

Research Article

Design of a Professional Sports Competition Adjudication System Based on Data Analysis and Action Recognition Algorithm

Qianfeng Lin ^{1,2} and **Jiashu Zou** ³

¹Department of Physical Education, Lanzhou University of Technology, Gansu, Lanzhou 730000, China

²Shanghai Sports College, Shanghai 200000, China

³Shanghai Lixin College of Finance and Accounting, Shanghai 200000, China

Correspondence should be addressed to Qianfeng Lin; 2021151077@sus.edu.cn and Jiashu Zou; 20170045@lixin.edu.cn

Received 9 March 2022; Revised 8 April 2022; Accepted 22 April 2022; Published 18 May 2022

Academic Editor: Sheng Bin

Copyright © 2022 Qianfeng Lin and Jiashu Zou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper uses data analysis and action recognition algorithms to conduct in-depth research and analysis of professional sports competition judging and designs a professional sports competition-assisted judging system for use in actual judging. In this paper, a wearable motion capture system based on an inertial sensing unit is developed and designed for kayaking technical motion monitoring to achieve the acquisition, analysis, and quantitative evaluation of kayaker motion data. To limit the gyroscope and pose estimation error, a gradient descent method is used for multisensor data fusion to achieve athlete pose update, and a quaternion-driven human skeletal vector model is proposed to reconstruct the kayaker's paddling technical movements. By calculating the angular sequences of the left shoulder, right shoulder, left elbow, and right elbow joints of the athlete's upper limbs and comparing them with the optical motion capture system, the results show that the motion capture system developed in this paper is comparable to the optical motion capture system in terms of measurement accuracy. It ultimately affects the result of pose estimation. Therefore, high-resolution networks and low-resolution networks can continuously maintain high-resolution features of the image by allowing each representation layer to repeatedly accept the representation information of other networks. A step matrix model is constructed to encode the multiscale global temporal information of action sequences, and action classification is achieved by calculating the response of the step matrix of test samples to the step matrix of each category of actions. The algorithm achieves 78.96%, 91.84%, and 91.18% accuracy of action classification on the Northwestern-UCLA database, MSRC-12 database, and CAD-60 database, respectively. The designed visual motion tracking system was applied to record the motion data of the experimental subjects in the fine motion assessment task and construct the motion assessment database. The experimental results show that the average error between the prediction results of the proposed action assessment method and the manual scoring is 1.83, and the automated assessment of fine movements is effectively realized.

1. Introduction

Fine classification and evaluation of human actions based on visual data refer to the prediction of action categories, temporal fine classification, and action quality scoring of human actions occurring in videos using computer vision technology, which belongs to the domain of computer vision. Action classification can be divided into the classification of action categories and localization of action occurrence times depending on the classification objectives [1]. Action assessment refers to the quantitative scoring of the occurring human actions, which can

also be regarded as the fine classification of the quality of action completion. Action classification is the basis of action evaluation, and, after determining the action category, it is possible to design evaluation algorithms to achieve a quantitative evaluation of human actions based on action characteristics. The study of fine classification and evaluation of human movement based on visual data can help solve three major problems in human movement understanding: distinguishing movement categories, locating movement occurrence time, and assessing movement completion quality [2]. The amount of computation decreased by 40.8%, but the accuracy increased

by 0.7%. Compared with the pretrained HRNet-32 * model, the accuracy decreased by only 0.3%. With the arrival of the fitness boom, many fitness problems have also entered the public eye. Due to unscientific exercise methods, athletes can suffer a variety of injuries while exercising, but these losses can be avoided if the tester is training in a correct movement pattern. For example, an athlete doing a deep squat does not realize that, because of his poor ankle flexibility, the curvature of his lumbar spine will bend when he squats to the bottom, so, during the usual training process, his lumbar spine's bending time will be more, and then it will cause back pain; this problem can be avoided.

Due to many participants, coaches have limited time to keep their eyes on each athlete, and it is difficult to accurately recapitulate the movement characteristics and technical defects of each athlete by memory in the technical summary stage at the end of training. Nowadays, various athlete training assistance systems using multisensor data fusion techniques have been widely used in the daily training monitoring of athletes for analyzing their skill performance, but most of these training assistance systems are expensive, require special experimental sites, and are inconvenient to use [3]. This paper intends to design a wearable motion capture system based on inertial sensors for kayak monitoring, which can collect technical movement data of kayakers conveniently and accurately without affecting the normal performance of athletes [4]. It will help the coaches to analyze the athletes' technical level characteristics and defects and make more targeted training plans, which will promote the athletes' training and technical level improvement.

We develop a system to detect the standardization of movements in the field of motion analysis. The best available detection network for human skeletal key point detection is constructed in a lightweight manner based on the analysis of the lightness of the convolutional neural network by capturing the action video with a common RGB camera. The data features in the standard action are then extracted and recorded in the library using this lightweight detection model, and the redundant frames in the training video are removed and aligned with the standard action frames for action alignment. The final realized action training system can compare and analyze the human action information in the training video with the data features in the standard library and give improvement advice, which can provide some help to the training of dance and sports industries, reduce the labor input, and improve the training efficiency of users. Therefore, it is necessary to propose a complete automated volleyball training video analysis process and system based on multiview training videos, combined with artificial intelligence technology. By realizing automatic analysis and statistics of volleyball training contents, it can reduce the degree of manual involvement in the training process and further improve the detection and statistical accuracy, making it possible to achieve rapid problem location in lengthy training videos.

2. Related Works

Among the systems that analyze the trajectory of the ball in the video, a better known one is the Hawkeye system

invented by computer experts in the UK [5]. The system consists of six to seven high-speed cameras connected to a computer that is placed around the field of play and they shoot from different angles. The computer then reads the video from each angle in real time, tracks and calculates the position of the ball from multiple angles at over 100 operations per second, and displays it as a VR animation [6]. The most important feature of this system is its high accuracy, which guarantees the detection of the ball's landing position within a 5 mm error range. With its high accuracy, it has become the official support system for major tennis tournaments, helping the umpires to determine the penalty situations that are difficult to define by the naked eye [7]. Therefore, first, it should provide a presentation layer operation page with clear logic, reasonable interface, and analysis function as the core and, second, provide various data-rich and intuitive visual display pages to provide users with a variety of training videos and display methods of analysis data. It is also necessary to provide reasonably designed data retrieval and background management pages. It is also possible to make a qualitative and quantitative analysis of human movement and to judge the risk of sports injury based on the degree of mechanical deviation of certain sports movements in the picture and to assist coaches in formulating adjustable training for movements [8].

Based on the human upper limb movement model, an autonomous rehabilitation assistance training system for the upper limb exoskeleton is designed based on the Kinect 2.0 sensor. The system can give corresponding training guidance and transmit the training status to the physician according to the different rehabilitation stages of the patient, which helps the patient's rehabilitation and facilitates the physician's training guidance [9]. Applying deep learning technology to the field of medical recovery, a method based on deep learning technology to identify brain wave signals of stroke patients is proposed, and a therapeutic rehabilitation system for stroke patients with hemiplegia of the hands is implemented in combination with other devices such as brain-machine devices and pneumatic rehabilitation gloves, which can provide independent rehabilitation training for the left and right hands of patients simultaneously [10]. To address the challenges of high complexity and inefficiency of existing networks, a novel convolutional neural network architecture, called Efficient Pose, is proposed. This network provides an efficient and scalable new model for single-person skeletal key point detection using a multiscale feature extractor and moving inverted bottleneck convolution, a computationally efficient detection block [11]. The first single-network whole-body skeletal key point detection algorithm was proposed to simultaneously locate key points on the body, face, hands, and feet. The method is improved based on Open Pose, but, unlike Open Pose, it does not require running a separate network for each candidate object such as hands and feet, resulting in a rapid increase in detection speed in human detection, especially in multi-person scenes [12]. Higher-handed is a bottom-up method, which uses multiresolution supervision and multiresolution aggregation methods to enhance the training and inference of the network, respectively, to solve the problem of scale

variation in multiperson pose estimation bottom-up, and to locate key points more accurately for small-scale people.

As a traditional motion analysis method, the video method is relatively straightforward and does not require the athlete to wear any special apparatus to obtain the athletes' game data, even in official games. However, this method is not efficient, and, in the subsequent analysis process, a coach needs to watch the athletes' video data frame by frame and judge the athletes' good or bad technical movements and skill level by experience. With the help of video analysis software, it is also necessary to put some colorful markers on the athletes in advance to make the video analysis software easy to identify the capture points; even so, due to the inherent quality problem of the video method, it still happens that important markers are not captured and there is a need to make up the points manually. For the setter training subjects, the focus is on whether the ball falls up to the standard and whether the incident ball trajectory affects the athlete's posture; for the spiking training subjects, the focus is on whether the ball falls up to the standard and the spiking speed; for the blocking and padding training subjects, the main concern is whether the ball crosses the net and whether it goes out of bounds. Limited by the video clarity, shooting frame rate, occlusion, shooting angle single, and other issues, this method is bound to have a large measurement error. The shooting range and distance of the camera are limited. In order to obtain a wide range of video data, multiple cameras must be set to shoot from multiple angles. The operation is cumbersome and requires a lot of labor cost.

3. Action Data Recognition Algorithm Design

According to the different data sources, action classification methods can be divided into RGB video-based action classification methods, depth image-based action classification methods, and skeletal data-based action classification methods. Compared with other modal data sources, 3D skeletal data has a simple structure and small data volume and is easier to extract low-dimensional effective features. Moreover, 3D skeletal data can effectively reduce the effects of lighting changes and background noise and achieve a clear description of human posture and motion state [13]. In addition, with the development of 3D motion capture technology and skeleton estimation algorithms, 3D skeletal data has become easier to obtain, either from motion capture devices such as Kinect or from RGB images or depth images through skeleton estimation algorithms. Therefore, this chapter focuses on the motion classification algorithm based on skeletal data.

The original skeletal sequence was firstly normalized in terms of viewpoint and scale so that the transformed skeletal sequence is robust to shooting angle, standing position, and skeletal size. Then a dynamic segmentation scheme is designed to segment the skeletal sequence into multiple skeletal segments, each of which consists of multiple consecutive frames of skeletal data with spatial similarity. The skeletal segments can be considered as the units composing human actions so that a complete action is segmented into multiple action units with smaller temporal granularity. All

action units composing the training sample are clustered according to their spatial features and local time-domain features to form a dictionary of key fragments, such that each skeletal sequence is encoded as an ordered sequence of words. Subsequently, the multiscale global time-domain information of the word sequences is encoded into the step matrix by the proposed multiscale time-step matrix model. In this way, the spatial and local temporal information of the skeletal sequence is encoded in the key segment descriptors, while the multiscale global temporal information is retained in the time step matrix. Finally, the action classification is achieved by calculating the responses of the step length matrix of the test samples and the step length matrix of each type of action, and the flowchart of the action classification algorithm proposed in this chapter is given in Figure 1.

Motion capture sensors use a variety of different principle methods to acquire skeletal data; for example, optical systems use stereo vision to triangulate the 3-dimensional spatial position of the target object; data acquisition devices use special markers attached to the surface of the object to acquire the corresponding spatial position; some systems use retroreflective materials by adjusting the threshold of the camera so that skin, walls, and so forth are ignored and only objects with markers emit light; some systems use one or more LEDs and use software to identify them by their relative positions. Such motion capture systems can acquire the exact position of human skeletal joint points and are mainly used for truth building of skeletal joint point position coordinates in the pose estimation database. Cross-validation can usually solve this problem very well. Cross-validation is sometimes called circular estimation. In this paper, the actual training set size is considered comprehensively, and 5-fold cross-validation is proposed; that is, the dataset is randomly divided into 5 subsets at the beginning, and 4 subsets are used for model training and a separate subset of samples is used for validation.

To reduce the sensitivity of the skeletal sequence to the shooting viewpoint, the original skeletal sequence needs to be transformed to a reference coordinate system determined by the skeletal sequence itself. As mentioned before, the skeletal frame-based coordinate transformation treats each frame separately and loses the relative motion information between frames. The skeletal sequence-based transformation preserves the interframe relative motion of the original skeletal data by performing the same coordinate transformation on all skeletal frames in the sequence. Therefore, the skeletal sequence-based transformation is used here to transform the coordinates of the joints of all frames in the original skeletal sequence to a unified new reference coordinate system through geometric transformations such as translation and rotation, and the coordinate transformation of the joints is given by

$$[x', y', z']^T = T(R_x^\alpha, R_y^\beta, R_z^\delta, d)[x, y, z, 1]^T, \quad (1)$$

where $[x', y', z']$ are the original joint point coordinates, $[x', y', z']$ are the joint point coordinates after transformation, and $T \in \mathfrak{R}^{2*2}$ is the translation-rotation transformation matrix, which is defined as follows:

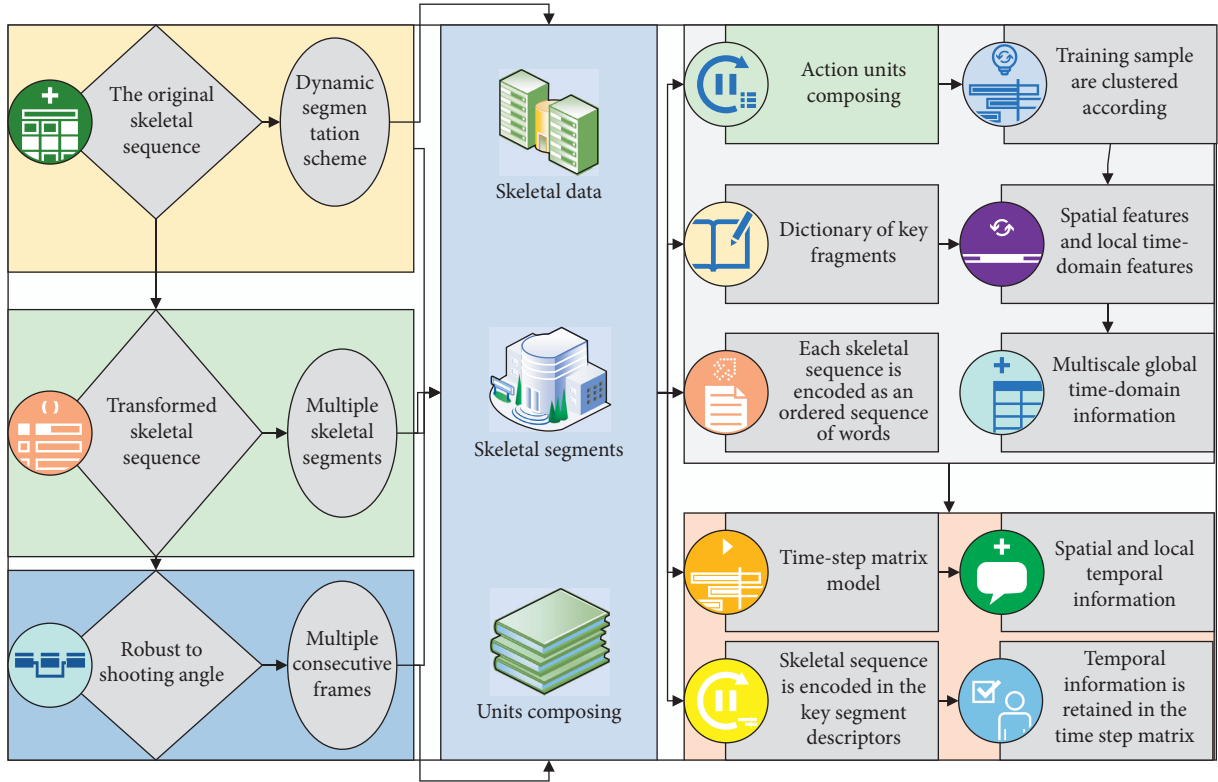


FIGURE 1: Framework diagram of human action classification algorithm.

$$T = \begin{bmatrix} R_x^\alpha & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_y^\alpha & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_z^\alpha & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_3 & d \\ 0 & 1 \end{bmatrix}. \quad (2)$$

The place where the three fundamental axes intersect is the ideal origin position; then, combined with the human skeletal information, the ideal origin position is the “hip joint center,” so the average of the “hip joint center” coordinates of all frames in the skeletal sequence is defined as the origin of the new coordinate system. This means that the original skeletal sequence is moved to a new coordinate system with its average “hip center” position as the origin [14]. Then, the value of the origin coordinate O is

$$O = \frac{1}{N} \sum_{n=1}^N J_2^n. \quad (3)$$

The first dynamic segmentation scheme is proposed to divide the normalized skeletal sequence into multiple skeletal segments automatically and then extract the spatial features and local time-domain features of the skeletal segments and cluster them to form a key segment dictionary and replace the skeletal segments in the skeletal sequence with the key segments closest to it in the key segment dictionary, so that the skeletal sequence is encoded as a word sequence.

Limited by the industrial design and processing process, the raw data collected by the MEMS sensor will inevitably have noise interference, which in turn affects the accuracy of the subsequent pose solution. The explained variance of each component is 64.0%, 30.3%, 2.9%, and 1.9%, respectively.

After PCA dimensionality reduction, the combined accuracies of decision tree, SVM, KNN, and Bagging are 95.9%, 97.8%, 97.4%, and 97.6%, respectively. Therefore, an error correction model of the sensor is needed in practical applications to constrain the influence of sensor measurement error on the subsequent attitude solution. The error model of the accelerometer is shown in the following equation:

$$a_i(t) = \hat{a}(t) - g_i(t) - b_{a,i}(t) - \varepsilon_{a,i}(t), \quad (4)$$

where $b_{a,i}(t)$ is the actual output value of the accelerometer in the i -axis at time t , $a_i(t)$ is the acceleration generated by the accelerometer in the i -axis when the accelerometer is subjected to an external force, which can be considered as the acceleration generated by the limb movement of the athlete in this paper, $g_i(t)$ is the component of gravitational acceleration in the i -axis at time t , and $b_{a,i}(t)$ and $\varepsilon_{a,i}(t)$ are the zero bias and Gaussian white noise of the accelerometer itself, respectively. MEMS accelerometer has good static performance with a small measurement bias.

The error model of a gyroscope is like that of an accelerometer, with the difference being that the output value of the gyroscope is not affected by gravity, and the error model of a gyroscope can be written in the following form.

Put the sensor at rest, calculate the average output w_i of each axis inside the gyroscope during the rest state, consider this value as the zero offsets of the gyroscope, and subtract this value from the actual measured value of the gyroscope at any later moment, which is used to get the true value of the gyroscope more accurately, as shown in the following equation:

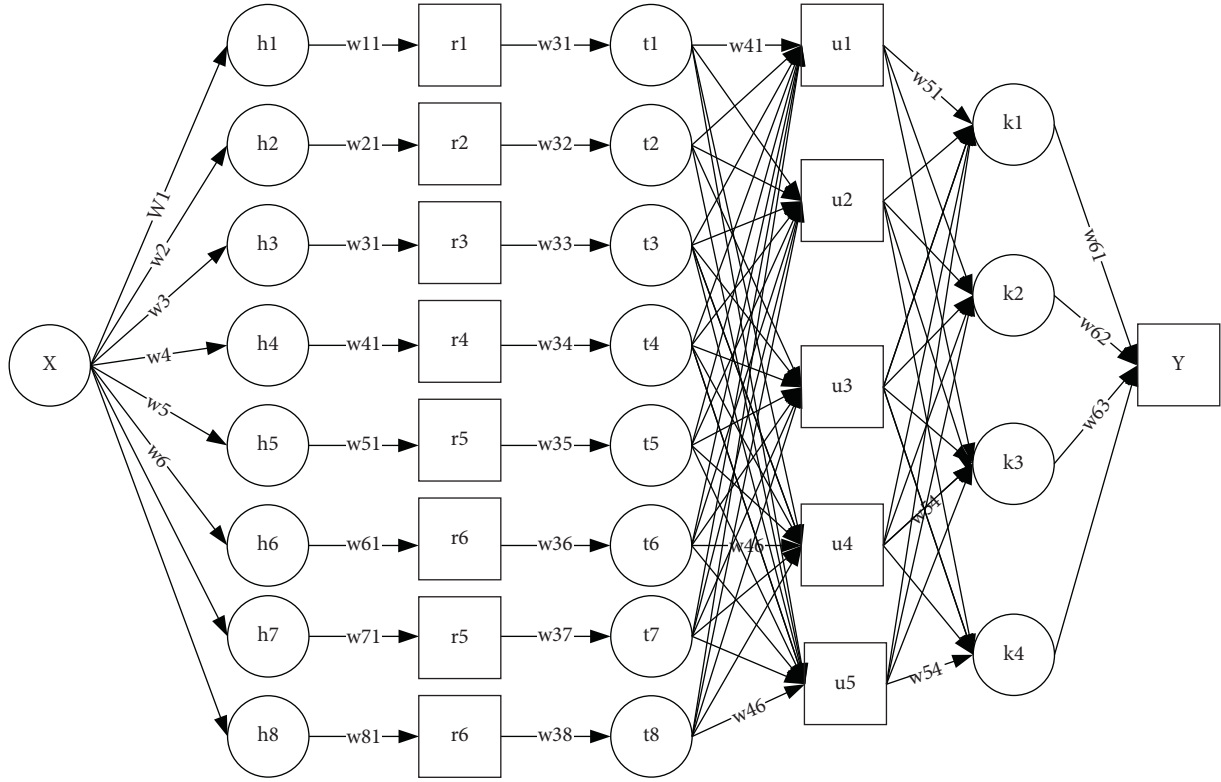


FIGURE 2: HRNet model structure.

$$\widehat{w}_i(t) = w_i(t) + \widehat{w}_i. \quad (5)$$

The magnetometer is susceptible to hard magnetic interference and soft magnetic interference in the environment, resulting in a large measurement error, and its simplified error model can be described by the following equation:

$$h = AH - B - \varepsilon, \quad (6)$$

where A is the soft magnetic interference term, B is the hard magnetic interference term, and ε is the Gaussian white noise in the environment. For the method of improving the perceptual field range by convolutional pooling operation, the quantization error is generated in the sampling if too deep cascades are used, and this problem is improved by the high-resolution network model that the system in this paper eventually uses in the personnel pose estimation task. A high-resolution network (HRNet) is one of the more advanced neural networks used in human pose estimation tasks and is a model of human pose estimation algorithm jointly researched by Microsoft Asia Research Institute and CSU. The structure of the network model is shown in Figure 2.

Most existing top-down methods generally use serial connections for feature information fusion. HDNet uses a parallel network structure to effectively solve the problem of image information loss caused by most existing network models during downsampling, that is, quantization errors caused by methods such as convolution and pooling that increase the scale of feature perceptual fields by reducing the resolution. High-resolution features have an important role in the pose estimation task [15]. It can be applied to data

collection of outdoor real boats and reconstruct the technical movements of athletes in 3-dimensional space through attitude calculation, which can provide coaches with movement information from multiple perspectives. The feature map loses image features after the downsampling and upsampling process, and the cascaded network structure will pass the resulting errors further and eventually affect the pose estimation results. Therefore, the high-resolution network and the low-resolution network achieve continuous maintenance of the high-resolution features of the image by allowing each representation layer to receive the representation information from other networks repeatedly.

The evaluation metrics of human skeletal key point models are currently available in two ways: one is the Percentage of Correct Key point (PCK) commonly used in single-person skeletal key point detection, and the other is the mean Average Precision (mAP) commonly used in multiperson skeletal key point detection, where, in competitions and academic papers, the PCK represents the percentage of the number of detected skeletal key points within a certain standardized distance from the true value. The normalized distance in the specific evaluation metric can be derived from the product of the longest distance of the human body and the threshold value σ , where σ is a decimal number in the range of $[0,1]$.

$$\text{PCK} @ \sigma = \frac{1}{N \sum_{t=1}^N \left(\left\| y_p^{t*} + y_p^t \right\|^2 / \left\| y_{\text{lip}}^{t*} - y_{\text{rsho}}^t \right\|^2 \geq \sigma \right)}. \quad (7)$$

Compared with HDNet, the number of parameters reduced by the model and the amount of computation are

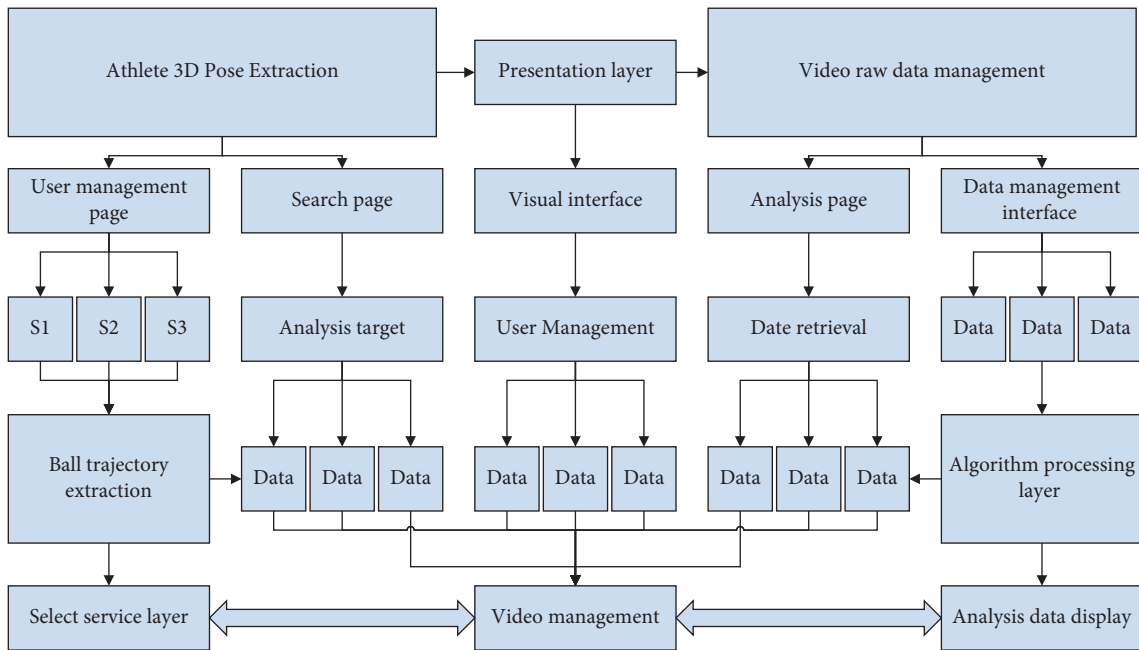


FIGURE 3: System logical architecture diagram.

41.7% and 40.8%, respectively, which are somewhat different from the theoretical reduction of 50.0% calculated according to (6) and (7), mainly considering the addition of the channel attention module at the end of the module, which has an impact on the number of parameters and the amount of computation of the model while enhancing the inter-channel feature fusion [16]. Compared with the HRNet-32 model without pretraining on the ImageNet dataset, HS-GattNet shows a 41.7% decrease in the number of parameters and a 40.8% decrease in the computation but a 0.7% increase in accuracy and a 0.3% decrease in accuracy compared with the pretrained HRNet-32 * model. The action unit is defined as a skeleton fragment composed of several consecutive skeleton frames with similar spatial structure, and the spatiotemporal features of the skeleton fragments are extracted for clustering to form a key fragment dictionary, and the skeleton sequence is represented as a word sequence. Moreover, the experimental results show that the HS-GattNet network maintains a similar distribution of other experimental measures, such as APM, APL, and AR, as the same underlying framework network, indicating that the robustness and generalization of the model are also better. The comparison with some SOTA models in terms of the number of parameters, computation, and accuracy shows that the HS-GattNet network has met the objective of reducing the complexity of the model while maintaining the accuracy of the model and improving the usefulness of the network.

4. Design of the Professional Sports Competition Evaluation System

The overall architecture design of the system is an important guiding part in the design and implementation of the whole

system, and the software logic architecture of a system has a key role in the overall system architecture design. In this paper, the system adopts B/S architecture from the actual situation of users, so the logical structure design of the volleyball training analysis system needs to ensure the design principles of integrity, hierarchy, and reliability of the architecture [17]. The system's software logical architecture determines the main structure, general features, and basic functions of the system. After the analysis of the system functional requirements, the logical architecture of the multiview volleyball training analysis system is shown in Figure 3.

According to the layered principle of system architecture design, the system is designed in a layered structure, and the logical architecture of the system is divided into five parts: performance layer, service layer, algorithm processing layer, data access layer, and data layer. If the tester is training in a correct movement pattern, for example, when an athlete is doing deep squats, he does not know that, due to the poor flexibility of his ankle joints, the curvature of his lumbar spine will bend when he squats to the bottom, so his lumbar spine will bend for more time during normal training, and then it will cause back pain; this problem can be avoided. The core function of the presentation layer is to provide users with a training video analysis operation interface and visual data display, and the secondary function is to provide data retrieval, management, and user management. Therefore, firstly, it should provide a logical and reasonable interface, with the analysis function as the core operation page of the performance layer, followed by a rich and intuitive visual display page of each data, providing users with a variety of training video and analysis data display methods, in addition to providing a reasonably designed data retrieval and background management page.

The service layer is designed based on the module division of the system. The services in this layer mainly realize the specific business logic of each module of the system, roughly by abstracting the underlying analysis algorithms and data, providing one or more service interfaces for the performance layer interface, and handling the requests of the performance layer [18]. For example, the data analysis module mainly obtains the annotation information of the users in the performance layer, calls the algorithms in the algorithm processing layer, completes the liaison between the layers, and provides automatic analysis services for the analysis modules in the performance layer; and, for the visual analysis module, it mainly liaises with the analysis data in the performance layer and the database, integrates all kinds of raw data and processing data, and provides the calling interface.

The data access layer is responsible for providing persistent data access to the service layer and storing the video data uploaded by users and the analysis data generated by the algorithm processing layer in the file system and database. Therefore, the data access layer realizes the liaison between the upper layer of the system and the database system and the file system. It is convenient for coaches to analyze the characteristics and defects of athletes' technical level and formulate more targeted training plans, which will promote the training of athletes and the improvement of their technical level. The data addition, deletion, modification, and checking are encapsulated in the data access layer, thus realizing the isolation of business logic and data storage layer and reducing the coupling degree of the system.

The database is mainly responsible for the storage of system data other than video files and analysis data files. The file system mainly specifies the organization of the basic files in the system including video data and analysis data. Good database design and unified file storage format can provide good underlying support for the data access layer, which has an important impact on the time delay and algorithm processing speed in all kinds of interactions in the system, as shown in Figure 4.

The visualization display module is mainly to display the data related to ball trajectory, personnel skeleton data, and other statistics generated by the automatic analysis algorithm of the data analysis module and to display them through charts and 3D animations so that coaches and athletes can easily locate the problems of each player in the training session. Combined with the research on coaches' training arrangement in actual training in the demand analysis, the visualization display should be designed according to different training subjects and the data visualization scheme of subsubjects.

In the five training subjects, the coach firstly focuses on the actual effect of the action, that is, whether the shot achieves the action goal, whether the serve is over the net or out of bounds, whether the second pass is passed to the designated position, whether the bucket is scored, whether the mat is saved successfully, and whether the block is blocked successfully [19]. Most of these action objectives are related to the ball trajectory associated with that action. For serve training subjects, the coach mainly focuses on the serve

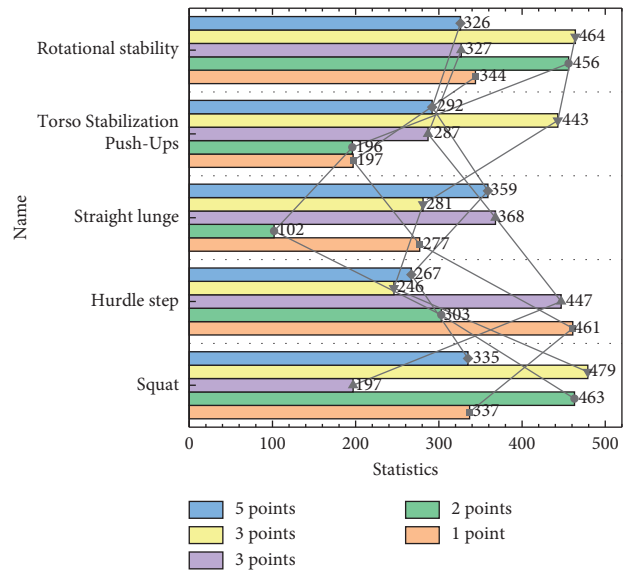


FIGURE 4: Number of individual action scores in the dataset.

ball speed, whether ball landing point is up to standard, and the height of the ball over the net; for second pass training subjects, the coach mainly focuses on whether the ball landing point is up to standard, and whether the trajectory of the incoming ball has an effect on the player's posture; for spike training subjects, the coach mainly focuses on whether the ball landing point is up to standard and the speed of the spike ball; for blocking and mat training subjects, the coach mainly focuses on whether the ball is over the net or out of bounds.

In the step length matrix, we focus on the contextual information of action units in the action sequence and record the contextual information as step length pairs [20]. Then a dynamic segmentation scheme is designed to segment the bone sequence into multiple bone fragments, each of which is composed of consecutive multiframe bone data with spatial similarity. However, in the repetitive database, not only the step length pairs within an action recorded but also the step length pairs between two consecutive identical actions are recorded in the step length matrix. Step pairs between consecutive actions are meaningless and can also become noisy, degrading the performance of the step matrix approach.

5. Results Analysis

5.1. Performance Analysis of Action Data Recognition Algorithm. A large training set may lead to model overfitting; that is, the trained model performs well for the training set data and poorly for the test set data. Conversely, a small training set may lead to underfitting of the model; that is, the model will underfit the data and perform poorly on the training set and equally poorly on the test set. A good machine learning model should find a balance between overfitting and underfitting so that the trained model performs well on the training set and predicts the test set well.

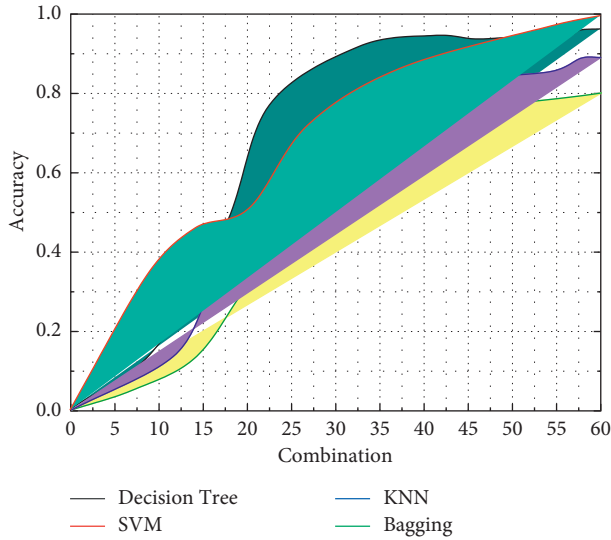


FIGURE 5: Comparison of algorithm accuracy between different combinations.

Cross-validation is usually a good solution to this problem, and cross-validation is sometimes called round-robin estimation. In this paper, considering the actual training set size, a 5-fold cross-validation is proposed; that is, the dataset is initially divided into 5 subsets randomly, and 4 subsets of data are used for model training, and a separate subset of samples is used for validation. The crossover is repeated 5 times, each subset is validated once, and the final prediction results are obtained by averaging the validation results of 5 times.

Nine different combinations of angle sequences were trained using four machine learning algorithms and the corresponding accuracies were calculated, as shown in Figure 5. The histogram shows that the decision tree model has the lowest accuracy and the support vector machine has the highest accuracy among the nine different combinations. Combination 1 and combination 2 have the lowest accuracy among the four algorithms, which shows that the training model does not work well using a single angle sequence, and the shoulder angle sequence provides more classification information than the elbow sequence, which can also be seen in the comparison between combination 7 and combination 8. By comparing the results of combination 5 and combination 6, the difference in accuracy using unilateral angle data is relatively large, which is related to the dominant hand of most athletes. Combinations 8 and 9 have the highest accuracy, which shows that the more data collection nodes are used, the more classification bases are provided to the classification model and the better the model performs in terms of accuracy.

From the perspective of Macro-PPV, combination 1 and combination 2 are the lowest; that is, the models trained with a single angle sequence generally have low combined prediction ability for each phase, while combination 8 and combination 9 still maintain good classification performance. In addition, the training time and prediction speed of

TABLE 1: Model classification results of combination 8 and combination 9 after PCA dimensionality reduction.

Combination	Model	Accuracy (%)	Macro-PPV (%)
Combination 8 (3 nodes)	Decision tree	99.1	85.7
	SVM	95.9	85.8
	KNN	95.1	86.2
	Ensemble	90.1	85.5
Combination 9 (5 nodes)	Decision tree	83.9	97.8
	SVM	99.8	88.3
	KNN	91.3	89.9
	Ensemble	80.5	90.3

the four machine learning models were counted in this paper for different combinations, as shown in Table 1. So skin, walls, and so forth are ignored, and only objects with markings glow; there are also systems that utilize one or more LEDs and use software to identify them by their relative positions. Among these nine combinations, the training time of the decision tree model is the shortest, and the training time of other algorithms is similar. The decision tree model has the fastest prediction speed, SVM and Bagging are the next fastest, and KNN has the slowest prediction speed.

Principal component analysis can be used to approximate the original data, which can be interpreted as discovering the inherent “basic structure” of the data. In this paper, PCA is used to reduce the dimensionality of combination 8 and combination 9, and the model is retrained. After training, four principal components were retained, and the components retained by PCA were sufficient to explain 98% of the variance. In combination 8, the explained variance of each component is 64.0%, 30.3%, 2.9%, and 1.9%, respectively. After PCA dimensionality reduction, the combined accuracies of the decision tree, SVM, KNN, and Bagging were 95.9%, 97.8%, 97.4%, and 97.6%, respectively. The explained variance of each component was 72.5%, 16.8%, 7.3%, and 3.0%, respectively, and, after dimensionality reduction, the combined accuracies of decision tree, SVM, KNN, and Bagging were 96.3%, 97.9%, 97.8%, and 97.6%, respectively.

Controlling the race rhythm and oar frequency is an important element of the leading tactics. At the start, athletes paddle with maximum force, and, after 10–20 seconds, both oar frequency and boat speed reach the maximum, and athletes must maintain this competitive state for more than 30 seconds. If the athlete continues to maintain this situation after the lead, it will cause great physical discomfort, so he/she needs to adjust the speed of rowing in the middle of the race according to the size of the lead he/she has, to control the oar frequency to decrease gently, and to reduce the boat speed to his/her comfortable range appropriately.

Then, in the skeletal sequence, the action unit is a skeletal fragment composed of several consecutive skeletal frames with similar spatial structure, and the skeletal fragment contains spatial information and local temporal information of human action. Replace the skeleton segment in the

skeleton sequence with its closest key segment in the key segment dictionary, so that the skeleton sequence is encoded as a word sequence. The normalized skeletal sequence is segmented into skeletal fragments representing decomposed actions according to the dynamic segmentation algorithm, and the spatiotemporal information of the skeletal fragments is extracted and clustered to form a key fragment dictionary so that the skeletal sequence is represented as a word sequence. The experimental results show that the key fragment descriptor is more effective than the single frame pose descriptor because it can represent not only the spatial information of the skeletal sequence but also the motion information in a short time.

5.2. Evaluation of System Analysis Results. Based on the consideration of simplifying the front-end calibration process, combined with the situation that coaches or other analysts may not be able to discern the identity of athletes in some field shooting angles, the automatic analysis algorithm should also identify the target athletes to be analyzed. Therefore, it is necessary to select the 3D skeleton of the personnel previously taken out and select some information as discriminative features to design a set of classifier-based recognition algorithms for the size of the personnel of a regular volleyball team of about 20 people.

To build a personnel recognition classifier, it is first necessary to construct a skeleton dataset of all players of the team. In the experiment, all the players of the volleyball team to which the video was taken were selected, and, from all the training videos in which these players participated, five training videos of different training subjects in which each player participated were selected as the basis for skeleton extraction, and the selection of training videos needed to ensure that the training players always appeared in the video screen. Based on the selected training videos, a video length of 300 frames is selected for each person for each video, and the dichotomous method is used so that the selected segments are divided as evenly as possible from the beginning to the end of the video, and, finally, the process of 3D pose extraction of the personnel is carried out to form 3D skeleton data of 1500 frames for each person. In the data preprocessing stage, the athletes' skeleton data are preprocessed and features are selected. The features are mainly described by the length and proportion of the skeleton, and a total of 12 features are selected, as shown in Figure 6.

The optical motion capture system is only suitable for data acquisition and analysis under indoor laboratory conditions and is difficult to be applied to outdoor scenes. Although it is more intuitive and clearer to use the video method to analyze athletes' technical movements, the video analysis method is limited by video clarity and shooting angle, and there are many inconveniences in actual use. The inertial sensor-based motion capture system developed in this paper for kayaking can be applied to outdoor real-boat data acquisition, and the technical movements of athletes can be reconstructed in 3-dimensional space through pose decomposition, which can provide coaches with action information from multiple perspectives.

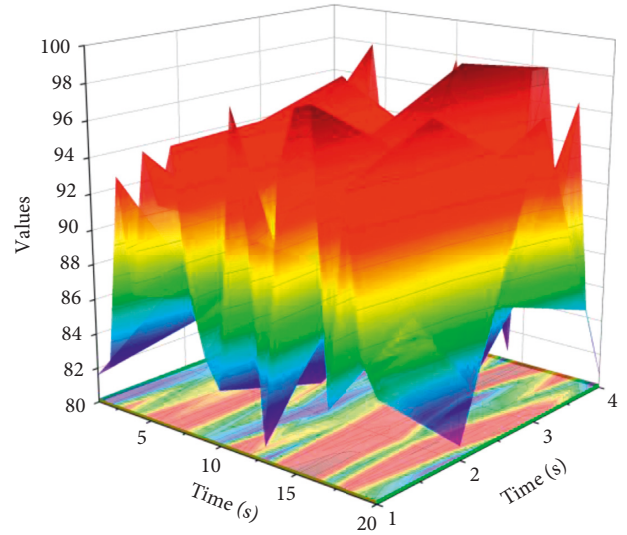


FIGURE 6: Upper limb movement trajectory.

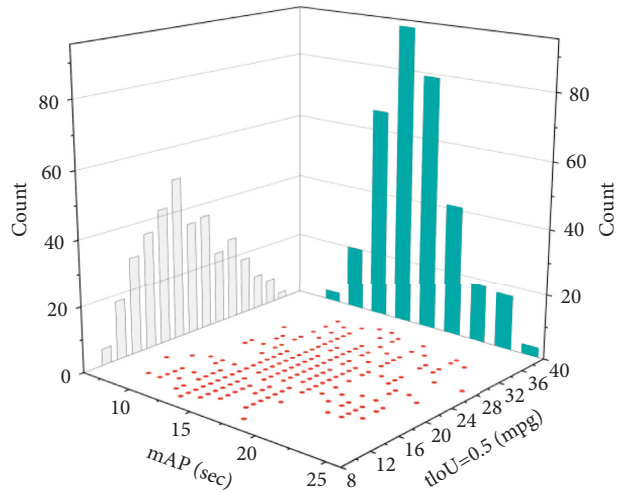


FIGURE 7: Review results.

The MAE obtained by migration learning is smaller than the MAE obtained by random initialization, which illustrates the effectiveness of applying migration learning in the training of progress label prediction network, and the subsequent results all refer to the results obtained by applying migration learning to the training network. This problem is improved by the high-resolution network model that our system finally uses in the task of human pose estimation. High-resolution network is one of the more advanced neural networks in the field of human pose estimation tasks. The MAE of the trained network on the test set is much smaller than the MAE when the progress labels are randomly generated, which indicates that the proposed 53-layer progress label prediction network does learn the progress label information of the image frames. However, the MAE on the test set is not infinitely close to 0 as we expect, as shown in Figure 7.

The PAS algorithm is more sensitive to the length of the progress label sequence, so the running time of the PAS

algorithm is greatly reduced to about 1/1667 of the original, and the downsampling operation effectively improves the running rate of the PAS algorithm. Overall, after downsampling, the average time for progressive motion detection of test video samples containing 6608 frames on average in the DFMAD-70 database is 10.95 s, which greatly improves the detection efficiency of the progressive motion detection algorithm and enables fast progressive motion detection.

6. Conclusion

This paper addresses the current problems in fine action classification and evaluation, designs a manual feature-based action classification method and a high-precision action detection method from the direction of action classification, action detection, and action evaluation, respectively, and proposes a complete fine action evaluation framework. Human action evaluation requires comparison and analysis of the training action video frames with the selected template action video frames. Action alignment can be a good solution to the frame alignment problem between different length action videos, but the common action alignment algorithm, DTW, has too high computational complexity, and it is unnecessary and time-consuming to detect the key points of the human skeleton and calculate the joint angles for all the captured training action videos. Therefore, this thesis uses a combination of keyframe extraction and segmented dynamic time regularization algorithm to align the training videos. The system realizes the functions of standard action library establishment, training action acquisition, action alignment, and posture analysis and can extract key action information from training action videos and compare it with the template information in the standard action library and give improvement suggestions, which can be used as an auxiliary training tool for learners to a certain extent.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Department of Physical Education, Lanzhou University of Technology.

References

- [1] W. Song, M. Xu, and Y. Dolma, "Design and implementation of beach sports big data analysis system based on computer technology," *Journal of Coastal Research*, vol. 94, no. SI, pp. 327–331, 2019.
- [2] E. E. Cust, A. J. Sweeting, K. Ball, and S. Robertson, "Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance," *Journal of Sports Sciences*, vol. 37, no. 5, pp. 568–600, 2019.
- [3] H. Li, "Analysis on the construction of sports match prediction model using neural network," *Soft Computing*, vol. 24, no. 11, pp. 8343–8353, 2020.
- [4] X. Meng, G. Zhang, S. Jia, X. Li, and X. Zhang, "Auxiliary criterion conversion via spatiotemporal semantic encoding and feature entropy for action recognition," *The Visual Computer*, vol. 37, no. 7, pp. 1673–1690, 2021.
- [5] Y. Li, "A new physical posture recognition method based on feature complement-oriented convolutional neural network," *Journal of Applied Science and Engineering*, vol. 24, no. 1, pp. 83–89, 2021.
- [6] C. Harper, M. Mafham, W. Herrington et al., "Comparison of the accuracy and completeness of records of serious vascular events in routinely collected data vs clinical trial-adjudicated direct follow-up data in the UK," *JAMA Network Open*, vol. 4, no. 12, p. e2139748, 2021.
- [7] L. Wu, Z. Li, Y. Xiang, M. Jian, and J. Shen, "Latent label mining for group activity recognition in basketball videos," *IET Image Processing*, vol. 15, no. 14, pp. 3487–3497, 2021.
- [8] R. Singh, A. Ravindra Bodhe, P. Kanuparthi, A. Ananthakrishnan, and J. Pitt, "From classification to definition: the changing nature of human adjudication," *IEEE Technology and Society Magazine*, vol. 38, no. 4, pp. 55–62, 2019.
- [9] L. Huang and G. Liu, "Functional motion detection based on artificial intelligence," *The Journal of Supercomputing*, vol. 78, no. 3, pp. 4290–4329, 2022.
- [10] M. Qi, Y. Wang, A. Li, and J. Luo, "Sports video captioning via attentive motion representation and group relationship modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 8, pp. 2617–2633, 2019.
- [11] J. McGrath, J. Neville, T. Stewart, and J. Cronin, "Upper body activity classification using an inertial measurement unit in court and field-based sports: a systematic review," *Proceedings of the Institution of Mechanical Engineers - Part P: Journal of Sports Engineering and Technology*, vol. 235, no. 2, pp. 83–95, 2021.
- [12] R. M. Chambers, T. J. Gabbett, and M. H. Cole, "Validity of a microsensor-based algorithm for detecting scrum events in rugby union," *International Journal of Sports Physiology and Performance*, vol. 14, no. 2, pp. 176–182, 2019.
- [13] P. Fussey, B. Davies, and M. Innes, "'Assisted' facial recognition and the reinvention of suspicion and discretion in digital policing," *British Journal of Criminology*, vol. 61, no. 2, pp. 325–344, 2021.
- [14] Y. Cao and H. Mao, "High-dimensional multi-objective optimization strategy based on directional search in decision space and sports training data simulation," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 159–173, 2022.
- [15] Z. Qiao and J. Yin, "Fuzzy deep medical diagnostic system: gray relation framework and the guiding functionalities for the professional sports club social responsibility," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1084–1090, 2020.
- [16] F. Kong and Y. Wang, "Design of computer interactive system for sports training based on artificial intelligence and improved support vector," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 5, pp. 6165–6175, 2019.
- [17] M. Du and X. Yuan, "A survey of competitive sports data visualization and visual analysis," *Journal of Visualization*, vol. 24, no. 1, pp. 47–67, 2021.

- [18] B. Bačić and P. A. Hume, “Computational intelligence for qualitative coaching diagnostics: automated assessment of tennis swings to improve performance and safety,” *Big Data*, vol. 6, no. 4, pp. 291–304, 2018.
- [19] N. Wagner, L. H. Sundhararajan, L. H. Son, and M. Joo, “Complex evolutionary artificial intelligence in cognitive digital twinning,” *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2013–2016, 2021.
- [20] A. M. Williams, P. R. Ford, and B. Drust, “Talent identification and development in soccer since the millennium,” *Journal of Sports Sciences*, vol. 38, no. 11-12, pp. 1199–1210, 2020.