

Cloud IoT-Oriented Neural Network-Based Taekwondo Teaching Scheme

Shuqun Qu^{1*}

Abstract

As Taekwondo is widely accepted, there is a rapid increase in Taekwondo learners. Traditional Taekwondo teaching mode is challenging to meet the increasing educational needs. In Taekwondo teaching, in addition to teaching the basic skills of combat, Taekwondo Poomsae teaching is also required. The quality of Taekwondo Poomsae learning is critical to the overall strength of Taekwondo. For a long time, the judgment of the quality of Poomsae action has relied on the manual evaluation of teachers, which could be more conducive to forming an objective and accurate quality score. Therefore, it is necessary to introduce in-depth learning to find new solutions to these problems. First, a multi-sensor data fusion method is proposed to collect Taekwondo Poomsae action. Then, a Taekwondo Poomsae expertise integrated multi-view feature extraction method is proposed. Finally, Convolutional Neural Network (CNN)-Mogrifier Long Short-Term Memory (LSTM) is proposed to train the generated Taekwondo Poomsae action scoring model, identifying whether Taekwondo Poomsae action meets the standard and improve Taekwondo teaching. The effect test of Taekwondo Poomsae action intelligent evaluation shows that the results of the proposed method follow the evaluation given by the teachers, and the discrimination of the scores is moderate, which indicates that the proposed method has good Taekwondo Poomsae action quality evaluation ability. Furthermore, the experimental results show that CNN-Mogrifier LSTM achieves high recognition accuracy of Taekwondo Poomsae, gives comprehensive evaluation and improvement suggestions for action completion quality, breaks away from the restrictions of teachers and venues, and realizes automatic and smart Taekwondo teaching.

Key Words: Taekwondo, Poomsae, Cloud IoT, CNN, Mogrifier LSTM.

I. INTRODUCTION

As Taekwondo became widely accepted by the public, the number of people learning Taekwondo increased rapidly. However, traditional face-to-face teaching has yet to meet people's needs [1-2]. Many problems have been exposed in the actual teaching, including: (i) Due to the enrollment expansion of colleges and universities, the increase of students taking Taekwondo courses, and the shortage of Taekwondo teachers, many colleges implement class teaching in the teaching arrangement, which directly leads to the difficulty in ensuring the learning quality of students. (ii) Traditional Taekwondo teaching relies heavily on coaches. Students need to complete the training to get effective movement instruction. Traditional Taekwondo teaching mode is challenging to cope with the increasing educational demand. The core of Taekwondo teaching and competition is the practical evaluation of movement completion quality [3]. In the Taekwondo Poomsae competition, movement completion quality significantly impacts the final score [4]. As an essential basic skill, Taekwondo Poomsae plays a

pivotal role in Taekwondo learning [5]. The quality of Poomsae learning is crucial to improve the overall strength of Taekwondo. Therefore, we use Taekwondo Poomsae as a representative movement for theoretical research to provide a feasible reference for other movements in Taekwondo.

Currently, Taekwondo teaches fundamental skills for actual combat and Taekwondo Poomsae. Poomsae in different chapters have unique movement characteristics, and the quality of movements is closely related to the student's physical coordination, balance, strength, speed, and rhythm [6]. Therefore, Poomsae is the core content of Taekwondo teaching and one of the primary evaluation items in various competitions. The judgment of the Poomsae action's quality for a long time depends on teachers' manual evaluation. The manual evaluation criteria mainly refer to three indicators: accuracy, proficiency, and expressiveness [7]. The accuracy score includes the accuracy of fundamental movements as well as the accuracy of each Poomsae movement. Proficiency is scored on a range of body movements, balance, speed, and power of movement—expressiveness

Manuscript received December 07, 2023; Revised December 21, 2023; Accepted January 29, 2024. (ID JMIS-23M-10-042)

Corresponding Author (*): Shuqun Qu, +86-13662142277, 13662142277@163.com

¹Department, Jilin Sport University, Changchun, China, 13662142277@163.com

rated for rigidity, slowness, change of pace, and the student's momentum. However, manual participation will be interfered with by many subjective factors, which are mainly reflected in (i) the movement rhythm changes quickly, the differences between the movements are slight, and the duration of completing a set of continuous movements are long, which all affect the accurate judgment of teachers. (ii) Quality evaluation is easily affected by teachers' knowledge and personal preferences. This utterly manual judgment brings many subjective factors to the final score. (iii) In Taekwondo teaching, only the teachers usually observe from a fixed angle, so the evaluation results are not comprehensive and accurate enough. (iv) Movement strength is one of the vital evaluation factors of movement quality [8]. However, movement strength cannot be measured by an accurate quantitative index. It can only be estimated by manual observation, which is not conducive to forming objective and accurate completion quality scores. It is not easy to solve the above difficulties and challenges effectively by combining traditional information collection with manual analysis. For example, although video recording and slow-motion analysis can be performed, it still belongs to the category of manual evaluation, which does not fundamentally solve the problem. Therefore, it is necessary to introduce the latest Artificial Intelligence (AI) to find new solutions to the above problems.

Recently, AI triggered by deep learning has swept many fields and achieved fruitful results [9-11]. In machine vision, natural language processing, and other fields, deep learning has demonstrated its advantages in processing big data and heterogeneous data from multiple sources. The application of deep learning technology in Taekwondo teaching, especially neural network, which uses a neural network to build innovative evaluation algorithms of Taekwondo movement quality, provides theoretical support for taekwondo teaching, can make up for the shortcomings in the existing Taekwondo teaching, effectively improve the supply level of the existing Taekwondo teaching, and promote students to reduce their dependence on teachers and sites. Improving the quality of Taekwondo teaching and students' learning efficiency is a solid support to promote the development of Taekwondo's innovative teaching, which has clear theoretical and practical value. Therefore, from the perspective of improving the automation and intelligence level of Taekwondo education, this study tries to integrate the professional knowledge of Taekwondo and introduce a neural network to promote the development of Taekwondo's innovative education.

The evaluation of Taekwondo Poomsae quality is to integrate the human body posture recognition based on a neural network with Taekwondo quality and perform the action feature extraction and scoring modeling. Traditional human motion recognition mainly adopts equipment that can be

used for video recording or photography, collects human video and photos, and uses the method of image processing for motion recognition [12-14]. With the development of technology, human action recognition is no longer limited to computer vision but has been extended to other fields. For example, they are using inertial sensors for recognition. By being equipped with sensors or cameras and other equipment, the movement data of students in Taekwondo Poomsae will be collected, including action, speed, and other information. Then, the collected Internet of Things (IoT) data is uploaded to the cloud for processing. This study uses cloud IoT model. Cloud IoT, or Cloud-based Internet of Things, is a model for delivering IoT services that utilizes the cloud for data storage, processing, and analysis. Instead of relying solely on local devices for data management and analytics, Cloud IoT enables devices to communicate with cloud servers, which can then process and analyze the data in real-time. In this model, IoT devices are connected to the internet and send data to cloud-based servers, where it is processed and analyzed using machine learning algorithms and other advanced analytics tools. The processed data can then be used to generate insights, trigger actions, or feed back into the device to optimize its performance. Cloud IoT offers several benefits compared to traditional IoT models, including scalability, flexibility, and cost-effectiveness, as well as the ability to support complex data processing requirements and integrate with other cloud-based services. The cloud uses big data technology to analyze and process data and extract the feature vector of Taekwondo Poomsae movements [15]. The neural network is used to train and recognize feature vectors and is constructed by deep learning technology to train and recognize Taekwondo Poomsae movements. The identified Taekwondo Poomsae movements are compared with the standard movements, and the performance of the movements is calculated. Through the above process, the real-time evaluation and feedback of students' Taekwondo Poomsae movements can be realized, and the teaching effect and students' self-correction ability can be improved. This approach has several benefits over traditional Taekwondo instruction. First, it provides real-time feedback to the student, which can help them correct mistakes and improve their performance faster. Second, the system can adapt to each student's individual needs and learning styles, providing a personalized learning experience. Third, using cloud computing and neural networks can provide insights into the performance of large groups of students, which can be used to improve the overall instruction and training programs.

Accordingly, the main contributions of this paper are organized as follows. (i) A multi-sensor data fusion method is proposed to collect Taekwondo Poomsae action. (ii) A Taekwondo Poomsae expertise integrated multi-view feature extraction method is proposed. (iii) CNN (convolu-

tional neural network)-Mogri LSTM (long short-term memory) is proposed to train the generated Taekwondo Poomsae action scoring model.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents a multi-sensor data fusion method. Section 4 studies the CNN-Mogri LSTM-based Taekwondo Poomsae action scoring model. Experimental results are reported in Section 5. Section 6 gives the conclusion of this paper.

II. RELATED WORK

2.1. Application of AI in Sports

Because of advancements in information and communication technologies, AI is increasingly used in sports [16-20]. There is a need to improve the precision of deep learning for use in diagnosing sports injuries. New possibilities for the growth of physical education arise when pattern recognition is combined with virtual reality technology, augmented reality technology, and mixed reality technology. Wearable devices with pattern recognition capabilities can track student progress in physical education [21-22]. Smart wearable technologies combined with VR allow for a more realistic and less risky learning environment in activities that need specific conditions, such as ice and snow sports and golf. Wearable devices and computer vision are the backbones of augmented reality, allowing teachers to capture real-time multi-view image data, collect students' three-dimensional action data, and provide instant feedback. To provide an immersive physical education experience, mixed reality uses wearable devices to interact with individuals and the environment. Wearable devices may be paired with virtual personal assistants to achieve human-computer interaction in PE through emotional computing and cloud computing and to develop personalized PE lesson plans. Furthermore, machine learning is often used in competitive sports, and its impact on PE may be seen mainly in two areas [23-25]. (i) Deep learning can be used to identify different types of exercise. Examples of artificial neural networks in this regard include assessing metabolic equivalents and identifying different activity categories (such as light activity, sports, strenuous exercise, and housework). Real-time muscle monitoring and feedback, fatigue prediction, and injury prevention are all possible thanks to computational modeling. (ii) Predicting future results is possible with deep learning. It offers a statistical base for hierarchical physical education by mining training and competition history to predict competition outcomes.

2.2. Human Action Recognition

Human action recognition through wearable sensors

mainly refers to the use of wearable sensors on the human body to collect data generated during the movement of the limbs and to analyze the sensor data to identify the current movement pattern of the wearer, such as climbing stairs, walking, and running [26]. Human action recognition usually includes four stages: data acquisition and preprocessing, feature extraction, classifier training, and recognition [27]. Typically, human action recognition methods are mainly machine- and deep learning-based methods. Machine learning-based methods must manually select relevant features from sensor data, such as mean, standard deviation, intragroup correlation coefficient, maximum value, minimum value, etc. Then use, manually selected features to train classifiers, such as support vector machines (SVM), linear discriminant analysis, random forests, etc. The recognition performance of the machine learning-based method depends mainly on the correlation degree of the features [28]. The recognition effect will be significantly reduced if the features cannot reflect the differences between the motion patterns. With the continuous development of deep learning technology, deep learning-based methods have gradually become the mainstream research direction of human action recognition [29]. Compared with traditional machine learning, deep learning can automatically extract features through the hidden layer, improving recognition accuracy, speed, and precision. Ronao and Cho [30] proposed using a convolutional neural network (CNN) to recognize patterns such as walking and going up and down stairs, and the effect is better than SVM. Wang and Liu [31] used memory characteristics and the storage function of long short-term memory to identify motion patterns. Ordenez and Roggen [32] proposed a hybrid model combining CNN and LSTM, using the output of CNN as the input of LSTM.

III. MULTI-SENSOR DATA FUSION

3.1. Data Collection

To collect the action data of Taekwondo Poomsae, it is necessary to establish the spatial coordinate system of students' actions. Due to the complexity of Taekwondo Poomsae, the actions of different body parts are very different. For example, the wrist, knee, shoulder, and other parts are complicated and changeable in Taekwondo Poomsae training, and the randomness of actions is high. The body sensor positions and data flow architecture are shown in Fig. 1.

As shown in the body sensor and data flow architecture (Fig. 1), sensors are positioned on the wrists, elbows, knees, and ankles. The motion data from these sensors flows into a cloud database for storage and analysis.

In this study, a triaxial digital acceleration sensor is selected to collect the gravity acceleration information of

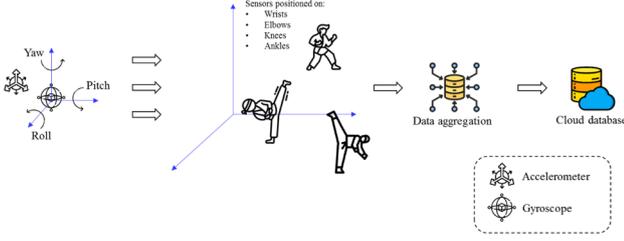


Fig. 1. Body sensor positions and data flow architecture.

Taekwondo Poomsae students, and a MEMS triaxial angular velocity sensor (gyroscope) is selected to collect the human body attitude angle information [33-34]. The angle collected by the sensor can be corrected in real-time to overcome the deviation of the single sensor in pose angle measurement through the fusion of the acceleration sensor and the gyroscope data.

When students do Taekwondo Poomsae action, \vec{G} is set as the vector measured by the acceleration sensor, whose acceleration values are \vec{G}_X , \vec{G}_Y and \vec{G}_Z in the direction of the three axes X , Y , and Z , respectively. The measured acceleration values need to be converted into angular values.

When the acceleration sensor is at rest, we have

$$\vec{G}^2 = \vec{G}_X^2 + \vec{G}_Y^2 + \vec{G}_Z^2. \quad (1)$$

Normalizing vector \vec{G} as follows.

$$\begin{cases} \vec{R}_X = \frac{\vec{G}_X}{|\vec{G}|} \\ \vec{R}_Y = \frac{\vec{G}_Y}{|\vec{G}|} \\ \vec{R}_Z = \frac{\vec{G}_Z}{|\vec{G}|} \end{cases}. \quad (2)$$

The vector value \vec{R}_{acc} in the direction of normalized gravity at the current time is defined as follows.

$$\vec{R}_{acc} = \sqrt{\vec{R}_X^2 + \vec{R}_Y^2 + \vec{R}_Z^2} = 1. \quad (3)$$

The angle $\theta_{acc,x}$, $\theta_{acc,y}$, and $\theta_{acc,z}$ of the gravity vector in the direction of the three axes can be calculated from the normalized vector value, and the angle can be calculated as follows.

$$\begin{cases} \theta_{acc,x} = \arcsin \frac{|\vec{R}_Y|}{\sqrt{\vec{R}_Y^2 + \vec{R}_Z^2}} \\ \theta_{acc,y} = \arcsin \frac{|\vec{R}_X|}{\sqrt{\vec{R}_X^2 + \vec{R}_Z^2}} \\ \theta_{acc,z} = \arcsin \frac{|\vec{R}_Y|}{\sqrt{\vec{R}_X^2 + \vec{R}_Y^2}} \end{cases} \quad (4)$$

Suppose that the angular velocity measured by the angular velocity sensor around axes X , Y , and Z are ω_x , ω_y ,

and ω_z respectively. It is also necessary to convert the measured angular velocity into the corresponding rotation angle. According to the current rotation angular velocity collected and the system's sampling period, the gyroscope's rotation angle around the three axes can be calculated as follows.

$$\begin{cases} \theta_{gy,x} = \int \omega_x dt \\ \theta_{gy,y} = \int \omega_y dt, \\ \theta_{gy,z} = \int \omega_z dt \end{cases} \quad (5)$$

where $\theta_{gy,x}$, $\theta_{gy,y}$, and $\theta_{gy,z}$ represent the rotation angles of the corresponding X , Y , and Z coordinate axes after calculation, and dt represents the sampling period of the gyroscope data.

The Kalman filter fuses the information collected by the acceleration sensor and gyroscope [35]. The measured value of the acceleration sensor is taken as the predicted value, the measured value of the gyroscope is taken as the observed value, and the observed value modifies the predicted value of the acceleration sensor as the output value. Gyroscope drift error b is estimated by acceleration sensor as state vector to obtain the state equation and observation equation.

$$\begin{cases} \begin{bmatrix} \theta' \\ b' \end{bmatrix} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} \theta \\ b \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \omega_{gy} + \begin{bmatrix} \omega_g \\ 0 \end{bmatrix} \\ \theta_{acc} = [1 \quad 0] \begin{bmatrix} \theta \\ b \end{bmatrix} + \omega_a \end{cases} \quad (6)$$

In equation (6), ω_{gy} represents the output angular velocity of the gyroscope with fixed deviation, θ_{acc} represents the angular value of the acceleration sensor obtained after processing, ω_g and ω_a are the measured noise of the gyroscope and the acceleration sensor respectively. For the convenience of calculation, they are independent of each other. Assuming that they are white noise and meet the normal distribution. Let the measurement noise be $\omega(k)$ and the sampling period be T_s . The equation of state and measurement are defined as follows.

$$\begin{cases} X(k) = \begin{pmatrix} 1 & T_s \\ 0 & 1 \end{pmatrix} X(k-1) + \begin{bmatrix} T_s \\ 0 \end{bmatrix} \omega_{gy}(k-1) + \begin{bmatrix} \omega_g(k)T_s \\ 0 \end{bmatrix} \\ V_i = [1 \quad 0]X(k) + \omega_a(k) \end{cases} \quad (7)$$

$$Kg(k) = \frac{P(k|k-1)\mathbf{H}^T}{\mathbf{H}P(k|k-1)\mathbf{H}^T + R(k)}. \quad (8)$$

In equation (8), $Kg(k)$ represents the Kalman increment at time k , $P(k|k-1)$ represents the covariance of the system since time $k-1$, \mathbf{H} represents the output matrix of the measurement, \mathbf{H}^T represents its transpose matrix, and $R(k)$ represents the covariance of the measurement noise. The pose angle of fusion is defined as follows.

$$\theta(k) = \theta_{gy}(k) + Kg(k) (\theta_{gy}(k) - \theta_{acc}(k)). \quad (9)$$

In equation (9), $\theta_{gy}(k)$ and $\theta_{acc}(k)$ are the pose angles output by gyroscope and acceleration sensor at time k respectively, and $\theta(k)$ is the output value of pose angle at time k after fusion, which is also the optimal output value of Kalman filter at this time. The covariance of the system state at time k is defined as follows.

$$P(k|k) = (1 - Kg(k)H)P(k|k-1). \quad (10)$$

Equations (6) to (10) are the calculation of Kalman filtering. Equations (8) and (10) are used to ensure the recursion and persistence of the filtering. When the system receives the angular velocity output of the gyroscope at time $k+1$, it returns to equation (5). At this time, the system enters the filtering operation at time $k+1$.

The sensor data details are shown in Table 1.

After fusion of the collected multi-sensor data, the error of angle measurement using acceleration sensor alone is overcome, and the Taekwondo Poomsae action data collected by students is closer to the true value.

3.2. Poomsae Integrated Feature Extraction

In studying Taekwondo Poomsae movements, it is necessary to extract such characteristics as the speed, strength, and key angle of limbs and torso. Combined with the teacher's professional knowledge, this method extracts features such as speed, strength, and angle after collecting human action data. Since a complete set of Poomsae actions includes two types: "action in movement" and "action at rest", the characteristics can be calculated separately. The first action in Poomsae Chapter One is illustrated here as an example to facilitate understanding. "Action in movement" is in change, with speed and strength as important reference indicators. There is a brief pause in "action at rest", which uses the angle of the body as an essential indicator.

The "action in movement" feature focuses on the speed and force of the body during movement. In the experiment, the acceleration of the same joint coordinate point is used to approximately replace the force. In Poomsae Chapter One, the two arms have the most extensive range of motion. We use multiple cameras to obtain multi-view videos of students' Taekwondo Poomsae actions. Therefore, the velocity and acceleration of the following coordinate points should be calculated mainly: left elbow, left wrist, right elbow, and right wrist.

Taking the left wrist as an example, the coordinates of

$t-1$ th frame is $wrist(X_{t-1}, Y_{t-1})$, and the coordinate of the t th frame is $wrist(X_t, Y_t)$. By calculating the Euclidean distance between the two coordinate points, the corresponding action distance $s = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$ can be obtained. The corresponding action time can be obtained by calculating the time difference between the two frames. For a camera with 30 frames per second, the time between frames is $1,000 \text{ ms} / 30 = 33.33 \text{ ms}$. From this, the speed and acceleration of the coordinate point between two frames can be obtained. That is, the speed between each frame is set as $v = s/t$, and the acceleration between each frame is set as $a = v/t$.

The "action at rest" feature focuses on the direction and angle of the body movement at the pause. Since the standard of "action at rest" should be judged according to the angle value of each limb and torso, several key angles are used as features: double elbow angle, arms with vertical angle, legs with vertical angle, and knees angle.

Taking the angle of the left elbow as an example. Set the coordinates of left elbow as $E(X_{elbow}, Y_{elbow})$, left wrist as $W(X_{wrist}, Y_{wrist})$ and left shoulder as $S(X_{shoulder}, Y_{shoulder})$. Vector EW can be extracted from the joint between elbow and wrist, and vector ES can be extracted from the joint between elbow and shoulder. The elbow angle is calculated as follows.

$$\cos(\alpha) = \cos(EW, ES) = \frac{EW \cdot ES}{|EW| \times |ES|} \quad (11)$$

According to Taekwondo training requirements, these "action at rest" have a typical angle value. Thus the threshold value can be set beforehand, and the difference between the action and the standard action can be obtained by comparing the action angle value and threshold value made by students in the video. For example, if the left arm needs to be at an angle of 45 degrees in the vertical direction and the angle of action of the student is between 42 and 48 degrees, then the movement is considered to be up to standard, and the feature is set to 1. If the angle of movement of students' limbs is between 39 and 42 degrees or between 48 and 50 degrees, the angle deviation of movement is large, which can be adjusted appropriately. The feature value is 0.6. If the angle value of students' actual actions deviates too much from the threshold, they directly set the dimension feature value to 0, indicating that the feature value is completely substandard.

Table 1. Summary of sensor data collected.

Sensor type	Body location	Data captured	Frequency
Accelerometer	Wrists, elbows	Acceleration along X/Y/Z axes	50 Hz
Gyroscope	Knees, ankles	Angular velocity around X/Y/Z axes	100 Hz

IV. CNN-MOGRIFIER LSTM-BASED TAE-KWONDO POOMSAE ACTION SCORING MODEL

A neural network is a computing model inspired by the human brain's network of biological neurons, referring to deep neural networks that have multiple layers and are capable of learning complex patterns from data through a training process. Cloud computing provides access to storage, computing resources, databases, and software applications over the internet, allowing for flexible scaling and efficient analytics. The IoT refers to the concept of connecting many physical devices and exchanging data over the internet. In this study, motion sensors attached to the body send data to cloud servers to be processed by deep neural networks. Specifically, CNN, which has convolutional layers to extract spatial patterns from images or sensor data, is utilized. The sequential nature of the motion data also lends itself to analysis using an LSTM network, a type of recurrent neural network well-suited for modeling time series data by remembering long-term dependencies.

With this high-level background establishing some key building blocks, we dive deeper into the technical details around the custom multi-sensor fusion methodology and tailored CNN-LSTM architecture developed to assess Taekwondo movements.

However, mining the dependence link before and after the action data is challenging, even if CNN can retrieve the locally relevant aspects of the data generated by sensors. Taekwondo Poomsae is composed of continuous actions, with a robust sequential relationship between the actions before and after. Thus, LSTM network in deep learning may be used to explore the consistency amongst students' activities, particularly alterations in speed, strength, and other characteristics. Poomsae action evaluation may benefit greatly from LSTM's gate mechanism, which efficiently addresses issues plaguing recurrent neural networks [36-37].

In this case, the structural properties of LSTM are the primary determinants of LSTM's use in mining Poomsae action coherence. (i) The new time step enters the LSTM neuron through the input gate, and the LSTM neuron computes the short-term memory output of the previous time step to permanently store the most important information from the current time step. When punching, for instance, there is a far greater variation in speed and strength at the wrists than there is at the shoulders. In the input door, the shoulder feature can be weakened, but the importance of the wrist feature can be highlighted more. (ii) The forget gate allows for selective forgetting of short-term memory generated in the prior time step. The most important pieces of information are the ones stored in long-term memory. When

the student punches, the points on the right side of his/her body will move more quickly and have a greater impact on the model's overall evaluation. Consequently, the whole left arm is weakened. Short-term memory (i.e., force, speed, angle, and other information at the latest moment) is output through the output gate while long-term memory (i.e., the general posture information of Poomsae action) is maintained, thereby reducing the impact of noise and features with little contribution.

The gates in LSTM include the input gate i_t , the forget gate f_t , and the output gate o_t . x_t is the input to the current node, h_{t-1} is the output of the previous node, and c_{t-1} is the state of the previous node. LSTM selectively memorizes the input x_t and h_{t-1} of the current node through the input gate i_t , determines how much information can be stored in the current node to generate a new state c_t , and then passes the forget gate f_t to the previous node's state c_{t-1} for selective forgetting. Only part of useful information is retained, and finally c_t is converted to the output h_t of the current node through the output gate o_t . The specific update process of LSTM is as follows.

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i). \quad (12)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f). \quad (13)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o). \quad (14)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c). \quad (15)$$

$$h_t = o_t \cdot \tanh(c_t). \quad (16)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t. \quad (17)$$

As can be seen from equation (17), c_t of the output state of the current node is weighted by the state of the previous node and the internal information of the current node. As long as f_t of the forget gate is not 0, LSTM can memorize the information of the previous node.

By using the aforementioned gate mechanism, LSTM is able to automatically filter out noise and extract meaningful temporal differences across frames. In a hypothetical conflict between two students, one would move at a constant tempo while the other would alternate between sluggish and quick bursts of activity. LSTM can figure out how much of a difference there is in terms of velocity and intensity between the two actions. As a consequence, there is a wide range of evaluation outcomes.

There is no direct connection between the input x_t of the current node and the output h_{t-1} of the previous node before it is sent to the node. Melis et al. [38] believe that x_t and h_{t-1} , which are independent of each other, would lose part of the temporal correlation of data, so they proposed Mogrifier LSTM based on LSTM. Mogrifier LSTM does not change the original structure of LSTM. Instead, x_t and

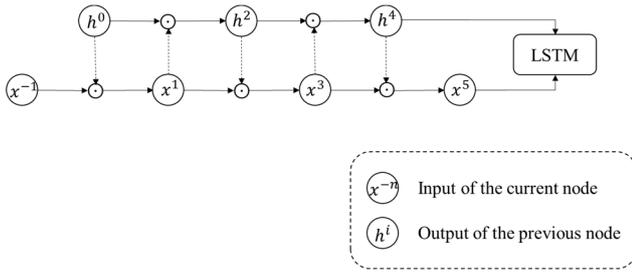


Fig. 2. Mogrifier LSTM structure.

h_{t-1} are first fully interacted with each other in the R -round according to equations (18) and (19) to enhance the main features and weaken the secondary features in the time series. Then, the updated x_t and h_{t-1} are sent to the current node to improve the ability of the model to extract data time correlation. Fig. 2 is the schematic diagram of Mogrifier LSTM when $R = 5$. Where x^{-1} is the input x_t of the current node, h^0 is the output h^{t-1} of the previous node, and the hyperparameter r is the number of rounds x_t and h_{t-1} interact. When $r = 0$, the Mogrifier LSTM degrades to LSTM. $i \in [1, \dots, r]$, when i is odd, x_t is updated; When i is even, h_{t-1} is updated, Q_i and R_i are the parameter matrix of the model. Reference [38] proves that in the text tasks, the performance of Mogrifier LSTM is obviously better than that of LSTM through experiments.

$$x^i = 2\sigma(Q^i h_{t-1}^{i-1}) x^{i-2}, \text{ for odd } i \in [1, \dots, r]. \quad (18)$$

$$h_{t-1}^i = 2\sigma(R^i x^{i-1}) h_{t-1}^{i-2}, \text{ for odd } i \in [1, \dots, r]. \quad (19)$$

Therefore, our approach employs Mogrifier LSTM to extract deep temporal characteristics of Poomsae data to generate an objective score, which relies on obtaining human bone nodes and extracting features. In particular, a model capable of efficiently extracting sequential aspects of students' activities may be generated by training Mogrifier LSTM. The time step of each sample is the information that makes up the Mogrifier LSTM layer's data. Its advantage lies in that the speed, strength, and angle feature changes in the time series can be extracted deeper. To mine more information from sensor data and improve recognition accuracy, this study uses Mogrifier LSTM for Taekwondo Poomsae action recognition. It proposes a CNN-Mogrifier LSTM recognition algorithm combined with CNN. The structure is shown in Fig. 3.

CNN-Mogrifier LSTM first uses a sliding window to divide the data collected by each sensor and filters and normalizes the original data in the window. Then, the local spatial features of the data are extracted through the two convolutional layers, and the features extracted from the convolutional layers are used as the input of the Mogrifier LSTM layer. By fully interacting with x_t and h_{t-1} , the

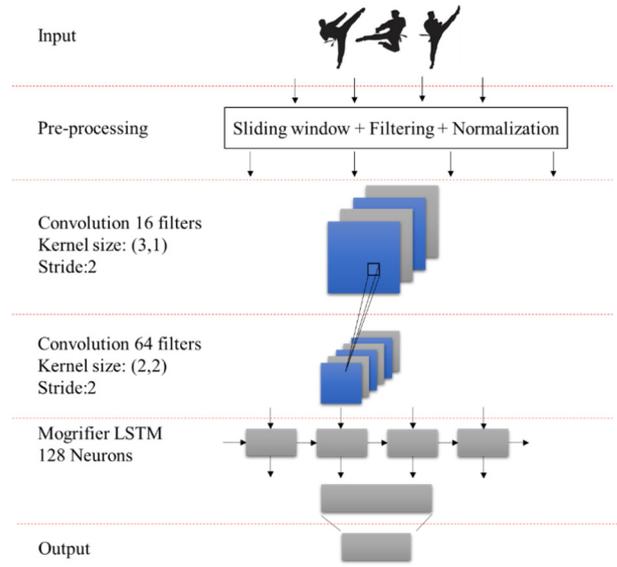


Fig. 3. CNN-Mogrifier LSTM structure.

dependence relationship between local space features is fully explored. Finally, the fully connected layer and Softmax function are used to identify the action mode of the current data. To include the information needed to recognize the action pattern, the window length needs to be at least 200 ms [39].

According to the output results of CNN-Mogrifier LSTM, this study proposes a new scoring metric. The last layer of the network is the fully connected layer that uses Softmax as the activation function. This activation function is often used in multi-classification to map the output of multiple neurons to the interval (0,1) and ensure that the sum is one. Suppose there is a vector V , V_i represents the i th element in V , then the Softmax of the element is calculated as follows.

$$S_i = \frac{e^i}{\sum_{j=1}^C e^j}, \quad (20)$$

where C is the number of elements in V . The value calculated by Softmax can be regarded as the probability of the occurrence of a certain classification. In this study, equation (20) is used as the scoring metric to obtain a probability relative to the standard action between the standard action and non-standard action. This probability represents the complete quality of this set of Poomsae actions and can be used as the final evaluation score.

Different from the subjective qualitative evaluation usually used in the existing Taekwondo teaching, the scoring metric shown in equation (20) covers the expression of key features of actions and includes the evaluation of the overall gesture and local action coherence. Therefore, it can provide a more objective and accurate evaluation of the quality of action completion.

V. EXPERIMENT AND RESULTS ANALYSIS

5.1. Effect Test of Taekwondo Poomsae Action Smart Evaluation

The test data of the effect test of Taekwondo Poomsae action innovative evaluation was completed by students from the physical education college of Jilin Sports University. Students make four sets of Taekwondo Poomsae actions with different grades for testing. The quality of the completed Taekwondo Poomsae actions is standard, sub-standard, non-standard, and completely non-standard. Professional teachers provide the judgment of these four grades. Since the video cannot be displayed in the paper, here we choose three kinds of "action at rest" for effect verification, as shown in Fig. 4. Among them, action videos labeled as "standard" to "completely non-standard" are collected from the four columns from left to right. It should be noted that the student's actions in the second column are not much different from those in the first column. Even though the pos-

tures of some actions are more standard (such as subfigure (10)), the strength of the student's actions is insufficient. They do not exert force according to the training requirements, so the teachers consider them "sub-standard". It can be seen that in subfigures (1) to (4), the angle between the four students' left arm and the vertical direction is different, which is consistent with the quality evaluation grade the teacher gave. There are similar differences in the angles of the elbow and shoulder and vertical direction shown in subfigures (5) to (10), and the angles of legs and vertical direction and knees shown in subfigures (9) to (11). This shows that the proposed method can accurately recognize the different actions of Taekwondo Poomsae. Moreover, there are apparent differences in the speed and strength of "action in movement".

To verify the effectiveness of the proposed method and each model, we constructed three comparison methods, and the results are shown in Table 2. To the best of our knowledge, there is no work using machine learning meth-

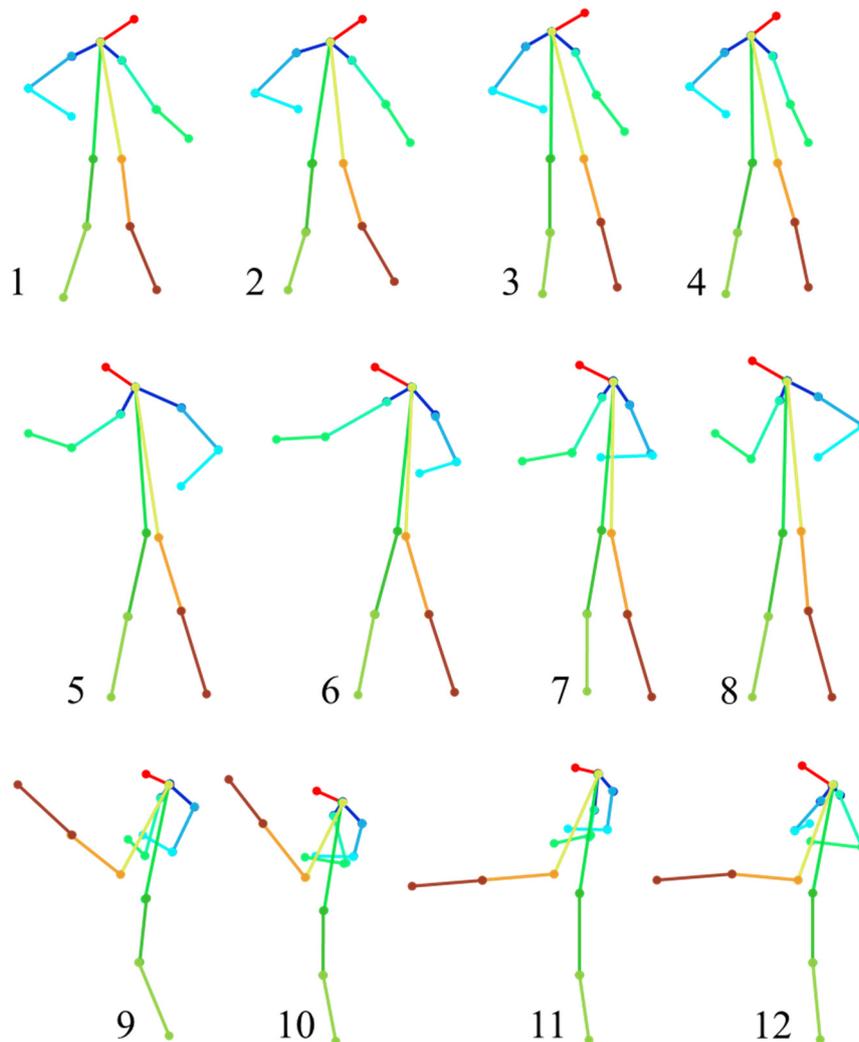


Fig. 4. Comparison of Taekwondo Poomsae action recognition effects of four groups of students with different quality levels under the frontal perspective.

Table 2. Smart evaluation comparison.

Test clip	The proposed	Method 1	Method 2	Method 3	Manual label
Clip 1	95	74	93	57	Standard
Clip 2	88	77	89	54	Sub-standard
Clip 3	66	69	35	46	Non-standard
Clip 4	47	68	22	50	Completely non-standard

ods to identify the quality of Taekwondo Poomsae actions. Therefore, it is impossible to compare existing methods, and we have to compare and verify by replacing or omitting part of the algorithm.

Method 1: To verify the effectiveness of the "Taekwondo Poomsae knowledge integrated feature extraction method", feature extraction is omitted. The vector coordinates of the 14 joint points of the human body in each frame are directly used as features and input into CNN-Mogrifier LSTM for training and scoring.

Method 2: To verify the accuracy of the "LSTM scoring model" for deep temporal feature mining, the final scoring model of the proposed method is replaced by BP neural network. The BP network uses three hidden layers, the output layer is two neurons, and the Softmax activation function is used for output.

Method 3: To verify the influence of the sample data collected under multi-view shooting on the final scoring results, a single camera is used to shoot from the front of the student in the video preprocessing and LSTM training steps. That is, the sample is collected from a single perspective. The rest of the various steps are the same as the method proposed in this study.

It can be seen from Table 2 that the evaluation scores of *method 1* are too concentrated and the discrimination is minimal, and there are error scores. This is due to the direct use of coordinate information as a feature vector. The features need to be more prominent, LSTM training takes a long time, and the loss function decreases slowly. However, CNN-Mogrifier LSTM takes less time to train, the loss function converges faster, and the final result is more reasonable. In *method 2*, the scores go to two extremes, and the separation between "standard" and "completely non-standard" is too large, resulting in an incomplete representation of the intermediate scores. The BP neural network is less sensitive to time series features than the LSTM network. Therefore, it is difficult to distinguish all test data effectively, and the model is not robust enough. The scores of *method 3* are too concentrated in the middle and lower reaches, and the discrimination needs to be more prominent. This is because the feature extraction of single-view videos

needs to be more apparent to extract the action features thoroughly, and thus cannot give an accurate score. In contrast, only the results of the proposed method match the evaluation given by the teacher. Simultaneously, the discrimination of the scores is moderate, indicating that the proposed method can evaluate the quality of Taekwondo Poomsae actions.

To quantitatively assess how the model output scores correlate to the subjective teacher evaluations of the Taekwondo Poomsae performance, threshold ranges were established as follows:

- Score ≥ 0.9 = Qualitative rating of "Excellent"
- Score 0.8–0.89 = Qualitative rating of "Good"
- Score 0.7–0.79 = Qualitative rating of "Fair"
- Score 0.6–0.69 = Qualitative rating of "Poor"
- Score < 0.6 = Qualitative rating of "Bad"

These threshold score ranges were determined based on consultation with domain experts and analysis of sample score distributions. For statistical validation, the F1-score and confusion matrices were examined. The high F1-score on the test dataset was achieved, providing evidence for strong agreement between the model rating and teacher rating for each performance.

The confusion matrices showed reliable score discrimination, with a high true positive rate for each qualitative rating bracket. Misclassifications predominantly occurred between adjacent rating categories, indicating that the output scores effectively capture differences in the Poomsae execution.

By establishing firm score rating thresholds aligned to subjective evaluations and applying standard statistical measures (F1 and confusion matrices) to quantify agreement, the correlation between the computed scores and human judgments can be concretely validated, increasing confidence in the model's ability to evaluate performance quality objectively.

Precision, recall, F1 metrics, and confusion matrices were leveraged to quantitatively evaluate model rating agreement compared to teacher ratings, using a test dataset encompassing 50 performances rated on a scale of 1 to 5 by both the model and teachers.

Fig. 5 presents the strong overall accuracy metrics achieved. Precision reflects how often teacher ratings match when the model predicts each rating, while recall quantifies sensitivity in correctly assigning all cases of a certain rating. F1 score combines both precision and recall. Across all ratings, high scores demonstrate excellent quantitative agreement.

The confusion matrices in Fig. 6 and Fig. 7 indicate over 80% match rates along the diagonals and fewer than 15% mismatches between adjacent rating categories (1 vs. 2, 2 vs. 3), which aligns well quantitatively.

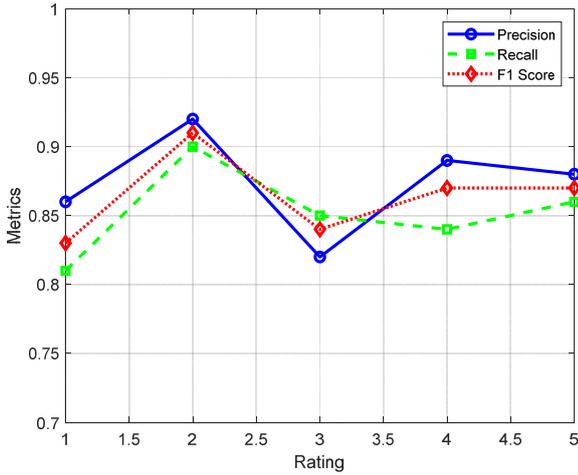


Fig. 5. Accuracy metrics.

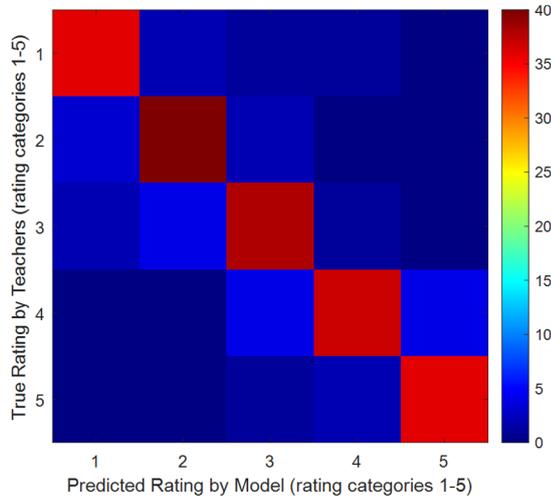


Fig. 6. Model confusion matrix.

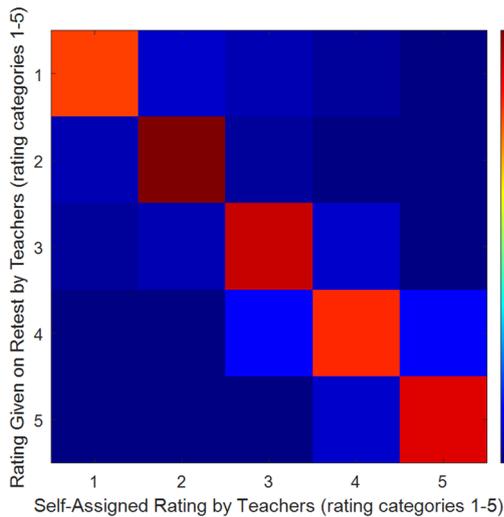


Fig. 7. Teacher confusion matrix.

The strong accuracy metrics and distribution alignment substantiate that the model evaluated Poomsae quantitatively on par with expert judgments, thereby demonstrating objective automated rating capabilities.

5.2. Recognition Effect Experiment

To verify the recognition effect of CNN-Mogrifier LSTM, this study selected ten taekwondo athletes for experiments. The Poomsae actions of the testers are jumping front kick, knee strike, roundhouse kick, hammerfist strike, side kick, and knifehand neck strike. A total of 27,040 samples were collected by the sliding window method, among which 18,319 samples (about 70%) were selected as the training set, and the remaining samples (about 30%) were selected as the test set for experiments. The number of samples for each action pattern in the dataset is shown in Table 3.

In this study, four algorithms including LSTM, Mogrifier LSTM, CNN-LSTM and the proposed CNN-Mogrifier LSTM were used for experiments to evaluate the recognition accuracy. The number of interaction rounds r between x_t and h_{t-1} in Mogrifier LSTM needs to be determined by experiments. Fig. 8 shows the recognition accuracy of Mogrifier LSTM under different values of r . It can be found that the accuracy is the highest when $r = 5$, so 5 is selected for the number of interactions r in Mogrifier LSTM and CNN-Mogrifier LSTM. Table 4 shows the recognition accuracy obtained by each algorithm.

The analysis of Table 4 shows that after five rounds of interaction between x_t and h_{t-1} , Mogrifier LSTM can

Table 3. Poomsae action recognition dataset.

Action	Number of samples	Training set	Test set
Jumping front kick	5,152	3,674	1,478
Knee strike	4,722	3,629	1,093
Roundhouse kick	4,243	2,913	1,330
Hammerfist strike	3,785	2,571	1,214
Side kick	4,762	3,347	1,415
Knifehand neck strike	4,376	3,056	1,320

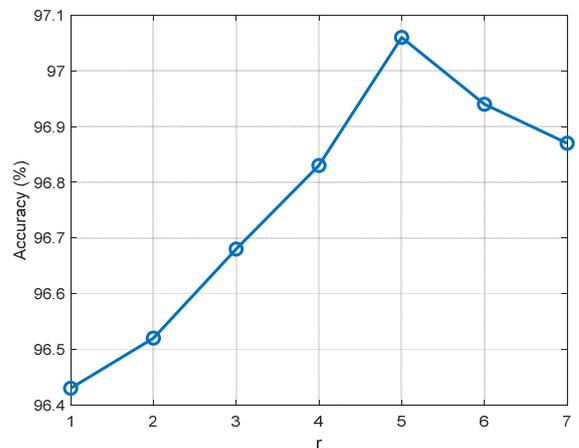


Fig. 8. Mogrifier LSTM recognition accuracy with different r .

Table 4. Poomsae action recognition accuracy.

LSTM	Mogrifier LSTM	CNN-LSTM	CNN-Mogrifier LSTM
93.66 (7,353/7,851)	97.16 (7,628/7,851)	97.43 (7,649/7,851)	98.09 (7,701/7,851)

better mine the dependence of time series data, and the recognition accuracy is significantly better than LSTM. Compared with only using LSTM, CNN-LSTM first extracts a large number of local spatial features of data through CNN, which can provide more effective input data for LSTM and improve the accuracy of recognition. However, compared with CNN-LSTM, the proposed CNN-Mogrifier LSTM makes full use of the advantages of Mogrifier LSTM and can mine the forward and backward dependencies of local spatial features better than LSTM. Experimental results show that CNN-Mogrifier LSTM has achieved high recognition accuracy of Taekwondo Poomsae actions.

VI. CONCLUSION

This paper studies the cloud IoT-oriented neural network-based Taekwondo teaching scheme. We use a technology-enabled approach to Taekwondo instruction that implements the IoT, cloud computing, and neural networks to enhance students' learning experience. We use sensors to collect data on students during Taekwondo training, and this Poomsae action data is uploaded to the cloud through the IoTz. Then, taking Taekwondo Poomsae actions as the research object, the CNN-Mogrifier LSTM algorithm is proposed for the intelligent evaluation of Taekwondo Poomsae actions. Accurate and robust evaluation results are provided by extracting and modeling multi-view features of Poomsae action videos. The CNN-Mogrifier LSTM algorithm can accurately identify the change characteristics of Poomsae actions. It can also effectively quantify the strength and coherence of actions, which can objectively score Poomsae actions. More importantly, it provides theoretical support for the quality evaluation of other actions in Taekwondo. The proposed method has good Taekwondo Poomsae action quality evaluation ability, as evidenced by the effect test of Taekwondo Poomsae action intelligent evaluation, which demonstrates that the results of the proposed method follow the evaluation provided by the teachers and the scores discrimination is moderate. The experimental results also demonstrate that CNN-Mogrifier LSTM breaks away from the constraints of teachers and venues to realize automatic and intelligent Taekwondo teaching. It also provides comprehensive evaluation and improvement

suggestions for the quality of action completion.

While this work focused explicitly on Taekwondo for feasibility analysis, the sensor-based motion capture and neural network rating approach could be adapted to other sports. For gymnastics, the methodology would need to customize sensor positioning and motion features to capture the nuances required for events like balance beam, vault, and floor exercises. Sensors need full body coverage to capture intricate details. Features around balance, height, rotation speed, and limb extension would be tailored specifically for judging proper technique and form. The model architecture of coupled CNN and LSTM networks could remain to identify spatiotemporal patterns in execution. However, the output layer ratings would be defined differently than the 1–5 scale used for Taekwondo - instead, outputting deductions on a 10-point scale as commonly used in gymnastic scoring. The model training would require an expansive dataset of sample gymnastics performances covering acceptable and faulty technique variations. While domain customization is needed, the core methodology of sensor-based data fusion, automated feature extraction, and deep neural network pattern rating can extend across sports and events with performance subjectivity concerns. The system outputs could provide athletes and coaches with objective, real-time feedback identifying areas for technique improvement personalized to the sport's unique scoring criteria.

Although promising results, a range of limitations and risks open opportunities for extending this exploratory research. While wearable motion sensors can achieve reasonable accuracy, susceptibility to environmental noise and calibration errors may impact the fidelity and reliability of collected data. Moreover, assessments were limited in scope to three specific forms, so additional evaluation across a more comprehensive set of Taekwondo techniques would strengthen generalizability. With a constrained annotated dataset available, overfitting risks may also arise during neural network training that could require further cross-validation and regularization strategies to mitigate. Significant practical barriers to field deployment and coaching adoption also exist, including effectively managing device hardware, cost issues, privacy concerns, model interpretability challenges, and facilitating non-technical end-user interaction. Since only near-term quantitative metrics were examined, follow-up work should also assess how longitudinal exposure could reshape trainee practices, retention, and pedagogical outcomes. Through holistically addressing this range of analytical risks, resource demands, broad applicability gaps, algorithmic vulnerabilities, interface obstacles, and influence time horizons, the field can pursue transformative translation of proposed methods beyond initial feasibility into robust and accessible real-world solutions.

REFERENCES

- [1] H. L. Liang, "Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model," *Journal of Electromagnetic Engineering and Science*, vol. 40, no. 2, pp. 3239-3250, Feb. 2021.
- [2] S. K. Kim, "A Study on the teaching methods and values of Chinese taekwondo masters," *Archives of BUDO*, vol. 16, pp. 203-210, Aug. 2020.
- [3] S. T. Chin, M. Su, C. Hong, J. L. Yu, Q. Ye, and M. Rahman, et al., "Research on Taekwondo teaching reform in colleges and universities based on nonlinear data prediction analysis," *Scientific Programming*, vol. 2022, no. 1464692, Jul. 2022.
- [4] Y. H. Chang, S. Meng, and H. Chao, "Innovative research on teaching method of taekwondo in college elective courses under the background of big data," *Applied Bionics and Biomechanics*, vol. 2022, no. 2329952, Jan. 2022.
- [5] C. H. Cho and H. J. Joo, "Motion recognition technology based remote Taekwondo Poomsae evaluation system," *Multimedia Tools and Applications*, vol. 75, pp. 13135-13148, Aug. 2015.
- [6] W. S. Jung, H. W. Moon, J. W. Kim, H. Y. Park, J. B. Park, and S. W. Choi, et al., "Analysis of cardiopulmonary function, energy metabolism, and exercise intensity and time according to the number of repetitions of Taekwondo Taegeuk Poomsae in Taek-wondo players," *Journal of Men's Health*, vol. 18, no. 4, pp. 1-10, Apr. 2022.
- [7] Q. G. Zeng, "Research on Taekwondo analysis Tactics based on artificial intelligence," *Argo Food Industry Hi-Tech*, vol. 28, pp. 2612-2616, Mar. 2017.
- [8] S. El-Ashker, H. Chaabene, and O. Prieske, "Maximal isokinetic elbow and knee flexor-extensor strength measures in combat sports athletes: The role of movement velocity and limb side," *BMC Sports Science, Medicine and Rehabilitation*, vol. 14, no. 40, Mar. 2022.
- [9] Y. Q. Yang, H. J. Lv, and N. Chen, "A survey on ensemble learning under the era of deep learning," *Artificial Intelligence Review*, vol. 56, no. 6, pp. 5545-5589, Nov. 2022.
- [10] A. Mumuni and F. Mumuni, "Robust appearance modeling for object detection and tracking: A survey of deep learning approaches," *Progress in Artificial Intelligence*, vol. 11, no. 21, pp. 279-313, Sep. 2022.
- [11] A. Mumuni and F. Mumuni, "Survey on the application of deep learning in the internet of things," *Telecommunication Systems*, vol. 79, no. 4, pp. 601-627, Apr. 2022.
- [12] F. Liang, Z. L. Zhang, X. Y. Li, and Z. Tong, "Lower limb action recognition with motion data of a human joint," *Arabian Journal for Science and Engineering*, vol. 41, no. 12, pp. 5111-5121, May 2016.
- [13] W. B. Zhao, "A concise tutorial on human motion tracking and recognition with microsoft kinect," *Science China Information Sciences*, vol. 59, no. 9, pp. 93101, Aug. 2016.
- [14] M. R. Keyvanpour, S. Vahidian, and M. Ramezani, "HMR-vid: A comparative analytical survey on human motion recognition in video data," *Multimedia Tools and Applications*, vol. 79, no. 43/44, pp. 31819-31863, Nov. 2020.
- [15] V. Varadarajan, V. Neelanarayanan, R. Doyle, I. F. Al-Shaikhli, and S. Groppe, "Emerging solutions in big data and cloud technologies for mobile networks," *Mobile Networks and Applications*, vol. 24, no. 3, pp. 1015-1017, Feb. 2019.
- [16] A. Liu, R. P. Mahapatra, and A. V. R. Mayuri, "Hybrid design for sports data visualization using AI and big data analytics," *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2969-2980, Oct. 2021.
- [17] E. Mazurova, W. Standaert, and E. Penttinen, "Paradoxical tensions related to AI-powered evaluation systems in competitive sports," *Information Systems Frontiers*, vol. 24, no. 3, pp. 897-922, Nov. 2021.
- [18] J. G. Claudino, D. O. Capanema, T. V. de Souza, J. C. Serrao, A. C. M. Pereira, and G. P. Nassis, "Current approaches to the use of artificial intelligence for injury risk evaluation and performance prediction in team sports: A systematic review," *Sports Medicine - Open*, vol. 5, no. 28, pp. 1-2, Jul. 2019.
- [19] S. Wei, K. L. Wang, and X. L. Li, "Design and implementation of college sports training system based on artificial intelligence," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 1, pp. 971-977, Jul. 2021.
- [20] L. Rigamonti, K. Estel, T. Gehlen, B. Wolfarth, J. B. Lawrence, and D. A. Back, "Use of artificial intelligence in sports medicine: A report of 5 fictional cases," *BMC Sports Science, Medicine and Rehabilitation*, vol. 13, no. 13, Feb. 2021.
- [21] N. V. Shah, R. Gold, Q. A. Dar, B. G. Diebo, C. B. Paulino, and Q. Naziri, "Smart technology and orthopaedic surgery: Current concepts regarding the impact of smartphones and wearable technology on our patients and practice," *Current Reviews in Musculoskeletal Medicine*, vol. 14, no. 6, pp. 378-391, Dec. 2021.
- [22] P. Ramasamy, E. Calderon-Sastre, G. Renganathan, S. Das, and Y. Kurita, "Soft actuators-based skill training wearables: A review on the interaction modes, feedback types, VR scenarios, sensors utilization and applications," *ROBOMECH Journal*, vol. 10, no. 1, Jan. 2023.

- [23] C. H. Wang and C. Du, "Optimization of physical education and training system based on machine learning and internet of things," *Neural Computing and Applications*, vol. 34, no. 1, pp. 9273-9288, Jul. 2021.
- [24] X. Wu, J. C. Zhou, M. X. Zheng, S. W. Chen, D. Wang, and J. Anajemba, et al., "Cloud-based deep learning-assisted system for diagnosis of sports injuries," *Journal of Cloud Computing*, vol. 11, no. 1, no. 82, Nov. 2023.
- [25] X. Ren, W. Ding, S. E. Crouter, Y. Mu, and R. Xie, "Activity recognition and intensity estimation in youth from accelerometer data aided by machine learning," *Applied Intelligence*, vol. 45, no. 2, pp. 512-529, Apr. 2016.
- [26] Q. H. Wu, Q. Huang, and X. Li, "Multimodal human action recognition based on spatio-temporal action representation recognition model," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16409-16430, Nov. 2022.
- [27] Y. Kong and Y. Fu, "Human action recognition and prediction: A survey," *International Journal of Computer Vision*, vol. 130, no. 5, pp. 1366-1401, Mar. 2022.
- [28] F. Kulsoom, S. Narejo, Z. Mehmood, H. N. Chaudhry, A. Butt, and A. K. Bashir, "A review of machine learning-based human activity recognition for diverse applications," *Neural Computing and Applications*, vol. 34, no. 21, pp. 18289-18324, Aug. 2022.
- [29] S. K. Ghosh, B. R. Mohan, and R. M. R. Guddeti, "Deep learning-based multi-view 3D-human action recognition using skeleton and depth data," *Multimedia Tools and Applications*, vol. 82, no. 13, pp. 19829-19851, Nov. 2022.
- [30] C. Han, L. Zhang, Y. Tang, W. B. Huang, F. H. Min, and J. He, "Human activity recognition using wearable sensors by heterogeneous convolutional neural networks," *Expert Systems with Applications*, vol. 198, no. 15, no. 166764, Mar. 2022.
- [31] L. K. Wang and R. Y. Liu, "Human activity recognition based on wearable sensor using hierarchical deep LSTM networks," *Circuits, Systems, and Signal Processing*, vol. 38, no. 2, pp. 837-856, Apr. 2019.
- [32] F. J. Ordonez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, no. 115, Jan. 2016.
- [33] Y. Zhang, "Track and field training state analysis based on acceleration sensor and deep learning," *Evolutionary Intelligence*, vol. 16, no. 4, pp. 1627-1636, Jan. 2023.
- [34] E. Ranjbar and A. A. Suratgar, "Composite adaptive controller design for 3-DOF MEMS vibratory gyroscopes capable of measuring angular velocity," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 43, no. 2, pp. 245-266, Jul. 2018.
- [35] B. Munoz-Barron, J. R. Rivera-Guillen, R. A. Osornio-Rios, and R. J. Romero-Troncoso, "Sensor fusion for joint kinematic estimation in serial robots using encoder, accelerometer and gyroscope," *Journal of Intelligent & Robotic Systems*, vol. 78, no. 3-4, pp. 529-540, Sep. 2014.
- [36] G. Rafiq, M. Rafiq, and G. S. Choi, "Video description: A comprehensive survey of deep learning approaches," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13293-13372, Nov. 2023.
- [37] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313-2324, Oct. 2019.
- [38] Y. Y. Fu, K. Liang, and J. Xu, "MLog: Mogrifier LSTM-based log anomaly detection approach using semantic representation," *IEEE Transactions on Services Computing*, vol. 16, pp. 3537-3549, Sep. 2023.
- [39] H. Giang Thi Thu, T. Nguyen Thanh, and T. Le Quy, "Dynamic sliding window and neighborhood LSTM-based model for stock price prediction," *SN Computer Science*, vol. 3, no. 3, no. 256, Apr. 2022.

AUTHOR



Shuqun Qu received her B.S. degree from Shanghai University of Sport in 1999, and received her M.S. degree from Liaoning Normal University in 2006. She is currently an Associate Professor at Jilin Sport University. Her research interests include Sport Education, IoT and computing model.

