Eng Int Syst (2025) 6: 721–732

Engineering
Intelligent Systems

Badminton Action Recognition Model Combining Artificial Neural Network Algorithm and DTW Algorithm

Yuanji Zhong[1], Jiangwei Yang[2]* and Wenhao Guo[1]

[1] *Department of Social Sports Guidance and Management, School of Physical Education and Art, Jiangxi University of Science and Technology, Ganzhou, 341000, China*
[2] *Department of Physical Education, Xiamen Institute of Technology, Xiamen, 361021, China*

Badminton is a sport that is susceptible to the influence of the racket and the player's attire, and it is difficult to recognize the action without the help of 3D-capturing technology. To accurately recognize the technical movements involved in badminton, this study proposes a badminton action recognition model that combines an artificial neural network algorithm and a dynamic time warping algorithm. The badminton player's posture can be estimated using the improved OpenPose algorithm, which can accurately capture the player's key skeletal nodes. Based on this, the mobile network V3 network architecture and dynamic time warping algorithm are used to recognize the action. The results revealed that the mobile network V3 model achieved significant performance in recognizing badminton movements, with an accuracy of 0.987, which was significantly higher than the results obtained with the visual geometric group network, the residual network and the mobile network V2 model. Moreover, the total number of parameters was significantly decreased by 4.7M. In addition, the precision of the model was improved by an average of 12.27%, outperforming the other three network models. The dynamic time warping algorithm also performed well in evaluating badminton technical movements. The results of the evaluation were significantly improved by introducing the weighting values. In the smash and split score, the score based on dynamic time warping algorithm was 84, while the score after introducing the weights reached 86, which was closer to the scoring value of the domain experts. The results demonstrated that the mobile network V3 model with the dynamic time warping algorithm used in the study was able to achieve accurate recognition of badminton movements with high computational efficiency. The study provides an efficient badminton action recognition and evaluation method, which helps to improve the scientific basis and effectiveness of athletes' training, and also provides a new technical means for sports analysis.

Keywords: Badminton action recognition; MobileNetV3; Improved OpenPose algorithm; DTW algorithm; Joint points

1. INTRODUCTION

Badminton, as a globally popular sport, not only has very high requirements for athletes' physical fitness, but also involves complex technical movements and tactical strategies[1–2]. With the development of sports science, the precise analysis of athletes' movements and the 'scientification' of training methods have become particularly important. In badminton training and competition, accurate recognition and evaluation of athletes' technical movements are important for improving training efficiency, optimizing technical details, and developing game strategies[3–4]. However, the traditional action recognition method relies on manual observation and video playback, which is both inefficient and difficult to quantified accurately[5–6]. Several scholars have undertaken research on the recognition of badminton actions. Yang et al. suggested an hourglass network model that includes a self-attention

*Corresponding author: Jiangwei Yang; E-mail: dhtw09@163.com

and convolution mechanism to detect the standardized hitting action used in badminton. The results indicated that the method achieved an average precision of 96.1% on badminton player's stance detection and 95.4% accuracy in classifying the other six hitting actions[7]. In an effort to improve the effectiveness of badminton players' training, Song et al. presented a temporal segmentation network-based badminton video action identification technique. The study predicted four key actions used by the player: forehand, backhand, top shot and pick. The results showed that the method had 91.2% recall and 91.6% accuracy[8]. Shi et al. developed a badminton swing system based on a single inertial sensor to improve the precision of recognizing the swing movements of badminton players. The study also used a two-layer classifier recognition model that included decision tree and random forest techniques. The model demonstrated potential for application in recognizing swing motions in badminton[9]. Chen proposed a badminton court action detection model based on multi-modal action recognition toolbox 2 (MMAction2) as well as Slow-Fast in order to reduce the risks caused by accidental collisions between badminton players. The findings revealed that the system was highly accurate in identifying a variety of badminton movements, including moving around the court and falling, in addition to the player's motions[10]. The above studies have achieved better results in detecting and analyzing badminton actions, but have not considered the temporal nature of the movements. Therefore, this current study proposes a badminton action recognition model based on the artificial neural network (ANN) algorithm and the dynamic time warping (DTW) algorithm. The innovation of the research is that it achieves the accurate recognition of badminton movements by improving the OpemPose algorithm, MobileNetV3 network architecture, and DTW algorithm. The aim of the study is to propose a means of accurately recognizing badminton movements while also evaluating the skill level of athletes' movements. The research can provide data support for the scientific training of badminton athletes and the prevention sports injuries.

2. METHODS AND MATERIALS

2.1 Badminton Player Pose Estimation Based on Improved OpenPose

Badminton is not only a sports activity; it also has important value and impact in the field of social sciences. First of all, badminton can improve the cognitive ability and thinking sensitivity of participants. It requires athletes to react quickly and make rapid decisions at high speeds, and the agile thinking acquired through training has a positive impact on other areas of daily life and learning. Second, because badminton is a two-player sport, it promotes communication, social interaction and cooperation skills. This is essential for improving interpersonal skills and teamwork. In addition, badminton has a positive impact on mental health by reducing stress, increasing self-confidence and strengthening resilience. In terms of social stratification, badminton is also regarded as an activity that can transcend social barriers and encourage communication among people of different backgrounds. It is not only a competitive sport, but also a platform for social and cultural exchange, which helps to build harmonious social relationships. As such, badminton has a place in the social sciences. It has an indispensable role to play in the psychological health of individuals, social interaction ability and the harmonious development of social structure.

Before recognizing the movements of a badminton player, the first step should be to obtain the key nodes of the individual's posture based on the badminton player's image[11–12]. This current study used the improved OpemPose algorithm for key node extraction for sportspersons. The input data of this algorithm is broader and compatible with various sources of different, pre-recorded as well as real-time images, while the output comprises the key skeletal nodes of the sportsperson. Combined with the body structure of the athlete, a skeleton map of the athlete can be created to determine the athlete's posture[13–14]. The main steps of the algorithm are described below. Firstly, the image or image data is input. Secondly, the partial affinity field as well as the key nodes are predicted. Subsequently, the relationship graph of limb connections is obtained by combining the partial affinity fields based on the confidence graph output from the neural network. Finally, the limb connection relationship graph is parsed based on greedy algorithm[15–16]. The network structure based on OpenPose is shown in Fig. 1.

As shown in Fig. 1, the network structure of OpenPose consists of two parts. The first part is based on the VGG19 architecture, whose main role is to extract features from the input image and image data to obtain the feature graph F. In the second part, the feature graph F is used to obtain the key nodes of the head, arms, and legs of the athlete, as well as the limbs and other parts that define the key nodes of the skeleton[17–18]. The second part consists of two parallel branches. Here, the confidence graph of the motors is generated in the first branch by assigning confidence to each key node to characterize the specific position of that node in the input image, while part of the affinity fields are generated in the second branch[19–20]. As the iterative process continues, the neural network is able to achieve more accurate predictions for key points as well as partial affinity fields. In the OpenPose neural network, the mathematical expression of skeletal joint point confidence is shown in Equation (1).

$$S^t = p^t(F, S^{t-1}, L^{t-1}),\ \forall t \geq 2 \tag{1}$$

where, S^t denotes confidence level. t denotes the time phase. p^t denotes reasoning from t time period. F denotes feature. L denotes the affinity domain. Its mathematical expression is shown in Equation (2).

$$L^t = \varphi^t(F, S^{t-1}, L^{t-1}),\ \forall t \geq 2 \tag{2}$$

where, φ represents the inference data for a certain time period. The size of the affinity domain value indicates the pixel fluency of the target subject's limbs. The confidence level of a joint point indicates the likelihood that a single joint point is in that position. To measure the difference between the true and predicted values, the L2 loss function (LF) is appended at the neural network's conclusion[21–22]. For badminton, the study introduces a null-weighted loss to show this discrepancy. The LF of the first branch at this point is shown in Equation (3).

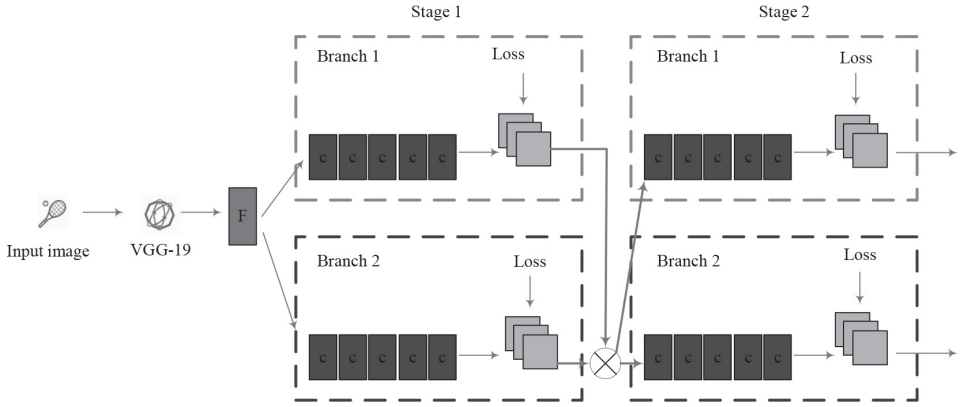


Figure 1 OpenPose network structure diagram.

Table 1 Badminton players' close nodes.

Serial number of key bone points	Body part	Serial number of key bone points	Body part
1	Nose	2	Neck
3	Right shoulder	4	Right elbow
5	Right wrist	6	Left shoulder
7	Left elbow	8	Left wrist
9	Right hip	10	Right knee
11	Right ankle	12	Left hip
13	Left knee	14	Left ankle
15	Oculus dexter	16	Oculus sinister
17	Right eat	18	Left eat

$$f_s^t = \sum_{j=1}^{J} \sum_{p} W(p) \cdot \|S_j^t(p) - S_j^*(p)\|_2^2 \tag{3}$$

where p denotes pixel. $W(p)$ denotes binary code. $S_j^*(p)$ denotes the true confidence level. The mathematical expression of the LF based on affinity domain branching is shown in Equation (4).

$$f_L^t = \sum_{c=1}^{C} \sum_{p} W(p) \cdot \|L_c^t(p) - L_c^*(p)\|_2^2 \tag{4}$$

where, $L_c^*(p)$ denotes the real affinity domain field. Based on the above two LFs, the study introduces a global objective function to optimize the gradient loss. Equation (5) provides a mathematical expression for it.

$$f = \sum_{t=1}^{T} (f_s^t + f_L^t) \tag{5}$$

where, f denotes the global objective function. For some specific affinity domain fields, the nodes of the same parts are spliced into the skeletal points of multiple people by greedy algorithm. To detect the joint points of badminton players, firstly, the study decomposes their body parts. The serial numbers for the joint points are shown in Table 1.

In Table 1, combined with the motion characteristics of badminton activities, the relevant feature points of the head are eliminated. After obtaining the RGB image sequence based on OpenCV, it is imported into OpenPose-bm and saved as text data, which is used to obtain the horizontal and vertical coordinates (HVCs) of individual nodes and then establish the corresponding two-dimensional coordinate axes. The mathematical expression for a single node is shown in Equation (6).

$$\begin{aligned} X^n(n, p) := (x^u(n, p), y^u(n, p) : x^u(n, p) \\ \in [0,\ W], y^u(n, p) \in [0,\ H]) \end{aligned} \tag{6}$$

where, n denotes the frame rate. p denotes the skeletal point. Since data may be lost during acquisition, the difference between the front and back frames are adjusted. The HVCs of key nodes are calculated with Equation (7).

$$\begin{cases} x_s = x_t + d[(x_t - x_l)/r] \\ y_s = y_t + d[(y_t - y_l)/r] \end{cases} \tag{7}$$

where, x_s and y_s denote the real-time HVCs, respectively. x_t and y_t denote the HVCs of the previous frame of a single joint point. x_l and y_l denote the HVCs of the frame after a single joint point. For the extraction of relevant data, the study selected the coordinate normalization method. The mathematical expression of coordinates is shown in Equation (8).

$$\begin{cases} x_s = x_v + a \\ y_s = y_v + b \end{cases} \tag{8}$$

where, a and b denote the translation distance based on the X and Y-axis.

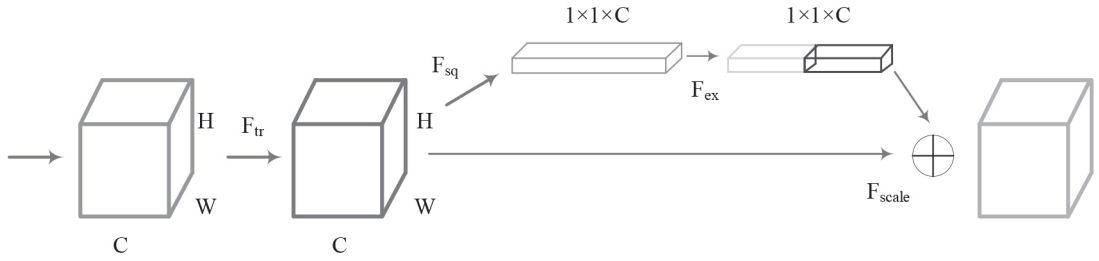


Figure 2 Schematic diagram based on SE.

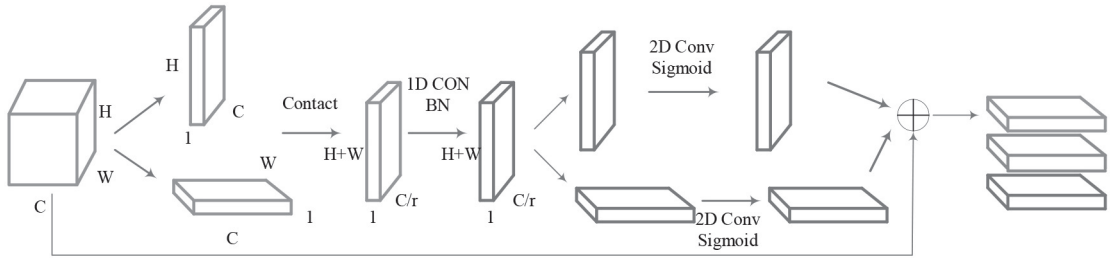


Figure 3 Diagram of CA attention mechanism module.

2.2 Badminton Action Recognition Based on MobileNetV3 Network Architecture

Based on the skeletal point information of badminton players obtained above, a neural network is used to classify the movements. ANN is an information processing system with many nodes that replicates the composition and operation of neural networks in the human brain. Connection weights are used to link the nodes to one another, and by varying the weights, the nodes can discover and identify patterns. There are many parameters and resource limitations in the badminton action recognition process, and the conventional ANN is constrained by the network topology and parameter count. Therefore, combined with the real-time and efficiency requirements of badminton action recognition, a lightweight convolutional neural network (CNN) model, MobileNet, is introduced in this study. It can effectively optimize the number of parameters and the degree of computing difficulty using point-by-point and deep convolution. In addition, MobileNetV2 introduces the concept of residual connection, which improves the performance of the network as a result of dimensionality reduction[23–24]. The study proposes the MobileNetV3 model based on this, which takes into account the concepts of residual connectivity of V2 as well as separation convolution of V1. In addition, the model coordinates the squeeze-and-excitation networks (SENet) module; i.e., the weights are optimized for individual feature maps (FMs)[25–26]. The schematic based on the SE module is shown in Fig. 2.

Since SE considers the attention only on the channel dimension, the study introduces the coordinate attention (CA) optimization model to address the problem that the pooling method used in the attention mechanism in the original MoboleNetV3 model has more difficulty acquiring the more complex overall data. The module based on the CA attention mechanism is shown in Fig. 3.

The first step of this attention model is to import the FM and pool this FM in X-direction as well as Y-direction. The mathematical expression of the FM based on both the X-direction and Y-direction is shown in Equation (9).

$$\begin{cases} z_e^h(h) = \frac{1}{2}\sum_{0\le t\le 2} x_e(h,i) \\ z_e^w(w) = \frac{1}{H}\sum_{0\le j\le H} x_e(j,w) \end{cases} \tag{9}$$

where, $z_e^h(h)$ and $z_e^w(w)$ denote the normalized FMs in the horizontal and vertical directions. Subsequently, the horizontal and vertical FMs are combined and then dimensionalized based on a 1×1 convolution kernel. The FMs are obtained by activation function after normalization operation. The related mathematical expression is shown in Equation (10).

$$f^t = \sigma(F_1([Z^h, Z^w])) \tag{10}$$

where, f^t denotes the transition map obtained after encoding the horizontal and vertical direction FMs, and continue to take convolutional action on the transition map. Therefore, the mathematical expression of the attention weights in the horizontal as well as vertical directions is Equation (11).

$$\begin{cases} g^h = \sigma(F_h(f^h)) \\ g^w = \sigma(F_w(f^w)) \end{cases} \tag{11}$$

where, g^h and g^w denote the transformation function. Finally, the obtained FM as well as the weight values are updated and corrected. That is, the residual form is used to connect the transformation function as well as the initial FM[27–28]. The mathematical expression of the CA-based FM is shown in Equation (12).

$$y^c = x^c \times g^h \times g^w \tag{12}$$

where, x^c and y^c represent the feature input and output maps, respectively. To make MobileV3 better able to accomplish the recognition of badminton movements, the study modifies the parameters on the original MobileV3 model. Table 2 lists the pertinent model parameters.

Table 2 Network structure parameters.

Input	Control	Attention mechanism	Non-linear activation function	Number of channels
224 × 224 × 3	Conv2D	/	HS	16
112 × 112 × 16	Bneck	CA, SE	RE	16
56 × 56 × 16	Bneck	CA, SE	RE	24
28 × 28 × 24	Bneck	CA, SE	RE	40
14 × 14 × 40	Bneck	CA, SE	HS	48
14 × 14 × 48	Bneck	CA, SE	HS	96
14 × 14 × 96	Bneck	CA, SE	HS	96
14 × 14 × 96	Bneck	CA, SE	HS	96
7 × 7 × 96	Bneck	CA, SE	HS	96
7 × 7 × 96	Conv2D	/	HS	244
7 × 7 × 244	Pool	/	/	/
1 × 1 × 244	NBN		/	640
1 × 1 × 640	NBN	/	/	5

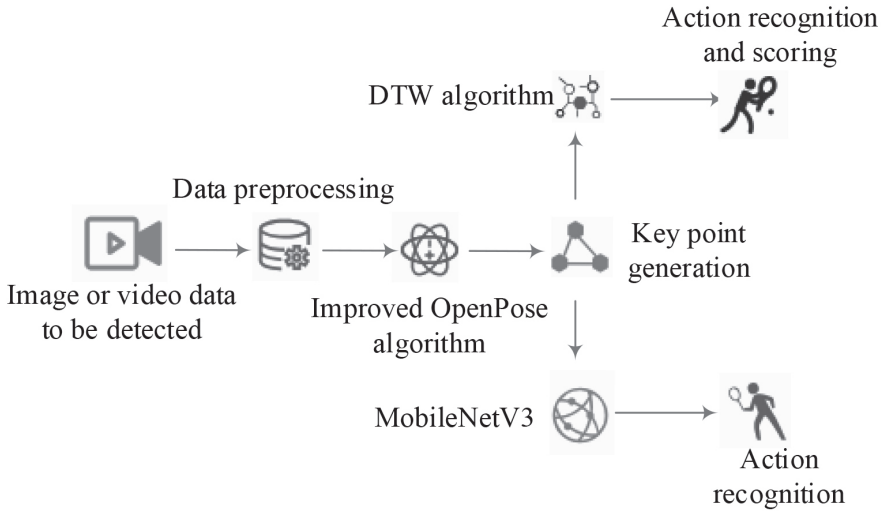


Figure 4 Badminton action recognition based on DTW algorithm.

2.3 Badminton Technical Movement Evaluation Based on DTW Algorithm

After recognizing badminton movements using the MobileNet network, the study applied the DTW algorithm to recognize and score badminton movements taking the time series into account. Given that badminton players have different levels of physical fitness, the input is scored for sequences that match the template movements. Therefore, the study proposes the DTW algorithm, which can characterize the similarity of individual valid badminton moves within a cycle through path information, and then evaluate the moves as a whole. The flow of badminton technical movement evaluation based on the DTW algorithm is shown in Fig. 4.

In Fig. 4, after the estimation of posture by OpenPose algorithm, the badminton action can be recognized by using the DTW algorithm in combination with the time sequence information. The time regularization function $O(n)$ is defined to represent the echo relationship between two sequences in the time axis. The action sequence of a professional badminton player is $D(D_1, D_2, D_3, \ldots, D_n)$. The action sequence of a non-professional sportsman is $Q(Q_1, Q_2, Q_3, \ldots, Q_m)$. The related regularization path planning is shown in Fig. 5.

In Fig. 5, the mathematical expression for the complete planning path and its elements are shown in Equation (13).

$$\begin{cases} G = \{g_1, g_2, g_3, \ldots, g_k\} \\ \max(n, m) \le k \le n + m \end{cases} \tag{13}$$

where, G denotes the regularized path. k denotes the element in the path. n and m denote the number of sequences. By comparing the size of the Euclidean distance to determine the similarity, a small value indicates high similarity. Therefore, it is sufficient to choose the path with the shortest length under the guaranteed constraints. The related mathematical expression is shown in Equation (14).

$$DTW(E, R) = \min\left\{\sqrt{\sum_{k-1}^{k} G_k/\lambda}\right\} \tag{14}$$

where, λ is mainly adjusted for paths with different lengths. Following the dynamic modification, Equation (15) is the recursive formula.

$$\gamma(i, j) = d(E_i, P_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{15}$$

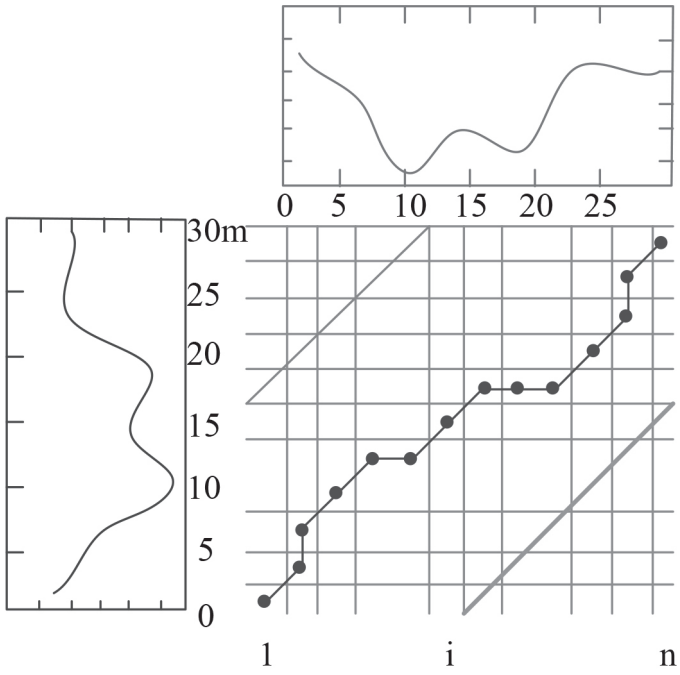


Figure 5 Regular path diagram.

where, γ denotes similarity. $d(E_i, P_j)$ denotes distance. Finally, the study introduces the concept of weight value to optimize the DTW algorithm for automatic classification by Mean Shift, while assigning weights to the number of categories. The mathematical expression based on weight values is shown in Equation (16).

$$\omega_e = \frac{h_c}{\sum_{z=0}^{d} h_s} \tag{16}$$

where, ω_c denotes the weight value of the cth variable. h_c denotes the number of categories of a single node after clustering. d denotes the number of joints. After combining the data of joints from different frames into sequences, the two sets of different sequences are merged and downscaled into one-dimensional coordinates. The integer is divided after taking the 2nd order paradigm for it. Lastly, the final badminton action score is obtained by combining the distance obtained by the DTW algorithm and converting it into a percentage.

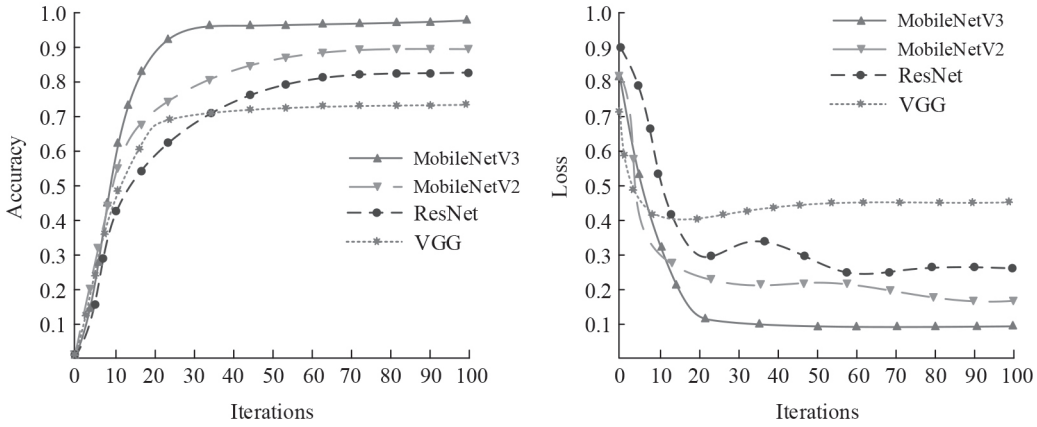
3. RESULTS

3.1 Performance Analysis Based on MobileNetV3 Network Architecture

The experimental environment for the study is described below. The learning framework is TensorFlow 2.0 developed by Google Inc. The graphics card is selected as GTX 1060 Ti, the gas pedal program is selected as CUDA 10.2, and Python 3.8 is selected for programming. The CPU is AMD Ryzen 7 5800HS and the running memory is 16GB. A homemade dataset is used and the ratio of training set to test set is set to 7:3. Stochastic gradient descent (SGD) optimization is used in the training process of the model. Meanwhile, the initial learning rate is set to 0.01 and the momentum is set to 0.8. The LF is selected as cross-entropy LF, the batch size is set to 64, and the Batch_size is set to 20. The classical CNN models VGG, ResNet, and MobileNetV2 were chosen for the comparison experiment. Fig. 6 displays the accuracy variation as LF curves according to the training set.

The comparison of the four distinct CNN accuracy curves is displayed in Fig. 6(a). The accuracy of all four CNNs increases with the number of iterations, and the MobileNetV3 network proposed in this study converges at 30 iterations. The accuracy converges to 0.985, which is significantly higher than the remaining three network models. The one with slightly lower accuracy than the MobileV3 model is MobileNetV2. It converges at about 70 iterations, and the final accuracy converges to 0.899, followed by ResNet and VGG after the accuracy convergence. Fig. 6(b) shows a comparison of the loss curves of four different CNNs. The results show that the loss curve of the MobileNetV3 network used in the study gradually decreases while there is no tendency to rebound. The MobileNetV3 model gradually converges to 0.998 after 20 iterations. As the number of iterations increases, the remaining three CNNs exhibit a certain amount of deterioration. Nonetheless, they all exhibit a rebound phenomenon. Also, the precision as well as recall curves of the four neural networks are compared. The results are shown in Fig. 7.

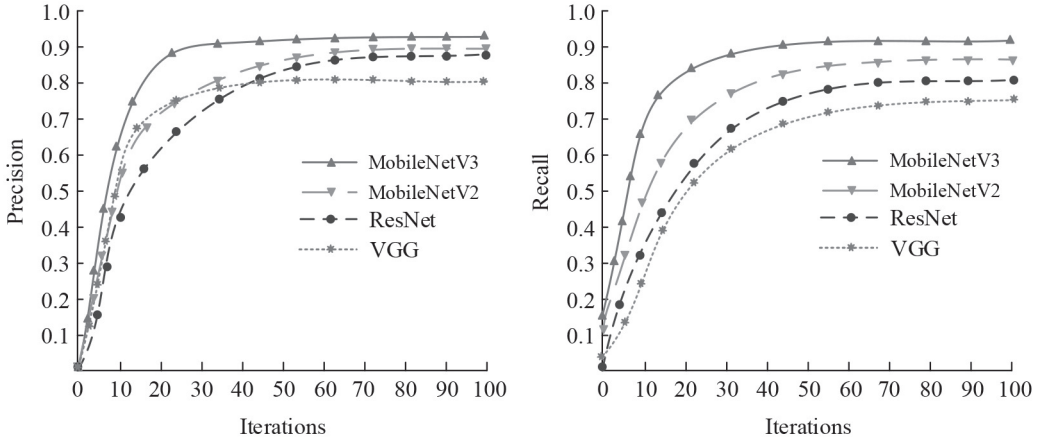
Fig. 7(a) allows comparison of the precision curves of the four different CNNs. The precision values of the four CNNs gradually increase as the iterations increases. The precision value of the MobileNetV3 network finally converges to around 0.921, which is significantly higher than those of the other three neural networks. The MobileNetV2 network precision value converges to 0.897, and the VGG network converges to the lowest precision value of 0.807. The recall comparison of four distinct CNNs is displayed in Fig. 7(b). The model's capacity to recognize positive samples is gauged by its recall.



(a) Comparison of accuracy curves of different convolutional neural networks

(b) Comparison of loss curves of different convolutional neural networks

Figure 6 Comparison of accuracy and loss curves of different convolutional neural networks.



(a) Comparison of precision curves of different convolutional neural networks

(b) Comparison of recall curves of different convolutional neural networks

Figure 7 Comparison of accuracy and recall rate curves of different convolutional networks.

Table 3 Comparison of performance indicators of different convolutional networks.

Performance index	VGG	ResNet	MobileNetV2	MobileNetV3
Accuracy	0.712	0.802	0.902	0.987
Precision	0.786	0.874	0.903	0.923
Recall	0.721	0.798	0.854	0.931
F1-score	0.741	0.821	0.898	0.951
Mean forecast time/ms	0.034	0.025	0.021	0.018

The results show that the recall of all four network models gradually increases as the number of iterations increases. Among them, the recall of the MobileNetV3 network model finally converges to 0.926, which is the highest value among all CNNs. This is followed by MobileNetV2, ResNet, and VGG network models. Also, the performance of each network model in relation to the test set is determined. Table 3 displays the performance metrics of the four neural networks according to the test set.

All of the performance metrics values of the MobileNetV3 model are the highest, as shown by the results presented in Table 3. Its accuracy, precision, recall, and F1-score are 0.987, 0.923, 0.931, and 0.951, respectively. The performance metrics have been improved by an average of 0.182, 0.069, 0.140, and 0.131. Meanwhile, the average recognition time of the MobileNetV3 network model is 0.018 ms, which is reduced by an average of 0.009 ms compared to the rest of the network models. Further comparisons are made between the parameters of the network models and the performance of the enhanced OpenPose algorithm applied in the study. Fig. 8 displays the findings.

Fig. 8(a) enables a comparison of the parameters and operations of the four neural networks. The results show that the total parameters of the MobileNetV3 model is 4.7M, and its value is significantly lower than that of the VGG model, which is 26.1M. Furthermore, its average precision is improved by an average of 12.27% compared with the remaining three network models. Fig. 8(b) shows the parameter comparison of the four algorithms. The outcomes indicate that the average precision of the improved OpenPose algorithm proposed in the

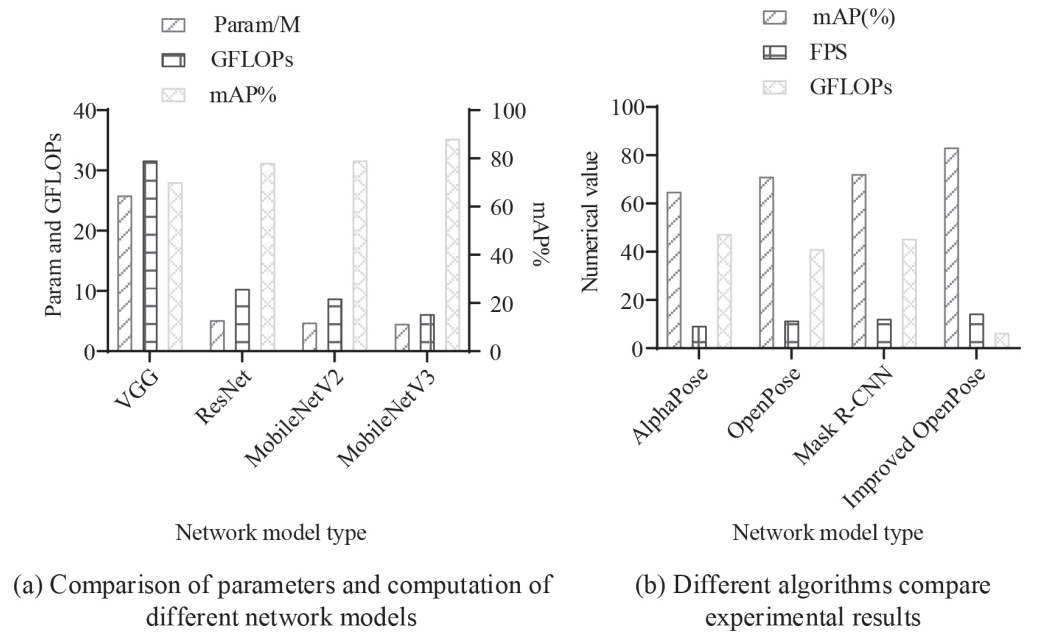


Figure 8 Comparison of network model parameters and experimental results.

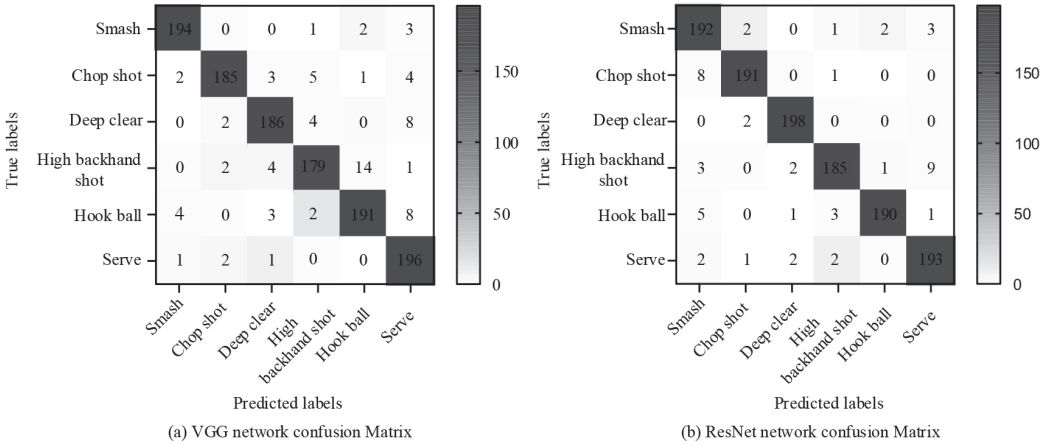


Figure 9 VGG network and ResNet network CM.

study is 83.6%, which is an average improvement of 13.8% compared to the remaining three algorithms. The improved OpenPose algorithm has a frame rate of 14.8, which is 3.4% higher than those of the remaining algorithms. Meanwhile, the floating-point operations' quantity of the model is only 6.8, which is significantly lower than that of the other three algorithms. This suggests that the technique can significantly reduce the model's parameter count. In conclusion, the enhanced OpenPose algorithm performs better and has a more straightforward network architecture. It can be concluded that the MobileNetV3 model is superior in terms of all performance metrics.

3.2 Badminton Action Recognition Based on MobileNetV3 Modeling

A homemade dataset is used, which consists of six common badminton maneuvers, i.e., smash, chop shot, deep clear, high backhand shot, hook ball, and serve a ball. In this case, there are 200 videos of each maneuver type and 100 network trainings are performed. The effect of neural network model on badminton action recognition is evaluated based on the four network models described above. First, the confusion matrix (CM) diagram based on VGG and ResNet network models is shown in Fig. 9.

Fig. 9(a) shows the CM of VGG neural network based on six badminton maneuvers. The results show that the VGG network is the worst at predicting the HIGH backhand shot, with a prediction precision of only 89.5%. The highest prediction precision is for the serve of a ball, with a value of 93%. The possible reason for this is that serving a ball is the beginning stage of badminton, when the joint features are more obvious. Fig. 9(b) shows the CM of ResNet network. According to the results, the ResNet network has the highest prediction precision for deep clear with a value of 99%. The worst prediction precision is for high backhand shot with a value of 92.5%. The confusion matrices of the MobileNetV2 network model and the MobileNetV3 network model are then compared. The results are shown in Fig. 10.

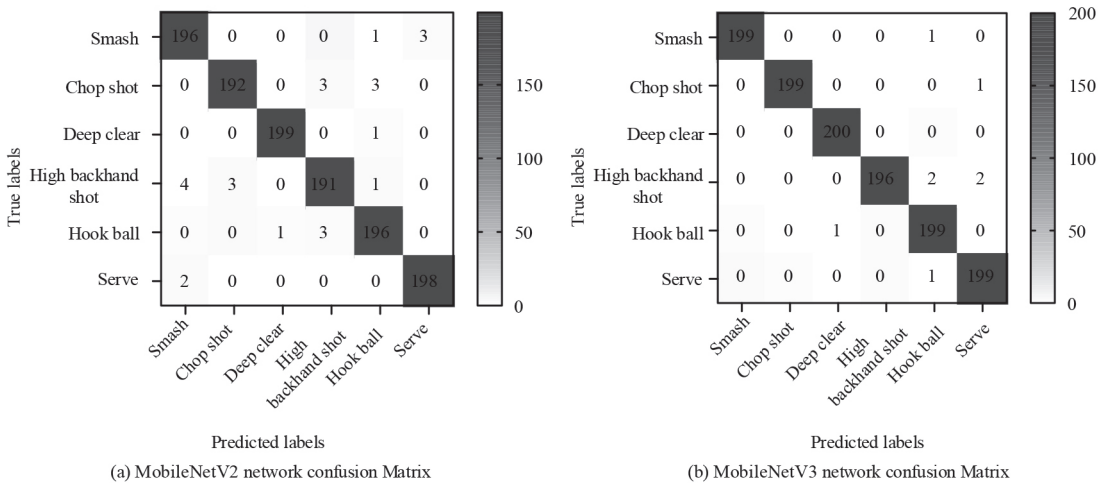


Figure 10 MobileNetV2 network and MobileNetV3 network CM.

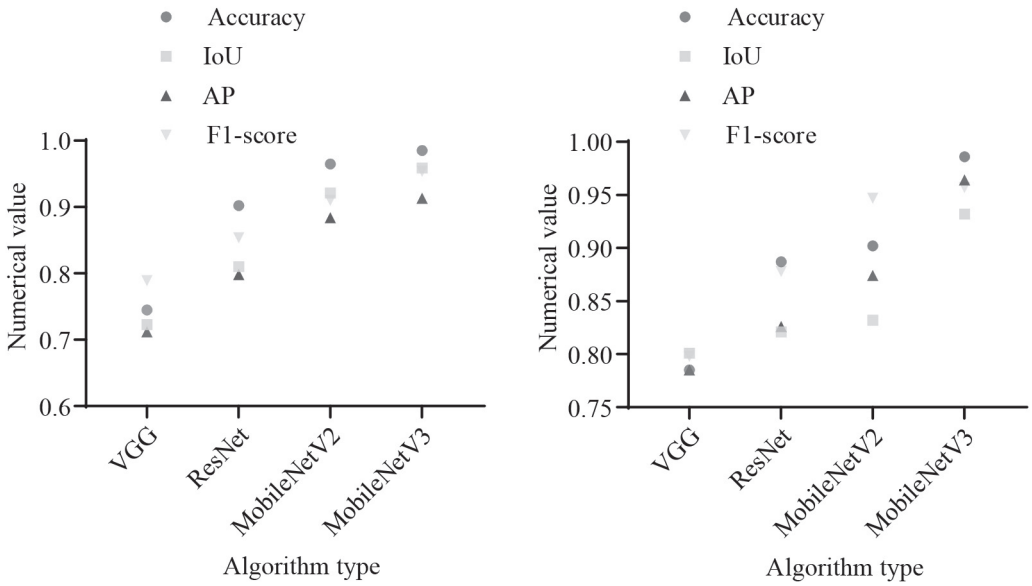


Figure 11 Comparison of recognition performance of different algorithms for smash and split.

Fig. 10(a) shows the CM of MobileNetV2 neural network for six different badminton moves. The results show that the MobileNetV2 neural network recognizes more than 190 correct numbers for each action. That is, the recognition precisions are all greater than 95%. Among them, the recognition precision for deep clear action is the highest, with a value of 99.5%. The lowest precision is for the high backhand shot with a value of 95.5%. Secondly, the MobileNetV2 neural network also has a high recognition precision value for serve action with a value of 99%. Fig. 10(b) shows the CM of MobileNetV3 neural network. The results show that the recognition precision of this neural network is greater than 97.5% for all six badminton moves. Among them, the recognition precision for deep clear is the highest, with a value of 100%. The recognition precision for high backhand shot is the lowest with a value of 98%. In conclusion, the MobileNetV3 has the highest recognition precision for the six neural networks. To address the problem that smash and chop shot are more closely similar, the study introduces intersection over union (loU) to evaluate the detection ability of the neural network model for the above two kinds of actions as well as the recognition effect. The results are shown in Fig. 11.

Fig. 11(a) enablesl the comparison of recognition performance metrics based on smash and split. The outcome displays that the recognition accuracy of the MobileNetV3 network model for smash and split is 0.985, which is improved by 0.114 on average compared with the rest of the network models. The IoU reflects the degree of overlap between the predicted and real bounding box. The MobileNetV3 network model has a loU of 0.959, which is improved by 0.141 on average compared to the rest of the network models. In addition, the AP and F1-score values of this network model are also in the highest numerical hierarchy, which are improved by 0.143 and 0.162 on average. The recognition

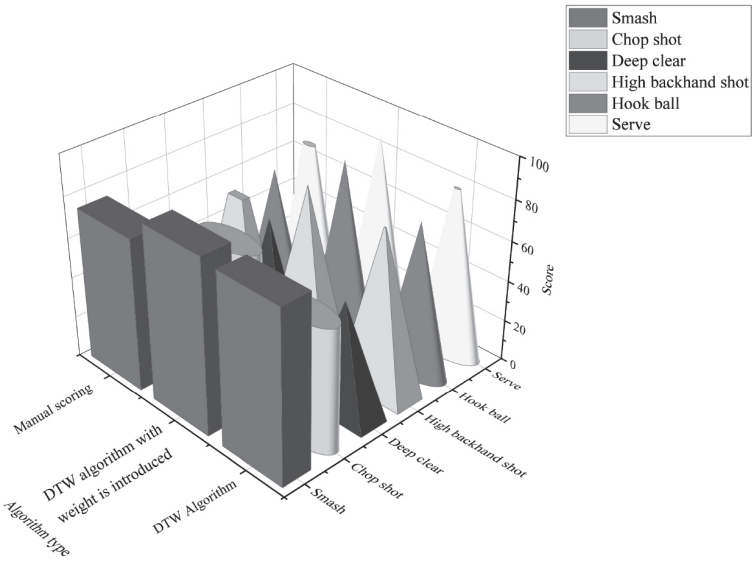


Figure 12 Comparison of scores of different badminton movements based on DTW algorithm.

Table 4 Final forehand action evaluation result based on DTW algorithm.

Action number	1	2	3	4
Action name	Preparatory phase	Beating phase	Batting phase	Random phase
Error key	16	38	47	36
Under 60 points	15	34	44	37
Over 60 points	16	35	41	32
Accuracy	100%	92.10%	87.20%	88.89%

performance metrics of several network models based on chop shot are compared in Fig. 11(b). The results show that every performance metric of the MobileNetV3 network model is at its maximum value. It has an average improvement of 0.128, 0.114, 0.129, and 0.142 compared with the remaining three network models. In conclusion, the MobileNetV3 network model is more effective for the more similar action recognition.

3.3 Badminton Movement Evaluation Based on DTW Algorithm

Firstly, the scores are calculated for the six specific badminton maneuvers mentioned above. The score values of different badminton maneuvers based on the DTW algorithm are shown in Fig. 12.

Fig. 12 shows the comparison of score values of different badminton maneuvers. Of the six badminton maneuvers, smash and split has the highest score value. The DTW algorithm, weight-based DTW algorithm, and manual scoring are 84, 86, and 75, respectively. The score values are all improved to some extent after the introduction of weighting values. In deep clear, the weight-based DTW algorithm improves the score value by 20 points compared to the original DTW algorithm. In chop shot ball, the score value is improved by 13 points. The results show that the introduction of the weights increases the attention to the joints corresponding to these movements, and the score values are also closer to those of the domain experts. The final study takes the forehand stroke deep clear in badminton as an example, and decomposes the related technical movements into four phases: preparatory, beating, batting, and random. Table 4 displays the findings for the final evaluation of forehand movement.

In Table 4, the overall accuracy of badminton movements assessed by the DTW algorithm with the introduction of weights are above 87%. In particular, the accuracy of the preparatory phase and the beating phase are 100% and 92.10%, respectively. The possible reason for this is that these two phases are characterized by simpler movements and less occlusion of the limbs. The preparatory phase has the lowest number of erroneous key points, and its value is significantly lower than the rest of the phases. In the batting phase, the number of erroneous key points and the number of scores below 60 and above 60 are at the maximum value level. This indicates that the features of this phase are more complex and at the same time the accuracy value is the smallest. The results show that in the swing and hitting phases, the greater the similarity of movements, the lower is the accuracy. However, the preparatory phase and the beating phase, due to the simplicity of the movements and the low difficulty coefficients, have fewer key points of error and higher accuracy.

4. DISCUSSION

In the experimental results, the accuracy of MobileNetV3 model reached 0.987, which was significantly higher than that of VGG, ResNet and MobileNetV2. Meanwhile, the average recognition time of MobileNetV3 was 0.018ms, which improved the performance by 0.009ms compared with

the other models, highlighting its computational efficiency. In addition, MobileNetV3 converged to 0.931 in recall and 0.951 in F1-score. In the action evaluation based on the DTW algorithm, the experimental results showed that the DTW algorithm was able to efficiently evaluate the similarity between badminton technical actions. The evaluation results of the DTW algorithm were also significantly improved by introducing weight values. In the smash and split score, the score based on the DTW algorithm was 84, whereas the score increased to 86 with the introduction of weights. This indicated that the introduction of weights increased the focus on the key nodes and thus improved the accuracy of the evaluation. In comparison, the manual score was 75, indicating that the research algorithm was able to better approximate the expert score and was more informative. In the accuracy analysis of the different actions, the accuracy in the preparatory phase reached 100%, while in the batting phase and random phase it was 87.20% and 88.89%, respectively. This indicated the effect of the complexity of movement on the recognition accuracy. The results for the batting are less accurate because this phase involved more movement details and rapid movement changes. Meanwhile, the comparative phase had relatively simple movement features and less limb occlusion, resulting in the lowest number of error critical points. To address this, Zheng et al. proposed a neural network based on robust video keying as well as improved I3D in an attempt to solve the problem of inaccurate recognition of badminton action time series by 2D network. The study introduced the DTW algorithm to explore different time sequences in order to recognize and score badminton action. The results indicated that the accuracy of badminton action recognition reached 91.5%. The results were basically consistent with this study on the recognition of badminton action by the DTW algorithm. Meanwhile, Zheng et al demonstrated the effectiveness of the I3D network in badminton action classification[29]. Sarwar et al. proposed a skeleton-based keyframe detection model for analyzing the movements of badminton players. Similar to the present study, Sarwar et al. also used a modified OpenPose algorithm to estimate the athlete's pose. The results indicated that the number of parameters of this algorithm was low and the complexity of the network structure constructed in this way was low. The results of the OpenPose algorithm based on the parameters and the floating-point operations in the study were consistent with the results of this present study. Meanwhile, the average absolute error of the key frame detection model constructed in this way was less than 0.168 seconds, and the absolute error was only 0.033 seconds. The results demonstrated the effectiveness of the improved OpenPose algorithm for badminton player pose recognition[30]. In summary, based on the MobileNetV3 neural network and the DTW algorithm, the proposed model could accurately recognize badminton movements. It also had a high recognition accuracy for different stages of badminton movements.

5. CONCLUSION

For accurate badminton action recognition, the study proposed a badminton action recognition model comprising the MobileNetV3 neural network and the DTW algorithm. The badminton player's pose was estimated by the improved OpenPose algorithm, and the badminton action was recognized using the MobileNetV3 network architecture and the DTW algorithm. The experimental results indicated that the model had high accuracy and computational efficiency in badminton action recognition. Meanwhile, the DTW algorithm could effectively assess the similarity between badminton technical movements, and the accuracy of the assessment was further improved by introducing weighting values. While the research has yielded certain findings, there are still some limitations to be addressed. Further improvements could be made to the recognition accuracy for highly complex and closely similar movements, such as those in the batting phase and random phase. Additionally, the efficacy of the model may be influenced by the magnitude and heterogeneity of the dataset. Consequently, it would be advantageous to collect further action data from badminton players in the future, thus increasing the size and diversity of the dataset. This would serve to enhance the model's generalizability and robustness. Furthermore, capturing action details during phases of rapid action change would facilitate more precise action recognition.

REFERENCES

1. D Suryadi, Y T J Samodra and U Gustian U. Problem-based learning model: Can it improve learning outcomes for long serve in badminton? Edu Sportivo: Indonesian Journal of Physical Education, vol. 4, no.1, pp. 29–36, March, 2023, DOI: 10.25299/es:ijope.2023.vol4(1).10987.
2. C Hebbi and H R Mamatha. "Comprehensive dataset building and recognition of isolated handwritten kannada characters using machine learning models". Artificial Intelligence and Applications. vol. 1. no. 3, pp 179–190. Nov, 2023. DOI: 10.47852/bonviewAIA3202624.
3. E Sepdanius, R M Rifki, and A Gemaini. "Development of Kinesthetic Movement Identification Instruments for Badminton". Physical Education Theory and Methodology, vol. 23. no. 3, pp 358–365. Nov, 2023. DOI: 10.17309/tmfv.2023.3.07.
4. E Edmizal, E Barlian and D Donie. "Biomechanical analysis of smash stroke in badminton: A Comparative Study of Elite and Recreational Players: a systematic review". Retos: nuevas tendencias en educación física, deporte y recreación. vol. 2. no. 57, pp 809–817. March, 2024. DOI: 10.1080/10255842.2023.2190439.
5. A. Guo, Y. Xu, and G. Li, "Exploring the role of multi-scale convolutional operators in behavior sequence recognition," J. Intell. Fuzzy Syst. , vol. 45, no. 4, pp. 6887–6896, April, 2023. DOI: 10.3233/jifs-231220.
6. E Edmizal, D Rahman and E Barlian. "Tactics and strategy analysis in professional badminton: insights from match data and performance metrics-a systematic review". Retos: nuevas tendencias en educación física, deporte y recreación. vol. 6. no. 61, pp 78–85. March, 2024. DOI: 10.1177/17540739241259567.
7. W Yang, M Jiang and X Fang. "A high-precision and efficient method for badminton action detection in sports using You Only Look Once with Hourglass Network". Engineering Applications of Artificial Intelligence. vol. 137. no. 2, pp 109177–109189. March, 2024. DOI: 10.1016/j.engappai.2024.109177.
8. Z Song and G Liu. "Research on badminton video action recognition method based on temporal segmentation network". Proceedings of the 2024 International Conference on Intelligent

Education and Computer Technology. vol. 4. no. 2, pp 356–360. June, 2024. DOI: 10.1145/3687311.3687375.
9. B Shi and B Lv. "Accuracy of badminton swing action recognition based on fractional time network". Journal of Intelligent & Fuzzy Systems. vol. 45. no. 2, pp 2409–2418. July, 2023. DOI: 10.3233/JIFS-231409.
10. Y H Chen. "Action Detection in Badminton Courts Using AVA Dataset and MMAction2 Architecture with Slow Fast Model". Highlights in Science, Engineering and Technology. vol. 85. no. 2, pp 783–789. Mar, 2023. DOI: 10.54097/qrh46274.
11. J Yang, C Li and D Hong "Real-Time Safety Behavior Detection Technology of Indoors Power Personnel Based on Human Key Points". Journal of Shanghai Jiaotong University (Science). vol. 29. no. 2, pp 309–315. Mar, 2024. DOI: 10.1109/TCSS.2024.3448400.
12. Y Zhao and F Dai. "Design of Mobile Application System for Recognition and Analysis of Dynamic Running Posture". CCF National Conference of Computer Applications. Singapore: Springer Nature Singapore. vol. 5. no. 3, pp 35–42. March, 2024. DOI: 10.2298/CSIS230610014P.
13. Y Saiki, T Kabata. "Reliability and validity of OpenPose for measuring hip-knee-ankle angle in patients with knee osteoarthritis".. vol. 13. no. 1, pp 3297–3215. Nov, 2024. DOI: 10.1038/s41598-023-30352-1.
14. P. Srihari, J. Harikiran, and C. V. S. Reddy, "Effective framework for human action recognition in thermal images using capsnet technique," J. Intell. Fuzzy Syst. , vol. 45, no. 6, pp. 11737–11755, June, 2023. DOI: 10.3233/jifs-230505.
15. C Lu, J Zheng and L Yin. "An improved iterated greedy algorithm for the distributed hybrid flowshop scheduling problem". Engineering Optimization. vol. 56. no. 5, pp 792–810. Nov, 2022. DOI: 10.1080/0305215X.2023.2198768.
16. N Lin and R Jiang "Lithological classification by hyperspectral images based on a two-layer XGBoost model, combined with a greedy algorithm". Remote Sensing. vol. 15. no. 15, pp 3764–3770. July, 2023. DOI: 10.3390/rs15153764.
17. M Shafapourtehrany, F Rezaie and C Jun. "Mapping post-earthquake landslide susceptibility using U-Net, VGG-16, VGG-19, and metaheuristic algorithms". Remote Sensing. vol. 15, no. 18, pp 4501–4512. September, 2023. DOI: 10.3390/rs15184501.
18. K Tyagi, S Vats and V Vashisht. "Implementing inception v3, VGG-16 and VGG-19 architectures of CNN for medicinal plant leaves identification and disease detection". Journal of Electrical Systems. vol. 20, no. 7, pp 2380–2388. September, 2023. DOI: 10.1109/STI53101.2021.9732592.
19. K R Mohith, A. S. Vickram, and P. R. Karthikeyan. "Classification of images using INCEPTION algorithm in comparison with VGG-19 algorithm to measure accuracy". vol. 2853, no. 1, pp 2563–2575. September, 2023. DOI: 10.1063/5.0198490.
20. A Şener and B Ergen. "Enhancing Brain Tumor Detection on MRI Images Using an Innovative VGG-19 Model-Based Approach. "Sakarya University Journal of Science. vol. 27, no. 5, pp 1128–1140. May, 2023. DOI: 10.1177/10732748231169149.
21. Y Kaya and E Gürsoy. "A MobileNet-based CNN model with a novel fine-tuning mechanism for COVID-19 infection detection". Soft Computing. vol. 27, no. 9, pp 5521–5535. June, 2023. DOI: 10.1007/s00500-022-07798-y.
22. R Vallabhaneni, S A Vaddadi, S Pillai "MobileNet based secured compliance through open web application security projects in cloud system". Indonesian Journal of Electrical Engineering and Computer Science. vol. 35, no. 3, pp 1661–1669. September, 2024. DOI: 10.11591/ijeecs.v35.i3.pp1661-1669.
23. L Jia, T Wang and Y Chen. "MobileNet-CA-YOLO: An improved YOLOv7 based on the MobileNetV3 and attention mechanism for Rice pests and diseases detection". Agriculture. vol. 13, no. 7, pp 1285–1289. September, 2023. DOI: 10.3390/agriculture13071285.
24. D Han, H X Zhou and T H Weng. "LMCA: a lightweight anomaly network traffic detection model integrating adjusted mobilenet and coordinate attention mechanism for IoT". Telecommunication Systems. vol. 84, no. 4, pp 549–564. October, 2023. DOI: 10.1007/s11235-023-01059-5.
25. R Chen, H Huang and Y Yu. "Rapid detection of multi-QR codes based on multistage stepwise discrimination and a compressed MobileNet". IEEE internet of things journal. vol. 10, no. 18, pp 15966–15979. September, 2023. DOI: 10.1109/JIOT.2023.3268636.
26. R O Ogundokun, S Misra and A O Akinrotimi. "MobileNet-SVM: A lightweight deep transfer learning model to diagnose BCH scans for IoMT-based imaging sensors". Sensors. vol. 23, no. 2, pp 656–671. December, 2022. DOI: 10.3390/s23020656.
27. M Z Sajid, I Qureshi, Q Abbas, M Albathan and K Shaheed "Mobile-Hr: An ophthalmologic-based classification system for diagnosis of hypertensive retinopathy using optimized MobileNet architecture". Diagnostics. vol. 13, no. 8, pp 1439–1452. December, 2023. DOI: 10.3390/diagnostics13081439.
28. K Sabanci. "Benchmarking of CNN models and mobilenet-BiLSTM approach to classification of tomato seed cultivars". Sustainability. vol.15, no.5, pp. 4443–4459, February, 2023, DOI: 10.3390/su15054443.
29. H N Zheng, H C Shen, Y S Zhang and H T Wang. "Badminton action recognition based on improved I3D convolutional neural network". SPIE, vol. 12717, no.2, pp. 776–784, July, 2023, DOI: 10.1117/12.2684732.
30. M A Sarwar, Y C Lin, Y A Daraghmi and Y L Li. "Skeleton Based Keyframe Detection Framework for Sports Action Analysis: Badminton Smash Case". IEEE Access, vol. 11, no.2, pp. 90891–90900, June, 2023, DOI: 10.1109/ACCESS.2023.3307620.