ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

A Study on the Design of Business English Education Courses and Learning Effectiveness Enhancement Mechanisms in Fujian's Private Universities Supported by Digital Intelligence Technologies

Yanfang Wang[1] *and Xiaodan Bao*[1,*]

[1] College of Humanities, Xiamen Huaxia University, Xiamen, Fujian, 361024, China

SUMMARY: *The advancement of digital intelligence technologies presents new opportunities for business English education. Their integration not only enriches course content but also introduces novel teaching methods such as online learning platforms and virtual reality. To further enhance teaching effectiveness, this paper designs a deep learning-based student behavior recognition and teaching effectiveness evaluation mechanism (LBREM). This mechanism first employs deep learning algorithms to analyze student behavior, then evaluates teaching effectiveness based on this analysis. Instructors adjust teaching approaches according to the evaluation outcomes, ultimately improving business English learning outcomes. The object detection model (YOLOv5s-Ghost-D4) and human keypoint detection model (AlphaPose) used in this mechanism demonstrate high recognition performance. When applying the LBREM method for learning effectiveness evaluation, the results show minimal deviation from manual evaluations. Based on the digital-intelligent teaching model and LBREM method, student learning outcomes showed significant improvement: business vocabulary size increased by 155.2%, oral fluency by 26.2%, listening comprehension rate by 29%, writing proficiency by 18.7%, and business adaptability by 20.8%.*

KEYWORDS: *LBREM; YOLOv5s-Ghost-D4; AlphaPose; Business English*

1 Introduction

With the rapid advancement of digital and intelligent technologies, the education sector is undergoing profound transformation[1]. As a critical component in cultivating professionals with specialized skills and global perspectives, English education urgently needs to adapt to this era and explore new classroom teaching strategies that integrate digital and intelligent technologies[2, 3]. Confronted with digitalization and intelligence, traditional English teaching models struggle to accommodate these new technologies and meet students' personalized, diverse learning needs[4, 5]. Supported by digital and intelligent technologies, educators can now collect and analyze student learning data—including academic performance, online learning behaviors, and classroom interactions—to enhance the interactivity, personalization, and effectiveness of vocational English instruction. This approach cultivates more high-caliber talents equipped with international competitiveness and cross-cultural communication skills[6, 7]. Business English education in Fujian's private universities has continuously undergone innovation, integration, and development. Digitalization has not only reshaped the teaching landscape but also driven transformative changes in business English instruction.

Against the backdrop of the digital and intelligent era, the abundance and sharing of

*daisychng8233@163.com
https://doi.org/10.65102/is2026645
Published: 30 April, 2026

teaching resources have brought unprecedented prosperity [8]. Pétursdóttir [9] conducted an experimental intervention involving eight teachers, some using digital teaching resources and others not. Through tests and interviews, it was found that the group utilizing digital learning resources achieved slightly better results than the group without such resources. Digital and intelligent technologies can construct diverse virtual teaching scenarios, catering to different instructional contexts for knowledge delivery and stimulating student interest [10]. Wang [11] utilized software like 3DS MAX, MAYA, and VRML to create 3D models and databases of artistic anatomy teaching specimens. By visualizing complex anatomical structures through 3D representations, student engagement increased and teaching quality improved. Tatar[12] conducted a study in high school biology classes, comparing two groups: one using digital cameras and one without. Results indicated that digital cameras helped stimulate student interest. Students using digital cameras in two biology experiments demonstrated greater initiative and better mastery of procedural skills. Gao[13] integrated interactive micro-course technology and digital tools into traditional classrooms to develop a novel digital mapping teaching model. Findings revealed that this approach stimulated student interest, enhanced practical skills, and significantly improved teaching outcomes. He[14] investigated an AI-based computer-assisted teaching system model featuring personalized instruction and collaborative learning. Findings revealed that this system encouraged learners to engage in deep learning, substantially boosted their motivation, and yielded remarkably significant learning outcomes.

The implementation of digital and intelligent teaching empowers students to autonomously select learning content and methods, facilitating more effective transformation of instructional materials and advancing teaching delivery [15]. Chen et al. [16] introduced digital mining technology into mining engineering curricula. By simulating mining environments through 3D digital technology, they helped students grasp the relationship between mining theory and practice while enhancing classroom engagement, demonstrating significant potential for improving teaching effectiveness and efficiency. Pérez-Marín [17] proposed a multimodal digital teaching approach embedded with the VARK (Visual, Auditory, Kinaesthetic, Reading/Writing) model. Human-computer interaction teaching experiments revealed that this method significantly improved both student performance and satisfaction, validating its effectiveness. Yan et al.[18] incorporated immersive virtual reality (IVR) technology to enhance teaching interactivity and immersion. A teaching experiment comparing experimental and control groups revealed positive impacts on students' academic performance, self-regulation abilities, and self-efficacy. Chen, X et al. [19] assessed the impact of teachers' attitudes and behavioral intentions toward iPad smart devices on teaching quality during the pandemic. Findings confirmed that positive attitudes and behaviors toward smart devices positively influence teaching quality. Gong [20] integrated artificial intelligence algorithms with virtual reality technology in digital media art creation instruction, comprehensively enhancing teaching experiences for both educators and students. Case studies demonstrated the pedagogical advantages of combining AI algorithms with VR, highlighting its strong practical value.

In the digital intelligence era, English classroom instruction faces unprecedented transformation [21]. Yang [22] developed a multimodal English teaching framework based on digital information technology, achieving a 42% increase in vocabulary acquisition efficiency and significantly improving students' spoken English proficiency at the 0.01 significance level. In their study, Moorhouse et al.[23] invited secondary English teachers to employ diverse asynchronous and synchronous digital technologies and pedagogical methods during instruction. Online survey and interview findings revealed that this approach enhanced students' English learning and facilitated assessment of learning outcomes. Huang et al.[24] proposed a blended learning platform integrating online and offline instruction based on digital intelligence technology. Experiments conducted in English classrooms at a vocational college demonstrated that the platform effectively optimized teaching models, enhanced interactive feedback between students and teachers, and improved students' English proficiency. Chen, F[25] developed a multi-interactive teaching model using digital information technology. Applied to English classroom teaching, this model improved teaching quality, enhanced course effectiveness, and boosted students' English application abilities.

This paper innovates business English course design based on digital intelligence technology. It proposes a deep learning-based student behavior recognition and teaching effectiveness evaluation mechanism (LBREM). This mechanism first acquires student learning behavior video data through an intelligent learning platform, then detects facial expressions and identifies learning behaviors within the videos, and finally evaluates teaching effectiveness based on the analysis of student learning behaviors. The study conducted teaching effectiveness evaluations using 2024-level Business English majors from a private university in Fujian Province as a case study. Empirical analysis was used to test the effectiveness of the digital-intelligent teaching model and the LBREM method in enhancing teaching outcomes.

2 Innovation in Business English Curriculum Design

Fujian now has 37 private higher education institutions. Marked by the establishment of Fujian South China Women's Vocational College in 1984, the development of private higher education in Fujian entered a period of vigorous growth. With the rapid advancement of digital and intelligent technologies, business English education in higher education faces unprecedented challenges and opportunities. Educators should effectively integrate digital and intelligent technologies into the curriculum design of business English programs to enhance teaching efficiency and student learning outcomes.

2.1 Curriculum Content Design Supported by Digital Intelligence Technologies

In modern business English curriculum design, integrating digital and intelligent technologies has emerged as an innovative trend. This integration not only enhances teaching methodologies but also enriches course content. A key innovation lies in utilizing digital and intelligent technologies to conduct case studies and simulate business environments.

Case studies have long been a vital teaching tool in business English instruction, but the application of digital and intelligent technologies makes these cases more vivid and interactive. Through online platforms, students can access real-world business cases from around the globe and engage in discussions with fellow students or industry experts via video conferencing and online forums. This approach not only enhances students' practical application skills but also deepens their understanding of the global business environment.

Additionally, simulating business environments serves as a key feature in integrating digital and intelligent technologies into business English courses. Through virtual reality or augmented reality technologies, students can immerse themselves in diverse business scenarios such as international conferences and commercial negotiations [26]. This immersive learning experience is highly effective in enhancing students' practical language skills and business acumen.

2.2 Reform of Course Structure and Assessment Methods

The integration of digital and intelligent technologies has also driven reforms in the structure and assessment methods of business English courses. In terms of course structure, more programs are adopting blended learning models that combine online and offline teaching activities. This approach offers greater flexibility, allowing students to learn at their own pace and according to their learning styles while retaining the advantages of face-to-face interaction. Regarding assessment methods, traditional exams and assignments are gradually being replaced by more diverse and dynamic evaluation approaches. For instance, project-based assessments enable students to demonstrate their ability to apply language and business knowledge in real or simulated commercial scenarios. Simultaneously, the application of digital and intelligent technologies facilitates continuous assessment, allowing instructors to track student progress and engagement in real time through learning management systems and provide immediate feedback. The integration of these technologies has unlocked unprecedented opportunities for innovation in business English curriculum design. These innovations not only enhance the appeal and practicality of the courses but also better equip students to navigate the globalized business environment.

3 Research on Mechanisms for Enhancing Business English Learning Outcomes

In classroom teaching, traditional evaluation methods primarily rely on expert classroom observations and assessment forms, requiring the aggregation of multiple observations to evaluate teaching effectiveness and quality. While this approach provides oversight for classroom instruction, the random nature of expert observations inevitably introduces subjective limitations and randomness into evaluation outcomes, making the process time-consuming and labor-intensive. In contrast, leveraging digital and intelligent technologies for instructional assessment enables analysis and evaluation of both the learning process and outcomes, providing a scientific basis for teaching decisions. Smart education, deeply integrating digital and intelligent technologies with learning, analyzes student behavior. This primarily relies on computer vision, deep learning, and other technologies and algorithms to intelligently identify and statistically track students' interactive classroom behaviors captured in videos. By evaluating classroom teaching effectiveness based on the statistics and distribution of these learning behaviors, it effectively addresses the challenge of assessing classroom teaching outcomes while simultaneously driving traditional classroom learning toward informatization and intelligence. Deep learning enables the automatic acquisition of pattern features and integrates feature learning into the model-building process, thereby reducing the incompleteness caused by manually designed features. Currently, certain deep learning-based applications have achieved recognition or classification performance surpassing existing algorithms under specific conditions. To address shortcomings in learning assessment, this study proposes a deep learning-based student behavior recognition and teaching effectiveness evaluation mechanism (LBREM). By analyzing and statistically evaluating student learning behaviors, it establishes a teaching effectiveness evaluation model to assess instructional outcomes, ultimately enhancing the effectiveness of Business English classroom instruction.

3.1 LBREM Operational Process

The design philosophy of LBREM involves first acquiring video data of students' learning behaviors through learning devices and platforms. Subsequently, it detects and identifies students' faces within the videos while simultaneously recognizing their facial expressions and learning behaviors. Finally, it analyzes students' learning behaviors to evaluate teaching effectiveness. Based on this approach, Figure 1 illustrates the operational workflow of LBREM.

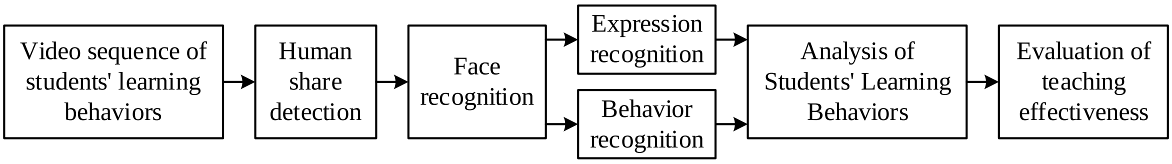


Figure 1: LBREM's operation process

3.2 Deep Learning-Based Student Behavior Analysis System

3.2.1 System Architecture

To ensure the system's future scalability, the architecture design should adhere to principles of modularity, high cohesion, and low coupling, thereby meeting requirements for expandability, stability, and maintainability. The overall system architecture is divided into four layers from bottom to top: data storage layer, data processing layer, business logic layer, and user interface layer. The architecture of the deep learning-based student classroom behavior system is illustrated in Figure 2.

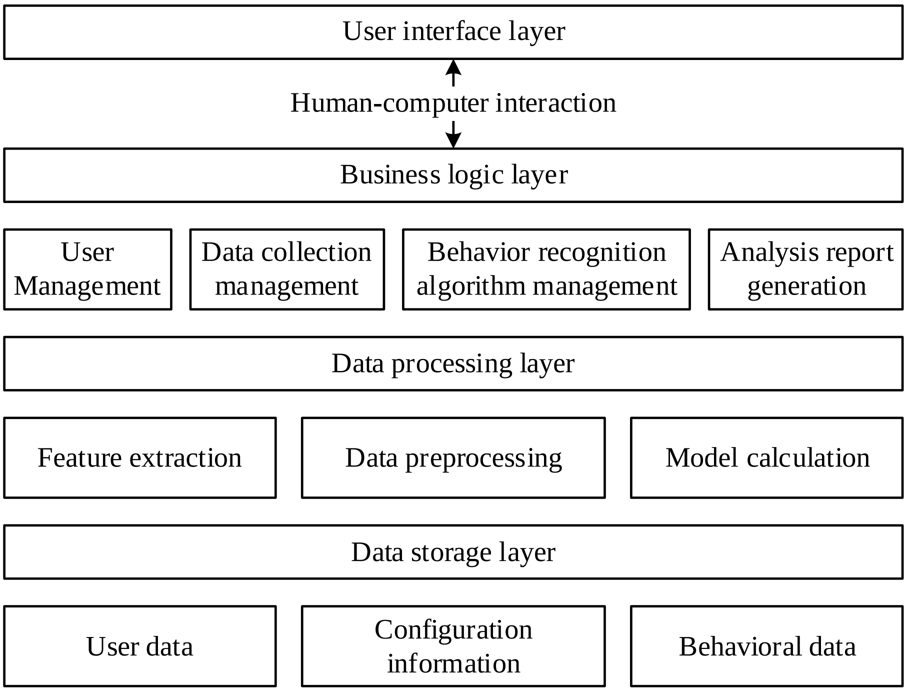


Figure 2: Student classroom behavior system architecture based on deep learning

(1) Data Storage Layer: Select databases (relational, non-relational, and distributed storage systems) based on data structure and access patterns, while ensuring data security, integrity, and efficient access.

(2) Data Processing Layer: Deep learning models perform computations at this layer, analyzing and processing video frames to accurately identify distinct student behaviors.

(3) Business Logic Layer: The user management module handles user registration, login, and permission allocation. The data collection module gathers student classroom behavior data from cameras, learning management systems, and other sources, performing preprocessing on raw data. The behavior recognition algorithm management module invokes and manages deep learning models to identify and analyze collected behaviors. The analysis report module generates corresponding reports based on recognition results and user requests, providing

insights for teachers, students, and administrators.

(4) User Interface Layer: Human-computer interaction interface.

3.2.2 Deep Learning Models

(1) Object Detection Models

Object detection models are currently advancing rapidly, with representative examples including YOLO, SSD, and Faster R-CNN. Compared to YOLOv4, YOLOv5 not only inherits v4's high performance but also achieves significantly faster inference speeds. On a Tesla P100 GPU, it reaches 140 FPS, whereas v4 only achieves 50 FPS [27]. Moreover, YOLOv5 maintains high detection accuracy while delivering fast operation speeds, meeting the real-time requirements for classroom behavior analysis. Considering the balance between real-time performance and accuracy in object detection models, this paper selects YOLOv5 as the target detection model.

To enhance detection performance, this paper modifies the YOLOv5s architecture by adding a feature fusion layer within its feature integration structure. This layer combines low-level features extracted by the backbone network with high-level features, feeding the merged feature map to the prediction layer. This ensures the prediction layer incorporates richer detail information. The modified network outputs four feature maps at distinct scales, hence named YOLOv5s_D4.

After adding a low-level feature fusion layer to the YOLOv5s network, the feature fusion pyramid is illustrated in Figure 3. An additional layer of feature maps with a scale of 160×160 is introduced at the base of the feature pyramid. The backbone network extracts low-level feature maps at a scale of 160×160. It then upsamples the intermediate layer feature maps (80×80 scale) from the backbone network. These upsampled feature maps are added to the 160×160 features extracted by the backbone network, followed by convolution. This process yields intermediate layer feature maps at a 160×160 scale. The resulting feature map is then convolved and output to the network's prediction layer via the feature output layer. The prediction layer of the network predicts the target of the image on the output of four feature maps of different scales, the eigenvalue of the high-level feature map with a smaller scale has a larger receptive field on the image, which is used to predict the large target in the image, and the eigenvalue of the low-level feature map with a larger scale has a smaller receptive field on the image, which is used to predict the small target in the image.

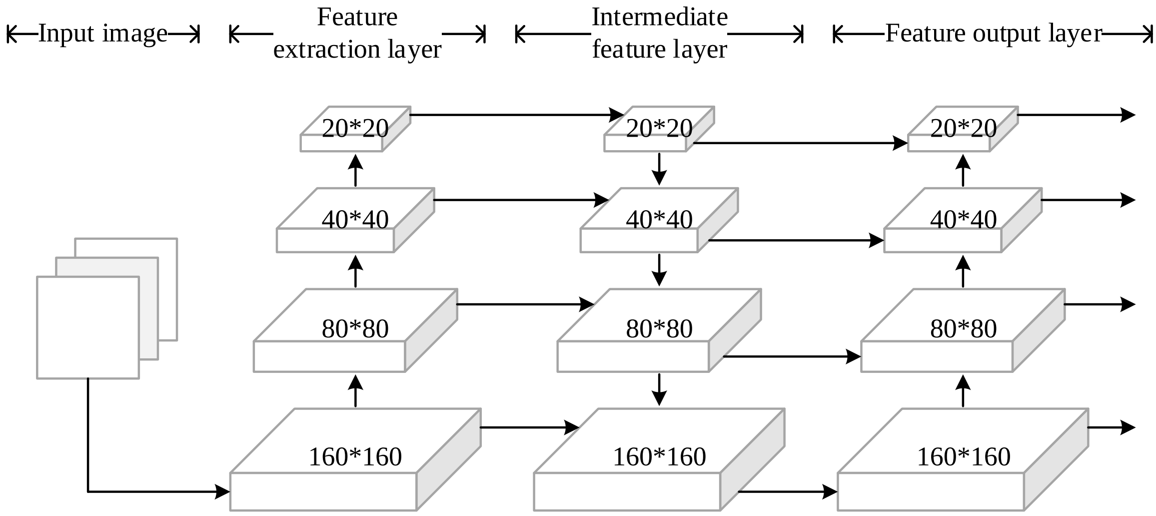


Figure 3: YOLOv5s_D4's feature fusion diagram

This paper explores the future deployment of neural networks on physical devices. In practical deployment scenarios, devices must possess sufficient computational power to ensure the network operates normally. When a network has numerous parameters and requires excessive computation, the demands on the deployment device increase significantly. If we can effectively reduce network parameters and computational load while maintaining stable network performance, we can lower the excessive requirements on deployment devices and save on computational resource consumption. During feature map generation at any layer of the YOLOv5s backbone network, each channel's feature map is produced using a large number of convolutional kernels. The extensive use of convolutional operations across all network layers results in excessive network parameters and computational overhead, demanding high computational power from deployment devices in practical applications. To mitigate these computational demands, this paper introduces a shadow module into the YOLOv5s architecture. By replacing conventional convolutional layers with shadow modules, the approach substantially reduces network parameters and computational overhead without compromising accuracy.

The feature map generation process using a standard convolutional layer is described by Equation (1):

$$Y = X \otimes f + b \tag{1}$$

Here, $X \in \mathbb{R}^{c \times h \times w}$ denotes data with c input channels and width and height of h and w respectively, while $\otimes$ represents the convolution operation. $f \in \mathbb{R}^{c \times k \times k \times n}$ denotes the convolution kernel of this layer's network, b denotes the bias term, $Y \in R^{h' \times w' \times n}$ denotes the feature map with n output channels and dimensions h' and w' respectively. Thus, the computational complexity of the convolution process for this network layer is $n \cdot h' \cdot w' \cdot c \cdot k \cdot k$. Typically, the number of convolution kernels n and the number of channels c are large values. Consequently, even a single layer of a neural network incurs significant computational overhead when generating feature maps.

The shadow module introduced in this paper generates feature maps through two steps, as illustrated in Figure 4.

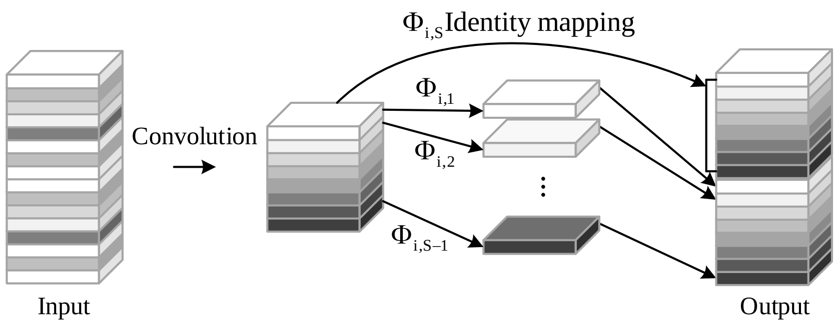


Figure 4: Schematic diagram of generating the feature map of the ghosting module

The first step generates a smaller number of feature maps through standard convolution operations. The feature maps produced in this step are referred to as intrinsic feature maps, as shown in Equation (2):

$$Y' = X \otimes f' \tag{2}$$

where $f^{'} \in R^{c\times k\times k\times m}$ denotes the number of convolutional kernels used, $Y^{'} \in R^{li\times w^{'}\times m}$ represents the output feature map, and $m \le n$. The computational cost of convolving the eigenfeature map in this step is $m\cdot h'\cdot w'\cdot c\cdot k\cdot k$.

The second step employs computationally efficient linear operations to generate additional feature maps from the eigenfeature map. The feature maps produced in this step are termed shadow feature maps. The linear operation is expressed as in Equation (3):

$$y_{ij} = \Phi_{i,j}(y_i^{'}), \quad \forall i = 1,...,m \quad j = 1,...,s \tag{3}$$

Here, $y_i^{'}$ denotes the i th feature map in $Y^{'}$, and $\Phi_{i,j}$ represents the linear operation generating the j th shadow feature map. The linear operation is performed on each channel, while $\Phi_{i,s}$ is the identity mapping for the eigenfeature map. This step ultimately yields $n = m\cdot s$ output feature maps. In the shadow module, linear operations can be performed using convolutional kernels of arbitrary size. assuming a convolution kernel size of $d\cdot d$. In this step, excluding the identity mapping, the computational cost is $m\cdot(s-1)\cdot h^{'}\cdot w^{'}\cdot d\cdot d$. Practically, $d\cdot d$ is comparable in magnitude to $k\cdot k$.

Through the above analysis, the process of generating feature maps in the original convolutional network layer is decomposed into two computationally less expensive steps. This reduces the computational cost of a single convolutional network layer by approximately s times, as shown in Equation (4):

$$\begin{aligned} r_s &= \frac{n\cdot h^{'}\cdot w^{'}\cdot c\cdot k\cdot k}{m\cdot h^{'}\cdot w^{'}\cdot c\cdot k\cdot k + m\cdot(s-1)\cdot h^{'}\cdot w^{'}\cdot d\cdot d} \\ &= \frac{n\cdot h^{'}\cdot w^{'}\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot h^{'}\cdot w^{'}\cdot c\cdot k\cdot k + \frac{n}{s}\cdot(s-1)\cdot h^{'}\cdot w^{'}\cdot d\cdot d} \\ &= \frac{c\cdot k\cdot k}{\frac{1}{s}\cdot c\cdot k\cdot k + \frac{s-1}{s}\cdot d\cdot d} \approx \frac{s\cdot c}{s+c-1} \approx s \end{aligned} \tag{4}$$

s denotes the number of ghost feature maps generated on the intrinsic feature map, where c is much larger than s.

This paper introduces a ghost module to compress the number of parameters in a single convolutional network layer, as shown in Equation (5):

$$r_c = \frac{n\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot c\cdot k\cdot k + \frac{n(s-1)}{s}\cdot d\cdot d} \approx s \tag{5}$$

The final detection network in this paper replaces the standard convolutional modules of the original network with ghost modules and adds a low-level feature fusion layer to the backbone. Therefore, this network is named YOLOv5s-Ghost-D4.

(2) Human Keypoint Detection Models

In practical applications, due to the large number of students in classroom settings, only multi-person keypoint detection algorithms such as OpenPose and AlphaPose [28] were selected. OpenPose offers advantages like fast processing speed and independence from environmental crowd density. However, it tends to exhibit misconnections between keypoints in real-world scenarios, particularly when two targets are in close proximity, making it unsuitable for classroom environments. AlphaPose, however, first performs object detection followed by keypoint detection on detected objects. This approach enables AlphaPose to better distinguish target behaviors in high-density environments. Therefore, this paper selects the AlphaPose algorithm as the human keypoint detection model.

(3) Model Training

1) Data Preprocessing

In the collected student classroom behavior dataset, issues such as incorrect camera angles, blurry footage, and insufficient lighting result in low-quality data or irrelevant samples. These samples negatively impact the model training process and must be filtered and removed to improve data quality. Consequently, data preprocessing operations are required, primarily encompassing data cleaning, data augmentation, and data annotation. The objective is to enhance the model's generalization capability, ensuring it learns features aligned with design requirements during training. Operations such as rotation, scaling, flipping, and color transformation simulate diverse shooting environments and scenarios to strengthen the model's generalization ability. Additionally, methods like cross-validation and confusion matrices can be employed to improve dataset quality and model generalization, ensuring the effectiveness of preprocessing. For student classroom behavior datasets, samples are annotated with corresponding behavioral categories, such as sitting normally, standing, or raising hands. Throughout this process, annotation quality must meet model training requirements, guaranteeing accuracy and consistency.

2) Model Training Steps

In deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) each possess distinct advantages. CNNs effectively extract image features and demonstrate outstanding performance in image processing. RNNs excel at capturing temporal information and exhibit significant strengths in handling sequential data. This design combines both models to meet the requirements of student classroom behavior analysis. The specific steps are as follows:

Step 1: Process input classroom video frames using a CNN model to extract image features.

Step 2: Input extracted image features into an RNN model. By incorporating temporal information, the RNN identifies student classroom behaviors. For instance, if a teacher delivers continuous instruction during a specific time interval, the temporal sequence will be encoded as a repetitive, sustained pattern like “2-2-2-2-4-4”.

Step 3: Based on the identified student behaviors, evaluate classroom teaching quality using the LBREM method (specific evaluation criteria will be described in Section 3.3.1).

Step 4: To address issues such as overfitting and slow convergence during model training, employ optimization techniques like regularization and learning rate adjustment to enhance the model's generalization capability and training efficiency.

Step 5: Model training and parameter tuning constitute a critical phase within the system. This process involves model learning, optimization, and adjustment to enable faster and more effective adaptation to the dataset, thereby improving the accuracy and robustness of student classroom behavior recognition. The steps for model training and parameter tuning are illustrated in Figure 5.

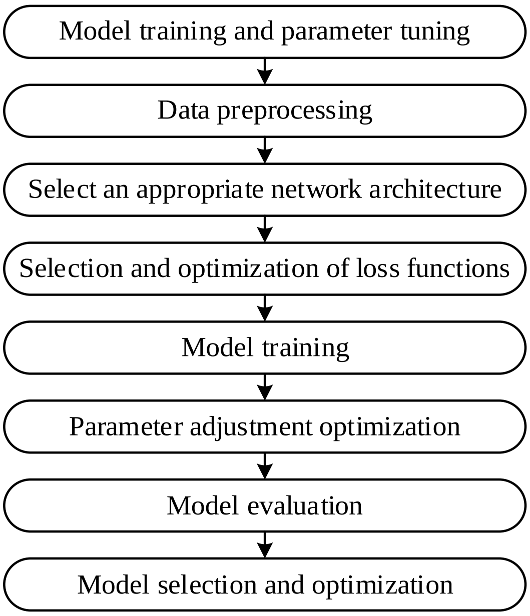


Figure 5: The steps of the model training and the reference

(4) Analysis of Experimental Results

1) Comparison of Facial Expression Recognition Performance

To validate the effectiveness of the adjusted YOLOv5 model, this section compares the detection performance of YOLOv5 and YOLOv5s-Ghost-D4 in multi-person classroom scenarios. The original YOLOv5 model and the adjusted YOLOv5s-Ghost-D4 model were implemented using the PyTorch deep learning framework. The training epochs were set to 200, with data split into training (60%), validation (20%), and testing (20%) sets. This experiment utilized a self-constructed dataset to compare algorithm performance before and after modification, employing detection accuracy and average detection accuracy as primary evaluation metrics.

Detection Accuracy:

$$P = \frac{TP}{TP + FP} \tag{6}$$

In the formula, precision P measures the proportion of instances correctly predicted as positive among all instances predicted as positive. TP represents the number of true positives, i.e., the number of samples correctly identified as positive by the model; FP represents the number of false positives, i.e., the number of samples incorrectly identified as positive by the model.

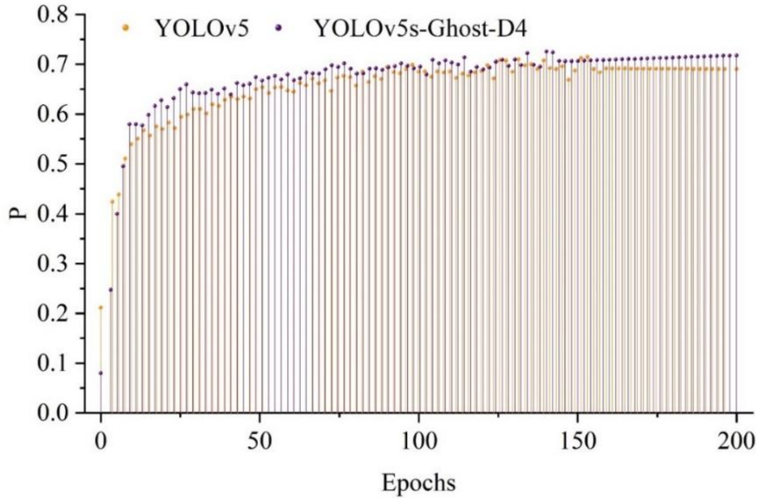
Average Detection Precision:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{7}$$

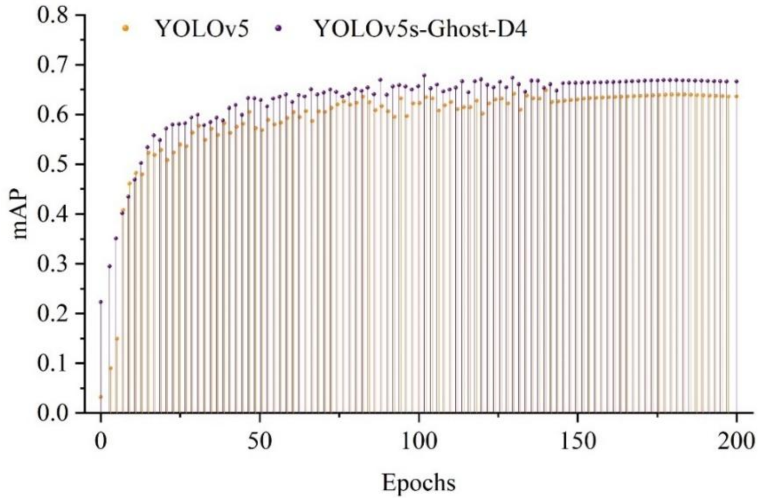
In the formula, mAP denotes the mean average precision, serving as a comprehensive metric to evaluate the overall detection accuracy of a detection network model. It represents the average of the mean detection accuracy across all detectable categories in the dataset. This metric assesses the object detection capability of multi-class classification models. Here, N

is the number of detection categories, and AP_i is the mean detection accuracy for the i-th category.

Figure 6 illustrates the detection accuracy and $mAP_{0.5}$ curves of the self-constructed dataset under both the original YOLOv5 model and the improved YOLOv5s-Ghost-D4 model. (a) and (b) represent detection accuracy and average detection accuracy, respectively. After approximately 150 training epochs, both detection accuracy and average detection accuracy of YOLOv5s-Ghost-D4 gradually surpass those of the original algorithm.



(a) P



(b) $mAP_{0.5}$

Figure 6: The comparison of the curve before and after the algorithm is improved

The detection results on the self-constructed dataset before and after the algorithm improvement are shown in Table 1. Compared to the pre-improvement version, the post-improvement algorithm achieved an accuracy of 83.9% on the self-constructed dataset, representing a 1.6% increase from the original 82.3%.

Table 1: Experimental data comparison table

Emotion	YOLOv5		YOLOv5s-Ghost-D4	
	P	$mAP_{0.5}$	P	$mAP_{0.5}$
All	0.823	0.858	0.839	0.883
Happy	0.823	0.859	0.837	0.887
Normal	0.813	0.857	0.84	0.877
Sad	0.833	0.857	0.839	0.884

The above experiments demonstrate that the improved algorithm has enhanced detection performance to a certain extent.

2) Comparison of Learning Behavior Recognition Performance

Using the same test dataset, we evaluated the recognition of students' classroom learning behaviors and compared accuracy (calculated as described above), recall, and F1 scores with those of the OpenPose recognition model.

Recall:

$$R = \frac{TP}{TP + FN} \tag{8}$$

In the formula, recall R reflects the proportion of all actual positive samples correctly identified by the model. TP similarly represents the number of true positives, while FN denotes the number of false negatives—that is, the number of positive samples incorrectly classified as negative by the model.

F1:

$$F1 = \frac{2 * P * R}{P + R} \tag{9}$$

$F1$ is the $F1$ score, which is the harmonic mean of precision and recall, used to comprehensively evaluate model performance. P represents precision, and R represents recall. The $F1$ score is the harmonic mean of precision and recall, providing a comprehensive metric for assessing model performance. The $F1$ score equally considers the importance of precision and recall; when both are high, the $F1$ score will also be correspondingly high.

Experimental results are shown in Table 2. The AlphaPose recognition model outperforms OpenPose in identifying three behaviors: sitting upright, standing up, and raising hands. AlphaPose achieves an average F1 score of 0.904, representing a 23.8 percentage point improvement over the OpenPose-based model.

Table 2: Behavior identification comparison results

	OpenPose			AlphaPose		
	P	R	F1	P	R	F1
Sit	0.666	0.738	0.738	0.922	0.91	0.953
Stand up	0.664	0.657	0.655	0.894	0.881	0.863
Hand up	0.657	0.685	0.605	0.844	0.813	0.895
Mean	0.662	0.693	0.666	0.887	0.868	0.904

The above experiments demonstrate that the AlphaPose recognition model can accurately identify students' learning behaviors.

3.3 Evaluation of Teaching Effectiveness Based on Student Learning Behavior Analysis

3.3.1 Method Design

LBREM categorizes student learning behaviors into four types: positive learning behaviors, relatively positive behaviors, neutral behaviors, and negative learning behaviors. Writing, raising hands, and standing up are classified as active learning behaviors. Smiling, maintaining forward attention, and interacting with the teacher are categorized as moderately active behaviors. Lowering the head, looking up, and sitting upright are neutral behaviors. Leaning on desks, looking around distractedly, and playing with mobile phones are identified as passive learning behaviors. Consequently, this study derived teaching effectiveness analysis indicators from the LBREM method, as shown in Table 3.

Table 3: Teaching effect analysis index

Classification of learning behavior	Corresponding action
Active behavior	Write, Hand up, Stand up
Positive behavior	Smile, Focus on the front, Communicate with teachers
Neutral behavior	Head, Rise, Sit
Negative behavior	Party desk, Left and right, Play phone
Teaching effect indicator	Quantitative processing
The classroom effect is good: ≥2	The classroom effect is good: 2
The classroom effect is better: 1.5~2	The classroom effect is better: 1
Classroom neutrality: 1~1.5	Classroom neutrality: 0
Classroom discipline: <1	Classroom discipline: -1

$$N_p = \sum_{k=1}^{F} f_p(k),\ p = 1,2,3,4 \tag{10}$$

Count the number of instances N_p for all learning behavior indicators according to formula (10), corresponding to active learning behavior, relatively active behavior, neutral behavior, and passive learning behavior, respectively. Here, $f_p(k)$ denotes the number of individuals exhibiting the corresponding learning behavior indicator in frame K, and $\sum_{r=1}^{i} f_p(k) = S$.

Based on learning behavior statistics, the TEM method for evaluating classroom teaching effectiveness is established. The evaluation is based on the ratio R between learning behavior indicators and the total number of learning behaviors. The calculation of R is shown in Formula (11):

$$R = \frac{\sum_{p=1}^{3} N_p}{N_4} \tag{11}$$

The classroom teaching effectiveness evaluation method can be defined as TEM(R), with the specific scoring rules shown in Formula (12):

$$TEM(R)=\begin{cases}2 & R\geq 2\\ 1 & 1.5\leq R<2\\ 0 & 1\leq R<1.5\\ -1 & R<1\end{cases} \tag{12}$$

Among these, TEM(R)=2 indicates excellent classroom learning outcomes, 1 indicates good outcomes, 0 indicates average outcomes, and -1 indicates poor outcomes.

3.3.2 Application Results Analysis

To validate the reliability of the LBREM method, the experiment selected 50 students from a private university in Fujian Province as research subjects. First, the learning efficiency of each subject was manually assessed using a four-point scale: excellent (2), good (1), average (0), and poor (-1). Subsequently, the LBREM method was applied for re-evaluation. The learning effectiveness assessments for all 50 subjects are presented in Figure 7. The results from both methods showed minimal discrepancy, with only three students exhibiting inconsistent evaluations. This indicates that the LBREM method possesses reasonable reliability and validity for teaching effectiveness assessment.

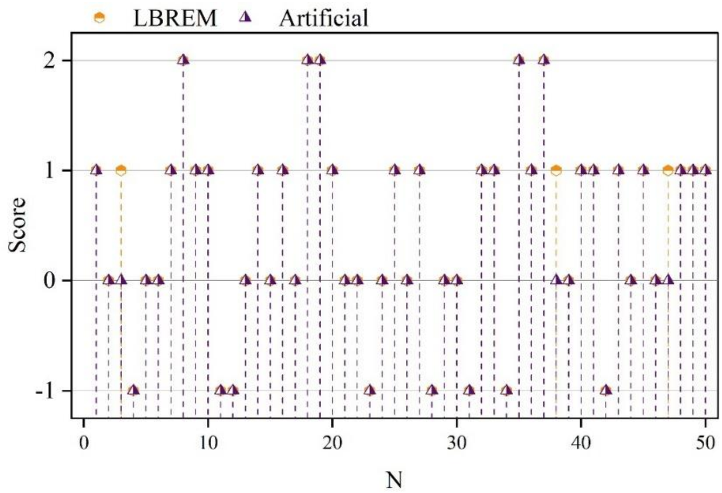


Figure 7: The subjects studied the effects of the assessment

4 Empirical Analysis

4.1 Experimental Protocol Design

To evaluate the impact of digital-intelligent course instruction and the LBREM method on learning outcomes, the experiment randomly selected 100 participants from the 2024 cohort of Business English majors at a private university in Fujian Province. Participants were divided into an experimental group and a control group, each comprising 50 students, based on similar academic performance. The experiment spanned one academic year. The experimental group received supplementary instruction using the digital-intelligence course designed in this paper and the LBREM method, while the control group maintained traditional teaching methods. Learning outcomes were evaluated based on learning engagement, learning quality, practical skills, and overall performance. A smart learning platform recorded learning data throughout the process to ensure the experiment was traceable and quantifiable.

4.2 Data Collection Methods

The experiment employed a multidimensional data collection approach, with the experimental group using an intelligent learning platform to record learning data in real time at a sampling frequency of 5 minutes per instance. This specifically included: online learning behavior data (12 fundamental metrics such as learning duration and resource access), oral practice data (acoustic features like phonemes, speech rate, and intonation), business practice data (behavioral features including facial expressions, gestures, and keyword usage), and classroom engagement data (changes in attention levels). Data collection reliability was ensured through a triple safeguard: 256-bit AES encryption guaranteed data security, distributed backup prevented data loss, and cross-validation ensured data accuracy. The experimental data collection achieved 99.7% coverage and 98.1% accuracy, providing reliable support for subsequent LBREM analysis. The control group collected baseline data via an intelligent attendance and assignment management system to ensure data consistency.

4.3 Comparative Analysis of Learning Outcomes

The experimental group demonstrated significant learning advantages under the digital-intelligence teaching model. A comparative analysis of learning outcomes between the experimental and control groups is presented in Table 4. Regarding learning engagement, the experimental group achieved a classroom participation rate of 96.1%, representing a 41.1% increase compared to the control group ($P < 0.05$). This substantial improvement primarily stemmed from the deep learning-based student behavior recognition and teaching effectiveness evaluation mechanism. This approach dynamically adjusted learning content in real-time based on students' learning states and progress, effectively boosting learning motivation. Regarding learning quality, the experimental group achieved an assignment excellence rate of 78.4% and a test pass rate of 96.4%, representing increases of 67.9% and 14.1% respectively compared to the control group ($P < 0.01$). The improvement in practical skills was particularly pronounced. The experimental group scored 88.9 points, 87.6 points, and 90.3 points in business negotiation, business writing, and business oral communication, respectively, representing increases of 19.8%, 19.3%, and 26.6% compared to the control group ($P < 0.01$). These findings indicate that the digital-intelligence teaching model effectively promotes the comprehensive development of students' practical abilities by providing diverse hands-on opportunities.

Table 4: The experimental group was compared with the control group

Evaluation index	Evaluation dimension	Experimental group (n=50)	Control group (n=50)	Lifting amplitude (%)	*P*
Learning participation	Classroom attendance (%)	96	95.4	0.6	<0.05
	Classroom engagement (%)	96.1	68.1	41.1	<0.05
	Job submission rate (%)	99	93.7	5.7	<0.05
Learning quality	Job excellence (%)	78.4	46.7	67.9	<0.01
	Test qualification rate (%)	96.4	84.5	14.1	<0.01
	Project completion (%)	91	76.7	18.6	<0.01
Practical ability	Business negotiation (points)	88.9	74.2	19.8	<0.01
	Business writing (points)	87.6	73.4	19.3	<0.01
	Business speaking (points)	90.3	71.3	26.6	<0.01
Integrated performance	Final evaluation (points)	91.2	79.2	15.2	<0.01
	Knowledge application (points)	88.9	73.5	21	<0.01
	Innovative ability (points)	87.2	69.9	24.7	<0.01

4.4 Measuring Capability Enhancement

Through quantitative analysis of the experimental group's competency gains, this study reveals the differentiated impact of digital-intelligent teaching models across various competency dimensions. Quantitative indicators of competency enhancement under the digital-intelligent teaching model are presented in Table 5. Business vocabulary increased from 1,521 words before the course to 3,881 words, representing a 155.2% improvement—850 words higher than the control group. This significant improvement stems from the vocabulary learning algorithm of the intelligent system, which dynamically adjusts review frequency based on the forgetting curve, thereby optimizing vocabulary retention. Synergistic effects were observed in the enhancement of speaking and listening skills. Speaking fluency scores rose from 73.2 to 92.4 (a 26.2% increase), while listening comprehension rates climbed from 71.5% to 92.2% (a 29% increase). In-depth analysis reveals this synergistic improvement primarily stems from the multimodal interaction design within VR virtual scenarios. Students simultaneously train listening and speaking skills in immersive environments, strengthening the connection between language input and output. Writing proficiency and business adaptability increased by 18.7% and 20.8%, respectively, exceeding the control group by 17 and 17.3 percentage points. Analysis of learning process data revealed that these improvements were closely linked to the system's real-time writing feedback and contextualized business training. The intelligent writing assessment system instantly identified writing issues and provided improvement suggestions, while diverse business simulation scenarios cultivated students' ability to adapt flexibly to different business contexts. Interestingly, the improvement in each skill was not evenly distributed. The improvement in business adaptability (20.8%) exceeded that in writing proficiency (18.7%), indicating the digital teaching model's unique advantage in cultivating practical application skills. Meanwhile, the increase in listening comprehension (29%) surpassed that in speaking fluency (26.2%). This disparity suggests that future teaching optimizations should enhance the targeted training of oral expression skills.

Table 5: Ability to improve quantitative indicators

Capacity dimension	Pretest	Posttest	Lifting ratio (%)	Control difference
Business vocabulary	1521	3881	155.2	850
Oral fluency (points/100)	73.2	92.4	26.2	14.4
Hearing comprehension rate (%)	71.5	92.2	29	17.5
Writing level (points/100)	78.1	92.7	18.7	17
Business resilience (points/100)	74.1	89.5	20.8	17.3

5 Conclusion

A business English teaching model based on digital intelligence technology was designed. Combined with YOLOv5s-Ghost-D4 and AlphaPose, a student behavior analysis system was developed to evaluate learning outcomes based on system analysis results. Compared to YOLOv5 and OpenPose, the YOLOv5s-Ghost-D4 and AlphaPose models demonstrate superior recognition performance. Teaching effectiveness evaluations based on system recognition results showed only 3 out of 50 subjects had inconsistent outcomes with manual assessments. This indicates the reliability of using this mechanism for teaching effectiveness evaluation.

The designed teaching model and evaluation mechanism were implemented in instruction and compared with traditional teaching approaches. Students in the experimental group demonstrated significant improvements in business English application skills, business negotiation abilities, and writing proficiency, achieving results markedly superior to those of the control group. These findings provide a replicable practical pathway for reforming business English instruction in higher education institutions.

References

[1] Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., ... & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. Education and information technologies, 28(6), 6695-6726.

[2] Indriani, C. L., Muth'im, A., & Febriyanti, E. R. (2024). English language learning through the use of digital technology: A literature review. Linguistic English Education and Art (LEEA) Journal, 7(2), 283-290.

[3] Bui, T. H. (2022). English teachers' integration of digital technologies in the classroom. International Journal of Educational Research Open, 3, 100204.

[4] Nawi, N. S. M., & Nor, N. A. A. M. N. (2023). The Challenges in the Teaching of English Literature: A systematic review. Journey: Journal of English Language and Pedagogy, 6(1), 130-147.

[5] Renau, M. (2016). A review of the traditional and current language teaching methods. International Journal of Innovation and Research in Educational Sciences, 3(2), 82-88.

[6] Sarker, M. N. I., Wu, M., Cao, Q., Alam, G. M., & Li, D. (2019). Leveraging digital technology for better learning and education: A systematic literature review. International Journal of Information and Education Technology, 9(7), 453-461.

[7] Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., ... & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. Education and information technologies, 28(6), 6695-6726.

[8] Xie, K., Di Tosto, G., Chen, S. B., & Vongkulluksn, V. W. (2018). A systematic review of design and technology components of educational digital resources. Computers & education, 127, 90-106.

[9] Pétursdóttir, S. (2012). The effectiveness of integrating existing digital learning resources into classroom teaching–an evaluation of the learning achievement. Nordic Studies in Science Education, 8(2), 150-161.

[10] Bao, Y., & Yunus, H. M. (2024). A Review on The Impact of Blended Learning, Learning Styles, and The Community of Inquiry Model on Student Interest in Learning in Higher Education. Pakistan Journal of Life & Social Sciences, 22(2).

[11] Wang, N. (2018). A Teaching Mode for Art Anatomy Based on Digital Virtual Technology. International Journal of Emerging Technologies in Learning, 13(8).

[12] Tatar, D., & Robinson, M. (2003). Use of the digital camera to increase student interest and learning in high school biology. Journal of science education and technology, 12(2),

89-95.

[13] Gao, N. (2018). Construction and implementation of teaching mode for digital mapping based on interactive micro-course technology. International Journal of Emerging Technologies in Learning (Online), 13(2), 21.

[14] He, C., & Sun, B. (2021). Application of artificial intelligence technology in computer aided art teaching. Computer-Aided Design & Applications, 18.

[15] Pinto, M., & Leite, C. (2020). Digital technologies in support of students learning in Higher Education: literature review. Digital education review, (37), 343-360.

[16] Chen, Z., Lang, L., Qi, X., & Geng, J. (2016). Digital mining technology-based teaching mode for mining engineering. International Journal of Emerging Technologies in Learning (Online), 11(10), 47.

[17] Pérez-Marín, D., Paredes-Velasco, M., & Pizarro, C. (2022). Multi-mode digital teaching and learning of human-computer interaction (HCI) using the VARK model during COVID-19. Educational Technology & Society, 25(1), 78-91.

[18] Yan, Y., Zheng, Y., & Ye, X. (2025). The impact of IVR-ADDIE-based digital storytelling teaching mode on students’ self-regulation ability and self-efficacy. Education and Information Technologies, 30(5), 6141-6162.

[19] Chen, X., & Chen, R. (2024). Examining the influence of the performance of mobile devices on teaching quality based on the technological acceptance mode. Library Hi Tech, 42(2), 670-695.

[20] Gong, Y. (2021). Application of virtual reality teaching method and artificial intelligence technology in digital media art creation. Ecological Informatics, 63, 101304.

[21] Xu, Z., Chen, Z., Eutsler, L., Geng, Z., & Kogut, A. (2020). A scoping review of digital game-based technology on English language learning. Educational Technology Research and Development, 68(3), 877-904.

[22] Yang, L. (2025). Research on Multi-Mode Teaching of College English Education Based on Digital Information Technology. International Journal of High Speed Electronics and Systems, 2540796.

[23] Moorhouse, B. L., & Wong, K. M. (2022). Blending asynchronous and synchronous digital technologies and instructional approaches to facilitate remote learning. Journal of Computers in Education, 9(1), 51-70.

[24] Huang, H., & Wang, J. (2022). Innovative research on collaborative design of blended english teaching in higher vocational colleges based on digital technology. Scientific Programming, 2022(1), 9982680.

[25] Chen, F. (2019). Study on the multivariant interactive teaching modes of college English under the information technology environment. Informatica, 43(3).

[26] Mu Yang. (2023). Gamification in Smart Classes: Implications for Teachers' Role Reform

and Teaching Mode Innovation. Curriculum and Teaching Methodology,6(8),

[27] Mei Bie, Quanle Liu, Huan Xu, Yan Gao & Xiangjiu Che. (2024). FEMFER: feature enhancement for multi-faces expression recognition in classroom images. Multimedia Tools and Applications, 83(2), 6183-6203.

[28] Lei Liu, Yeguo Sun, Yinyin Li & Yihong Liu. (2025). A hybrid human fall detection method based on modified YOLOv8s and AlphaPose. Scientific Reports,15(1),2636-2636.