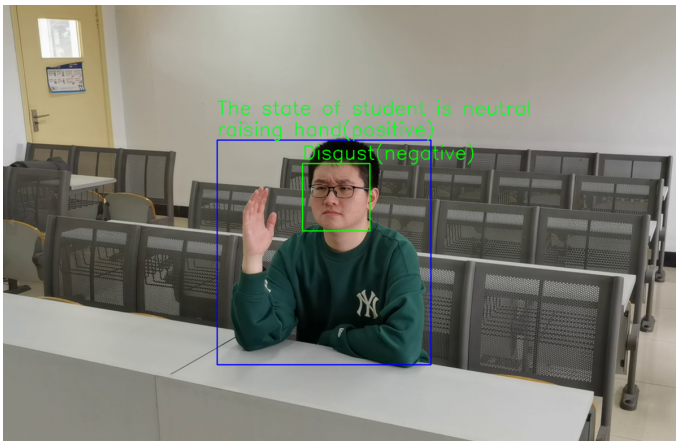

Fig. 7: Example of a student engagement state using the proposed method

Table 6: The weighted average fusion calculations for Figure 7

	Classifications	**Softmax output probabilities**	**Confidence level**	**Single data source results**
Actions	playing phone	0.077%	0.076%	0.892
	raising hand	98.2%	97.7%	
	sitting	1.56%	1.54%	
	sleeping	0.17%	0.16%	
	standing	0.013%	0.013%	
	writing	0.012%	0.012%	
Emotion	disgust	95.2%	84.8%	0.126
	happy	0.069%	0.068%	
	neutral	4.72%	4.57%	
Fusion results	0.547			

The test experiment findings in Table 6 show that the highest probability in the Softmax layer of the behavior recognition model corresponds to the action of hand-raising, which also has the highest confidence level. Similarly, in the facial emotion model, disgust is associated with the highest likelihood and level of confidence. Therefore, the most likely interpretations of the student's actions and emotions are hand-raising and disgust. The weighted calculation results in a fusion state value of 0.547, falling within the neutral range $[0.35, 0.7]$. Therefore, the decision-making system categorizes the student's engagement condition as neutral.

The weighted average fusion method is used to combine students' behavior (action) and facial emotion. There are six student behaviors: playing with a cell phone, raising hands, sitting, sleeping, standing, and writing, and three types of student emotions: disgusted, happy, and neutral. This results in a total of 18 potential behavior-emotion pairings. However, because a student's face is completely covered while sleeping, their emotion cannot be identified. As a result, only behavioral information is used to determine the sleeping condition. Taking this into consideration, a total of 16 comparison experiments need to be established. The "Positive," "Neutral," and "Negative" state determination methods were employed to classify the 16 comparison experiments. The weighted average fusion was determined from the behavioral and emotional states. The outcome of the experiments is shown in Table 7.

In order to assess the accuracy of the fusion result presented in Table 7, a manual evaluation of the 16 different student states was performed. A questionnaire was developed and distributed to individuals who had viewed the classroom video dataset. The manual evaluation findings were examined using the fuzzy logic method and the maximal affiliation principle to identify the state with the highest degree of affiliation. The results are shown in Table 8.

The accuracy of the behavior (action)-only evaluation, facial emotion-only evaluation, and weighted average fusion-based method combining behavior and emotion were evaluated. The manual evaluation results were used as the stan-

Table 7: Experimental results

Experimental information	**Behavior-based outcomes**	**Emotion-based outcomes**	**Combined behavioral and expressive outcomes**	**Fusion results**
Sitting/Happy	Neutral	Positive	Neutral	0.45
Sitting/Neutral	Neutral	Neutral	Neutral	0.41
Sitting/Disgust	Neutral	Negative	Negative	0.33
Standing/Happy	Positive	Positive	Positive	0.71
Standing/Neutral	Positive	Neutral	Positive	0.75
Standing/Disgust	Positive	Negative	Neutral	0.61
Playing phone/Happy	Negative	Positive	Negative	0.23
Playing phone/Neutral	Negative	Neutral	Negative	0.28
Playing phone/Disgust	Negative	Negative	Negative	0.11
Raising hand/Happy	Positive	Positive	Positive	0.71
Raising hand/Neutral	Positive	Neutral	Positive	0.76
Raising hand/Disgust	Positive	Negative	Neutral	0.62
Writing/Happy	Positive	Positive	Neutral	0.67
Writing/Neutral	Positive	Neutral	Positive	0.71
Writing/Disgust	Positive	Negative	Neutral	0.48
Sleeping	Negative	/	Negative	0

Table 8: Results of manual evaluation

Experimental information	**State**	**Degree of affiliation**
Sitting/Happy	Positive	0.63
Sitting/Neutral	Neutral	0.6
Sitting/Disgust	Negative	0.77
Standing/Happy	Positive	0.87
Standing/Neutral	Positive	0.73
Standing/Disgust	Neutral	0.43
Playing phone/Happy	Negative	1
Playing phone/Neutral	Negative	0.97
Playing phone/Disgust	Negative	1
Raising hand/Happy	Positive	0.93
Raising hand/Neutral	Positive	0.93
Raising hand/Disgust	Neutral	0.4
Writing/Happy	Positive	0.6
Writing/Neutral	Positive	0.77
Writing/Disgust	Neutral	0.4
Sleeping	Negative	1

dard. The behavior (action)-only method achieved an accuracy of 68.75%, while the facial emotion-only based method had an accuracy of 43.75%. The combined behavior and emotion-based method achieved an accuracy of 87.5%.

These findings demonstrate that the combined method is more accurate than strategies relying on either behavior (action) or facial emotion for student engagement state. This research emphasizes the benefits of the fusion method for improving accuracy and decision credibility, as well as validating the usefulness of combining information from different sources. Furthermore, the accuracy of the facial emotion-only method is 25% lower than that of the behavior (action)-only method, supporting the rationale for giving slightly more weight to behavior information (δ_1) than to emotion information (δ_2), as discussed in previous studies.

5 Conclusion

The growing use of smart classrooms in education requires improved methods for assessing student engagement using multisensory fusion techniques. In our study, a decision-level weighted average fusion approach that combines student behavior (action) recognition and facial emotion recognition to evaluate engagement in the classroom was introduced. A custom-made student behavior (action) dataset was created to train the ResNet34 model for behavior (action) recognition, achieving a classification accuracy of 97.5% across six behavior (action) categories. In addition, the Emotion-Domestic Expression Recognition dataset was used to train the MobileNetV2 model for emotion recognition, with an accuracy of 97.3%. Classroom engagement levels were determined by using a weighted average decision-level fusion method to combine behavioral and emotional data. The evaluation involved comparing the engagement metrics derived from behavior (action)-only, emotion-only, and combined behavior (action)-emotion recognition methods against manual evaluation results. The behavior(action)-only method achieved an accuracy of 68.75%, while the emotion-only method yielded 43.75%. Importantly, the combined approach significantly outperformed both, with an accuracy of 87.5%. These results emphasize the superiority of the combined method over single-modality approaches in accurately assessing student engagement.

Bibliography

[1] J. Li and E. Xue, "Dynamic interaction between student learning behaviour and learning environment: Meta-analysis of student engagement and its influencing factors," *Behavioral Sciences*, vol. 13, no. 1, p. 59, 2023.

[2] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Review of educational research*, vol. 74, no. 1, pp. 59–109, 2004.

[3] A. L. Reschly and S. L. Christenson, *Handbook of research on student engagement*. Springer, 2022.

[4] G. Eisele, H. Vachon, G. Lafit, P. Kuppens, M. Houben, I. Myin-Germeys, and W. Viechtbauer, "The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population," *Assessment*, vol. 29, no. 2, pp. 136–151, 2022.

[5] S. Gupta, P. Kumar, and R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 11365–11394, 2023.

[6] A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2132–2143, 2022.

[7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[8] csdn.net, "Emotion-domestic (Asia) expression recognition dataset," 2024. Accessed: March 24, 2024.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[10] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," in *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, pp. 775–782, World Scientific, 1996.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.