IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 3, June, 2025, pp. 1721-1732
Journal homepage: http://iieta.org/journals/ts

A Multidimensional Image Feature-Based Method for Classifying Student Group Learning Behaviors and Its Educational Intervention Strategies

Shiqiu Dong[1, 2], Xiao Gao[3*], Qiuying Li[4], Yanli Wang[2], Jiayi Feng[5], Shunxia He[6], Xinxin Chun[6]

[1] Graduate School, Gachon University, Seongnam 13120, Republic of Korea
[2] Department of Clinical Medicine, Heilongjiang Nursing College, Harbin 150086, China
[3] The Fourth Affiliated Hospital of Heilongjiang University of Chinese Medicine, Harbin 150077, China
[4] The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, China
[5] College of Economics and Management, Northeast Agricultural University, Harbin 150030, China
[6] Graduate School, Heilongjiang University of Chinese Medicine, Harbin 150040, China

Corresponding Author Email: arthur2557@126.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ts.420341

Received: 20 December 2024
Revised: 3 May 2025
Accepted: 21 May 2025
Available online: 30 June 2025

Keywords:
multidimensional image feature extraction, student group learning behavior classification, educational intervention strategies, feature fusion, classification model

ABSTRACT

With the deep integration of educational informatization and intelligent technology, multidimensional image data collected by smart devices has become a rich resource for analyzing student group learning behaviors. Accurate classification of these behaviors is essential for optimizing teaching strategies and enhancing educational quality. However, existing research faces three major limitations: (1) reliance on single image features, overlooking the association between local structural features such as body movements and complex learning environments; (2) simplistic feature fusion methods that fail to account for the correlation and varying importance of multidimensional features, thereby limiting classification accuracy; and (3) a lack of systematic development of educational intervention strategies, hindering the practical application of behavioral analysis. To address these issues, this study proposes a multidimensional image feature extraction method for classifying student group learning behaviors. The method integrates local structural features, globally weighted local phase quantization (LPQ) index structure features, scene features, color features, and image information entropy to construct a comprehensive feature representation framework. A high-efficiency classification model is developed in tandem with targeted educational intervention strategies, forming a complete framework of "feature extraction–behavior classification–intervention implementation." The research outcomes are expected to significantly improve the accuracy of learning behavior classification, provide robust data support for personalized teaching and optimized learning guidance, and promote the deep integration of behavior analysis techniques with practical educational interventions in the context of educational informatization.

1. INTRODUCTION

With the rapid development of educational informatization, classroom teaching scenarios are becoming increasingly complex and diverse [1-4], and students' learning behaviors are showing diversified characteristics. With the widespread application of intelligent devices [5-7] in the education field, such as smart cameras, interactive whiteboards, etc., a large amount of multidimensional image data containing students' learning behaviors can be collected in real time. These image data contain rich information, such as students' body movements, facial expressions, classroom participation, etc., providing a massive data source for in-depth analysis of student group learning behaviors. Accurate recognition and classification of students' learning behaviors are of great practical significance for understanding the learning process, optimizing teaching strategies, and improving education quality.

Conducting classification research on student group learning behaviors helps educators accurately grasp students' learning states and needs [8-10]. By analyzing different types of learning behaviors, such as attentive listening, active interaction, and distraction, it can provide a basis for personalized education and realize teaching in accordance with students' aptitudes. At the same time, accurate classification of learning behaviors can provide scientific support for the formulation of educational intervention strategies, and adopt corresponding guidance and assistance measures for students with different learning behaviors, so as to improve learning efficiency and outcomes. In addition, this research can also provide technical support for the development of educational informatization, promote the development and application of intelligent education systems, and facilitate the modernization transformation of education and teaching.

At present, research on student learning behavior classification has achieved certain results, but there are still many deficiencies. Some studies only use single image features, such as color features or local texture features [11-14], making it difficult to comprehensively describe students' complex learning behavior scenarios. For example, literature [15] only uses local texture features to represent human behavior, ignoring important local structural features such as body movements, resulting in low classification accuracy. Some other studies, although integrating multiple features, have deficiencies in feature fusion methods and fail to fully utilize the advantages of multidimensional features [16-19]. Literature [20] proposed a behavior classification method based on simple weighted fusion, without considering the correlation and importance differences between different features, resulting in limited improvement in classification performance. In addition, most of the existing studies lack systematic research on educational intervention strategies, and fail to effectively combine learning behavior classification results with specific educational intervention measures, making it difficult to realize the transformation from behavior analysis to practical application.

This paper mainly conducts research in two aspects. On the one hand, it proposes a multidimensional image feature extraction method for student group learning behavior classification, specifically including local structural features, globally weighted LPQ index structural features, scene features, color features, and image information entropy features. By comprehensively extracting these features, students' learning behavior scenarios can be described more comprehensively and accurately. On the other hand, this study investigates student group learning behavior classification methods and their educational intervention strategies. An efficient classification model is constructed to accurately classify students' learning behaviors, and targeted educational intervention strategies are formulated based on the classification results, such as personalized teaching programs, learning guidance methods, etc. The value of this study lies in improving the accuracy and reliability of student learning behavior classification through the fusion extraction of multidimensional image features and efficient classification methods, providing educators with more accurate information on student learning behaviors. Meanwhile, the proposed educational intervention strategies can be directly applied in actual teaching to help teachers better guide students' learning and improve learning outcomes and education quality. In addition, this study provides new ideas and methods for the analysis and intervention of students' learning behaviors under the background of educational informatization, with important theoretical significance and practical application value.

2. MULTIDIMENSIONAL IMAGE FEATURE EXTRACTION METHOD FOR STUDENT GROUP LEARNING BEHAVIOR CLASSIFICATION

This paper selects five aspects for feature extraction from multidimensional images oriented to student group learning behavior classification: local structural features, globally weighted LPQ index structural features, scene features, color features, and image information entropy features. Student group learning behavior classification requires the precise capture of both individual behavior details and group interaction patterns, and the above five types of features correspond to different representation dimensions of learning behaviors, enabling systematic coverage of key information related to learning states in classroom scenarios. Local structural features can effectively extract micro-expression details such as students' body movements and facial expressions, and these local dynamic features are the core basis for judging individual concentration and participation. The globally weighted LPQ index structural features, by constructing spatial structure relation models, can describe global patterns such as the spatial distribution and interaction frequency of student groups in classrooms, providing support for distinguishing group behaviors like collaborative learning and passive listening. Scene features and color features focus on environmental context. Scene features can eliminate background interferences such as classroom lighting and seating arrangement, highlighting the physical scene where learning behaviors occur. Color features in the HSV color space can capture visual stimuli such as textbook color and screen content, which influence students' attention. Image information entropy features essentially reflect the complexity level of image content and can serve as a quantitative indicator for judging the depth of learning behaviors. Figure 1 presents the proposed multidimensional image feature extraction framework for student group learning behavior classification.

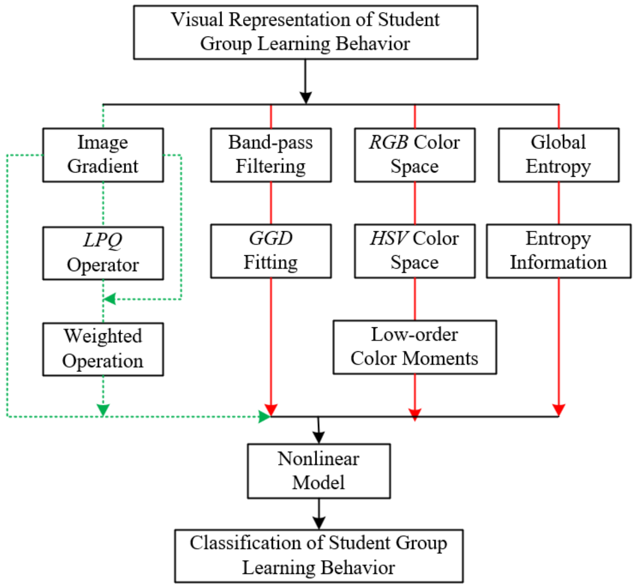


Figure 1. Framework of multidimensional image feature extraction method for student group learning behavior classification

In this study, the categories of multidimensional images for feature extraction closely revolve around the visual representation needs of student group learning behaviors, and mainly include the following five types: First, individual behavior detail images, such as close-up images of students' facial micro-expressions and body movements, are used to extract gradient-based local structural features to precisely capture micro-level behavior signals such as concentration and emotional state. Second, group interaction scene images, including panoramic or mid-range images showing students' spatial distribution and body orientation relations, are used to construct globally weighted LPQ index structural features based on gradient and texture information to depict the spatial correlation patterns of group behaviors such as group discussions and teacher-student interactions. Third, environmental context images, covering overall classroom layout, lighting conditions, etc., are processed by division normalization to form scene features, eliminating environmental noise interference in behavior analysis. Fourth, visual stimulus images, such as textbook cover colors, electronic screen contents, and teaching tool colors, are local region images from which color features are extracted based on HSV color space to analyze the impact of visual elements on students' attention allocation. Fifth, cognitive load images, including note-taking trajectories, experiment operation processes, and teaching tool usage states, reflect the complexity of behaviors. Image information entropy features are calculated via global entropy to quantify the cognitive input depth of students in specific learning tasks. These multidimensional image categories correspond to different observation scales such as micro-level individual behaviors, meso-level group interactions, and macro-level environmental context, together forming an image data source that covers the "subject–interaction–environment" triadic elements of learning behaviors.

2.1 Local structural feature extraction

Image gradients, as a quantitative representation of grayscale variation rate, play a key role in capturing local structural changes in pixel neighborhoods through mathematical differential operations, which have a direct mapping relationship with micro-level details of students' individual learning behaviors. In classroom scenarios, students' behavioral features such as facial micro-expressions and body movements essentially manifest as grayscale gradient variations in local regions of the image. Table 1 presents common gradient operator convolution tables. This paper selects the Sobel operator to perform convolution on the image in order to effectively extract the direction and magnitude information of these local structures: the gradient direction reflects the spatial distribution trend of behavior details, while the gradient magnitude quantifies the clarity of those details. This process of converting behavior details into gradient signals lays the physical foundation for characterizing the dynamic variation of learning behaviors through statistical gradient features in subsequent steps.

Table 1. Common gradient operator convolution table

	Scharr	Sobel	Prewitt
Horizontal Direction	$\begin{bmatrix} -3 & 0 & +3 \\ -10 & 0 & +10 \\ -3 & 0 & +3 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix}$
Vertical Direction	$\begin{bmatrix} -3 & 0 & +3 \\ 0 & 0 & 0 \\ -3 & +10 & +3 \end{bmatrix}$	$\begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$	$\begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$

This paper chooses the Sobel operator to compute image gradients due to its dual advantages in noise robustness and directional sensitivity. This operator performs weighted convolution on the center pixel and its 4-neighbor or 8-neighbor pixels, which suppresses random noise interference while more precisely capturing the true gradient direction of edges. Specifically, the gradient components in the horizontal and vertical directions are first calculated, then synthesized to obtain gradient magnitude and gradient direction, forming a complete gradient map. On this basis, in response to the specific needs of student behavior classification, the mean value of the gradient map and the standard deviation of the vertical gradient map are extracted as core features: the gradient mean reflects the overall grayscale variation intensity in the local region and can be used to distinguish high-dynamic and low-dynamic behaviors; the vertical gradient standard deviation characterizes the dispersion degree of grayscale variation in the vertical direction and is suitable for describing behavioral differences in vertical dimensions such as head posture and note-taking.

Specifically, suppose the input image is represented by U, and the partial derivatives in the horizontal and vertical directions are represented by H_g and H_n, with the convolution operation denoted by *, then in this paper, the image gradient is calculated as follows:

$$|H| = |H_g| + |H_n| \tag{1}$$

$$H_g = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * U \tag{2}$$

$$H_n = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * U \tag{3}$$

Assuming the gradient map and the vertical gradient map are denoted as H and H_B, the mean operation is denoted as L, and the standard deviation operation is denoted as δ. The mean of the gradient map and the standard deviation of the vertical gradient map, as local structural features, can be calculated as follows:

$$MEAN(H) = L(H) \tag{4}$$

$$STANDARA(H_b) = \delta(H_b) \tag{5}$$

Local structural features based on gradient information directly serve the accurate classification of students' learning behaviors by quantifying the "discernibility" and "regularity" of behavior details. In individual behavior analysis, regions with lower gradient mean and smaller vertical gradient standard deviation usually correspond to "passive listening" states, characterized by smooth facial expressions and small body movement amplitudes; whereas regions with higher gradient mean and more scattered distribution in vertical gradient direction may indicate gestural communication in "group discussion" or coordinated physical actions in "experimental operations". By fusing these features with global phase quantization features, scene features, and other multidimensional signals, a feature system covering "micro behavior details – meso spatial relationships – macro environmental context" can be constructed, effectively addressing the partial representation problem of complex behaviors caused by single features.

2.2 Globally weighted LPQ index structural feature extraction

The classification of student group learning behaviors not only relies on individual micro actions, but also requires capturing spatial structural patterns of group interactions, such as encirclement postures during group discussions and orientation distributions during teacher-student Q&A. The design of globally weighted LPQ index structural features centers on constructing a feature system capable of describing the spatial correlation of group behaviors by integrating texture operators and gradient information. The LPQ operator is robust against monotonic grayscale transformations based on the Fourier phase spectrum, making it suitable for handling grayscale fluctuation problems in classroom scenarios caused by lighting variation or equipment differences. It encodes texture patterns through phase relationships of pixels in a 3×3 neighborhood and can effectively capture structural features such as edges and corners, which serve as visual carriers of key information such as body orientation and spatial distribution in group interactions. By combining gradient information, the description of "lines" and "planes" and other macro structures in the image can be further enhanced, forming a quantitative representation of the spatial layout where group behavior occurs and compensating for the contrast sensitivity weakness of single texture operators. The following formula gives the calculation of the texture S under the above distribution pattern:

$$S=o\begin{pmatrix} h_0,h_1-h_0,h_2-h_0,h_3-h_0,h_4 \\ -h_0,h_5-h_0,h_6-h_0,h_7-h_0,h_8-h_0 \end{pmatrix} \tag{6}$$

After eliminating unimportant information in the formula, we have:

$$S=\sum_{u=1}^{8} t\left(h_u-h_0\right)2^{u-1} \tag{7}$$

where,

$$t\left(a\right)=\begin{cases} 1, a\geq 0 \\ 0, a<0 \end{cases} \tag{8}$$

The expression of the LPQ operator used in texture calculation is as follows:

$$LPQ=\begin{cases} \sum_{u=1}^{8} t\left(h_u-h_0\right), if\ I\left(S\right)\leq 2 \\ 9, otherwise \end{cases} \tag{9}$$

The specific extraction process consists of three core steps: First, the Prewitt operator is used to calculate the horizontal and vertical gradient components of the panoramic image, and a gradient map is synthesized to highlight edge and contour information, providing a structural basis for subsequent texture analysis. Assuming the input panoramic image is denoted as P, the partial derivatives in the horizontal and vertical directions are denoted as O_g and O_n, and the convolution operation is denoted by *. The gradient calculation formula is as follows:

$$H_p=\sqrt{\left(P*O_g\right)^2+\left(P*O_n\right)^2} \tag{10}$$

Second, the LPQ operator is applied on the gradient map. Each pixel's 3×3 neighborhood is phase-quantized and encoded into a 256-dimensional local phase pattern. Rotational invariance is achieved through uniformity measurement, ensuring adaptability of the features to group posture variations. The LPQ operator is applied on the obtained gradient map to generate the weighted LPQ index map. Assuming the pixel grayscale value is denoted as h_u, the texture operator with rotational invariance is denoted as S, and the uniformity measurement pattern is denoted as $I(*)$, the formula is:

$$GLPQ=\begin{cases} \sum_{u=1}^{8} t\left(h_u-h_0\right), if\ I\left(S\right)\leq 2 \\ 9, otherwise \end{cases} \tag{11}$$

The function $t(.)$ in the above formula is defined as:

$$t\left(h_u-h_0\right)=\begin{cases} 1, h_u-h_0\geq 0 \\ 0, h_u-h_0<0 \end{cases} \tag{12}$$

Finally, a gradient magnitude weighting mechanism is introduced. Gradient values of pixels with the same LPQ pattern are accumulated and then normalized by the global gradient sum to form the weighted LPQ index map. This weighting operation essentially assigns importance values to different texture patterns: LPQ patterns in high-gradient regions are given higher weights, thereby highlighting key structural features in group interactions and suppressing noise or irrelevant background interference. Assuming the pixel values of the input image are denoted as V, the possible weighted LPQ index patterns are denoted as j, and the gradient magnitude of each pixel is denoted as q_u, the formulas are:

$$GLPQ\left(j\right)=\sum_{u=1}^{V}\mu_u d\left(GLPQ, j\right) \tag{13}$$

$$d\left(GLPQ, j\right)=\begin{cases} 1, GLPQ=j \\ 0, otherwise \end{cases} \tag{14}$$

The final extracted 256-dimensional feature vector not only preserves the spatial distribution patterns of group behaviors but also enhances the influence of visually salient regions through gradient weighting, providing the classification model with highly discriminative global structural signals.

The globally weighted LPQ index structural features, through the dual mechanism of "texture pattern encoding + gradient weight modulation", precisely depict the spatial structural patterns of student group interactions, serving as a bridge between individual behaviors and group patterns. In classroom scenarios, this feature can effectively distinguish the organizational forms of different behavioral categories. In specific cases, "group discussion" behaviors correspond to LPQ pattern distributions that are more dispersed with concentrated high-gradient weights, while "passive listening" behaviors show concentrated LPQ patterns with low gradient weights. By integrating with gradient-based local structural features, scene features, and other multidimensional features, a three-dimensional representation system of "micro individual – meso group – macro scene" can be constructed, addressing the limitation of traditional methods relying on a single perspective in group behavior analysis. When the local gradient features of a region indicate high individual activity, and the globally weighted LPQ features show aggregation of multi-directional phase patterns, it can be determined as "collaborative learning" behavior. This provides quantitative indicators such as "group interaction intensity" and "spatial collaboration efficiency" for educational intervention strategies, assisting teachers in designing targeted group tasks or adjusting classroom layout, and realizing the effective transformation from behavior classification to practical application.

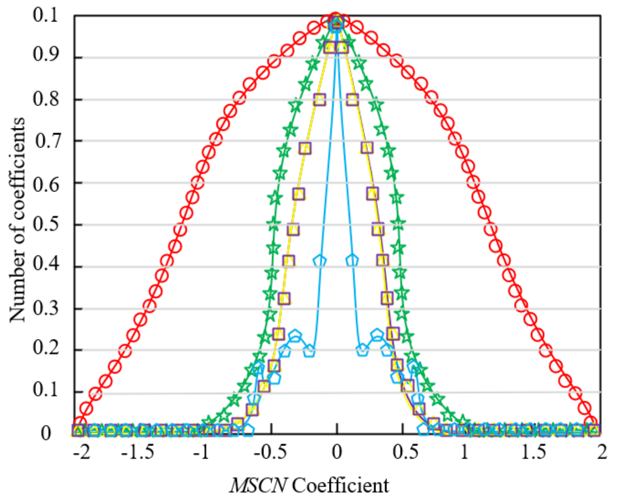
2.3 Scene feature extraction

Student group learning behaviors occur in complex physical classroom environments. Environmental factors such as classroom lighting intensity, seating layout, and placement of teaching tools directly affect the visual presentation of images, leading to deviations in image features of the same behavior under different scenes. The core goal of scene feature extraction based on divisive normalization is to eliminate the interference of environmental noise and highlight the essential characteristics of learning behaviors themselves. Its theoretical basis comes from the Natural Scene Statistics (NSS) theory. The local brightness and contrast of natural images have regular statistical characteristics, but environmental factors destroy such characteristics, forming "distortion" effects. Through local mean subtraction and divisive normalization operations, the brightness and contrast of images can be standardized, so that images from different scenes become comparable in a unified feature space, providing a stable environmental baseline for subsequent behavior classification.

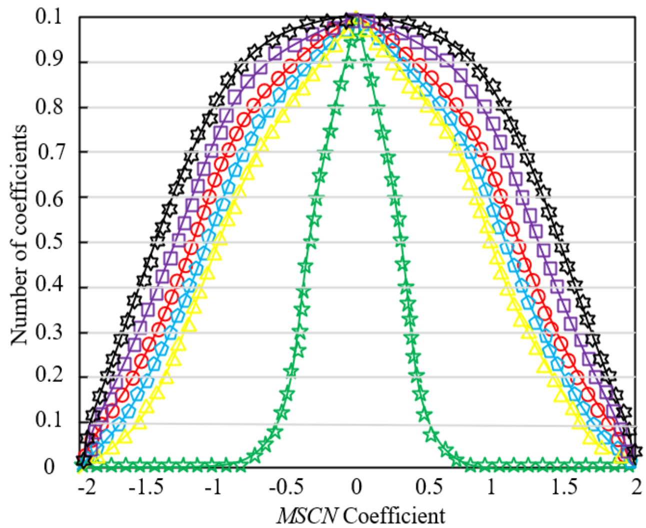
The specific extraction process follows a technical route of "noise suppression — feature normalization — statistical modeling": First, local mean subtraction is performed on the input panoramic image to eliminate brightness shift caused by uneven environmental lighting, so that the brightness of local image regions is distributed around zero mean. Assuming the distorted panoramic image is denoted as P, the normalized panoramic image is denoted as $\hat{P}$, spatial coordinates are represented by l, v, the local mean and standard deviation of the input image are represented by $\omega(l,v)$, $\delta(l,v)$, and the constant is denoted by Z, the formula is:

$$\hat{P}(l,v)=\frac{P(l,v)-\omega(l,v)}{\delta(l,v)+Z} \tag{15}$$

Then, through divisive normalization, brightness is scaled by local standard deviation to suppress contrast differences in different regions, generating the Mean Subtracted Contrast Normalized coefficient (MSCN). This coefficient simulates the human visual system's perception characteristics of contrast change and converts the local structural information of the image into a feature representation insensitive to environmental changes. Figure 2 shows the histogram of MSCN coefficients of visual scene images for student group learning behaviors. As can be seen from the figure, the distribution of MSCN coefficients approximates a zero-mean Generalized Gaussian Distribution (GGD), and environmental noise in classroom scenes will cause significant changes in the shape parameters and variance of GGD. Therefore, by extracting GGD parameters at two scales, the complexity and regularity of the scene can be captured from a multi-resolution perspective, forming a quantitative representation of environmental factors such as classroom layout and lighting conditions. This normalization process not only preserves the spatial structure where the behavior occurs but also filters out irrelevant environmental variables, enabling the subsequent classification model to focus on the feature differences of student behaviors themselves.



(a) Distorted scenes of different types



(b) Scenes with different distortion levels

Figure 2. Histogram of MSCN coefficients for student group learning behavior visual scene images

$$GGD(a,\beta,\delta^2)=\frac{\beta}{2\alpha\Xi(1/\beta)}\exp\left(-\left(\frac{|a|}{\alpha}\right)^{\beta}\right) \tag{16}$$

where,

$$\alpha=\delta\sqrt{\frac{\Xi\left(3/\beta\right)}{\Xi\left(1/\beta\right)}} \tag{17}$$

Assuming the shape parameter is represented by β, the variance by δ^2, and the Gamma function by $\Xi(\cdot)$:

$$\Xi(\beta)=\int_0^{\infty} s^{\beta-1}r^{-s}fs,\beta>0 \tag{18}$$

Scene features based on divisive normalization essentially build an "environmental context coordinate system" for student learning behaviors, solving the problem of interference from environmental noise in traditional behavior classification methods. In actual classrooms, different learning behaviors have different dependencies on scene conditions. In specific cases, "experimental operation" behavior is usually accompanied by a complex scene with teaching tools arranged, while in "theoretical lecture" scenarios, students face a single direction. By integrating scene features with local structural features and globally weighted LPQ features, a complete feature system that includes "behavior subject — interaction pattern — environmental context" can be constructed. When the scene feature of a certain region shows a high contrast normalization value and the local gradient feature shows a high magnitude, it can be jointly judged as "exploratory learning" behavior, avoiding misrecognition caused by dim lighting. When the scene feature shows low variance and the global LPQ feature presents aggregation of unidirectional phase patterns, it can be judged as "teacher-dominated lecture" behavior.

2.4 Color feature extraction

In student group learning scenarios, color information is a core element of visual perception, which directly affects students' attention distribution and emotional states. The characteristics of the HSV color space are highly compatible with the human visual system's perception logic of color. Compared with the RGB color space, the HSV space decomposes color into three independent components: hue, saturation, and brightness. Among them, hue represents color type using angular degrees, saturation reflects color purity, and brightness describes brightness level. This decomposition method can more intuitively capture key features of visual stimuli in classroom scenes, such as the color type of teaching tools, the vividness of the projection screen, and changes in classroom lighting. These factors are closely related to students' learning behaviors. The selection of the HSV space, especially the H channel, as the core of color feature extraction is based on the high sensitivity of humans to hue changes: studies have shown that hue changes are more easily perceived than changes in brightness or saturation. Key visual elements in the classroom often convey information through hue differences, so the H channel can effectively capture core color signals affecting learning behavior.

The specific extraction process follows the technical procedure of "space transformation — component separation — feature quantization": First, the original RGB image is converted into the HSV color space, and the H channel image is separated. This channel represents the color type of each pixel using continuous angular values ranging from 0° to 360°, forming an intuitive mapping of color distribution in the classroom scene. The transformation formulas from the RGB color space to the HSV color space are as follows:

$$g = \begin{cases} 0^{\circ}, \text{if } MAX = MIN \\ 60^{\circ} \times \dfrac{h-y}{MAX-MIN} + 0^{\circ}, \text{ if } MAX = e \text{ AND } h \geq y \\ 60^{\circ} \times \dfrac{y-e}{MAX-MIN} + 120^{\circ}, \text{ if } MAX = h \\ 60^{\circ} \times \dfrac{e-h}{MAX-MIN} + 240^{\circ}, \text{ if } MAX = y \end{cases} \tag{19}$$

$$t = \begin{cases} 0, \text{if } MAX = 0 \\ \dfrac{MAX-MIN}{MAX} = 1 - \dfrac{MIN}{MAX}, \text{otherwise} \end{cases} \tag{20}$$

$$n = MAX \tag{21}$$

Subsequently, the mean of the H channel is extracted using the color moment theory as the core feature. Color moments describe image color distribution through low-order statistical moments, where the mean reflects the global color tendency and can effectively summarize the dominant tone of the scene. Different distortion types can cause significant changes in the histogram distribution of the H channel, and the mean of the H channel can sensitively capture such distribution differences, forming a robust representation of color information. Assuming the H channel image is represented by U, the mean operation is denoted by L, the color feature formula is:

$$MEAN(U) = L(U) \tag{22}$$

Color features based on the HSV color space essentially convert color information in classroom scenes into quantifiable behavioral influencing factors, solving the problem in traditional methods of ignoring the role of visual stimuli in learning behavior. In practical applications, the mean of the H channel can reveal the color environment characteristics corresponding to different learning behaviors. In specific cases, "theoretical lecture" scenes are dominated by a single background color, and the H channel mean is concentrated and stable. When color features are integrated with local structural features and scene features, a more complete visual representation system can be constructed. When the H channel mean shows high dispersion and the local gradient feature shows high magnitude, the behavior can be judged as "group collaborative learning," reflecting the use of colorful teaching tools and active interaction. If the H channel mean is abnormal and the scene feature shows low contrast, it may indicate a risk of "attention distraction," providing a basis for teachers to adjust the color parameters of teaching media.

2.5 Image information content feature extraction

Shannon information entropy, as a core indicator for measuring the uncertainty of a system, essentially quantifies the degree of disorder in information distribution. This highly corresponds to the correlation between "behavioral complexity — information richness" in student group learning scenarios. Figure 3 shows the multi-dimensional image information entropy features of different distortion types and levels. In classroom images, global entropy can effectively represent the overall complexity of the image content by calculating the probability distribution of all pixel grayscale values. When students are in a "passive listening" state, the pixel grayscale value distribution in the image is relatively concentrated, corresponding to a lower global entropy value. In contrast, in active scenes such as "group discussion" and "experimental operation," body movements and the use of teaching tools lead to drastic changes in pixel grayscale values, resulting in significantly higher global entropy values.

The choice of global entropy rather than local entropy is due to the high resolution and complex scene characteristics of panoramic images. Global entropy can capture the comprehensive information of multi-subject interaction in the classroom environment from an overall perspective, avoiding the defect of local entropy losing macro behavioral patterns due to its focus on details. This provides a global information baseline covering “individual actions — group interactions — scene complexity” for learning behavior classification. The extraction process of global entropy follows the technical path of "probability modeling — information quantification — complexity representation": First, the frequency of occurrence of each grayscale value in the image is counted to construct a grayscale probability distribution function. Then, based on the Shannon entropy formula, the global entropy value is calculated. A larger value indicates that the grayscale distribution in the image is more dispersed and the information content is richer. Taking student experimental operations as an example, test tubes, instruments, gestures, and other elements have significant grayscale differences, resulting in a balanced grayscale probability distribution and a higher global entropy value. In contrast, in a teacher-centered lecture scene, students face the same direction with few body movements, and the grayscale is concentrated in background and uniform posture regions, resulting in a more concentrated grayscale probability distribution and a lower global entropy value. Different distortion types and levels cause monotonic changes in global entropy values, proving its sensitivity in response to image content complexity. This global statistics-based feature extraction method not only retains the overall information load of learning behaviors in classroom scenes but also efficiently compresses complex image data through a single-dimensional vector, providing a lightweight quantification index of "cognitive load" for subsequent classification models. Assuming the pixel value is denoted by v, the probability density is denoted by $o(v)$, the specific formula is:

$$R_U = -\sum_v o(v)\log_2 o(v) \tag{23}$$

The image information quantity feature based on global entropy essentially transforms the visual complexity of classroom scenes into a computable “behavioral activity index,” solving the problem in traditional methods of ignoring the differences in cognitive depth of learning behaviors. In practical classification, global entropy values can effectively distinguish the information density of different behavior categories: low entropy corresponds to “passive reception” type behaviors, indicating that students are in a relatively static state with low interaction and low movement; high entropy corresponds to “active construction” type behaviors, reflecting dynamic interactive scenes involving multiple subjects and modalities. When combined with local structure features and color features, a three-dimensional feature system of “micro detail — macro complexity — visual environment” can be constructed. In specific cases, if a region shows high global entropy, high amplitude in local gradient features, and large dispersion in H channel mean, it can be jointly judged as “experimental inquiry” behavior, accurately identifying students’ high cognitive engagement in complex operations. If the global entropy value is abnormally low and the scene feature shows low contrast, it may indicate a “distraction risk,” providing a data basis for teachers to adjust teaching pace or optimize environmental lighting.

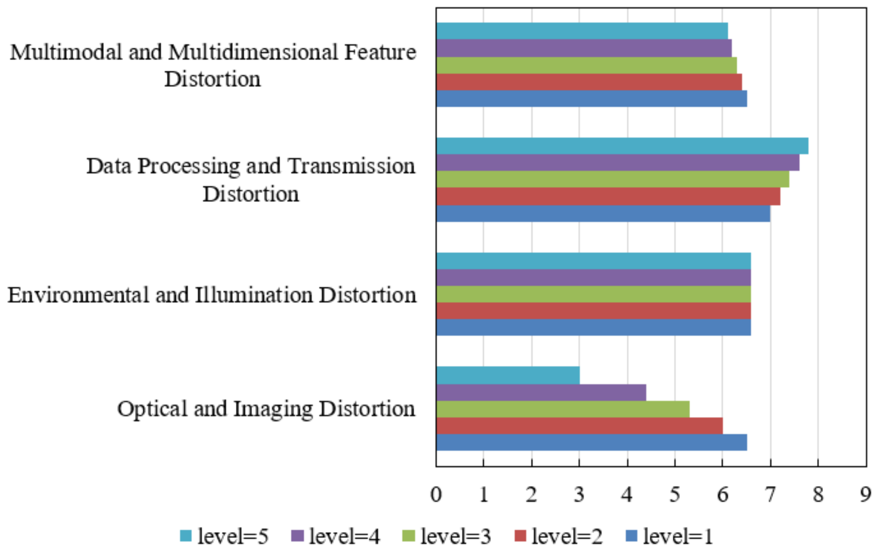


Figure 3. Multidimensional image information entropy features of different distortion types and levels

3. STUDENT GROUP LEARNING BEHAVIOR CLASSIFICATION AND EDUCATIONAL INTERVENTION STRATEGIES

Based on the previously extracted five types of image features, student group learning behavior classification is realized through the technical path of “feature encoding — model training — category determination.” First, the five types of features are dimensionally integrated: local structure features characterize individual action details; global weighted LPQ features describe the spatial structure of group interaction; scene features eliminate environmental noise and provide scene context; color features capture the core signals of visual stimuli; image information quantity features quantify scene complexity. Dimensional redundancy is reduced through principal component analysis or feature selection algorithms, forming a composite feature vector including micro behaviors, meso-level interactions, and macro environments.

In the construction of the classification model, for the high-dimensional and high-distinction feature space, nonlinear models such as support vector classification and random forests are used to solve the problem that traditional linear models are insufficient in fitting complex behavior patterns. During the training process, labeled behavior samples and their multi-dimensional features are input into the model. Parameters are optimized through cross-validation so that the model learns the decision boundaries of different behavior categories in the feature space. In specific cases, “experimental operation” behavior is characterized in the feature space by: high local gradient standard deviation, dispersed global LPQ patterns with high gradient weighting values, large variance of scene features, high dispersion of H channel mean, and high global entropy; while “distracted or absent-minded” behavior corresponds to combinations such as low local gradient mean, concentrated global LPQ pattern, low contrast of scene features, and low global entropy. Through model inference, the behavior category of unknown samples can be accurately determined.

The formulation of educational intervention strategies closely relies on the results of learning behavior classification, forming a closed-loop mechanism of “behavior diagnosis — demand analysis — strategy matching,” providing differentiated schemes for different behavior categories and their feature causes:

(1) Active Interaction Behavior

When the classification model determines a behavior with high entropy, high gradient amplitude, and dispersed LPQ pattern, it indicates that the student is in an active collaboration or deep operation state. Intervention strategies focus on reinforcing positive behavior: ① Environment adaptation: Optimize spatial layout according to scene features, such as retaining multi-colored teaching tools areas needed for experimental operations, increasing the enclosing seat distance for group discussions; ② Resource supply: Design visual guidance schemes based on color features, using high-contrast color annotations to highlight key operation steps or discussion topics to enhance attention focus; ③ Process support: Use global LPQ features to identify weak links in interaction and push targeted collaboration tools to improve group interaction efficiency.

(2) Passive Reception Behavior

For behaviors with low entropy, low gradient standard deviation, and concentrated LPQ pattern, it is necessary to identify whether it is an effective learning state: ① State diagnosis: If scene features show high contrast and color features are balanced, it is judged as a normal focused state and the existing teaching rhythm is maintained; ② Activation intervention: If accompanied by low variance of scene features or abnormal color features, increase dynamic visual stimuli to raise global entropy, combine local gradient features to monitor changes in body movements, and adjust the explanation method in real time; ③ Personalized guidance: For students with low local gradient mean but high vertical gradient standard deviation in note-taking behaviors, push structured note-taking tools to transform low-interaction behaviors into efficient knowledge construction processes.

(3) Abnormal Distraction Behavior

When it is detected that the local gradient mean approaches zero, global entropy suddenly drops, and LPQ pattern concentrates in non-teaching areas, the intervention mechanism is triggered: ① Environment optimization: Identify interference sources based on scene features, and improve scene features by automatically adjusting classroom light color temperature or shielding irrelevant color stimuli; ② Behavior guidance: Design attention anchors using color features, locate the spatial position of distracted students using global LPQ features, and push short prompts in a directional manner to reconstruct visual focus; ③ Task adaptation: Increase micro-interaction tasks based on image information quantity features to improve scene complexity and student participation, transforming low-information scenes into moderately complex effective learning states.

The core value of educational intervention lies in forming a closed-loop optimization through real-time feedback of multi-dimensional features: ① Feature monitoring: Establish a real-time image feature acquisition system for classrooms, continuously monitor the dynamic changes of indicators such as local gradient and global entropy, and identify the critical points of behavior category transitions; ② Strategy iteration: Use the prediction error of the classification model to optimize feature weights in reverse. For example, if the classification accuracy of “experimental operation” is significantly affected by scene features, increase the decision weight of GGD parameters in the model; ③ Effect evaluation: Quantify the effectiveness of strategies by comparing feature distributions before and after intervention, forming an upward spiral system of “classification — intervention — evaluation — improvement.”

4. EXPERIMENTAL RESULTS AND ANALYSIS

From the performance comparison on the SCBDataset5 dataset shown in Table 2, it can be seen that the proposed algorithm performs outstandingly in the classification tasks of the three types of learning behaviors. The classification accuracy for active interaction behavior reaches 0.9238, which is close to advanced methods such as 3DCNN and far exceeds traditional methods, reflecting the precise depiction of group interaction scenes by multidimensional features. In the classification of passive reception behavior, the proposed algorithm leads all comparison algorithms with an accuracy of 0.7793, benefiting from the effective description of static scenes by scene features and color features, which can distinguish between focused listening and low-interaction states. The classification accuracy for abnormal distraction behavior is 0.9328, higher than 3DCNN and others, indicating that the combination of global weighted LPQ and image information quantity features can sensitively capture the abrupt complexity changes of scenes when attention is distracted, achieving high-precision recognition. Traditional methods rely on single features and lack sufficient description of group interaction and scene environment, resulting in low classification accuracy for active interaction and distraction behaviors. Deep learning methods, although having advantages in spatiotemporal features, are slightly inferior in the classification of passive reception behaviors due to insufficient integration of knowledge such as color and environment normalization in educational scenes. The proposed algorithm integrates five-dimensional features to form a “micro-meso-macro” three-dimensional system, which not only utilizes the capabilities of deep learning but also incorporates educational prior knowledge. Excellent performance is achieved in all three behavior classifications, verifying the scientificity of multidimensional feature extraction.

Table 2. Performance comparison on SCBDataset5 image dataset

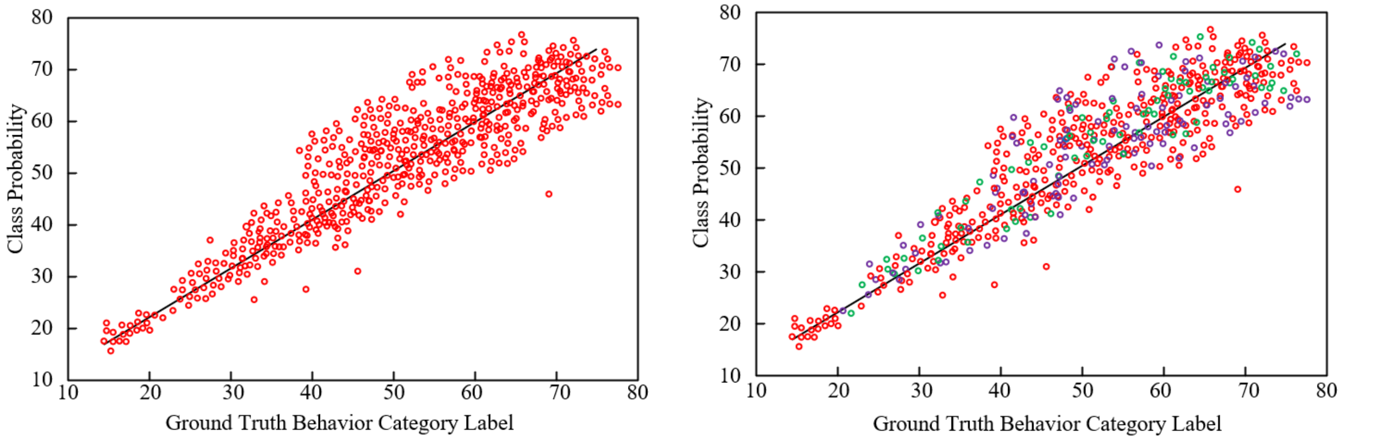
Algorithm	Active Interaction Behavior	Passive Reception Behavior	Abnormal Distraction Behavior
HOG+SVM	0.7574	0.5546	0.2434
LBP + Random Forest	0.6678	0.4789	0.6895
Optical Flow + Hidden Markov Model	0.8237	0.6392	0.8234
STIP + Bag of Words	0.8879	0.7238	0.9218
CNN Classification	0.8467	0.6478	0.8594
3DCNN	0.9378	0.7319	0.9183
Two-stream Network	0.9237	0.7349	0.9128
Spatial Transformer	0.8127	0.6129	0.8237
CBAM	0.7896	0.5898	0.8222
LSTM-CNN	0.9238	0.7392	0.9128
GRU-CNN	0.9232	0.7568	0.9289
Temporal Segment Network	0.8213	0.6128	0.8328
Nonlocal Network	0.9258	0.7185	0.9238
Proposed Algorithm	0.9238	0.7793	0.9328

Table 3. Performance comparison on EduAction dataset image dataset

Algorithm	Active Interaction Behavior	Passive Reception Behavior	Abnormal Distraction Behavior
HOG+SVM	0.5574	0.3846	0.5734
LBP + Random Forest	0.8678	0.6489	0.8895
Optical Flow + Hidden Markov Model	0.8237	0.6392	0.8234
STIP + Bag of Words	0.9279	0.7238	0.9218
CNN Classification	0.3367	0.2178	0.3294
3DCNN	0.9378	0.7319	0.9183
Two-stream Network	0.9137	0.7349	0.9128
Spatial Transformer	0.5227	0.3429	0.5237
CBAM	0.5196	0.3498	0.5222
LSTM-CNN	0.5138	0.3392	0.5128
GRU-CNN	0.9112	0.7668	0.9289
Temporal Segment Network	0.8113	0.6228	0.8228
Nonlocal Network	0.9118	0.7345	0.9218
Proposed Algorithm	0.9438	0.7793	0.9328

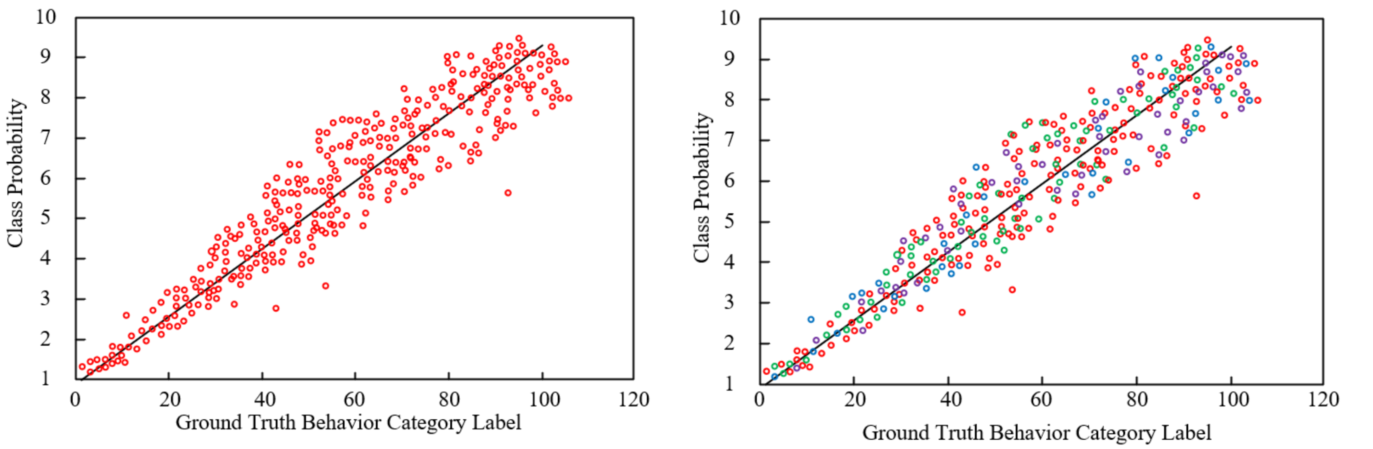
On the EduActionDataset image dataset shown in Table 3, the proposed algorithm demonstrates significant advantages in classification performance. The classification accuracy for active interaction behavior reaches 0.9438, which is 0.6% higher than 3DCNN and surpasses methods such as STIP+Bag of Words, indicating that multidimensional features provide more refined depiction of group dynamic interaction scenes. In the classification of passive reception behavior, the proposed algorithm leads all comparison models with an accuracy of 0.7793, benefiting from the synergistic effect of scene features and color features, effectively distinguishing "focused listening" from "low-interaction distraction," solving the problem of insufficient semantic understanding of static behaviors in traditional methods. The classification accuracy for abnormal distraction behavior is 0.9328, higher than 3DCNN and Nonlocal Network, verifying the high sensitivity of global weighted LPQ and image information quantity features to attention-diverted scenes. The combination of low entropy values, concentrated LPQ patterns, and abnormal scene features precisely identifies attention deviation, reflecting the targeted nature of feature design. Compared with traditional methods, the proposed algorithm integrates multidimensional features, modeling individual actions, group interaction, and environmental context uniformly, significantly improving adaptability to complex educational scenes. Compared with deep learning models, the proposed algorithm incorporates educational prior knowledge in feature design, avoiding semantic bias of purely data-driven models. In passive reception behaviors, traditional CNN performs poorly due to ignoring the educational semantics of scene and color, while the proposed algorithm filters environmental interference through scene features and captures the rationality of visual stimuli using color features, achieving accurate distinction between "effective low interaction" and "ineffective low interaction," verifying the scientificity and educational scene adaptability of the feature system.

Figures 4 and 5 visually present the classification performance of the proposed method on SCBDataset5 and EduActionDataset. In the left plot, class probabilities show a strong linear positive correlation with real behavior labels, and red scatter points are densely distributed near the fitted line, indicating high consistency between model predictions and actual labels. Taking SCBDataset5 as an example, the predicted probability of active interaction behavior increases steadily with the real labels, showing no obvious dispersion, verifying the accurate depiction of dynamic interaction scenes by multidimensional features. For passive reception behavior, the slope of the fitted line is close to 1, reflecting that the collaborative effect of scene and color features effectively eliminates environmental interference, and the classification error in static low-interaction scenes is minimal. In the right plot, multi-colored scatter points still closely surround the line, showing strong robustness of the method against distortions such as illumination and compression. The phase quantization of global weighted LPQ and the complexity quantification of global entropy ensure that attention distraction recognition is not affected by image quality, ensuring stable performance across distortion scenarios. This stability stems from the complementarity of multidimensional features: local gradient is noise-resistant, scene GGD normalizes the environment, H channel captures semantics, and global entropy quantifies complexity—together forming a distortion "immunity" mechanism. For example, in SCBDataset5, active interaction behavior under high distortion still fits well, proving the robustness of LPQ phase patterns against grayscale variations. When combined with gradient weighting to enhance edges, it can accurately recognize interaction patterns. The visualization results verify the reliability of the method in complex educational scenes and provide technical support for real-time classroom analysis.



(a) Overall performance (b) Performance under different distortion levels

Figure 4. Fitted scatter plot of algorithm performance on SCBDataset5



(a) Overall performance (b) Performance under different distortion levels

Figure 5. Fitted scatter plot of algorithm performance on EduAction dataset

Table 4. AUC values of different algorithms for classifying different behaviors

	Active Interaction Behavior		Passive Reception Behavior		Abnormal Distraction Behavior	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Without Local Structure Feature Extraction	0.8234	0.0489	0.8374	0.0427	0.8234	0.0223
Without Global Weighted LPQ Index Structure Feature Extraction	0.9178	0.0267	0.8790	0.0429	0.8879	0.0225
Without Scene Feature Extraction	0.8145	0.0519	0.8094	0.0337	0.8094	0.0283
Without Color Feature Extraction	0.8897	0.0287	0.8796	0.0348	0.8876	0.0224
Without Image Information Quantity Feature Extraction	0.9228	0.0297	0.8830	0.0425	0.8914	0.0255
Complete Algorithm	0.9378	0.0145	0.9237	0.0123	0.9128	0.0189

Table 4 clearly shows through feature ablation experiments the key roles of each feature dimension in classification performance. The complete algorithm achieves an AUC mean of 0.9378 for active interaction behavior, which is an 11.44% increase compared to removing local structure features, indicating that local structure features are crucial for capturing individual action details and distinguishing dynamic interaction scenes. After removing the global weighted LPQ features, the AUC for active interaction behavior drops to 0.9178, a 2.00% decrease, verifying that group spatial structure features provide complementary descriptions of interaction patterns, i.e., single local features cannot fully depict the global layout of multi-subject interactions. For passive reception behavior, the complete algorithm achieves an AUC mean of 0.9237. After removing scene features, it sharply drops to 0.8094, a decrease of 11.43%, highlighting the normalization effect of scene features on environmental noise. Uneven classroom lighting and messy seat arrangements, etc., are filtered through scene features, allowing precise differentiation between “focused listening” and “mind-wandering” in low-interaction behavior. The ablation of color features leads to a 3.41% drop in AUC for passive reception behavior, reflecting that the color semantics of teaching materials directly affect student attention and serve as a core visual cue for behavior classification in static scenes. The ablation experiments quantitatively verify the effectiveness of the multidimensional feature system: each feature dimension is indispensable, and together they achieve comprehensive visual element analysis and educational semantic embedding of student learning behavior. The experimental data show that the complete algorithm significantly outperforms single-dimension ablation methods in AUC values, providing a solid technical foundation for accurate classification and educational intervention, and highlighting the core value of multidimensional feature extraction in educational scenarios.

5. CONCLUSION

This paper, focusing on student group learning behavior classification, constructs a multidimensional feature system including local structure, global weighted LPQ, scene, color, and image information quantity, realizing a full visual element analysis of learning behavior. Experimental results show that the method outperforms comparison algorithms in classifying active interaction, passive reception, and abnormal distraction behaviors across two datasets, verifying the complementarity of features and the adaptability to educational scenarios. Through fitted scatter plots and ablation experiments, it is further demonstrated that the method has strong robustness against distortion scenes and irreplaceability of feature dimensions: scene features filter environmental noise, color features capture visual semantics, and global entropy quantifies complexity—together enhancing the educational semantic interpretability of classification. Based on the classification results, stratified intervention strategies are formed, deeply integrating behavior analysis with educational practice. Data-driven support for personalized teaching is provided, forming a closed-loop of “recognition–intervention–evaluation,” significantly enhancing the scientificity of classroom management and providing a systematic technical framework for intelligent education.

Despite these breakthroughs, the paper still has limitations such as high computational complexity, limited coverage of behavior categories, and reliance on manual interpretation for intervention deployment. Future research can be deepened from three aspects: (1) Lightweight optimization: adopt depthwise separable convolution and attention mechanisms to compress the model, and combine edge-cloud collaboration architecture to improve real-time inference performance to meet millisecond-level classroom response requirements. (2) Multimodal fusion: integrate audio and physiological signals to build a “visual–auditory–physiological” multidimensional model, enhance the comprehensiveness of behavior description, and expand recognition capabilities for complex behavior states. (3) Adaptive intervention: introduce reinforcement learning to dynamically adjust strategies, develop visualization interfaces and intelligent recommendation systems, reduce the threshold for technical application, and promote the evolution of intervention from rule-driven to intelligent adaptation. In addition, improving cross-scenario generalization ability will enhance the universality of the method and provide continuous innovative technical support for the precise improvement of education quality. These directions continue the advantages of the proposed feature system, deepen the integration of image analysis and educational intervention, and promote the large-scale application of intelligent education.

ACKNOWLEDGMENT

This research was funded by the Humanities and Social Sciences Research Planning Project of Heilongjiang Province (Grant No.: 23SHD138); Education Science Planning Project of Heilongjiang Province (Grant No.: ZJB1421222); Heilongjiang Nursing College Level Project (Grant No.: 202402007).

REFERENCES

[1] Dawson, M.R., Lignugaris/Kraft, B. (2017). Meaningful practice: Generalizing foundation teaching skills from TLE TeachLivE™ to the classroom. Teacher Education and Special Education, 40(1): 26-50. https://doi.org/10.1177/0888406416664184

[2] Zamarro, G., Engberg, J., Saavedra, J.E., Steele, J. (2015). Disentangling disadvantage: Can we distinguish good teaching from classroom composition? Journal of research on educational effectiveness, 8(1): 84-111. https://doi.org/10.1080/19345747.2014.972601

[3] Nguyen, N.Q., Lee, K.W., Szabo, C.Z., Nguyen, D.N.P. (2021). Implementing the flipped classroom model in teaching a translation module in Vietnam. Translation & Interpreting: The International Journal of Translation and Interpreting Research, 13(2): 109-135. https://doi.org/10.12807/ti.113202.2021.a07

[4] Turner, S., Cole, L.G. (2017). Using high-fidelity simulation scenarios in the classroom to engage learners. Creative Nursing, 23(1): 35-41. https://doi.org/10.1891/1078-4535.23.1.35

[5] Ghafurian, M., Ellard, C., Dautenhahn, K. (2023). An investigation into the use of smart home devices, user preferences, and impact during COVID-19. Computers in Human Behavior Reports, 11: 100300. https://doi.org/10.1016/j.chbr.2023.100300

[6] Barradah, R. K., Mohamed, E.Y., Sami, W., Alsahly, R.J., et al. (2020). Medical students’ perceptions towards smart devices and their relationship with academic performance. Journal of Research in Medical and Dental Science, 8(4): 83-88.

[7] Onishi, R., Tone, H., Kubota, M., Chino, N., Maruyama, F. (2023). Associating parental efficacy with the utility of smart devices: A cross-sectional study of their role in alleviating maternal parenting concerns among infants aged 6–11 months. Children, 10(9): 1437. https://doi.org/10.3390/children10091437

[8] Som, A., Kim, S., Lopez-Prado, B., Dhamija, S., Alozie, N., Tamrakar, A. (2021). Automated student group collaboration assessment and recommendation system

using individual role and behavioral cues. Frontiers in Computer Science, 3: 728801. https://doi.org/10.3389/fcomp.2021.728801
[9] Liu, X., Liu, J., Demmans Epp, C., Cui, Y. (2025). Exploring the effect of parental involvement on student engagement and academic performance using process data from learning management system. Educational Technology Research and Development, 73(2): 1071-1092. https://doi.org/10.1007/s11423-024-10440-3
[10] Abouelenein, Y.A.M., Selim, S.A.S., Aldosemani, T.I. (2025). Impact of an adaptive environment based on learning analytics on pre-service science teacher behavior and self-regulation. Smart Learning Environments, 12(1): 8. https://doi.org/10.1186/s40561-024-00340-7
[11] Wang, J., Fan, Y., Li, N. (2017). Combining fine texture and coarse color features for color texture classification. Journal of Electronic Imaging, 26(6): 063027. https://doi.org/10.1117/1.JEI.26.6.063027
[12] Kayhan, N., Fekri-Ershad, S. (2021). Content based image retrieval based on weighted fusion of texture and color features derived from modified local binary patterns and local neighborhood difference patterns. Multimedia Tools and Applications, 80(21): 32763-32790. https://doi.org/10.1007/s11042-021-11217-z
[13] Kavitha, J.C., Suruliandi, A. (2018). Feature extraction using dominant local texture-color patterns (DLTCP) and classification of color images. Journal of Medical Systems, 42(11): 220. https://doi.org/10.1007/s10916-018-1067-6
[14] Hosny, K.M., Magdy, T., Lashin, N.A. (2021). Improved color texture recognition using multi-channel orthogonal moments and local binary pattern. Multimedia Tools and Applications, 80(9): 13179-13194. https://doi.org/10.1007/s11042-020-10444-0
[15] Tournier, C., Forde, C.G. (2024). Food oral processing and eating behavior from infancy to childhood: Evidence on the role of food texture in the development of healthy eating behavior. Critical Reviews in Food Science and Nutrition, 64(26): 9554-9567. https://doi.org/10.1080/10408398.2023.2214227
[16] Jisi, A., Yin, S. (2021). A new feature fusion network for student behavior recognition in education. Journal of Applied Science and Engineering, 24(2): 133-140. https://doi.org/10.6180/jase.202104_24(2).0002
[17] Ben-Younes, H., Zablocki, É., Pérez, P., Cord, M. (2022). Driving behavior explanation with multi-level fusion. Pattern Recognition, 123: 108421. https://doi.org/10.1016/j.patcog.2021.108421
[18] Kan, L., Wang, M. (2025). behavior anomaly detection based on multi-modal feature fusion and its application in English teaching. Journal of Applied Science and Engineering, 28(9): 1657-1666. https://doi.org/10.6180/jase.202509_28(9).0002
[19] Nguyen, V.A., Kong, S.G. (2023). Multimodal feature fusion for illumination-invariant recognition of abnormal human behaviors. Information Fusion, 100: 101949. https://doi.org/10.1016/j.inffus.2023.101949
[20] Hao, Z., Li, Z., Dang, X., Ma, Z., Liu, G. (2022). MM-LMF: A low-rank multimodal fusion dangerous driving behavior recognition method based on FMCW signals. Electronics, 11(22): 3800. https://doi.org/10.3390/electronics11223800