SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 638.
https://doi.org/10.62617/mcb638

Article

Recognition of foot footwork based on dual model convolutional neural network-driven biomechanics patterns

Weihong Shen

Department of Arts and Sports, Henan Technical College of Construction, Zhengzhou 450000, China; shenweihong2023@163.com

CITATION

Shen W. Recognition of foot footwork based on dual model convolutional neural network-driven biomechanics patterns. Molecular & Cellular Biomechanics. 2025; 22(2): 638. https://doi.org/10.62617/mcb638

ARTICLE INFO

Received: 25 October 2024
Accepted: 5 November 2024
Available online: 23 January 2025

Abstract: Today, with the increasing popularity of football, more and more scholars will focus on using the digital and systematic management methods of football to further improve the safety and effectiveness of football. Similar to how cells within an organism operate in a highly coordinated and biomechanically regulated manner, football players' movements also involve complex biomechanical processes. Establishing a football data analysis system and guiding athletes is like understanding and modulating the molecular interactions and mechanical forces within cells to optimize their function. However, the current such systems are mostly based on video monitoring technology, and their actual operation process is limited by the deployment environment and expensive. In order to realize the widespread popularity of football movement analysis, this paper uses intelligent wearable devices, based on dual model convolutional neural network (DMCNN) for football players virtual step (behind), step (puskash), push progress (sliding), test (inside and outside cycling) and jump (Ronaldo), and by adjusting the convolutional core size and convolution step parameters optimize neural network performance. The resulting algorithm model, which outperforms the K nearest neighbor (KNN) and support vector machine (SVM) algorithms, provides a more accurate understanding of football players' biomechanical patterns, just as advanced cell molecular biomechanics techniques offer deeper insights into cellular behavior and function, potentially leading to more refined training regimens and injury prevention strategies in football.

Keywords: dual convolutional neural network; football footwork; action recognition; and experimental analysis; biomechanical patterns

1. Introduction

Football is a popular competitive sport, and improving the competitive level of athletes is crucial. World Cup players have professional training plans for their physical fitness, including physical training and skill training. Core strength training can improve shooting rates, and physical training helps prevent injuries. Among these, football footwork training is especially important as modern football games require better footwork for rapid attack and defense conversion, physical control, coordination, and tactical system formation.

In recent years, with the wide application of video surveillance, smart home and unmanned driving, behavior recognition has gradually become a research hotspot in the field of computer vision [1–4]. In the context of human movement recognition, a popular research direction [5], two main methods for action recognition using artificial intelligence technology exist.

The robust recognition and assessment of human actions are crucial in human-robot interaction (HRI) domains [6–9]. One is action recognition based on video images [10]. It has several limitations: (1) The detection and extraction of targets in the image sequence demand complex image processing; (2) It requires infrastructure support such as installing video recorders in venues and cannot perform motion recognition anywhere; (3) Cameras may capture other objects, interfering with the recognition effect; (4) It is highly dependent on lighting, and poor lighting affects identification accuracy; (5) Cameras recording human activities raise personal privacy concerns; (6) High-definition cameras are expensive and not universally applicable.

The other method is action recognition based on wearable sensors, which has become a preferred choice for sports, especially football, due to its many advantages. These include: (1) Adaptability to different movement scenarios and ease of implementation; (2) The ability to be worn on various parts of the athlete's body to obtain real-time movement data; (3) Lightweight, portable, comfortable, and easy-to-control features; (4) The data comes from the target player and is not affected by other data during action recognition.

Regarding the application of wearable devices in football sports analysis, significant progress has been made. Wearable global positioning systems have been widely used to quantify the external exercise load of students in campus football training and competitions. They can accurately record various data such as running distance, speed, and acceleration. Portable heart rate tracking devices monitor the internal exercise load, providing real-time information about athletes' exertion levels and injury susceptibility. Some advanced systems integrate player movement data with video analysis, enabling coaches to make more informed decisions. Smart shoes with built-in sensors can generate professional data related to footwork, and wearable devices also play a role in injury prevention and the transfer market. Future developments may lead to more miniaturized, intelligent, and integrated devices, although challenges such as data security and device accuracy remain.

2. Related work

At present, there are two main methods for action recognition using artificial intelligence technology. One scheme is action recognition based on video images, and the other is action recognition based on wearable sensors. Action recognition based on video image has the following disadvantages: 1) The detection and extraction of targets in the image sequence requires a series of complex processing and processing of images [11]; 2) The technology requires infrastructure support, the installation of video recorders in venues, and no motion recognition anytime, anywhere [12]; 3) Besides capturing the athletes, the camera also captures other objects, interfering with the final recognition effect [13]; 4) Highly dependent on lighting, if the light is too dark, will affect the identification accuracy [14]; 5) Cameras video record human activities and involve personal privacy issues [15]; 6) For the desired recognition effect, the system requires a high-definition camera, but it is too expensive to be universal [16].

In recent years, technology based on wearable sensors have become the best choice to identify sports, especially football [17,18]. Using wearable technology for motion recognition has the following advantages:1) Can be adapted to different movement scenarios, easy to implement [19]; 2) It can be worn in various parts of the athlete's body and obtain real-time movement data from the sensors [20]; 3) Lightweight, portable, comfortable, and easy to control features [21]; 4) The data comes from the target player and is not affected by other data when identifying the action [22]. To sum up, the integration of wearable sensors and artificial intelligence technology has become the main application scheme of a large number of action analysis and recognition systems, and is also a hot research direction in the field of action recognition in the future.

The scheme based on wearable sensor collects data through various sensors, processes and analyzes the data accordingly, and then uses various classification algorithms to realize action recognition [23]. Some scholars have proposed a sensor system based on friction electrification and adopted the K-nearest neighbor algorithm to identify five daily activities from the extracted features [24]. Other scholars have used deep convolutional neural networks to identify daily activities [25]. The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions [26]. The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions [27].

The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions. The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions. The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions. The literature proposes a wristband-based sensor system to identify actions in football games that contains two-stage classifiers that effectively avoid interference from erroneous actions.

3. Footstep recognition of football players based on dual model convolutional neural network

3.1. Overview of action recognition based on wearable devices

3.1.1. Action identification, classification and process

Based on wearable sensors action recognition scheme is according to people different movement scene, choose the appropriate sensor such as the most common pressure gauge and accelerometer, they will be installed in the important parts of people's body movement, when people move sensor will generate signals, collect effective data information from each position of the body. Because the body movement sensor also move, so need to the sensor data processing such as noise reduction and jitter, for different movement events to establish the corresponding model, extract the key features of important parts of the body or generate advanced features, finally we set up good model training, implementation in different movement scenarios of low-level simple action and advanced complex action. **Figure 1** depicts the low-level and higher-level classification for action recognition.

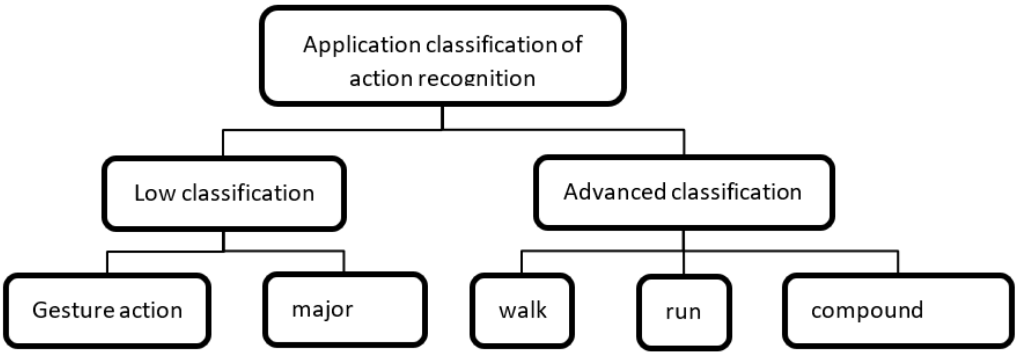

Figure 1. Application classification of action recognition.

Based on wearable sensors action recognition scheme is according to people different movement scene, choose the appropriate sensor such as the most common pressure gauge and accelerometer, they will be installed in the important parts of people's body movement, when people move sensor will generate signals, collect effective data information from each position of the body. Because the body movement sensor also move, so need to the sensor data processing such as noise reduction and jitter, for different movement events to establish the corresponding model, extract the key features of important parts of the body or generate advanced features, finally we set up good model training, implementation in different movement scenarios of low-level simple action and advanced complex action. **Figure 2** depicts the low-level and higher-level classification for action recognition.

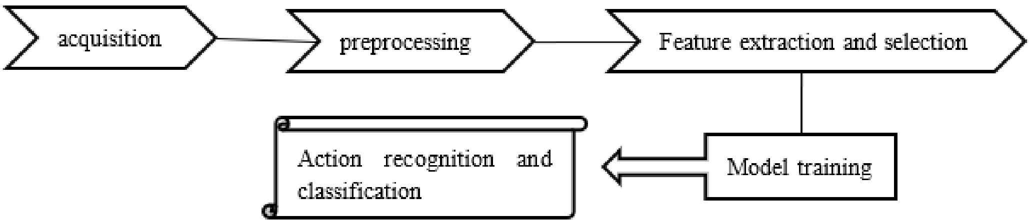

Figure 2. Action recognition process.

3.1.2. Data processing and segmentation

Data collection, as the first step in the action recognition process, is actually not simple, but is limited by many conditions. For different motion scenarios, we can not choose the sensors at will. First of all, we should evaluate the performance and versatility of the sensors. Because many of the actions in daily life are uncertain, the selected sensor combination should also adapt to various test schemes. In addition, the size, weight, signal, and life of the sensor should also be considered. Wearable sensors commonly used in daily life include health monitoring sensors, inertial sensors and cameras, among which inertial sensors such as barometers, accelerometers and angular speedometers are the most commonly used, which can qualitatively measure pressure, displacement and rotation. After choosing the sensor, we need to arrange the sensor according to different motion scenes and recognition tasks. Because various parts of the human body have different ability to respond to various movements, so being placed in different parts will produce different performance. When deploying sensors, the same type of sensor can be placed in multiple parts of the body, or multiple types of sensors can be installed in the same part to establish a basic sports movement recognition network. In short, the high complexity of the sensor layout is conducive to improving robustness.

Based on wearable sensors action recognition scheme is according to people different movement scene, choose the appropriate sensor such as the most common pressure gauge and accelerometer, they will be installed in the important parts of people's body movement, when people move sensor will generate signals, collect effective data information from each position of the body. Because the body movement sensor also move, so need to the sensor data processing such as noise reduction and jitter, for different movement events to establish the corresponding model, extract the key features of important parts of the body or generate advanced features, finally we set up good model training, implementation in different movement scenarios of low-level simple action and advanced complex action. **Figure 1** depicts the low-level and higher-level classification for action recognition. Several common normalization methods are log function transformation, atan function transformation, min-max normalization, and *Z*-score normalization. Several common normalization methods are log function transformation, atan function transformation, min-max normalization, and *Z*-score normalization.

$$Y^{*} = \frac{Y - \min}{\text{max-min}} \tag{1}$$

where, the vector *Y* is the raw data, and min and max represent the minimum and maximum values in *Y*, respectively. The *Z*-score standardization is to process the standard deviation and mean value of the raw data, and its conversion function is:

$$Y' = \frac{Y - \overline{Y}}{T} \tag{2}$$

After completing the above steps, the final step is to segment the raw sensor data. Because in different motion scenarios, the data collected by the sensor at a certain period of time may contain multiple action segments of the subject. In order to subsequently add action tags and send the data to the model for training, a complete piece of data needs to be divided into multiple segments. In experiments, the sliding window technique is often used to achieve data segmentation, completing the specific required data segmentation by defining the distance of the sliding step length and the width of the sliding window. **Figure 3** shows the operation of the sliding window, where gray lines represent the original data collected and white lines represent the data segments generated by the sliding window segmentation. The sliding window width is 3 and the sliding step length is 2, creating overlap between two adjacent windows, so there may be the same sampling point in different data segments.

The sliding window technique plays a vital role in segmenting the collected sensor data into portions that are amenable to effective analysis by the model. The window width serves as a determinant of the quantity of data encompassed within each segment. A wider window width would incorporate more data points, while a narrower one would include fewer. In the context of this study, a window width of 3 was selected. This choice was not arbitrary but was based on a comprehensive understanding of the football footwork data. Through prior knowledge acquisition and an initial exploration of the data, it was determined that a window width of 3 could effectively capture a substantial amount of pertinent information regarding each footwork action. This is attributed to the fact that it encompasses a short sequence of data points. These data points are crucial as they have the capacity to represent the essential characteristics of a specific foot movement. For instance, during a footwork action like a quick step or a turn, there are distinct phases such as acceleration and deceleration. The window width of 3 is sufficient to cover these phases and provide a comprehensive snapshot of the movement, thereby enabling the model to better understand and analyze the footwork.

The sliding step length, on the other hand, is a critical factor in controlling the overlap between adjacent segments. This overlap is of great significance as it influences the way the model interprets the continuous flow of the footwork data. A step length of 2 was chosen for this study. The rationale behind this selection lies in its ability to ensure an optimal level of overlap between segments. When the step length is set to 2, there is a notable overlap between adjacent windows. This overlap is beneficial in multiple ways. It allows for a more accurate capture of any transitions or continuations in the footwork actions. For example, when a player transitions from one type of footwork to another, such as from a normal step to a quick step, the overlap provided by a step length of 2 ensures that the data related to this transition is not lost. Instead, it is captured within the overlapping regions of the adjacent segments. This overlap also contributes to a smoother representation of the continuous nature of the movement. It mimics the natural fluidity of the footwork and reduces the likelihood of losing important information that might occur at the boundaries of the segments. If there were no overlap or an insufficient overlap, there would be a risk of data discontinuity, which could lead to inaccurate analysis and recognition of the footwork by the model.

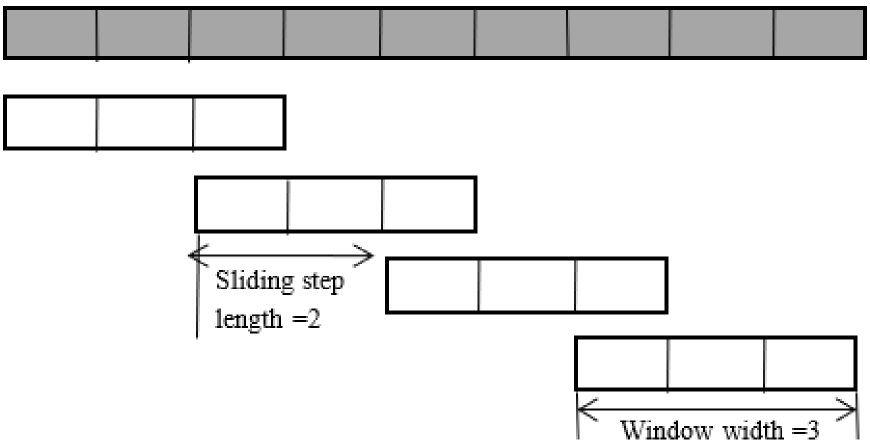

Figure 3. Data segmentation.

In this study, we use the wavelet transform for noise reduction. The wavelet transform uses a set of basis functions called wavelets. The choice of wavelet basis function is crucial for effective noise removal. Here, we choose the Daubechies wavelet family as our wavelet basis function. The Daubechies wavelets are a family of orthogonal wavelets that have good localization properties in both the time and frequency domains.

The wavelet transform of a signal $x(t)$ is given by:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\phi(\frac{t-b}{a})\, dt \tag{3}$$

where a is the scale parameter and b is the translation parameter, and is the wavelet function.

For the Daubechies wavelets, the wavelet function ϕ(t) is defined by a set of coefficients. The number of coefficients determines the order of the Daubechies wavelet. In this study, we choose the Daubechies-4 wavelet, which has 4 coefficients. The Daubechies-4 wavelet function is given by:

$$\phi(t) = \sum_{k=0}^{3} h_k \varphi(2t-k) \tag{4}$$

where h_k are the coefficients of the Daubechies-4 wavelet and ϕ(t) is the scaling function.

The scaling function ϕ(t) is related to the wavelet function ϕ(t) by the following equation:

$$\varphi(t) = \sqrt{2} \sum_{k=0}^{3} gk\varphi(2t-k) \tag{5}$$

where *gk* are the coefficients of the scaling function.

In addition to choosing the wavelet basis function, we also need to determine the decomposition level. The decomposition level determines the number of times the signal is decomposed into different frequency components. A higher decomposition level means more detailed frequency analysis but also more computational complexity. In this study, we choose a decomposition level of 3. This means that the signal is decomposed into 8 different frequency components ($2^3 = 8$).

3.2. Footballer footwork recognition

3.2.1. Football footwork data processing

After completing the above steps, the final step is to segment the raw sensor data. Because in different motion scenarios, the data collected by the sensor at a certain period of time may contain multiple action segments of the subject. In order to subsequently add action tags and send the data to the model for training, a complete piece of data needs to be divided into multiple segments. In experiments, the sliding window technique is often used to achieve data segmentation, completing the specific required data segmentation by defining the distance of the sliding step length and the width of the sliding window. **Figure 3** shows the operation of the sliding window, where gray lines represent the original data collected and white lines represent the data segments generated by the sliding window segmentation. The sliding window width is 3 and the sliding step length is 2, creating overlap between two adjacent windows, so there may be the same sampling point in different data segments (**Table 1**).

Table 1. Sample basic information.

Football Players No.	age	height	weight
01	25	173	65
02	26	176	73
03	23	180	75

The purpose of the experiment is to identify the footwork of football players, through the research a lot of football training data and watch football training video, the experiment selected five classic football footwork (**Figure 4**): virtual sway (behind), turn (sparkas), fast step (sliding), test step (inside and outside bike) and jump (Ronaldo ball).The purpose of the experiment is to identify the footwork of football players, through the research a lot of football training data and watch football training video, the experiment selected five classic football footwork: virtual sway (behind), turn (sparkas), fast step (sliding), test step (inside and outside bike) and jump (Ronaldo ball).

(a) False swing

(b) turn round

(c) quick attack

(d) Temptiv

(e) jump

Figure 4. Five classic football footwork styles.

The purpose of the experiment is to identify the footwork of football players, through the research a lot of football training data and watch football training video, the experiment selected five classic football footwork: virtual sway (behind), turn (sparkas), fast step (sliding), test step (inside and outside bike) and jump (Ronaldo ball).

In order to eliminate the influence of the noise generated during the football footwork data acquisition on the recognition effect, this experiment uses the wavelet transform to reduce the noise of the original data of the five football footwork. The following two formulas represent its scale function and its wavelet function respectively:

$$\phi_{j,k}(t) = 2^{j/2}\phi_k(2^j t - k)/\psi_{j,k}(t) = 2^{j/2}\psi_k(2^j t - k) \quad (6)$$

In addition, due to the huge numerical gap in acceleration and angular velocity in the collected football footwork data and the different measurement units, this paper, *Z*-score standardized operation is used to process the raw sensor data. In addition, due to the huge numerical gap in acceleration and angular velocity in the collected football footwork data and the different measurement units, this paper, *Z*-score standardized operation is used to process the raw sensor data. After processing the football footwork data, in order to facilitate the subsequent experimental data analysis and the classification of football footwork, we added a label to each football footwork, with $f_1, \ldots, f_5$ representing side virtual swing, turn, fast break, test and jump, as shown in **Table 2**.

Table 2. Labels of football footwork.

football footwork	labels
False swing	f_1
Turn round	f_2
quick attack	f_3
Temptiv	f_4
jump	f_5

Finally, for the subsequent model training of football footwork recognition, data segmentation is needed on the data set composed of five football footwork actions, and this data set is divided into the training set and test set for football footwork recognition in a ratio of 7:3.

3.2.2. Extraction of football footwork features based on deep learning

CNN is used to identify football footwork, which can realize the automatic extraction of football footwork data features, and improve the efficiency of feature extraction in this experiment. CNN is usually composed of many neural network layers, which are mainly extracted through the convolutional layer, and then pools the layer to rescreen the features. Convolutional layer is the convolution operation through the convolution kernel, which can also be understood as the "filter operation" on image processing. In fact, it is given a square matrix and input matrix, and finally output a matrix. The calculation formula of output matrix size is as follows:

$$O = (I - F + 2P)/S + 1 \quad (7)$$

Convolutional layer is the convolution operation through the convolution kernel, which can also be understood as the "filter operation" on image processing. In fact, it is given a square matrix and input matrix, and finally output a matrix. The calculation formula of output matrix size is as follows:

$$f(x,y) \times g(x,y) = \int_{\phi_1=-\infty}^{\infty} \int_{\phi_2=-\infty}^{\infty} f(\phi_1,\phi_2) \times g(x-\phi_1, y-\phi_2)\, d\phi_1 d\phi_2 \quad (8)$$

Convolutional layer is the convolution operation through the convolution kernel, which can also be understood as the "filter operation" on image processing. In fact, it is given a square matrix and input matrix, and finally output a matrix.

3.3. Identification of football players' footwork based on dual-model convolution

CNN was also a deep learning method to identify and classify five football footwork. In the previous paper, the role and working principle of convolutional layer and pooling layer in feature extraction and selection are explained. In addition to these two parts, the CNN also has the full connection layer and the two important components of the softmax. **Figure 5** illustrates the CNN architecture applied to the football step identification of this experiment, and the model successively consists of an input layer, a first convolutional layer, a pooling layer, a second convolutional layer, and a softmax layer. This experiment first set the filter size of the first convolutional layer to 60 to extract the low-level features of the five football footwork data, then set the maximum convolution kernel size and the convolutional step to 2 to select the features of the five football footwork data; again, take the output of the pooling layer as the input to the second convolutional layer and set the filter size to 6 to extract the advanced features of the five football footwork data.

After obtaining the five high-level features of the five soccer footwork data, these high-level features were mapped to the label space of the tanh function in a nonlinear mapping of the football stepping data through a fully connected layer with 1000 neurons. The fully connected layer is similar to the convolutional layer, using point multiplication, combining weight, and bias b to give the score that the sample to be identified is one of the five football steps:

$$f(x) = \omega x + a \tag{9}$$

Finally, the softmax layer outputs the probability distribution of the five football footwork actions. For the input of this layer, the probability of the sample to be identified as type *i* can be expressed as:

$$Pi = \frac{e^{x_i}}{\sum_{m=1}^{N} e^{x_m}} \tag{10}$$

N represents the number of football footwork types, and the N value is 5. In order to minimize the loss function, the gradient descent algorithm was used to train the football footwork data, and the parameters were constantly updated through iteration. The batch size was 10 and the learning rate was 0.0001 for the CNN for football footwork recognition on TensorFlow.

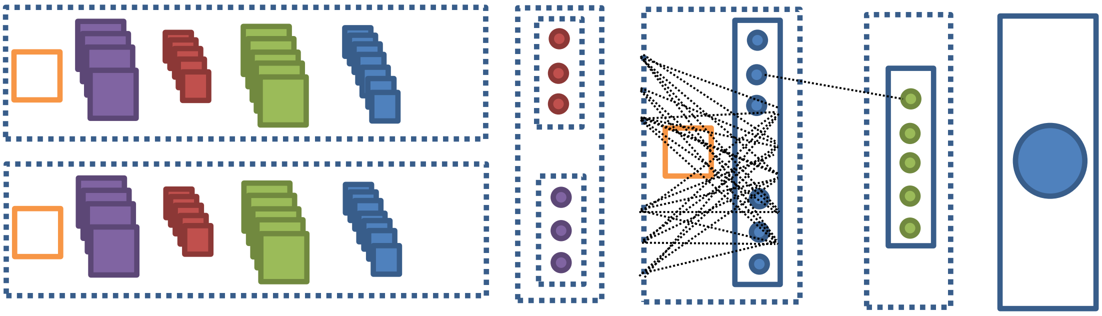

Figure 5. Basic architecture of foot recognition.
Based on dual model convolutional neural network.

Such a single network model may show good identification performance against small sample data. To prevent the phenomenon of overfitting, methods of increasing the number of samples are usually used. However, this approach will reduce the computational efficiency of the model. To solve this problem, this paper tries to integrate multiple networks on the basis of CNN, and proposes dual-model convolutional neural networks to realize the identification of five football steps. **Figure 5** shows the basic architecture of DMCNN, and the specific implementation steps of DMCNN are described below.

First, unlike the identification of football footwork by the same CNN implementation described above for inputting acceleration data and angular speed data. DMCNN inputs the acceleration data of football footwork actions into a CNN as a base classifier, and then inputs the angular speed data of football footwork actions into a CNN as a base classifier. For these two CNN, this experiment set the convolution kernel size to 60, the convolution kernel size of 20 and the size of 2, while the second convolutional layer to 60. Secondly, the above two CNN applied a fully connected layer containing 1000 neurons to combine the output of the second convolutional layer, that is, advanced football footwork features, and output “highly purified” football footwork features, which together constitute a new set of feature vectors. Thirdly, in order to reduce the feature vector of football step samples, principal component analysis was used and n-components was set to 200, so that the first 200 components contained important feature information. Then, the above features were integrated by a fully connected layer containing 200 neurons to obtain the football footwork features required for the final part of this experiment. Finally, the softmax layer outputs the probability distribution after five football footwork actions.

4. Experiment and application

In the majority of action recognition studies, metrics such as accuracy, miscalculation, recall, and accuracy are used to evaluate the performance of an action recognition system. Because the football footwork data set collected in this experiment is relatively balanced, the accuracy was used as a performance index to evaluate the football footwork identification. In the CNN and DMCNN models used for football footwork recognition, the identification accuracy also has some influence. Without changing the structure of CNN super-parameter step network, this experiment obtained the best combination of football step parameters by adjusting the size and the size of the convolution step size. **Table 3** shows the recognition accuracy of the five football footwork methods in this case.

Table 3. Identification accuracy (CNN).

football footwork	accuracy rate
False swing	100%
Turn round	100%
quick attack	100%
Temptiv	96.4%
jump	98.1%

In addition, the DMCNN model proposed in this experiment slightly improves the accuracy of the CNN model, improves the computational efficiency of the whole network, and strengthens the generalization ability and robustness of this experimental model. As shown in **Table 4**, the correct identification of the four football steps was realized under this model, and only the jump steps were not fully identified.

In addition, the DMCNN model proposed in this experiment slightly improves the accuracy of the CNN model, improves the computational efficiency of the whole network, and strengthens the generalization ability and robustness of this experimental model. As shown in **Table 3**, the correct identification of the four football steps was realized under this model, and only the jump steps were not fully identified.

Table 4. Identification accuracy (DMCNN).

Football footwork	Accuracy rate
False swing	100%
Turn round	100%
quick attack	100%
Temptiv	96.4%
jump	99.8%

In the LSTM model experiment, the football footwork dataset was also used. LSTM is a type of recurrent neural network suitable for sequential data. The data was prepared in a way that was compatible with the LSTM model's requirements. The model was trained with appropriate hyperparameters and then evaluated on the test set. The accuracy values for different football footwork actions are shown in **Table 5**.

Table 5. Identification accuracy (LSTM).

Football Footwork	Accuracy Rate
False swing	92.5%
Turn round	93.2%
quick attack	94.1%
Temptiv	88.7%
jump	90.3%

The GRU model experiment was carried out in a similar manner. The football footwork dataset was used, and the GRU model, which is another type of recurrent neural network, was trained and tested. The GRU model has a different structure compared to LSTM, and its performance was evaluated on the test set. The results are presented in **Table 6**.

Table 6. Identification accuracy (GRU).

Football footwork	Accuracy rate
False swing	93.8%
Turn round	94.5%
quick attack	95.2%
Temptiv	90.1%
jump	92.6%

From **Tables 3** and **4**, it can be observed that the DMCNN model shows a similar or better performance compared to the CNN model. In most of the football footwork actions, the accuracy rates are comparable. However, for the "jump" action, the DMCNN model achieves an accuracy of 99.8%, which is higher than the 98.1% of the CNN model. This indicates that the DMCNN model may have better performance in capturing the characteristics of some complex actions.

When compared with the LSTM and GRU models (**Tables 5** and **6**), it is evident that both the CNN and DMCNN models have higher accuracy rates for all football footwork actions. For example, in the "False swing" action, both CNN and DMCNN achieve 100% accuracy, while LSTM has 92.5% and GRU has 93.8%. This shows that in the task of football footwork recognition, the convolutional neural network-based methods (CNN and DMCNN) have better accuracy than the recurrent neural network-based LSTM and GRU.

CNN and DMCNN can automatically learn and extract local and high-level features from the football footwork data through their convolutional and pooling layers. This automatic feature extraction ability enables the models to better adapt to different types of football footwork actions. In contrast, LSTM and GRU may not be able to capture spatial feature information as effectively as CNN and DMCNN when dealing with such data.

The DMCNN model further improves its performance by integrating multiple network structures. It can process acceleration data and angular velocity data simultaneously and obtain a more comprehensive and accurate feature representation by fusing the outputs of different networks. In comparison, the structures of LSTM and GRU are relatively simple and may have limitations when dealing with complex football footwork data.

Generally, CNN and DMCNN may have an advantage in computational efficiency. CNN can perform convolution operations in parallel through the sliding of the convolutional kernel, which significantly increases the computational speed. Although the DMCNN integrates multiple networks, it can also maintain a high computational efficiency while ensuring accuracy through reasonable design and parameter adjustment. In contrast, due to their recurrent structures, LSTM and GRU need to calculate one time step at a time when processing sequential data, resulting in relatively low computational efficiency.

In actual experiments, the computational efficiency can be evaluated by measuring the time required for each model to process the entire football footwork dataset. If the experimental results show that the computational times of CNN and DMCNN are significantly shorter than those of LSTM and GRU, it further proves their advantage in computational efficiency. This computational efficiency advantage makes CNN and DMCNN more suitable for real-time recognition and analysis of football footwork in practical applications.

5. Real-Time implementation of the proposed system

5.1. Computational requirements

In the real-time implementation of the proposed system, computational requirements play a crucial role. The system, especially when using complex neural network models like CNN and DMCNN, demands significant computational resources. Training these models requires high-performance computing devices with sufficient processing power and memory. For example, during the training process, a large number of matrix multiplications and convolutions are involved. These operations need to be executed quickly to ensure timely convergence of the model. If the computing resources are insufficient, the training time will be significantly prolonged, and in some cases, the model may not converge to an optimal solution.

In practical applications, such as real-time football footwork recognition during a game, the system needs to process incoming data in a short period. This requires a balance between model complexity and computational efficiency. While more complex models may offer better accuracy, they also increase the computational burden. Therefore, optimizing the model structure and parameters to reduce computational requirements without sacrificing accuracy is essential. This may involve techniques such as pruning neural network connections, using quantization methods to reduce the precision of data representation, and choosing appropriate batch sizes during training.

5.2. Potential latency issues

Latency is another critical aspect to consider in the real-time implementation of the system. Latency refers to the delay between the input of data (e.g., the occurrence of a football footwork action) and the output of the recognition result. In a fast-paced sports environment like football, even a small latency can have a significant impact on the effectiveness of the system.

There are several factors contributing to latency. Firstly, the data acquisition process from wearable devices may introduce some delay. The sensors need to collect and transmit the data, and any communication or processing bottlenecks in this stage can cause latency. Secondly, the computational complexity of the model itself can lead to processing delays. As mentioned earlier, complex neural network models require more time to process the data, resulting in a longer latency. Thirdly, the integration of the system with other components, such as display devices or decision-making systems, may also introduce additional latency.

To address latency issues, several strategies can be employed. Improving the data acquisition process by using faster sensors and more efficient communication protocols can reduce the initial delay. Optimizing the model to reduce its computational complexity, as discussed in the previous section, can also help minimize processing time. Additionally, careful design of the system architecture to ensure seamless integration between different components can mitigate integration-related latency.

5.3. Trade-offs and future directions

In considering the real-time implementation of the system, there are several trade-offs that need to be carefully balanced. One of the main trade-offs is between accuracy and computational efficiency/latency. As we have seen, increasing the complexity of the model can improve accuracy but at the cost of higher computational requirements and potentially greater latency. On the other hand, simplifying the model to reduce these factors may lead to a decrease in accuracy.

Future research directions should focus on finding better ways to optimize this trade-off. This may involve developing new algorithms or architectures that can achieve high accuracy with lower computational complexity. For example, exploring the use of lightweight neural network architectures specifically designed for real-time applications a promising direction. Additionally, advancements in hardware technology, such as the development of more powerful and energy-efficient processors, can also provide opportunities to improve the real-time performance of the system.

Another important aspect for future research is to further investigate and address the sources of latency. This may include studying the performance characteristics of different sensors and communication protocols in more detail, as well as exploring new techniques for reducing the computational latency of neural network models. By continuously improving the real-time implementation aspects of the system, we can ensure its wider applicability and effectiveness in practical applications such as football footwork recognition.

6. Conclusion

The research in this paper focuses on the footwork identification of football players. The intelligent device equipped with multiple sensors is introduced, along with its components and related functions. The preprocessing and segmentation method of football footwork data, the feature extraction and selection link, and the process of identifying five football footwork methods using KNN, SVM, CNN, and DMCNN are described in detail. The experimental results show that KNN has poor classification performance, while SVM and CNN perform well in football footwork recognition, achieving more than 97%. The proposed DMCNN model obtains the best football footwork recognition effect and improves the calculation efficiency and generalization ability of the model.

However, when considering the practical application of the algorithm, several challenges and limitations need to be addressed. Firstly, the performance of the algorithm may be affected by the complexity and variability of real-world football scenarios. Different playing surfaces, weather conditions, and player physical states can introduce uncertainties and errors in footwork recognition. Secondly, the algorithm's computational requirements may pose a challenge in real-time applications. As the volume of data and the complexity of the model increase, ensuring timely and accurate recognition while maintaining reasonable computational resources becomes crucial. Thirdly, the generalization ability of the algorithm across different football leagues and player populations needs to be further investigated. There may be differences in playing styles and footwork patterns among various regions and levels of play, which could impact the algorithm's performance. Future research could focus on addressing these challenges by exploring more robust feature extraction methods, optimizing the algorithm's computational efficiency, and conducting extensive tests across diverse football datasets to enhance its generalization ability.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Khan M A, Javed K, Saba T. Human Action Recognition using Fusion of Multiview and Deep Features: An Application to Video Surveillance. Multimedia Tools and Applications, 2020.
2. Mingkang Z, Xianling L, University J, et al. Human Action Recognition Algorithm Based on Bi-LSTM-Attention Model. Laser & Optoelectronics Progress, 2019, 56(15):1503-1512.
3. Knudson, D. Introduction to Biomechanics of Human Movement. In: Fundamentals of Biomechanics. Cham. 2021. https://doi.org/10.1007/978-3-030-51838-7_1
4. Tai WH, Zhang R, Zhao L. Cutting-Edge Research in Sports Biomechanics: From Basic Science to Applied Technology. Bioengineering, 2023, 10(6), 668. https://doi.org/10.3390/bioengineering10060668
5. Shun L, Ring G, Jianguo W. Human Action Recognition Method Based on Key Frame and Skeleton Information. Transducer and Microsystem Technologies, 2019:26-30.
6. Baveye Y, Chamaret C, Dellandrea E, et al. Affective Video Content Analysis: A Mul- tidisciplinary Insight. IEEE Transactions on Affective Computing, 2017:1.
7. Meng S, Wang T, Liu L. Monitoring Continuous State Violation in Datacenters: Exploring the Time Dimension. 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010), Long Beach, CA, USA, 2010, pp. 968-979, doi: 10.1109/ICDE.2010.5447923.
8. Parisi G I. Human Action Recognition and Assessment via Deep Neural Network Self-Organization. In: Noceti, N., Sciutti, A., Rea, F. (eds) Modelling Human Motion. Springer, Cham. https://doi.org/10.1007/978-3-030-46732-6_10
9. Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. Sensors, 22(4), 1476. https://doi.org/10.3390/s22041476
10. Wang, X., Yu, H., Kold, S., Rahbek, O., Bai, S. (2023). Wearable sensors for activity monitoring and motion control: A review. Biomimetic Intelligence and Robotics, 3(1), 100089, https://doi.org/10.1016/j.birob.2023.100089
11. Badii C, Bellini P, Difino A, et al. Smart City IoT Platform Respecting GDPR Privacy and Security Aspects. IEEE Access, 2020, (99):1.
12. Wangbin C, Shuoqing L, Shanshan T, et al. Design of Smart Home Fireproofing and Security System Based on Single-chip Microcomputer. Natural Ences, 2019, 32(1):39-42.
13. Kong L, Huang D, Qin J, et al. A Joint Framework for Athlete Tracking and Action Recognition in Sports Videos. IEEE Transactions on Circuits and Systems for Video Technology, 2019:11.

14. Gunnar, Johansson. Visual Perception of Biological Motion and A Model for Its Analysis. Attention, Perception & Psychophysics, 1973, 14:201-211.
15. Ke Y, Sukthankar R, Hebert M. Spatio-temporal Shape and Flow Correlation for Action Recognition. Ke, Y., Sukthankar, R., & Hebert, M. (2007). Spatio-temporal Shape and Flow Correlation for Action Recognition. 2007 IEEE Conference on Computer Vision and Pattern Recognition, 1-8.
16. Tran S D, Dams L S. Event modeling andrecognition using markov logic networks. In: Forsyth, D., Torr, P., Zisserman, A. (eds) Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol 5303. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-88688-4_45
17. Shimotashiro Y, Shinya M. Quantification in shooting precision for preferred and non-preferred foot in college soccer players using the 95% equal confidence ellipse. Front Sports Act Living. 2024;6:1434096. Published 2024 Sep 13. doi:10.3389/fspor.2024.1434096
18. Bertozzi F, Rabello R, Zago M, Esposito F, Sforza C. Foot dominance and ball approach angle affect whole-body instep kick kinematics in soccer players. Sports Biomech. Published online August 22, 2022. doi:10.1080/14763141.2022.2110514
19. Laptev, Lindeberg. Space-time Interest Points. Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 2003, pp. 432-439 vol.1, doi: 10.1109/ICCV.2003.1238378
20. Wang J, Domingos P M. Hybrid Markov Logic Networks. Machine Learning, 2008, 62(1-2):1106-1111.
21. Wan S, Qi L, Xu X, et al. Deep Learning Models for Real-time Human Activity Recognition with Smartphones. Mobile Networks and Applications, 2020, 25(2):743-755.
22. Mutegeki R, Han D S. A CNN-LSTM Approach to Human Activity Recognition[C]// ICAIIC.2020:362-366.
23. Hamilton W L, Ying R, Leskovec J. Inductive Representation Learning on Large Graphs. Neural Information Processing Systems, 2017:1024-1034.
24. Wang J, Yang Y, Mao J, et al. CNN-RNN: A Unified Framework for Multi-label Image Classi-fication. 2016:2285-2294.
25. Gers F A, Schmh Huber J, Cummins F. Learning to Forget: Continual Prediction with LSTM. Neural Computation, 2000, 12(10):2451-2471.
26. Zhaowei Q, Yuan W, Xiaoru W. Transfer learning based hierarchical attention neuralnet-work for sentiment analysis. Journal of Computer Applications, 2018:3054-3062.
27. Belardinelli A, Pirri F. A biologically plausible robot attention model, based on space andtime. Cognitive Processing, 2006, 7(1):11-14.