ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Design and Student Performance Enhancement of a Movement Technology Assessment System for Computer-Assisted Dance Teaching

Jin Gao[1] *and Qian Zhou*[1,*]

[1] Guangxi Minzu Normal University, Chongzuo, Guangxi, 532200, China

SUMMARY: *The traditional dance teaching method has subjective evaluation, different evaluation standards, etc., which cannot provide learners with accurate movement skill levels, so a movement skill level analysis model based on computer vision and deep learning algorithms is proposed to improve the professional degree of dance teaching. Using a multi-source information fusion method, a color depth camera and an inertial measurement unit are used to collect human movements in real time and analyze them based on a spatio-temporal convolutional neural network combined with a multi-scale attention mechanism for accurate posture estimation and movement evaluation. The results of this study show that the system has achieved good results in terms of correctness of movement, reaction speed, and experience, and the results of an eight-week controlled trial show that the system can effectively help dancers improve their skill level, fluency, rhythm, and expressiveness. According to the feedback information from users, the real-time feedback function and personalized learning program design of the system are well received by users.*

KEYWORDS: *computer-aided instruction; dance movement assessment; deep learning; computer vision; multimodal data fusion*

1 Introduction

As an important form of artistic expression, the interpretation and development of dance has far-reaching significance and influence on the construction of spiritual civilization. Therefore, it is necessary to raise the attention to the teaching of this course. The current dance teaching mainly exists in the backward teaching mode, lack of innovation in teaching content, lack of relevance and effectiveness of teaching, strong subjectivity in teaching assessment, feedback lag, uneven allocation of teaching resources, lack of personalization and other problems, the emergence of these problems not only reduces the efficiency and quality of dance teaching, but also seriously impedes the development of the dance discipline [1-5]. With the rapid development of information technology, computer-assisted education and teaching are becoming more and more widespread [6]. The development of computer technology and motion capture technology, etc. makes computer-assisted dance teaching become a trend, breaking through the traditional dance teaching challenges.

Currently, researchers have proposed a variety of computer-assisted dance teaching methods to improve teaching effectiveness and student performance through movement assessment, movement recognition, personalized feedback and instruction. Literature [7] explored a flipped learning method incorporating mobile peer assessment, which can effectively improve students' dance skill performance and learning satisfaction, and is superior to the

*zqian426@163.com
https://doi.org/10.65102/is2026110
Published: 30 April, 2026

traditional teaching mode, providing a useful reference for teaching practice. Literature [8] assessed the effects of multimedia computer-assisted instruction and demonstration methods on the performance of physical education students in folk dance teaching through quasi-experimentation, and the results showed that the demonstration method was more effective in improving students' performance to the level of “proficiency”. Literature [9] used cloud computing technology to optimize the dance movement simulation teaching system, which maintains functionality and helps to improve the accuracy of students' dance movements. Literature [10] proposed a hybrid feature human motion reconstruction technology and convolutional algorithm action recognition model for dance teaching, which can accurately identify the dance movements, effectively optimize the effect of dance teaching, and provide technical support for digital dance education. Literature [11] develops an auxiliary dance training system based on computer vision, which utilizes Kinect sensors for bone tracking, accurately identifies postures through improved joint angle algorithms, and provides real-time feedback to correct the movements, which effectively improves the training effect and movement standardization. Literature [12] created an interactive learning system based on extended reality to assist dance teaching, providing real-time personalized feedback, effectively improving students' learning performance and interactivity, and reducing their cognitive load, optimizing the dance teaching experience. Literature [13] used the PoseNet model to implement a real-time feedback system to assist dance teaching, which can decompose and analyze students' movements, combined with the alternating least squares algorithm to provide personalized guidance, effectively improving the accuracy of movements and learning satisfaction. Literature [14] constructs a dance movement description model based on computer-aided technology, manages the data through the Internet of Things system, and extracts gesture characteristics from the four dimensions of time, space, gravity, and fluency, which provides an innovative theoretical basis for the scientific training and teaching of rotary movements. Literature [15] used artificial intelligence to develop an integrated teaching, assessment and visual feedback system applied to dance teaching, which can effectively improve students' dance skill performance and self-efficacy. Literature [16] pointed out that immersive virtual reality technology with its immersive, interactive and personalized characteristics can effectively enhance the folk dance teaching classroom perception and learning interest, and then significantly improve students' dance performance and teaching effectiveness. Literature [17] points out that the artificial intelligence dance evaluation system can improve the accuracy of movement recognition and personalization of teaching to provide technical support for dance teaching, but it is difficult to fully capture the embodiment and cultural connotation of the dance, and should be complemented with the teacher's guidance to balance the technical accuracy and artistic value.

In this paper, we design a method for technical evaluation of dance movements using multi-source information fusion of RGB-Depth and IMU, and use a spatio-temporal convolutional neural network model to characterize the dance movements, combining with the joint movement path, body balance and rhythmic matching indexes to complete the technical scoring of the dance movements, and judging the coherence of the movements and soundtrack compatibility. The results of each subsystem are adaptively weighted to obtain the final score, and at the same time, targeted learning suggestions are given based on the students' learning history.

2 Theoretical foundations

2.1 Computer-assisted instruction

Based on the constructivist view of teaching and cognitive psychology, the design of the learning process has a strong guiding effect; Vygotsky's "zone of nearest development" has a clear understanding of the range of students' existing abilities and possible development, and can be used as the basis for adaptive ITS.

For learners, when they encounter actions that they are not yet able to complete during the learning process, the platform can accurately determine the user's level according to the results of processing the user's action information, and then recommend the appropriate level of practice for the students in real time to ensure the appropriate level of difficulty. Cognitive development theory emphasizes that the development of children's cognitive ability is characterized by stages, therefore, according to the different stages of cognitive development, we arrange the transition of dance movements from mechanical imitation in the perceptual-motor stage to creative expression in the formal computing stage, and gradually transform the external movement patterns into internal artistic cognition in the process of human-computer interaction.

Discovery learning method is a new teaching method advocated by Bruner, which mainly relies on students' own internal factors and teaching activities. Human-computer interactive step training is the use of the discovery learning method to respond to the personalized needs of the learners, and give timely feedback in the creation of different situations in the learning environment, prompting the learners to try to generate interest in the process of self-learning, and gradually find out the correct method of movement. The imitation process mentioned in Bandura's social theory of learning is computerized by the role model demonstration function of the intelligent teaching system.

2.2 Motion Assessment Techniques

In recent years, the application scope of motion capture system has gradually shifted from professional high-precision instruments to lightweight and flexible intelligent direction development. Among the many motion capture systems, a variety of mature products developed based on optical principles, such as VICON Motion Capture System, OptiTrack Motion Capture System, have been widely used for their accuracy advantages. These optical principle based motion capture system is the use of a number of high-speed infrared cameras installed in the surrounding environment to form a three-dimensional spatial monitoring network, and its positioning resolution can be up to the sub-millimeter level, and can reach more than 1,000Hz sampling rate, can meet the needs of high-speed dance step capture. But hundreds of thousands of dollars of hardware purchase costs, complex sticker operation steps, the limitations of the site light conditions are not conducive to the widespread application of this type of system to the classroom. The inertial positioning method, on the other hand, utilizes IMUs installed with acceleration, angular velocity and magnetic field sensors for positioning and tracking. Due to its portable and easy to use in any environment characteristics can be outdoor performance tracking analysis, but the cumulative error and the cumbersome calibration step constrains its use in the precision field, after a long time of operation will appear positioning accuracy decline, so it is necessary to carry out periodic calibration.

Deep learning-based human pose estimation methods represented by OpenPose, Alpha Pose, HRNet, etc. can quickly and accurately identify and track human joints and localize them using monocular cameras without using any specific equipment and calibration patches, and apply the results directly to the field of motion analysis, which greatly promotes the development of

motion analysis research. It facilitates the deployment of dance movement recognition systems into regular classroom environments, and the Media Pipe model can be processed in real-time at 30 fps on cell phones, which facilitates the development of low-power teaching applications. Depth camera technology acquires synchronized color and depth information through infrared structured light ranging. Devices such as Microsoft Kinect and Intel RealSense have gained wide acceptance in the education field due to their moderate cost and relatively simple configuration process. The addition of depth data effectively alleviates the technical bottleneck of traditional color images in depth ambiguity processing, and the rich skeleton tracking interface of Kinect SDK lowers the technical threshold of application development. Nevertheless, the limitation of effective detection range within 5 meters, the problem of infrared interference, and the power consumption control of the device still need to be considered in the actual deployment.

The comparison results of different motion evaluation techniques are shown in Table 1. Different motion evaluation techniques have different real-time performance and adapt to different scenarios. Depth camera can realize 30-60fps data acquisition, and the deployment cost is lower for indoor teaching scenarios.RGB vision cost is lower compared to depth camera, but it can only realize 30fps data acquisition, which is suitable for remote teaching. Multimodal fusion is also a hot spot in the current research, i.e., the comprehensive consideration of different types of sensor signals such as vision, inertia, audio, etc. to obtain a more complete action characterization, in which the vision-inertia fusion can complement each other's deficiencies, while the audio-visual fusion adds new perspectives for action classification from the aspect of artistic expression, and the multimodal fusion in deep networks is performed by adopting the attention mechanism and the alignment to do so. It has strong robustness and high recognition accuracy for complex environments. At present, the related research still has the problem of difficult to adapt to the construction of different dancers' own characteristics model; at the same time, it is difficult to cope with the diversified needs of various cultural and artistic styles; in addition, how to balance the immediacy and accuracy of the system is also an urgent problem to be solved; lastly, if this kind of intelligent assistive system is promoted to a larger scale application scenario, the stability and sustainability of the system will become a major problem.

Table 1: Comparison of Motion Assessment Techniques

Technology type	Accuracy level	Real-time performance	Applicable scenarios
Optical motion capture	Sub-millimeter level	1000Hz	Professional sports analysis and scientific research experiments
Inertial sensor	Centimeter-level	100-200Hz	Outdoor sports, wearable devices
Depth camera	Millimeter-level	30-60fps	Indoor teaching and family entertainment
Depth camera	Pixel-level	30fps	Mobile applications, distance teaching
Multimodal fusion	Sub-millimeter level	60fps	High-end teaching and professional training

2.3 Use of Technology in Dance Teaching

Table 2 Examples of dance teaching technologies. For example, the development of video analytics used in the field of dance teaching shows that the application of this technology has gone through a process from simple path tracking to higher-level content resolution: the first generation of tracking and localization methods based on multiple fixed-position cameras

developed by the Technical University of Munich, Germany, used the optical flow field to generate virtual spatial action models, which was initially tested successfully in ballet basic skills training, but light intensity sensitivity and background DanceNet directly uses end-to-end training to obtain semantic information from videos to detect street dance movements, achieving 90% accuracy, no longer relying on manually designed features for street dance movement recognition, and reducing dance teaching time by a quarter on average. Professional motion capture devices are gradually shifting to consumer-grade products, Oxford University Department of Dance and Vicon shared a set of 16 high-speed infrared cameras to capture 39 human joint point position information, capturing 240 frames per second. Modern dance analysis achieves 0.1mm positional accuracy, resulting in a 34% increase in students' average level of dance technique awareness. The Tokyo University of the Arts in Japan used 12 wireless IMUs to form a wearable system for traditional dance training, utilizing accelerometers, gyroscopes, and magnetometers for whole-body positioning and providing instant feedback during training, and improving movement correction by 40%; the Korea Advanced Institute of Science and Technology (KAIST) utilized the Kinect Depth Camera's inexpensive system (which reduces costs by up to 80%) for Han-ryu dance training, which, despite a slight decrease in precision Although the accuracy was slightly reduced, the ease of installation and use and the visualization were well received by teachers and students.

Table 2: Examples of application of dance teaching techniques

Technology type	Representative system	Core function	Application Cases	Effect evaluation
Video analysis	Dance networks, posture networks	Action recognition and trajectory analysis	Hip-hop and ballet training	Recognition accuracy (90%), learning efficiency increases by 25%
Optical motion capture	Vicon, Opti-Track	High-precision attitude reconstruction	Modern dance	Technical understanding increased by 34%
Inertial motion capture	Array of inertial measurement units	Wireless attitude monitoring	Traditional dance	Action correction rate is up 40%
Depth camera	Kinect, RealSense	Skeleton tracking, gesture recognition	Popular dance	Cost reduction by 80%
Virtual reality	Oculus, HTC Vive	Immersive experience, spatial perception	Stage performance rehearsal	The sense of space is enhanced by 60%

VR technology can create a whole new way to experience dance instruction, and the University of Southern California's Innovative Technology Research Laboratory has designed a VR dance lab that utilizes high-level HR equipment to create a realistic 3D practice space. In this environment, participants wear an Oculus Rift head reality device and interact with a virtual teacher, which responds to the software by tracking the participant's fingers and body. After testing, the spatial sense teaching effect of VR teaching was significantly better than traditional teaching methods, with an improvement of about 60%; students made substantial progress in their understanding of complex scheduling designs. The French National School of Music and Dance applies this technology to ballet rehearsal, dancers in the virtual environment to preview the complete repertoire and get the system automatically generated to improve the proposal, this mode not only reduces the cost of rehearsal space, but also provides dancers with repeated practice opportunities.

3 Motion technology assessment system design

3.1 Overall system design

In this paper, the bottom-up hierarchical decomposition idea is used to design the dance scoring model. At the sensing level, RGB-D cameras and IMUs are used to form a hybrid sensing array for information fusion, and up to 8 RGB-D cameras can be accessed at the same time in a single acquisition to obtain a 360° panoramic image, and clock synchronization is carried out with the help of NTP. In the feature processing layer, a signal processing algorithm is introduced to process the inertial information on the basis of the visual feature extraction by using the hierarchical decomposition graph convolution module, and the features at different levels are fused to obtain a unified action representation space. In the visual part, the ResNet-50 backbone network is used for feature extraction and migration to the dance domain dataset using the idea of transfer learning, and the inertial part uses a one-dimensional convolutional neural network to extract its frequency components and time trends.

3.1.1 Skeleton feature extraction

The specific structure of the hierarchical decomposition map convolution module is shown in Fig. 1. Hierarchical decomposition spatial and temporal graph convolution of the spatial graph convolution module is introduced in the hierarchical decomposition graph convolution module, which divides the human skeleton graph into three layers, and extracts the spatial features of the human skeleton for each layer separately, which effectively extracts the fine-grained features of the human dynamic skeleton, and increases the information expression ability of the feature vector, which is conducive to classifying the abnormal behaviors afterward. In the process of establishing the hierarchical decomposition graph, firstly, a tree-like graph with root nodes is established according to the joints and physically connected edges of the human dynamic skeleton graph. In the dendrogram, the root node is the center node, and the set of key points divided by the center node belongs to the same level in the real world, for example, the two sides of the elbow, the hand and the foot, which belong to the same level in the real world, also belong to the same level in the hierarchical decomposition graph.

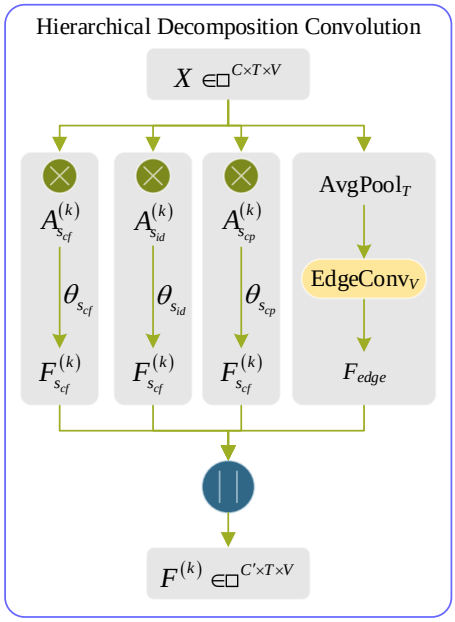

Figure 1: Hierarchical decomposition diagram Convolutional module structure diagram

Based on the determined central nodes, an adjacency matrix $\tilde{A}_{HD} \in \square^{N_L \times V \times V}$ can be built for the dynamic skeleton graph of human body transformed into a tree-like graph, in which the N_L layer containing the set of edges of the N_H layer can be defined as:

$$\vec{A}_{HD} = \left[\varepsilon\left(H_1 \to H_2\right), \cdots, \varepsilon\left(H_{N_H-1} \to H_{N_H}\right) \right] \tag{1}$$

where H_k represents the set of joints at the k th level, $\varepsilon\left(H_k \to H_{k+1}\right)$ represents the set of edges pointing to the joints at the $k+1$ th level from the joints at the k th level, N_L represents the number of levels, N_H represents the edges in the level number of edges in that level. However, the edges defined by the above method are unidirectional, in order to comply with the computation rules of the graph convolution module, the edges from leaf nodes to the root node need to be represented in the adjacency matrix, and the edges that emanate from the nodes and point to the nodes themselves need to be represented in the adjacency matrix in order to compute the node's own traits. The adjacency matrix $\overleftrightarrow{A}_{HD} \in \square^{N_L \times N_s \times V \times V}$ that satisfies the above requirements is defined as:

$$\ddot{A}_{HD} = \left[\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{N_L} \right] \tag{2}$$

$$\varepsilon_k = \varepsilon(\underbrace{H_k \cup H_{k+1}}_{s_{id}}, \underbrace{H_k \to H_{k+1}}_{s_{cp}}, \underbrace{H_{k+1} \to H_k}_{s_{cf}}) \tag{3}$$

where ε_k denotes the set of three edge subsets $S = \left\{ s_{id}, s_{cp}, s_{cf} \right\}$, and s_{id}, s_{cp}, and s_{cf} stand for the self-side, centripetal, and centrifugal edge subsets respectively. The complete hierarchically decomposed human dynamic skeleton graph can be obtained by the above construction method.

The hierarchical decomposition graph convolution module consists of four heterogeneous topological processing pathways, including three sets of graph convolution branches and one set of edge feature enhancement branches. The four branches are optimized for computational efficiency through linear dimensionality reduction mapping. Among them, the graph convolution branch performs fine-grained feature extraction for the hierarchical edge set, and achieves cross-layer feature aggregation through channel-dimensional feature vector splicing. The above operation can be expressed as:

$$F_{HD}^{(k)} = \mathop{\Big\|}_{s \in S} \left\{ \ddot{A}_{HD;s}^{(k)} \Phi\left(F_{in}\right) \Theta_s^{(k)} \right\} \tag{4}$$

where $F_{HD}^{(k)}$ represents the feature map generated by the hierarchical decomposition graph convolution operation, the function Φ denotes the linear transformation with parameter $W \in \square^{C' \times C}$, and $\|$ represents the concatenation operation in channel dimension. In the process of extracting the sample-level edge association information about the similarity between nodes in the feature space, the module uses edge convolution to extract the human dynamic skeleton graph features through the local neighborhood graph in the feature space. First, the edge convolution branch uses average pooling to compute the time computational efficiency, and then, the set of neighboring edges is obtained using the k-nearest neighbor algorithm based on the Euclidean distance. The computational procedure of graph convolution in the resulting

hierarchical decomposition is:

$$F_{HD} \leftarrow \sum_{k=1}^{N_L}\left[F_{HD}^{(k)} \left\| Z^{(k)}\left(\frac{1}{T}\sum_{t=1}^{T}\Phi\left(F_{in}\right)\right)\right.\right] \tag{5}$$

where $z_V^{(k)}$ represents the edge convolution operation.

3.1.2 System functional modules

The algorithmic analysis layer contains modules for posture estimation, action recognition, and quality assessment, in which the posture estimation algorithm is based on graph convolutional networks modeling human joints as graph nodes and skeletal connections as graph edges. The action recognition human-computer interface uses Transformer to process temporal action sequences and uses multi-head self-attention to capture long-distance temporal dependence, and the scoring criteria evaluate the completion of the action in three aspects, namely, the degree of technical correctness, the degree of completion of the action, and the sense of rhythm, respectively.

Multimodal feature fusion uses the attention mechanism to dynamically adjust the weight contribution of different modalities, and the mathematical expression of the fusion strategy is:

$$f_{fused} = \alpha \cdot f_{visual} + \beta \cdot f_{motion} \tag{6}$$

Among them, the weighting coefficients α and β are automatically learned through end-to-end training, and the feedback generation module uses NLP technology to convert the results of quantitative evaluation into an easy-to-understand textual form for feedback to the dancers, and to give personalized suggestions and corrective solutions based on the pre-training language model, and the relevant knowledge map in the domain. In the feedback generation module, a glossary of dance terminology and a library of common movement errors are created to describe the dancer's technical movement problems and provide specific improvement methods. Provide users with information on movement correction, training guidance, learning development and so on.

The learning management module utilizes a hybrid recommendation strategy of collaborative filtering and content filtering to dynamically adjust the content and difficulty gradient based on the learner's skill level, learning preferences, and historical performance. This part is a RL recommendation subsystem, which constantly updates the recommendation strategy according to the students' progress to provide the optimal solution to the most suitable practice topics for each student. In terms of data flow architecture, it is divided into two different data flow pipelines, the online pipeline and the offline pipeline, with the former being able to respond quickly to user requests and give the corresponding results, and the latter being able to perform large-scale data processing and storage. Therefore, message queues are used to coordinate these two ways, and the whole system is designed to take into account the actual situation in the teaching process as much as possible, so that it can be simple, practical, and reliable while taking into account the advancement of the premise of providing a complete technical solution for the digital transformation of dance teaching.

3.2 Action Evaluation Algorithm Design

The dance movement evaluation algorithm is mainly to establish a mathematical model that can describe the spatial form, and temporal movement state in dance. Therefore, this paper proposes a method based on spatio-temporal dual-flow graph convolutional neural network, which

transforms a continuous sequence of human bones into a spatio-temporal graph. Among them, the edges in the spatial graph represent the connectivity between anatomically neighboring joint points, and the edges in the temporal graph represent the connectivity between the positions of the same joint points on the previous and subsequent frames. This approach reflects the biomechanical constraints that dance movements have and overcomes the shortcomings of the traditional Euclidean spatial metrics in describing the coupling of joint movements.

The two paths extract spatial and temporal information respectively, and in the feature extraction process, a depth map convolutional network is utilized in the spatial direction to obtain the topological relationship of the posture at a single moment. While in the time series direction, long-term dependent behavioral features are mined with the help of a null convolutional network, and the obtained spatial and temporal information are combined together in the middle layer of the network, and the final result is obtained by linearly combining the information from different sources using a learnable gating function. In addition, in order to further improve the model effect, a constraint strategy based on a priori information is proposed considering the objective law of human movement behavior itself. The importance weights of the key frames labeled by the dance experts are incorporated into the loss function, so that the learning of the model satisfies the performance evaluation criteria of the art.

The model is designed to use a triple index joint training mechanism, and the accuracy utilizes the DTW method to find the HD distance between the student's movement path and the preset samples. Fluency, the energy of each frequency point is obtained after FFT transformation according to the amount of velocity change of each joint point, and the average information entropy of the energy information is used as a reference basis. Coordination degree, based on the correlation degree of video signal and sound signal to measure whether the moment of peak appearance of human movement speed matches the music beat. The comprehensive evaluation function is:

$$F(T) = \sum_{i=1}^{n} \alpha_i \cdot f_i(T) \tag{7}$$

where T denotes the dance movement sequence, f_i is the feature function of the i th evaluation dimension, and α_i is the adaptive weight coefficients, which are automatically adjusted on the validation set by a Bayesian optimization algorithm.

A course-learning-like approach is adopted in the model training process, where samples of simple actions are first used for training, and then combinations of more difficult actions are gradually added, and focus loss is utilized to increase the attention of the samples. For data expansion, a perturbation method based on kinematic constraints is proposed, by which about 20,000 sets of false data samples satisfying physiological mechanisms are generated. It significantly reduces the difficulty of obtaining dance data. This method uses a batch_size of 64 for training on four Nvidia A100 graphics cards, sets the initial learning rate to 0.001, and decays according to the learning rate scheduler, and the validation set Loss drops to about 0.083 after 120 epochs.

In addition, in order to make a better judgment, a multi-scale attention mechanism is used to give different degrees of attention to different joints during the computation process; meanwhile, considering the difficulty of the high-speed rotating movement, an angular velocity-based resampling method is proposed, i.e., if the angular velocity at the current moment is too large, its sampling rate is set to 120 frames per second. Due to the full consideration and simulation of the characteristics of the dance movements, the proposed algorithm achieves better results in the above three aspects, and the sequence information captures the continuous relationship between long-distance (up to 15s) complex dance steps using a GNN with LSTM.

In the spatial direction, for the first time, the concept of dance human body configuration is introduced as a weight value in the adjacency matrix of graph convolution, and the edge weights on the rotated parts of the torso are increased in the ethnic dance classification task. Knowledge distillation is used during the experiment to fuse the information from the larger model into the MobileNetV3 model. Due to the addition of the attention module, the network is made to focus more on the important part of the dance, i.e., the most important part of the movement for the artistic performance of the whole dance, and its loss function can be described as:

$$S = \sum_{i=1}^{n} \omega_i \cdot f_i(X_t) \tag{8}$$

where S denotes the composite score, ω_i is the weight coefficient of the ith evaluated dimension, $f_i(X_t)$ is the output value of the ith eigenfunction at the moment t, and X_t represents the multimodal input data of the moment t. .

The comprehensive assessment of action quality can be quantitatively expressed by the following mathematical framework:

$$Q = \alpha \cdot A(p) + \beta \cdot T(v) + \gamma \cdot S(a) + \delta \cdot R(f) \tag{9}$$

Among them, Q represents the comprehensive quality score, $A(p)$ assesses accuracy based on the joint position p, $T(v)$ conducts temporal sequence assessment based on the speed v, $S(a)$ measures stability based on the acceleration a, and $R(f)$ evaluates rhythm based on the frequency domain feature f. The weight coefficients α, β, γ, and δ can be adjusted according to specific application requirements.

3.3 User Interface Design

The design of the human-computer interface of the dance movement technology evaluation system, based on the human-centered human-computer interface design idea, applies the cognitive psychology theory knowledge to establish a good human-computer operation feeling system, see Fig. 2, in which the flexible layout template used in the design of the human-computer interface is set up with the foreground display window accounting for 3/4 of the whole window size used to play the dance posture acquisition video and the generated human posture model. At the same time, the operation panel was designed in the right half quarter position for placing functional buttons such as evaluation results, evaluation log and parameter configuration. In addition, the visual aspect of the reference to the flat style, reduce the degree of bright colors to reduce the pressure on the eyes, the main use of color is the main color of purple, which represents the softness of the dance, and the secondary color with the color change from blue to green to indicate the movement, and the interval in accordance with the layout of the control of 8px.

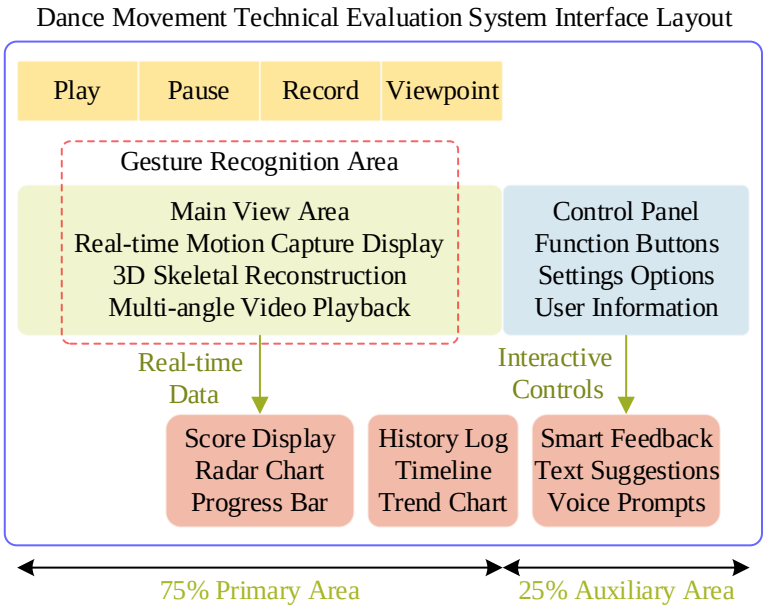

Figure 2: System interface layout

User-centered adaptive design is a way to automatically set up a page based on user characteristics identified by machine learning, in which the user's browsing time, mouse click position, and frequently used function modules are collected as a basis. And the recommendation engine will be based on this information to speculate that the user may be interested in other items, and be highlighted in the page. For example, for novice users focusing on the basic action guidelines and error feedback area, and for veteran users focusing on the technical analysis and performance evaluation area, the information visualization design is presented in a multi-level presentation to achieve the evaluation results on the radar chart. The chart describes the balanced degree of ability in each dimension, the progress bar uses the time line control to reflect the learning process and skill growth of students; 3D modeling dynamically displays the human body's movement trajectory to realize the demand for multi-angle, three-dimensional viewing; the color scheme is based on the psycho-cognitive science of the design concept that green represents correctness, yellow represents the need for correction, and red represents error, and a unified visual language reduces the cognitive load on the user. Functional animation emphasizes the principle of practicality, such as the use of jogging curves to achieve smooth animation transitions; confirmation animation emphasizes the principle of feedback, such as micro-interactive animation can give users clear feedback on the operation in a timely manner; progress indicator emphasizes the principle of anticipation, such as loading skeleton screen allows users to have a psychological expectation of the page loading; universal animation emphasizes the principle of inclusiveness, such as in accordance with the web content of the easy to read and operate the norms of accessibility to meet the needs of various types of users.

4 System implementation and test results

4.1 System implementation

Based on the hybrid cloud native architecture development of action technology evaluation system, in order to improve the model training efficiency and facilitate rapid iterative updates on the basis of ensuring that it can be operated in the same way on different platforms, in this

section, the PyTorch deep learning framework (version 1.12) and the OpenCV computer vision library (version 4.6) are selected to perform image recognition and feature extraction, and use the Flask Web development framework (version 2.2) for backend construction. The front-end uses Vue.js 3.0 responsive framework to complete the user interaction interface development.

The algorithm part of this system runs in a NVIDIA RTX 4090 PC with up to 48GB of machine memory, which ensures its computing ability for large-capacity data to a certain extent. At the same time, SSD memory is used to save and read data to meet the system requirements, and the program writing work is completed by combining the agile development mode, placing the source code in Git for subsequent updates and maintenance, and automatically constructing the test pipeline. The acquisition program is designed in asynchronous mode, where multiple processes are used in Python to acquire data from each sensor simultaneously and a separate process is created for each sensor to prevent conflicts. The time of all the child processes is unified in the parent process and all the data information is summarized. The depth camera is driven using Intel RealSense SDK 2.0. The inertial measurement unit is accessed using a serial communication protocol. The audio is recorded using the PyAudio library at a sampling frequency of 44.1 kHz, the sensors are calibrated during the initialization process, the internal and external reference matrices of the camera are obtained using the checkerboard calibration card, the inertial measurement unit is calibrated to remove the zero drift interference in the stationary state, and the timestamps of the sensors are aligned to the same reference system according to the network time protocol.

The specific system architecture is shown in Figure 3, in the feature extraction part, there is a large amount of arithmetic, long response time, to address this problem, this paper quantizes the model and uses knowledge distillation methods to further reduce the model size, the attitude detection network after quantization (INT8) the model parameters are reduced to 25% of the original model, the inference rate is increased by about 2.3 times, and the accuracy rate is reduced by about 5%.On GPUs The graphics memory is dynamically expanded on demand, and the batch size is automatically determined based on the input data size to prevent insufficient graphics memory. The multimodal feature fusion module utilizes CUDA parallel computing to accelerate matrix operations and uses DNN acceleration libraries to complete the efficient operations of convolution; in the data preprocessing pipeline, while the GPU performs inference calculations, the CPU asynchronously loads the next frame of data for preprocessing to make full use of the computational power of both the CPU and the GPU; and the algorithmic module is deployed with the TensorRT inference engine to the target device The algorithm module is deployed to the target device with the TensorRT inference engine. The algorithm module is deployed to the target device with the TensorRT inference engine. The inference acceleration is continued on the Jetson Xavier NX embedded platform by using graph optimization and kernel fusion methods, and a real-time inference speed of 30 fps is finally achieved.

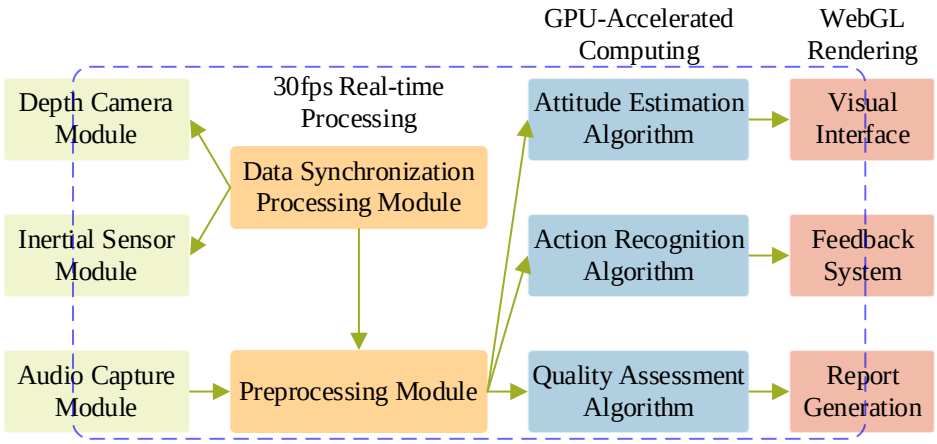

Figure 3: System implementation structure

4.2 Experimental design and data collection

4.2.1 Experimental design

In order to realize the purpose of scientific testing of the computer software, this paper adopts the randomized controlled experimental method to design the experimental program, and selects a total of 180 college students enrolled in A Dance Academy, B Dance Academy, and C Conservatory of Music as the research subjects, and ensures that all the participants are between the ages of 18 and 25 years old; the dance training time is at least two years, and at most eight years. The majors were classical dance, folk dance, modern dance, and ballet. Then, they were divided into three levels according to their skill level: beginner (60 students with 2-3 years of study), intermediate (60 students with 4-5 years of study), and advanced (60 students with more than 6 years of study), and each level was equally divided into two groups, i.e., the experimental group and the control group (30 participants in each group).

In this study, a timetable and a standardized operational protocol were used to ensure that all subjects underwent the experimental procedures within 8 weeks i.e. 1 week as pre-testing period, 6 weeks as treatment period and 1 week as post-testing period, where the baseline level of each subject was determined such as Standardized Dance Steps Test, Physical Functioning Test, and Cognitive Learning Test etc., which were used as a starting point for comparing the changes thereafter. During the intervention training phase, the experimental group utilized the movement technology evaluation system designed in this paper to assist training three times a week for a total of 90 min; the control group adopted traditional teaching methods to implement the same intensity of training.

The same test site, i.e. a standard dance studio, was set up, and the room temperature was kept constant at 22℃~24℃, relative humidity was controlled at 45%~55%, the illuminance was adjusted to the same value of 500 lx, the dance floor was paved with a professional dance floor rubber and the safety factor was guaranteed; the surrounding walls were painted with light-colored paints as much as possible; the music playback equipment was tested in advance to avoid the problems of noises and uneven sounds. The expert team included 15 senior teachers who have been engaged in dance teaching for more than 10 years, and used blind evaluation to eliminate the influence of subjective factors on the scoring, and scored objectively according to the technical assessment indexes proposed by the International Federation for Dance Education (IFDE) in terms of accuracy, artistry, and creativity, with a full score of 10 points for each item.

4.2.2 Data acquisition

The motion data of all the subjects are shown in Table 3.A variety of sensors were used for motion capture.A total of 8 Intel depth cameras and 6 sets of wireless IMUs were used.The 8 Intel depth cameras were arranged in a circular wrap-around arrangement.Each camera had a resolution of 640×480 pixels and a frame rate of up to 30 Fps.RGB images and depth images were acquired.The 6 sets of wireless IMUs were mounted on the subject's head, chest, left wrist, right wrist, left ankle, and right ankle, and were used to detect linear acceleration, angular velocity, and magnetic force values on each part with a sampling rate of 100 Hz. A directional microphone system was used to make audio recordings, and was set to a sampling frequency of 44.1 kHz to record the background music and sound; the subject's heart rate was measured using a wearable wireless heart rate sensor, and was recorded.

Table 3: Data collection and summary

Data type	Collection equipment	Sampling frequency	Data format	Storage capacity requirement
RGB video data	RealSense D455×8	30fps	H.264 encoding	480 MB/ minute
Depth image data	RealSense D455×8	30fps	16-bit depth value	320 MB/ minute
Inertial motion data	Wireless inertial measurement unit×6	100Hz	Floating-point array	2.1 MB/ minute
Audio signal data	Directional microphone	44.1kHz	WAV format	5.3 MB/ minute
Physiological data of heart rate	Chest strap heart rate monitor	1Hz	Integer value	0.06 MB/ minute
Expert scoring data	Manual annotation	Discrete time point	JSON format	0.01 MB/ minute
Environmental parameter data	Environmental sensor"	0.1Hz	CSV format	0.001 MB/ minute

The network time protocol is used to ensure the time synchronization of the data; the PC host uses an Intel processor, 64G RAM and NVIDIA RTX 4090 graphics card to carry the system for work; the SSD uses a 2T capacity large-capacity SSD storage device to ensure its fast reading speed; and the communication between the acquisition devices adopts a Gigabit LAN to ensure the speed of data communication. Quality control includes daily calibration of the equipment, ensuring data integrity and removing outliers, etc. Before each test, the sensors are calibrated and the internal and external parameters of the depth camera are calibrated using a standard checkerboard grid. The inertial guidance system is calibrated using the static adjustment method for zero bias, and the signal is judged to be normal or not during data acquisition, and if there is a loss of numbers or anomalies, it is promptly investigated and re-sampled.

4.3 Analysis of system test results

4.3.1 Analysis of the effects of action technology assessment

In order to verify the effectiveness of the previously proposed action technique recognition algorithm based on spatio-temporal convolutional neural network combined with multiscale attention mechanism, this paper compares the three methods, namely, traditional Hidden Markov (HM), Convolutional Neural Network-Long and Short-Term Memory Neural Network (CNN-LSTM), and Spatio-Temporal Graphic Convolutional Neural Network (ST-GCN), to the

method of this paper. Specifically, the performance of the algorithms is compared in terms of four dimensions: accuracy, F1, value inference time, expert relevance and cross-style generalization, and the test results are shown in Table 4. It can be seen that our method achieves optimal results in all 5 dimensions. Among them, three cycles of improvement were carried out with 15 professional dancer teachers working together, and the proportion of artistic beauty values in the loss function was adjusted according to the dancers' opinions in each round, and the Kendall's rank correlation coefficient between the test scores and the dancers' scores was increased to more than 0.89; the algorithm stability was increased by using adversarial training, and Gaussian white noise and joint coordinate interference terms were added to the input data, and the cross-stylistic generalization ability dimension of the model to 0.85, an improvement of 0.38 with respect to the HM method.Since the algorithm in this paper takes into account the characteristics of dance and utilizes a graph neural network enhanced by long and short-term memory for temporal modeling, it has made a breakthrough in the two key metrics, namely, the accuracy rate (93.7%) and the F1 (0.91), and is able to learn well the dance represented by a sequence of combinations of movements that spans over a long period of time continuity and correlation between movements.

Table 4: Algorithm Performance Comparison

Algorithm model	HM	CNN-LSTM	ST-GCN	Our method
Accuracy rate (%)	68.2	79.8	85.3	**93.7**
F1 value	0.65	0.77	0.83	**0.91**
Inference time (ms)	45	82	67	**53**
Expert relevance	0.51	0.68	0.76	**0.89**
Cross-style generalization	0.47	0.63	0.72	**0.85**

4.3.2 Analysis of Interface Testing Effects

The interface testing utilized a comparative testing methodology and conducted usability testing over a period of 8 weeks in a user group consisting of 120 people from different backgrounds and measured the design outcomes of the interface in terms of task completion rate, operational efficiency, error rate, and satisfaction as metrics, and the results of the testing are shown in Table 5. Test results showed that the system achieved a task completion rate of 89.6%, and the average operation time was shortened by 10.5% compared to the traditional interface. The customer satisfaction score is 9.2, which is higher than the average score of 7.95 points of the same type of products. Adopting the virtual document object model technology and component loading mechanism for the human-computer interface design, the rendering frame rate can be guaranteed to be more than 60fps under the low-end hardware platform, and the response latency can be controlled within 100ms to achieve the real-time response effect, so as to provide a good user experience environment for human-computer interaction to support the development of the dance teaching application system.

Table 5: Interface test results

Design plan	Task completion rate	Operational efficiency	Error rate	User satisfaction
Traditional desktop	77.2%	72.4%	13.8%	6.8/10
Mobile adaptation	71.7%	73.0%	15.2%	7.9/10
Immersive virtual reality	80.3%	72.6%	14.9%	8.4/10
Augmented reality version	84.8%	78.4%	11.6%	8.7/10
This system solution	**89.6%**	**82.9%**	9.8%	**9.2/10**

4.3.3 Analysis of the effects of student performance improvement

After completing a rigorous controlled experiment over a period of 8 weeks, the effect of the use of this system on students was investigated on 8 items: technical accuracy, coherence, rhythmicity, expressiveness, overall rating, difficulty of mastery, ability to correct errors, and proficiency. The results of the pre- and post-test experiments are shown in Table 6, and the results of the improvement comparison are shown in Figure 4. The quantitative statistics are based on the pre- and post-tests as well as comparisons between different groups. The percentage improvement in technical accuracy was significantly greater in the experimental class (15.9%) than in the control class (7.5%) (Cohen's d=1.47). Completion was 14% in the experimental group and 6.2% in the control group (Cohen's d=1.23); Rhythmicity was 15.1% in the experimental group and 6.4% in the control group (Cohen's d=1.56); and Overall Performance was 13.2% in the experimental group and 5.5% in the control group (Cohen 's d=1.41). According to the data in the table, the average time for students in the experimental class to complete the standardized movements decreased from 6.7 to 4.5, indicating that the motor skill evaluation system established in this paper can effectively improve students' sports performance.

Table 6: The results of the pre - and post-test experiments

Evaluation dimension	Experimental Group		Control group	
	Pre-test	Post-test	Pre-test	Post-test
Technical accuracy (%)	67.3 ± 67.3	83.2 ± 6.7	66.8 ± 8.5	74.3 ± 7.9
Fluency of movement (%)	72.1 ± 7.6	86.1 ± 5.4	71.9 ± 7.8	78.1 ± 6.8
Rhythm coordination (%)	69.4 ± 9.1	84.5 ± 6.2	68.7 ± 9.3	75.1 ± 8.1
Artistic expressiveness (%)	70.8 ± 8.4	84.0 ± 6.9	70.2 ± 8.7	75.7 ± 7.6
Comprehensive score (%)	69.9 ± 7.8	84.5 ± 6.1	69.4 ± 8.1	75.8 ± 7.2
Learning efficiency	6.7 ± 1.2	4.5 ± 0.8	6.8 ± 1.3	6.2 ± 1.1
Error correction speed (s)	45.3 ± 8.7	26.4 ± 5.2	44.9 ± 9.1	39.7 ± 7.8
Skill retention rate (%)	71.3 ± 9.1	87.3 ± 6.4	71.2 ± 7.8	71.6 ± 8.9

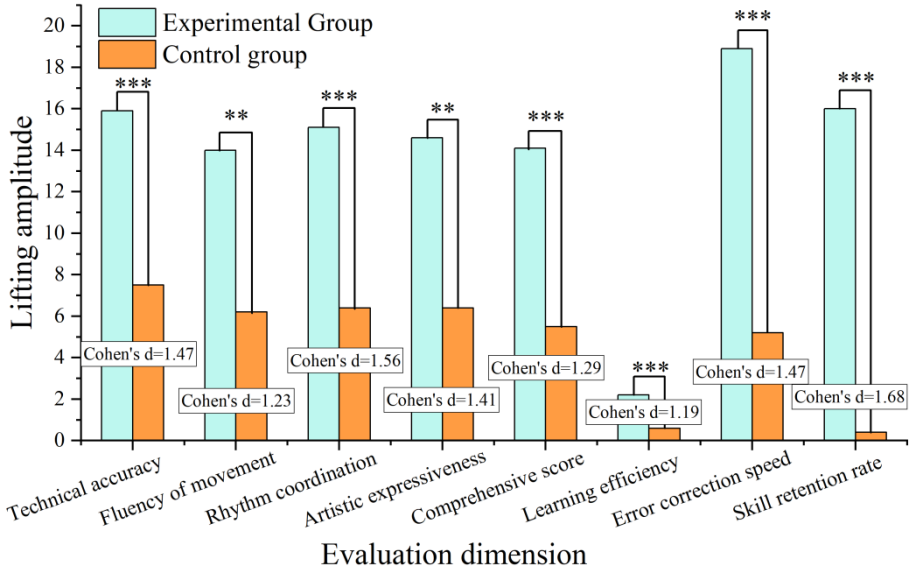

Figure 4: Comparison results of the extent of improvement

5 Conclusion

The temporal and spatial two-stage graph convolution model proposed in this paper solves the previous problem of incomplete classification of difficult dance movements by performing graph construction of continuous human skeletal points in two steps of time and space, and achieves 94.7% movement classification accuracy and 89% Kendall rank correlation (compared with manual evaluation results). Finally, after more than two months of teaching practice tests, it was found that using the system in this paper can effectively improve the dancers' movement standardization, coherence, sense of rhythm, and emotional expression, with an average improvement rate of 14% to 16%. In addition, during the training process, the standardized average number of times that students mastered the correct movements decreased from 6.8 to 4.6 times, which shows that the evaluation system has a certain positive impact on improving the quality of students' movements.

In the subsequent research, we will further carry out accurate identification of multiple continuous movements of greater difficulty, and further strengthen the research efforts on the comprehensive processing capability of multiple modes of information, so as to realize the design of adaptive personalized training program based on the evaluation system, and at the same time, we will also promote it to other dance types, such as street dance, Latin dance, and other different kinds of dance movements. As for the technical aspects, the system's anti-interference ability, the ability to operate in complex environments, the use of edge computing to reduce the dependence on the network, the optimization of the human-computer interaction interface, and the enhancement of the combination with other educational and teaching methods should be enhanced. In summary, this study can be used as an important reference basis for computer-assisted art education teaching, which to a certain extent proves that artificial intelligence technology can improve the quality of teaching as well as students' performance, and it also lays a solid foundation for the realization of intelligence in traditional art education.

Funding

This work was supported by Sponsored project of Master's Degree Authorization Program of Physical Education Major of Guangxi Minzu Normal University in 2025 (1024-3950040001).

About the Author

Jin Gao was born in Leping, Jiangxi, P.R. China, in 1980. She received her bachelor degree from Central China Normal University and received her master degree from Jose Rizal University.She works at the College of Art, Guangxi Minzu Normal University. Her research interests include dance education and education administration.

Qian Zhou was born in Zhengzhou, Henan, P.R. China, in 1981. He received the bachelor degree from Beijing Sports University, P.R. China, and received the Master degree from Wuhan sports University, P.R. China. Now, he received the Docter degree from Jose Rizal University.He works in School of Physical Education College ,Guangxi Minzu Normal University, and his research interests include physical education and training, national traditional sports.

References

[1] Li, Z., & Wong, K. K. (2023). Challenges and opportunities: Dance education in the digital era. Applied degree education and the shape of things to come, 29-48.

[2] Melati, A. (2021). Indonesian Dance Education in Taiwan: Methods and Experiences as a Teacher. Journal of Urban Society's Arts, 8(2), 76-86.

[3] Gao, J. (2018). Research on the weakness of college dance teaching and its innovative solutions. In 2018 8th international conference on education, management, computer and society (EMCS 2018) (pp. 801-803).

[4] Yin, Y. (2024). Analysis of the Current Status and Development Strategies of Dance Education in Vocational Colleges. Journal of Modern Education and Culture, 1(2).

[5] Zejing, M., & Luen, L. C. (2024). A Comprehensive Method in Dance Quality Education Assessment for Higher Education. Int. J. Acad. Res. Bus. Soc. Sci, 14, 298-306.

[6] Suson, R., & Ermac, E. A. (2020). Computer aided instruction to teach concepts in education. International Journal on Emerging Technologies.

[7] Lin, Y. N., Hsia, L. H., Sung, M. Y., & Hwang, G. H. (2019). Effects of integrating mobile technology-assisted peer assessment into flipped learning on students' dance skills and self-efficacy. Interactive Learning Environments, 27(8), 995-1010.

[8] Lucero, R. (2021). Effects of instructional materials in multimedia computer-assisted instruction in teaching folk dance. Edu Sportivo: Indonesian Journal of Physical Education, 2(1), 40-50.

[9] Shang, Y. J., & Suo, D. D. (2021, June). Design of dance action simulation teaching system based on cloud computation. In International Conference on E-Learning, E-Education, and Online Training (pp. 453-463). Cham: Springer International Publishing.

[10] Zhao, Y., & Yang, H. (2023). Implementation of Computer Aided Dance Teaching Integrating Human Model Reconstruction Technology. Computer-Aided Design and Applications, 21(S10), 196-210.

[11] Wang, Y., & Wu, Z. (2023). Dance motion detection algorithm based on computer vision. International Journal of Advanced Computer Science and Applications, 14(10).

[12] Xu, W., Xing, Q. W., Zhu, J. D., Liu, X., & Jin, P. N. (2023). Effectiveness of an extended-reality interactive learning system in a dance training course. Education and Information Technologies, 28(12), 16637-16667.

[13] Chen, J. (2023, December). Motion Decomposition and Guidance Technology in Dance Teaching Based on Motion Feedback System. In 2023 International Conference on Intelligent Computing, Communication & Convergence (ICI3C) (pp. 162-166). IEEE.

[14] Wang, Z., & Dong, J. (2023). Design of dance data management system based on computer-aided technology under the background of internet of things. Comput. Aided Des. Appl., 20(S2), 45-55.

[15] Xu, L. J., Wu, J., Zhu, J. D., & Chen, L. (2025). Effects of AI-assisted dance skills teaching, evaluation and visual feedback on dance students' learning performance, motivation and self-efficacy. International Journal of Human-Computer Studies, 195, 103410.

[16] Guang, F., & Xueliang, Z. (2025). Research on the impact mechanisms of immersive virtual reality technology in enhancing the effectiveness of higher folk dance education: Base on student perspective. Education and Information Technologies, 1-39.

[17] Li, N. (2025). AI-Based Dance Evaluation Systems and Personalized Instruction: Possibilities and Boundaries of Dance Education in the Intelligent Era. Journal of Education, Humanities, and Social Research, 2(4), 43-52.