Traitement du Signal
Vol. 41, No. 5, October, 2024, pp. 2287-2296
Journal homepage: http://iieta.org/journals/ts

Intelligent Analysis and Optimization of Adaptability in Interdisciplinary Learning Environments Using Image Recognition Technology

Yinling Wang[1*], Lei Yu[1,2], Fang Wang[3], Haining Gao[1], Ali Mazhar[2], Ali Khan Imran[2]

[1] Henan International Joint Laboratory of Internet of Things for New Energy, Huanghuai University, Zhumadian 463000, China
[2] Department of Computer Science, COMSATS University Islamabad, Abbottabad 22060, Pakistan
[3] School of Smart Energy and Environment, Zhongyuan University of Technology, Zhengzhou 450007, China

Corresponding Author Email: wangyinling@huanghuai.edu.cn

https://doi.org/10.18280/ts.410506

Received: 6 April 2024
Revised: 5 September 2024
Accepted: 26 September 2024
Available online: 31 October 2024

Keywords:
interdisciplinary learning environment, adaptability assessment, image recognition technology, intelligent analysis, behavior detection algorithm

ABSTRACT

Driven by global educational reforms, interdisciplinary learning has become a key approach to cultivating well-rounded, innovative talent. However, effectively assessing students' adaptability in interdisciplinary learning environments remains a significant challenge in educational research. With the rapid development of image recognition technology, behavior-based intelligent analysis offers new opportunities for adaptability assessment by capturing students' behavioral performance in real-time and dynamically. Traditional approaches, such as surveys and interviews, are limited by subjectivity and inefficiency, making them insufficient for the precise, real-time analysis required in interdisciplinary settings. This study defines the key behavioral indicators of adaptability in interdisciplinary learning environments, develops an algorithm for detecting student behavior, and evaluates adaptability based on the detected results. An intelligent evaluation system is constructed to provide educators with objective data support, thereby enhancing teaching effectiveness and optimizing interdisciplinary learning environments.

1. INTRODUCTION

In the context of global educational reform, interdisciplinary learning has gradually become an important approach to cultivating innovative talents [1-4]. Interdisciplinary learning environments break the boundaries of traditional disciplines, requiring students to possess abilities such as integrating diverse knowledge, collaborating with others, and engaging in self-directed exploration [5-8]. However, how to assess students' adaptability in such complex environments remains a significant challenge in educational research. With the development of artificial intelligence and computer vision technologies, the use of image recognition technology to analyze students' behavior in interdisciplinary learning environments provides new technical means for adaptability assessment [9-11]. This approach not only helps educators monitor students' learning status in real-time but also provides data support for improving teaching effectiveness.

Research on adaptability in interdisciplinary learning environments holds important educational significance. First, understanding students' adaptability can help educators identify their strengths and weaknesses in the learning process, allowing them to adjust teaching strategies to promote students' overall development. Second, studying adaptability in interdisciplinary environments can support personalized education and precise teaching, helping students overcome challenges in interdisciplinary learning and enhancing their self-directed learning ability and innovative mindset [12-16]. Therefore, constructing an intelligent adaptability analysis system plays an essential role in optimizing interdisciplinary learning environments and improving students' learning outcomes.

Although some research has been conducted on adaptability in interdisciplinary learning, most of it relies on traditional methods such as questionnaires and interviews. These methods are time-consuming and labor-intensive, and the results are susceptible to bias from subjective factors [17, 18]. Moreover, existing applications of image recognition technology mostly focus on limited areas of behavior recognition, lacking a comprehensive analysis of multidimensional behavioral characteristics in complex learning environments [19-24]. Thus, current methods cannot fully meet the needs of adaptability assessment in interdisciplinary environments, making it difficult to achieve real-time, dynamic, and precise analysis of students' behavior.

This study focuses on three main aspects. First, based on the characteristics of interdisciplinary learning environments, it defines key behaviors that reflect students' adaptability and constructs a clear system of behavioral indicators. Second, it develops an adaptability behavior detection algorithm to identify specific behavioral characteristics through image recognition technology. Finally, it conducts a comprehensive evaluation of students' adaptability in interdisciplinary environments based on the behavior detection results. This research provides essential data support for intelligent analysis and optimization in interdisciplinary learning environments, aiming to offer educators scientific, real-time methods for assessing students' adaptability to support personalized teaching and enhance learning outcomes.

2. DEFINITION AND VECTORIZATION OF KEY BEHAVIORS FOR ADAPTABILITY IN INTERDISCIPLINARY LEARNING ENVIRONMENTS

With the increasing adoption of interdisciplinary teaching methods in modern education, students must integrate and apply knowledge from multiple fields, placing new demands on their adaptability in learning behaviors. Unlike traditional disciplines, learning scenarios in interdisciplinary environments are often more complex, requiring students to employ various cognitive skills and emotional regulation abilities. In such environments, image recognition technology can capture details such as body language, facial expressions, and attention levels, enabling the analysis of students' adaptive behaviors and revealing their emotional fluctuations and engagement states in interdisciplinary settings. This provides educators with real-time, intuitive data to assess students' adaptability and adjust teaching strategies promptly.

The key behaviors that indicate students' adaptability in interdisciplinary learning environments, which can be recognized through image technology, include the following: (1) Active Participation: Interdisciplinary learning environments require students to demonstrate active participation. Image recognition technology can identify students' positions, interaction frequency, and interaction styles during group activities. For example, highly adaptive students typically exhibit positive body language in group discussions, such as leaning toward the group center, frequent eye contact, and hand gestures. These behaviors can be captured by cameras and analyzed using behavioral analysis algorithms to determine whether a student is actively engaged in group discussions. (2) Hands-on Experimentation and Field Investigation: Many interdisciplinary learning activities involve hands-on experiments and fieldwork, requiring significant manual operation and object handling. Image recognition technology can evaluate students' engagement by capturing hand movements and operational frequency. For instance, hand-tracking technology can identify whether students are actively participating in experimental tasks or merely observing. Metrics such as operation frequency, precision of hand movements, and participation time help assess students' hands-on ability and enthusiasm. (3) Self-directed Learning and Time Management: Students who adapt well to interdisciplinary environments often possess strong self-directed learning abilities and effective time management skills. These behaviors can be evaluated by monitoring students' actions during independent study sessions. Image recognition technology, combined with facial recognition and emotion detection, can observe students' focus and emotional changes while studying independently. Facial features and expressions reflect concentration and emotional states, providing data to determine whether students have good self-directed learning habits. (4) Cross-cultural Communication and Expression: Interdisciplinary learning requires students to have cross-cultural communication and expression abilities. These can be assessed by observing students' performance during cross-cultural exchanges. Image technology can capture students' body language, facial expressions, and interaction patterns when communicating with peers from different cultural backgrounds. Analyzing these behaviors helps determine whether students can engage comfortably in cross-cultural communication and whether they understand and respect cultural differences.

To facilitate the capture and analysis of students' spatiotemporal behaviors in interdisciplinary environments, this study vectorizes the collected image data, transforming continuous behavioral changes into vectors suitable for input into network models. This process can be likened to extracting specific spatial and temporal information from a dynamic scene and converting it into a series of feature vectors. These feature vectors, combined with position vectors, serve as model inputs to identify students' adaptability changes. See Figure 1 for details.

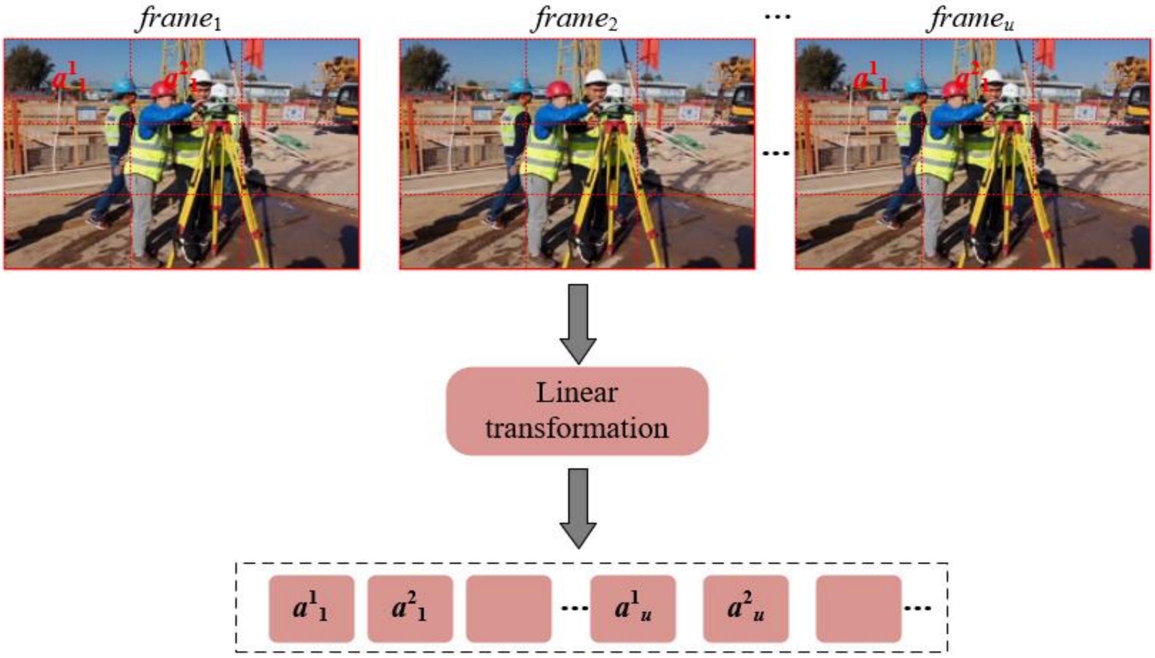


Figure 1. Vectorization of image data in interdisciplinary learning

In interdisciplinary learning environments, students' behavioral adaptability exhibits varying characteristics over time and across different contexts. This change can be captured through a series of sequentially collected image frames. Let the input collected image segments be denoted as $A \in R^{S\times G\times Q\times Z}$, where S represents the number of frames, G and Q represent the height and width of each frame, respectively, and Z denotes the number of color channels. Each frame can be viewed as a snapshot of a state, and by combining these frames, a sequence of behavioral changes at different time points can be obtained. Therefore, to effectively input the collected image information into the model, the images need to be divided into smaller segments, with each segment representing a refined behavioral characteristic.

In each frame, the image is divided into $T\times T$ sized blocks, resulting in V image blocks. These blocks represent the behavioral details of the student at different locations within that frame, and each block is flattened into a one-dimensional vector $a^t{}_s$, where t denotes the spatial position of the image block and s indicates the temporal position of the frame. The purpose of this segmentation is to further abstract the local behavioral features of the student and combine them to present subtle changes in both space and time. The length of the flattened vector is $M=Z\times O\times O$, namely the number of pixels per block. This method of segmentation and flattening allows for the extraction of local behavioral information while preserving the spatiotemporal structure of the collected image frames.

After obtaining the block vectors, each block vector undergoes a linear transformation to adapt to the input requirements of the network. Additionally, to help the model better understand the positional attributes of these vectors, spatiotemporal position vectors are introduced. The purpose of the spatiotemporal position vectors is to add positional information to each image block, allowing the model to differentiate between different frames and their spatial locations. Specifically, two position vectors r_t and r_s can be defined using a separable spatiotemporal approach, where r_t is a learnable spatial position vector representing the two-dimensional position of each image block, while r_s is a temporal position vector indicating the position of each frame on the time axis. By summing the position vectors and image block vectors, a feature vector with positional information is obtained, ultimately generating the input vector for the model:

$$\hat{a}_s^t = Ra_s^t + r_t + r_s \tag{1}$$

3. INTERDISCIPLINARY LEARNING ENVIRONMENT ADAPTABILITY BEHAVIOR DETECTION ALGORITHM

3.1 Encoder

Figure 2 illustrates the principles of the adaptability behavior detection algorithm for interdisciplinary learning environments. In the algorithm for detecting students' adaptability behaviors in interdisciplinary settings, a multi-scale separable spatiotemporal feature encoder has been designed for more efficient and accurate capture of students' spatiotemporal behavioral characteristics. The core of this encoder lies in the independent processing of temporal and spatial dimension information within student behaviors, thereby reducing the computational burden associated with global self-attention. Students' behavioral adaptability typically encompasses variations in bodily actions and trends in continuous behaviors, namely subtle spatial positioning information and temporal sequence information. In the multi-scale separable spatiotemporal feature encoder, the attention mechanisms for space and time are separated, implemented through two submodules: Temporal Pooling Attention (TPA) and Spatial Pooling Attention (SPA).

In the separable spatiotemporal pooling attention module, the input data $a^\wedge$ is a sequence of vectors with M dimensions of F, specifically representing the encoded features of a sequence of S frames of collected image segments at a spatial resolution of $G\times Q$. This module is divided into the SPA and TPA submodules, which are used to extract spatial and temporal features from interdisciplinary behavioral data to reveal potential spatiotemporal adaptability changes in student behavior.

The function of the SPA submodule is to compute the attention weights between various spatial positions within the same time frame, thereby extracting multi-scale spatial features within each frame. As shown in Figure 3, the SPA processes all input vector sequences within each frame, obtaining the corresponding query, key, and value matrices through linear transformations. Unlike traditional self-attention mechanisms, SPA reduces computational complexity and extracts multi-scale features by applying pooling operations to downsample the query, key, and value matrices. This pooling process reduces the spatial dimensions, allowing for multi-scale integration of feature information from each frame, enabling the model to focus on more prominent local features. The pooling results are then used to calculate attention through dot products, producing a pooled spatial attention matrix that captures the relationships among spatial blocks. The pooled spatial attention matrix effectively represents the multi-scale spatial structural information within the frame, forming multi-scale spatial features for subsequent temporal feature extraction. Assuming layer normalization is denoted by LN, the learnable tensor of dimensions $F\times F$ is represented by Q_{Ws}, Q_{Js}, and Q_{Ns}, the spatial position variable is denoted by r_t. The specific transformation process can be represented as:

$$W_s = Q_{W_s} LN\left(\hat{A}_s + r_t\right) \tag{2}$$

$$J_s = Q_{J_s} LN\left(\hat{A}_s + r_t\right) \tag{3}$$

$$N_s = Q_{N_s} LN\left(\hat{A}_s + r_t\right) \tag{4}$$

Assuming the scale transformation factor is represented by $1/(f)^{1/2}$, the pooling operation can be expressed as:

$$\hat{W}_s = POOL\left(W_s; O_W\right) \tag{5}$$

$$\hat{J}_s = POOL\left(J_s; O_J\right) \tag{6}$$

$$\hat{N}_s = POOL\left(N_s; O_N\right) \tag{7}$$

The attention calculation formula for the pooled results is given by:

$$ATT\left(W_s, J_s, N_s\right) = \text{softmax}\left(\frac{\hat{W}_s \hat{J}_s{}^S}{\sqrt{f}}\right)\hat{N}_s \tag{8}$$

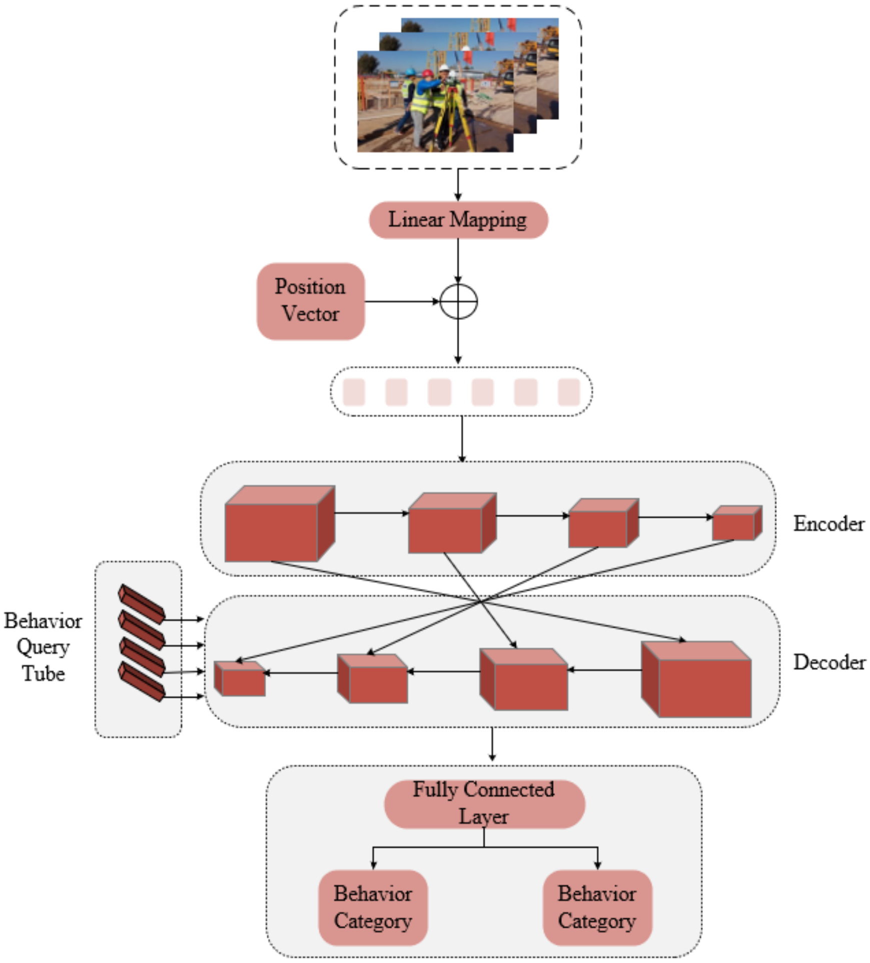


Figure 2. Principle of the adaptive behavior detection algorithm in interdisciplinary learning environments

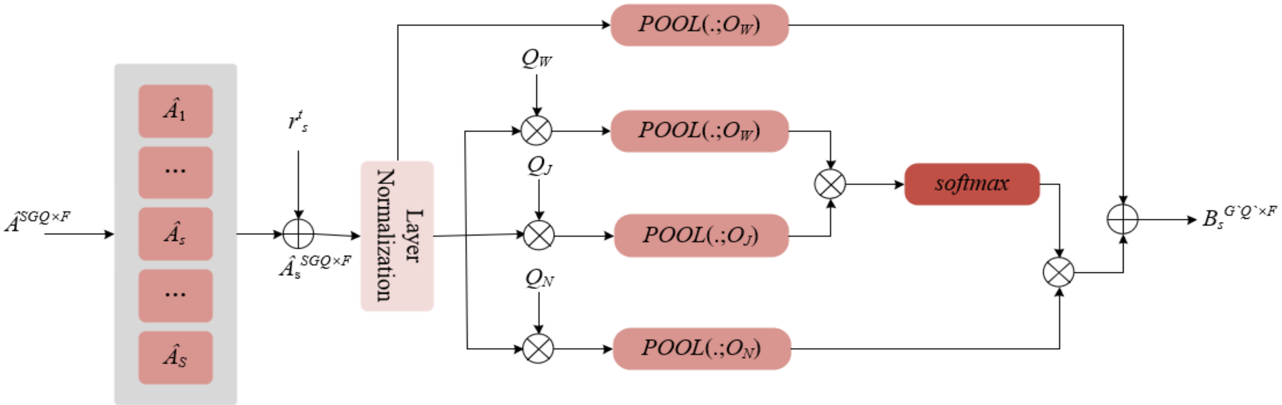


Figure 3. Structure of the SPA module

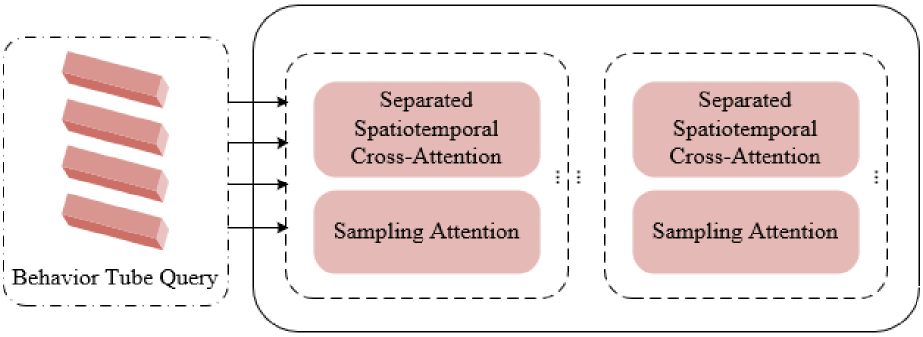

Behavior Tube Query
Separated Spatiotemporal Cross-Attention
Sampling Attention
Separated Spatiotemporal Cross-Attention
Sampling Attention

Figure 4. Decoder structure

The pooling residual connection operation calculation formula is as follows:

$$B_s = ATT\left(W_s, J_s, N_s\right) + POOL\left(LN\left(\hat{A}_s + r_t\right); O_W\right) \tag{9}$$

The role of the TPA is to extract multi-scale behavioral features along the temporal dimension from different frames, detecting changes in student behavior over time. Specifically, TPA takes the output $B \in R^{M' \times F}$ from the SPA module as input and performs temporal pooling operations on the features of the same spatial position in a chronological order. By pooling along the temporal dimension, TPA reduces the dimensionality of the temporal feature representation and captures behavioral changes over long sequences, eliminating subtle noise. As a result, the features extracted by TPA better reflect the ongoing adaptation process of students in interdisciplinary learning environments, identifying behavioral patterns and trends as students adapt to interdisciplinary tasks. Through dot product calculations of the pooled temporal attention, the output $C \in R^{M'' \times F}$ represents the attention feature matrix after temporal pooling, where $M''=S' \times G' \times Q'$, which serves as the output of the final multi-scale spatiotemporal encoder. Assuming $W_t=Q_{Wt}MV(B_t+r_s)$, $J_t=Q_{Jt}MV(B_t+r_s)$, and $N_t=Q_{Nt}MV(B_t+r_s)$, with all vectors at temporal position t represented by B_t and the temporal position vector represented by r_s, the formula is:

$$\begin{aligned} C_t = \mathrm{softmax}\left(\frac{POOL\left(W_T; O_W\right) POOL\left(J_T; O_J\right)^S}{\sqrt{f}} \right) POOL\left(N_T; O_N\right) \\ + POOL\left(LN\left(B_T + r_s\right); O_W\right) \end{aligned} \tag{10}$$

In interdisciplinary learning environments, students' behavioral adaptability often manifests as alternating changes across different time periods and spatial locations. For instance, in an interdisciplinary task combining scientific experimentation with artistic creation, students may initially exhibit fluctuations in attention over short periods, while over time, their behavior evolves into a global adaptability to the entire task process. The multi-scale spatiotemporal feature encoder needs to capture the behavioral patterns of students transitioning from local adaptation to global adaptation through these gradual changes, aiding in the intelligent analysis of their adaptability status. In the initial phase, the encoder focuses on local detail features, with more refined temporal and spatial feature representations, that is, larger feature dimensions S, G, Q, indicating that more detailed information is preserved in both time and space, while the number of feature channels F is smaller, representing weaker expressive ability. This stage primarily captures local changes in student behavior, such as micro-behaviors or short-term emotional fluctuations. As the model progresses, spatiotemporal features gradually become more abstract, with the encoder placing greater emphasis on global behavioral patterns; the feature dimensions S, G, Q decrease, indicating global changes in time and space, while the number of feature channels F increases, providing a stronger representational capability.

The structure of the multi-scale spatiotemporal feature encoder consists of multiple stacked multi-scale spatiotemporal attention modules, each extracting multi-scale features of student behavior through the separable spatiotemporal pooling attention mechanism. Each module contains several separable spatiotemporal pooling attention mechanism modules, each performing attention calculations and pooling operations for the spatial and temporal dimensions separately, gradually extracting student behavioral features across different temporal and spatial scales. Specifically, for the input A_{uk} of the k-th separable spatiotemporal pooling attention mechanism module in the u-th module, assuming spatial pooling attention is represented by SPA, temporal pooling attention by TPA, and multilayer perceptron by MLP, the calculation process for the output C_{uk} is as follows:

$$B_{uk} = TPA\left(SPA\left(A_{uk}\right)\right) \tag{11}$$

$$C_{uk} = MLP\left(LN\left(B_{uk}\right)\right) + B_{uk} \tag{12}$$

3.2 Decoder

In the student adaptability behavior detection algorithm for interdisciplinary learning environments, the introduction of the multi-scale spatiotemporal feature decoder aims to restore more abstract global features to specific spatiotemporal detail features, thereby better analyzing and understanding students' adaptive behaviors. Unlike the encoder, the decoder achieves a multi-layered, fine-grained analysis of students' adaptive behaviors through a stepwise decoding process that combines feature representations at different scales. The structure is shown in Figure 4.

In this multi-scale spatiotemporal feature decoder, a behavior micro-tube query mechanism is introduced, generating a set of dynamic micro-tube queries for each specific adaptive behavior, which are used to track and decode students' adaptation details during interdisciplinary learning processes. For example, in interdisciplinary tasks, students may need to switch between “analysis” and “synthesis” cognitive modes; this behavioral change will be gradually decoded and recorded by specific micro-tube queries, forming a dynamic trajectory that includes the adaptation process, thereby helping the model better understand students' adaptive behavior characteristics.

Here, the behavior micro-tube query $W \in R^{V \times Z}$ is a set of representation vectors used to capture students' adaptive behaviors in interdisciplinary learning contexts. Here, V represents the number of queries, indicating the possible instances of adaptive behaviors to be detected, and Z represents the feature dimension of each query, which includes both spatial and temporal dimension information of student behaviors. Specifically, for each query W_u, it can be understood as a behavior micro-tube that contains the spatial coordinates of that behavior in time and space, allowing for the gradual recording and tracking of students' adaptation processes across different times and subject tasks.

The introduction of a behavior micro-tube spatiotemporal self-attention module in the decoder further enhances the extraction and decoding effect of spatiotemporal features in behavior detection. In the spatial dimension, the attention mechanism focuses on the relationships between different behavior self-labels at the same point in time, capturing the feature interactions among these behavior self-labels to identify the details of students' adaptive behaviors in a particular subject task. In the temporal dimension, the attention mechanism focuses on the feature interactions of the same human target at different time points. This mechanism can track changes in students' behaviors in interdisciplinary contexts, such as the transition from a divergent thinking mode in the literature subject to a concentrated thinking state in mathematical logic; this interdisciplinary adaptive behavior can be captured and interpreted through temporal feature associations. During the feature decoding process, the behavior micro-tube spatiotemporal self-attention module employs upsampling operations instead of pooling operations to elevate the spatiotemporal resolution of micro-tube queries to align with the features of the encoder. Due to the multi-scale feature distribution of adaptive behaviors in different subject environments, directly merging the original low-resolution micro-tube features with encoder features may result in a loss of detail. Through upsampling, micro-tube queries can be accurately aligned with the multi-scale features of the encoder at a higher resolution, ensuring that the details of adaptive behaviors are preserved within the micro-tube queries, thereby providing higher spatiotemporal decoding accuracy.

Given that behavioral features in different subject environments often exhibit different granularities and patterns in both time and space, the algorithm specifically employs consistent granularity attention, or scale-consistent attention, during the decoding process to achieve precise spatiotemporal feature fusion for features at the same scale.

In this module, the input behavior micro-tube features and the spatiotemporal features from the encoder are subjected to spatial cross-attention (SCA) and temporal cross-attention (TCA) calculations at each scale. This separable spatiotemporal attention mechanism allows for feature fusion along the spatial and temporal dimensions, avoiding potential feature entanglement that may occur when processing both dimensions simultaneously. Additionally, to ensure the scale consistency between behavior micro-tube features and encoder spatiotemporal features, the module adheres to a scale alignment strategy during the feature fusion process, satisfying the requirement $u+k=M+1$, where M is the number of different spatiotemporal scales possessed by both the encoder and decoder.

After completing the spatial and temporal cross-attention calculations, the spatiotemporal cross-attention module incorporates fully connected layers and residual connections to further fuse behavior micro-tube features with encoder features, yielding the output $C^u{}_{XS}$ of the spatiotemporal cross-attention module. The calculation process can be expressed as:

$$B_{XS}^{u} = TCA\left(SCA\left(A_{XS}^{u}, A_{rv}^{k}\right), A_{rv}^{k}\right) \tag{13}$$

$$C_{XS}^{u} = MLP\left(LN\left(B_{XS}^{u}\right)\right) + B_{XS}^{u} \tag{14}$$

3.3 Prediction head and loss function

The primary objective of the prediction head is to simultaneously predict behavior categories and spatiotemporal locations. In the student adaptability behavior detection task within interdisciplinary learning environments, this means not only identifying the types of adaptive behaviors exhibited by students in different subject contexts but also accurately localizing these behaviors in both temporal and spatial dimensions. To achieve this, the prediction head employs a strategy similar to spatiotemporal behavior detection, using continuous bounding box coordinates to represent the spatiotemporal locations of behaviors. These bounding boxes correspond to the start and end times of the behaviors in the temporal dimension and to specific regions of behavior features in the spatial dimension. In the prediction head, the prediction of behavior micro-tubes is treated as a set matching problem. The core of this mechanism lies in optimally matching the predicted behavior micro-tubes from the network model with the true behavior micro-tubes input into the model. In practical implementation, the set matching mechanism considers two aspects of loss: first, the classification loss of adaptive behaviors, which is the difference between the predicted behavior category and the actual category; second, the bounding box loss, which measures the error between the predicted spatiotemporal location of behavior micro-tubes and the true location. This comprehensive loss optimization strategy ensures that the prediction head can accurately localize students' adaptive behaviors in both time and space, enhancing the overall generalization ability of the model.

The input to the prediction head is denoted as D_{fr}, which is the output feature from the last layer of the decoder. In the prediction head, these features undergo linear mapping through fully connected layers to produce the predicted values for behavior categories and spatiotemporal locations. Specifically, in the context of adaptive behavior detection in interdisciplinary learning environments, the categories of adaptive behaviors may exhibit significant diversity due to differences in disciplines. For example, cognitive behaviors may be displayed in scientific environments, while social behaviors may manifest in humanities contexts. Through the optimization of cross-entropy loss, the model can accurately capture the variations in adaptive behavior categories within interdisciplinary situations. Assuming V represents the number of behavior micro-tubes and Z the total number of behavior categories, with DZ representing the linear mapping from the fully connected layer, the calculation formulas are given as:

$$\hat{B}_{CL} = FC_{CL}\left(D_{fr}\right) \tag{15}$$

$$M_{CL} = CE\left(\hat{B}_{CL}, B_{CL}\right) \tag{16}$$

The bounding box regression task aims to determine the specific spatiotemporal locations of adaptive behaviors. Unlike traditional object detection tasks that first predict all possible bounding boxes and then determine the final bounding box through methods like non-maximum suppression, this algorithm's prediction head directly outputs the final target bounding box $B^{\wedge}{}_{CO}$ to enhance the accuracy and efficiency of spatiotemporal localization. Let η_{M1} and η_{IO} represent the weights for L_1 loss and IoU loss, respectively. The specific bounding box regression loss consists of a weighted sum of L_1 loss and IoU loss:

$$\hat{B}_{BO} = FC_{BO}\left(D_{fr}\right) \tag{17}$$

$$\begin{aligned} M_{CO} &= \eta_{m_1} M_1\left(B_{CO}, \hat{B}_{CO}\right) \\ &+ \eta_{IO} M_{IO}\left(B_{CO}, \hat{B}_{CO}\right) \end{aligned} \tag{18}$$

In the student interdisciplinary learning environment, the appearance and disappearance of adaptive behaviors may span considerable time periods. For example, when students transition from one subject task to another, their behavior may experience a brief adaptation phase. The behavior switching module can effectively detect this transitional state through the optimization of binary cross-entropy loss:

$$\hat{B}_{xt} = FC_{xt}\left(D_{fr}\right) \tag{19}$$

$$M_{xt} = BCE\left(\hat{B}_{xt}, B_{xt}\right) \tag{20}$$

To comprehensively optimize the classification loss, bounding box regression loss, and behavior switching loss, the algorithm adopts a weighted sum of multi-task losses as the final optimization objective. Assuming the weights of the various loss functions are represented by η_1, η_2, η_3, and η_4, and $B^{\wedge}_{MA}$ and B are a pair of optimal matches, the calculation formula is given by:

$$\begin{aligned} M = {} & \eta_1 M_{CL}\left(\hat{B}^{MA}_{CL}, B_{CL}\right) + \eta_2 M_1\left(\hat{B}^{MA}_{CL}, B_{CL}\right) \\ & + \eta_3 M_{IO}\left(\hat{B}^{MA}_{CO}, B_{CO}\right) + \eta_4 M_{xt}\left(\hat{B}^{MA}_{CO}, B_{xt}\right) \end{aligned} \tag{21}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

As shown in Table 1, different settings of spatiotemporal scales have a significant impact on the detection performance measured by Frame-mAP@0.5 and Video-mAP@0.2. Without spatial or temporal multi-scale settings, the Frame-mAP@0.5 is 28.9, and Video-mAP@0.2 is 25.6. When only spatial multi-scale is introduced, Frame-mAP@0.5 and Video-mAP@0.2 increase to 31.2 and 25.7, respectively. Introducing only temporal multi-scale further improves these metrics to 31.5 and 28.5. The combination of both spatial and temporal multi-scale settings achieves the highest values of 32.6 for Frame-mAP@0.5 and 28.9 for Video-mAP@0.2. This trend indicates that the integrated application of spatial and temporal multi-scales enhances the accuracy and precision of behavior detection. The use of image recognition technology for detecting adaptive behaviors in interdisciplinary learning environments effectively improves the accuracy of student behavior detection and provides detailed support for behavior feature recognition and adaptability assessment. Optimizing the spatiotemporal scales not only enhances detection precision but also offers efficient technical support for the comprehensive evaluation of student adaptive behaviors in interdisciplinary settings.

Table 2 illustrates the influence of different numbers of image frames on Frame-mAP@0.5 and Video-mAP@0.28. As the number of image frames increases, detection performance progressively improves. At 8 frames, Frame-mAP@0.5 and Video-mAP@0.2 are 24.6 and 21.5, respectively. When the number of frames increases to 15, Frame-mAP@0.5 rises to 31.5, with Video-mAP@0.2 also increasing to 25.9. Further increasing the frame count to 31 leads to even higher values of 32.8 and 26.8. At the maximum frame count of 65, Frame-mAP@0.5 and Video-mAP@0.2 reach peak values of 32.9 and 27.5. This demonstrates that increasing the number of image frames significantly enhances the accuracy and comprehensiveness of behavior detection, particularly in video behavior recognition. The image recognition technology employed, combined with varying frame settings, effectively validates the efficacy of intelligent analysis in assessing students' adaptability in interdisciplinary contexts. By increasing the number of image frames, the model can more accurately capture student behavior features and better reflect key adaptive behaviors in interdisciplinary learning, providing technical support for constructing an accurate behavior feature indicator system. The experimental results indicate that a higher number of frames improves the detection model's performance in multidimensional scenarios, thereby providing a solid technical foundation for assessing student adaptability.

Table 3 demonstrates the impact of different resolution settings on Frame-mAP@0.5 and Video-mAP@0.2 detection performance. As the resolution increases, detection performance improves. At the lowest resolution of 105×108, Frame-mAP@0.5 and Video-mAP@0.2 are 31.2 and 25.6, respectively. When the resolution increases to 231×216, Frame-mAP@0.5 rises to 32.5, and Video-mAP@0.2 increases to 26.7. At the highest resolution of 451×435, Frame-mAP@0.5 reaches 32.9, and Video-mAP@0.2 reaches 28.9. This trend indicates that higher resolutions lead to greater accuracy and comprehensiveness in detection, particularly in video analysis. By adjusting the input resolution, this study validates the effectiveness of intelligent analysis of adaptability in interdisciplinary learning environments using image recognition technology. Higher resolutions significantly enhance the ability to capture student behavior features, thereby greatly improving the accuracy and reliability of behavior detection. This improvement allows for more detailed and accurate behavioral data support in constructing adaptability assessment systems for interdisciplinary environments, optimizing the intelligent analysis process of student adaptive behaviors.

This research employs image recognition technology to detect and analyze students' adaptive behaviors in interdisciplinary learning environments, proposing an intelligent assessment method for interdisciplinary adaptability. To evaluate the effectiveness of this method more intuitively, we used a confusion matrix to present detection results for different behavior categories, as shown in Figure 5. The confusion matrix clearly illustrates the model's performance in predicting various adaptive behaviors, with the vertical axis representing true behavior features and the horizontal axis representing predicted behavior features.

The figure shows four types of behavior, among which active participation behavior is manifested in the interdisciplinary learning environment as students actively engage in the learning process, take on responsibilities, and participate in discussions. The behavioral characteristics include asking questions, expressing opinions, actively applying for task division, and providing feedback. These behaviors indicate students' interest in the learning content and their proactive attitude toward seeking knowledge. In addition, active participation is also reflected in students' sense of cooperation, such as participating in group decision-making, encouraging others to engage, and sharing learning outcomes. These characteristics demonstrate that students are not only focused on their individual learning but also dedicated to achieving collective goals, showcasing the team spirit and cooperation skills required for interdisciplinary learning.

Hands-on operation behavior reflects students' adaptability in practical and operational aspects. In the interdisciplinary learning environment, students deepen their understanding and application of knowledge through hands-on practice. For example, operating experimental equipment, assembling objects, constructing models, designing experimental procedures, and conducting field studies are all manifestations of students completing specific tasks through hands-on operation. This type of behavior reflects students' ability to apply theory to practical problems, as well as the problem-solving skills developed through hands-on activities. Through hands-on operation, students can better understand complex interdisciplinary concepts and deepen their mastery of knowledge through practice.

Table 1. Comparison of experimental results with different spatiotemporal scale settings

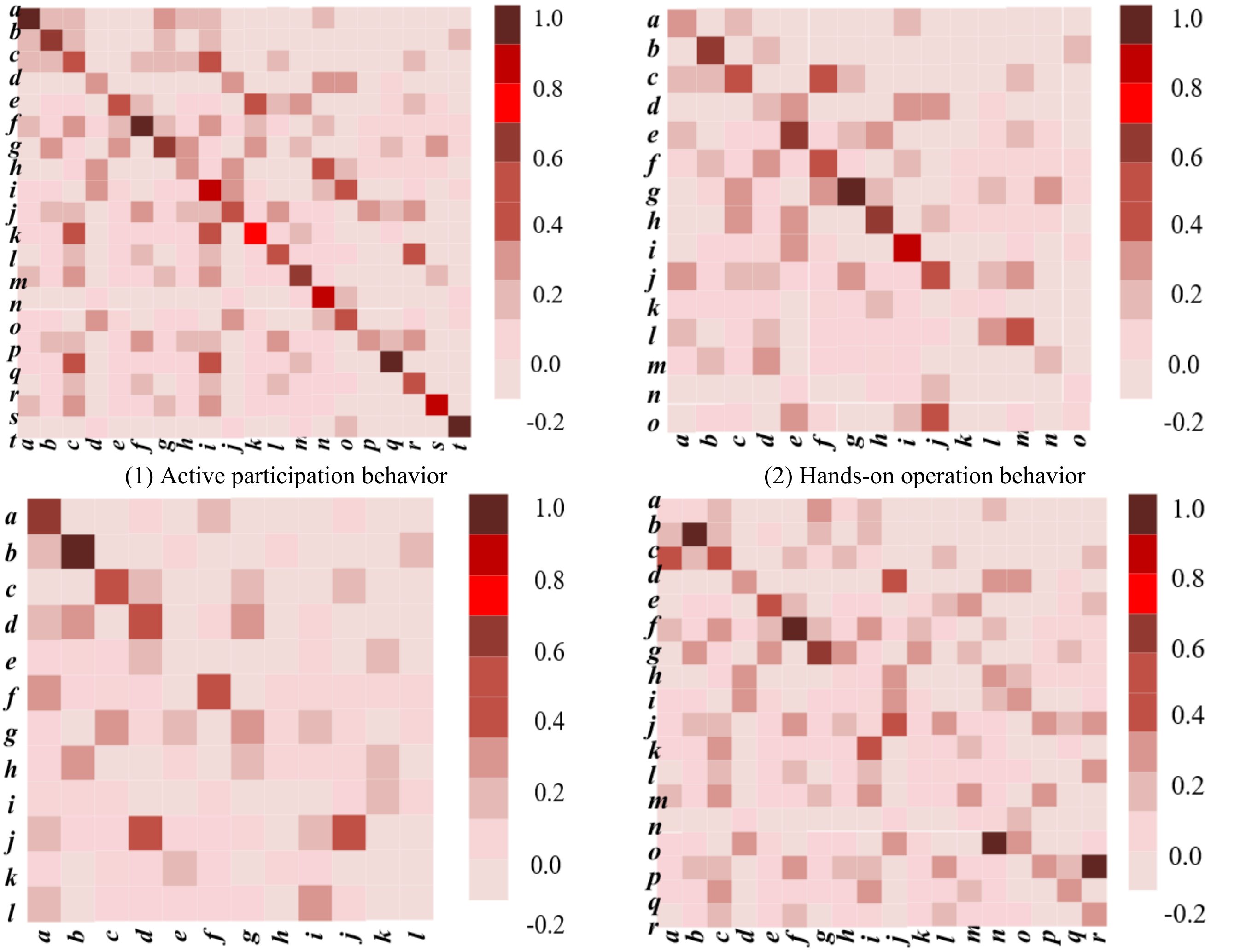
Spatial Multi-Scale	Temporal Multi-Scale	*Frame-mAP@0.5*	*Video-mAP@0.2*
-	-	28.9	25.6
√	-	31.2	25.7
-	√	31.5	28.5
√	√	32.6	28.9

Table 2. Comparison of experimental results with different numbers of input "Interdisciplinary" learning image frames

Number of Image Frames	*Frame-mAP@0.5*	*Video-mAP@0.2*
8	24.6	21.5
15	31.5	25.9
31	32.8	26.8
65	32.9	27.5

Table 3. Comparison of experimental results with different input resolutions

Resolution	*Frame-mAP@0.5*	*Video-mAP@0.2*
105×108	31.2	25.6
231×216	32.5	26.7
451×435	32.9	28.9



(1) Active participation behavior

(2) Hands-on operation behavior

(3) Focused learning behavior

(4) Communication and expression behavior

Figure 5. Confusion matrix of different behavior types

Focused learning behavior emphasizes students' ability to maintain effective concentration during the learning process. This type of behavior includes sustained reading of materials, attentive listening, note-taking, in-depth thinking, and self-reflection. In an interdisciplinary learning environment, students face a large amount of knowledge from different fields, requiring them to concentrate and absorb information in a timely manner. Therefore, focused learning performance is core to students' adaptation to interdisciplinary study. By maintaining focus, students can better understand the connections between subjects and apply their knowledge for interdisciplinary analysis and problem-solving. Additionally, this behavior is also reflected in self-management, such as enhancing learning efficiency through self-questioning, self-testing, and adjusting study states.

Interdisciplinary communication and expression behavior reflects students' ability to clearly and effectively convey their thoughts and understanding in different disciplinary contexts.

Performance characteristics include clearly expressing viewpoints, listening to others, using interdisciplinary terminology, summarizing others' viewpoints, and explaining ideas with diagrams. This type of behavior not only indicates that students possess strong expressive abilities but also demonstrates their capability to translate and apply knowledge across different subjects. Interdisciplinary communication requires students to understand and use various professional languages and forms of expression, collaborate with others, and exchange information to promote collective learning outcomes. These behaviors are crucial for cooperation and innovation in interdisciplinary study, helping to cultivate students' integrated thinking and communication skills.

For active participation behavior, Figure 5 shows a very high degree of match between predictions and actual behaviors, indicating that the model can effectively capture the characteristics of active participation exhibited by students in interdisciplinary environments. This result is crucial for assessing the capabilities of proactive exploration and autonomous learning emphasized in interdisciplinary learning settings. The prediction results for hands-on operation behavior are also satisfactory, demonstrating that image recognition technology can effectively monitor students' practical skills, which are often involved in experimental operations and practical tasks in interdisciplinary projects. The high prediction accuracy for focused learning behavior reflects the model's effectiveness in assessing students' levels of concentration during the learning process; this behavior detection can help teachers understand students' learning states. The excellent prediction results for interdisciplinary communication and expression behavior reflect the model's ability to detect students' interdisciplinary communication skills, which are core requirements for collaboration in interdisciplinary learning.

Based on the above analysis, the image recognition technology proposed in this paper has demonstrated excellent performance in detecting adaptive behaviors in interdisciplinary environments, particularly in the prediction of key behaviors. This method enhances the accuracy of recognizing student behaviors and the depth of assessment in complex interdisciplinary settings through the optimization of behavior detection algorithms and the construction of a behavior feature indicator system. From these results, we can conclude that using image recognition technology for intelligent behavior analysis can not only help teachers gain a more comprehensive understanding of students' adaptability in interdisciplinary learning environments but also provide data support for teaching improvement, contributing to optimized instructional design and enhanced student learning outcomes. This research offers new insights into intelligent evaluation methods in the field of interdisciplinary education and demonstrates the effectiveness and feasibility of this approach in practical applications.

5. CONCLUSION

This study focuses on the intelligent evaluation of students' adaptive behaviors in interdisciplinary learning environments through image recognition technology, providing a novel technical pathway for adaptive analysis in interdisciplinary education. The research consists of three main parts: First, based on the complex characteristics of interdisciplinary learning, key adaptive behaviors required of students—such as active participation, hands-on operation, focused learning, and interdisciplinary communication—are defined. A corresponding behavioral feature indicator system is constructed to provide a standardized theoretical framework for behavior recognition. Second, the study develops an adaptive behavior detection algorithm that utilizes image recognition technology to monitor students' behavioral performance in real time, accurately identifying behavioral features. Experiments with different frame rates and resolutions further optimize the precision of the image recognition model. Finally, based on the results of behavior detection, a comprehensive adaptive assessment system is constructed, providing valuable analytical data and references for personalized support and instructional improvements in interdisciplinary education.

The experimental results indicate that the proposed detection method performs well under various input settings. The experiments with different spatial and temporal scales validate the impact of varying frame rates and resolutions on detection performance, revealing that higher frame rates and resolutions significantly enhance the accuracy of adaptive behavior recognition. Furthermore, the confusion matrix for testing different behavior categories demonstrates that the model exhibits high accuracy and stability in detecting active participation, hands-on operation, focused learning, and interdisciplinary communication behaviors. These results suggest that the proposed method can reliably identify student behaviors in interdisciplinary learning environments, offering targeted feedback to teachers and enhancing the quality and efficiency of interdisciplinary instruction.

The value of this research lies in advancing adaptive assessment in the field of interdisciplinary education through intelligent analytical techniques, addressing the gap in existing studies regarding the intelligent detection and evaluation of student adaptive behaviors. However, certain limitations remain, such as the model's performance showing some confusion among specific behavior categories due to experimental constraints. Additionally, as the content of interdisciplinary education evolves dynamically, the behavioral feature indicator system may need continual updates to meet the demands of different educational contexts. Future research could further expand upon this method by incorporating additional behavioral dimensions into the detection system and optimizing the model's ability to recognize complex behaviors. Specifically, in collaborative learning scenarios, exploring multimodal data fusion—such as integrating voice and text information—could enhance the accuracy and comprehensiveness of the detection system. Moreover, future efforts could focus on developing adaptive indicator systems for dynamically updated behavioral features, supporting more personalized assessment and interventions in interdisciplinary education.

ACKNOWLEDGMENT

This research was funded by the Collaborative Education Project of the Ministry of Education (Grant No.: 230802936290321), Natural Science Foundation in Henan Province (Grant No.: 222300420239), Research and Practice Project on Higher Education Teaching Reform in Henan Province (Grant No.: 2023SJGLX092Y, 2021SJGLX536, 2024SJGLX0172, 2024SJGLX0481). The Postgraduate Education Reform and Quality Improvement Project in Henan

Province (Grant No.: YJS2021JD08, YJS2023KC15), Program for Innovative Research Team in University of Henan Province (Grant No.: 24IRTSTHN020), and the Youth Project of National Scientific Research Project Cultivation Fund in Henan Province (Grant No.: 2023GGJS157).

REFERENCES

[1] Ider, S., Okumuslar, M. (2024). The opinions of theology faculty undergraduates and graduates on interdisciplinary learning. Eskıyeni, 54: 1003-1025. https://doi.org/10.37697/eskiyeni.1465912

[2] Budiman, A., Nopembri, S., Andika, R. (2024). Interdisciplinary physical education: Implementation and insights of Indonesian PE teachers. Retos: Nuevas Tendencias en Educación Física, Deporte y Recreación, 57: 607-615. http://doi.org/10.47197/retos.v57.105910

[3] Huang, X.M., Zhu, P.H., Ma, J. (2023). A transfer learning approach to interdisciplinary document classification with keyword-based explanation. Scientometrics, 128(12): 6449-6469. http://doi.org/10.1007/s11192-023-04825-z

[4] Routhe, H.W., Holgaard, J.E., Kolmos, A. (2023). Experienced learning outcomes for interdisciplinary projects in engineering education. IEEE Transactions on Education, 66(5): 487-499. https://doi.org/10.1109/TE.2023.3284835

[5] Kasch, J., Schutjens, V.A.J.M., Rebel, K.T. (2023). Distance and presence in interdisciplinary online learning: A challenge-based learning course on sustainable cities of the future. Journal of Integrative Environmental Sciences, 20(1): 2185261. https://doi.org/10.1080/1943815X.2023.2185261

[6] Dallmann, A.A. (2021). Reflecting on 50 years: The university without walls and integrative interdisciplinary learning. Journal of Adult and Continuing Education, 27(2): 341-359. http://doi.org/10.1177/14779714211019046

[7] Hu, K., Jin, J.L., Ding, Y.W. (2023). Overview of behavior recognition based on deep learning. Artificial Intelligence Review, 56(3): 1833-1865. https://doi.org/10.1007/s10462-022-10210-8

[8] Liu, B.J., Wang, Z.M. (2024). Research on human behaviour recognition method of sports images based on machine learning. International Journal of Bio-Inspired Computation, 23(2): 99-110. https://doi.org/10.1504/ijbic.2024.136728

[9] Ma, R.J. (2024). An online learning behaviour recognition method based on tag set correlation learning. International Journal of Biometrics, 16(3-4): 350-363. https://doi.org/10.1504/IJBM.2024.138232

[10] Lin, J.X., Li, J.M., Chen, J. (2022). An analysis of English classroom behavior by intelligent image recognition in IoT. International Journal of System Assurance Engineering and Management, 13: 1063-1071. https://doi.org/10.1007/s13198-021-01327-0

[11] Zhang, F.X., Wang, F.F. (2024). Study on abnormal behaviour recognition of MOOC online English learning based on multi-dimensional data mining. International Journal of Continuing Engineering Education and Life-Long Learning, 34(1): 111-122. https://doi.org/10.1504/IJCEELL.2024.135225

[12] Rezaei, F., Yazdi, M. (2021). Real-time crowd behavior recognition in surveillance videos based on deep learning methods. Journal of Real-Time Image Processing, 18(5): 1669-1679. https://doi.org/10.1007/s11554-021-01116-9

[13] Li, T.X., Li, T.K., Han, D. (2023). Classification and recognition of goat movement behavior based on SL-WOA-XGBoost. Electronics, 12(16): 3506. https://doi.org/10.3390/electronics12163506

[14] Lu, W.B., Zhao, Y.Q., Tang, J.X. (2023). MammalClub: An annotated wild mammal dataset for species recognition, individual identification, and behavior recognition. Electronics, 12(21): 4506. https://doi.org/10.3390/electronics12214506

[15] Lu, C.K. (2021). Multifeature fusion human motion behavior recognition algorithm using deep reinforcement learning. Mobile Information Systems, 2021(1): 2199930. https://doi.org/10.1155/2021/2199930

[16] Xu, P.F., Sulaiman, N.A.A., Li, S.P. (2024). A study of falling behavior recognition of the elderly based on deep learning. Signal Image and Video Processing, 18(10): 7383-7394. https://doi.org/10.1007/s11760-024-03401-z

[17] Lewis, S., Burley, K. (2024). Online interdisciplinary work-integrated learning: An undergraduate course review. International Journal of Work-Integrated Learning, 25(3): 417-432. https://files.eric.ed.gov/fulltext/EJ1441978.pdf.

[18] Hains-Wesson, R., Ji, K.Y. (2021). An interdisciplinary, short-term mobility, work-integrated learning experiment: Education for change. Issues in Educational Research, 31(3): 800-815. https://www.proquest.com/docview/2702191026?forcedol=true&pq-origsite=primo.

[19] Zavalevskyi, Y., Khokhlina, O., Chupryna, O. (2023). Project based STEM activities as an effective educational technology in the context of blended learning. Amazonia Investiga, 12(67): 152-161. http://doi.org/10.34069/AI/2023.67.07.14

[20] Schmid, U., Wrede, B. (2022). What is missing in XAI so far? An interdisciplinary perspective. Kunstliche Intelligenz, 36(3-4): 303-315. https://doi.org/10.1007/s13218-022-00786-2

[21] Qattawi, A., Alafaghani, A., Jaman, M.S. (2021). A multidisciplinary engineering capstone design course: A case study for design-based approach. International Journal of Mechanical Engineering Education, 49(3): 223-241. http://doi.org/10.1177/0306419019882622

[22] Van, D.T.H., Khang, N.D., Thi, H.H.Q. (2022). The impacts of fears of COVID-19 on university students' adaptability in online learning. Frontiers in Education, 7: 851422. https://doi.org/10.3389/feduc.2022.851422

[23] Stan, M.M., Topala, I.R., Cazan, A.M. (2022). Predictors of learning engagement in the context of online learning during the COVID-19 Pandemic. Frontiers in Psychology, 13: 867122. https://doi.org/10.3389/fpsyg.2022.867122

[24] Kar, S.P., Das, A.K., Mandal, J.K. (2024). Assessment of learning parameters for students' adaptability in online education using machine learning and explainable AI. Education and Information Technologies, 29(6): 7553-7568. https://doi.org/10.1007/s10639-023-12111-x