International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3318-3326

https://doi.org/10.70517/ijhsa463280

Design of a framework for skill assessment and performance analysis of soccer players based on multimodal data fusion

Changwei Chen[1,*], Chanjuan Liu[1] and Xiaowen Song[2]
[1] Basic Department of Qilu Institute of Technology, Jinan, Shandong, 250200, China
[2] Baotou Medical College, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia, 014040, China
Corresponding authors: (e-mail: 13345125783@163.com).

Abstract The author in order to realize the skill assessment of soccer players and analyze the athletes' performance on the field. Based on the extraction of multimodal data of soccer players and the fusion of multistream data with adaptive multiscale differential graph convolutional network, the athlete action recognition model is constructed. Next, the PC-CNN skill assessment model is constructed to assess and analyze the skills of soccer players. The performance performance of the assessment model in this paper is tested through comparative experiments. Finally, analyze the performance of the soccer team in the shooting mode and shooting area. The accuracy of this paper's PC-CNN assessment model is more than 90% in the assessment of three levels of skills of soccer players, and the accuracy of the assessment of high-level skills reaches 100%, which is more accurate than other models. Among the multimodal data of athletes' physiology, joint movements had the greatest influence on the assessment accuracy of the assessment model, which reached 78.17%.The soccer teams A, B, and C had the highest success rate in the item of penalty kick shooting skill, with success rates of 75%, 66.7%, and 66.7%, respectively. Teams A and C had the highest success rate in Zone 1.Team B had the highest success rate in Zone 4.

Index Terms action recognition model, skill assessment model, multimodal data fusion, performance analysis

I. Introduction

Soccer is a sport that is popular in almost all countries of the world and is practiced and enjoyed by a very large number of young people in every country. Skill technique and athletic performance are considered to be the two key components of soccer, where players not only achieve technical accuracy but also maintain a high level of athletic performance at all times in order to fulfill the training and competition requirements of modern soccer [1]-[3]. Since soccer is a same-field confrontational sport, athletes need to complete the required technical movements in high-speed movement or intense confrontation, which puts very high requirements on the standardization of athletes' technical movements [4], [5]. At the same time, monitoring the real-time performance of the athletes can determine the shortcomings and deficiencies of the athletes, which is very important for the improvement of their own competitive ability [6], [7]. Therefore, an evaluation framework that can respond to athletes' skills and techniques and sports performance can provide theoretical and practical support for the scientific and rational development and selection of soccer players [8].

The arrival of the big data era has brought new opportunities and challenges for solving the above problems, and the huge amount of educational data and sports data makes it possible to portray the growth of athletes from a multidimensional perspective [9], [10]. Using big data analysis technology, mining the intrinsic connection between data and establishing data-driven assessment models can realize the accurate portrait and dynamic assessment of soccer players' sports quality, providing powerful data support and intellectual support for scientific selection, personalized training, potential prediction and so on [11]-[14].

The article firstly constructs an action recognition model for athletes through adaptive multi-scale differential graph convolutional network, and performs multi-stream data fusion in the recognition model to construct an action recognition model based on multi-modal data (AMD-GCN). Through the action recognition model, the limb data such as joint movement and eye tracking of soccer players during the movement process are collected, which are used as the prerequisite factors for skill assessment. The PC-CNN network model is established, and then the multimodal data fusion model and data analysis methods are used to establish the skill evaluation model, so as to construct the skill assessment model based on CNN and decision fusion. After testing the performance of the evaluation model, the performance of the soccer player in terms of re-shooting mode and success rate of the shooting area is analyzed through the model data analysis method.

II. Action recognition model based on multimodal data (AMD-GCN)

II. A. Preparatory knowledge

II. A. 1) Skeleton sequence characterization

The structure of the human skeleton can be represented as a network graph, with human joints as nodes and bones as edges. Usually a skeleton containing N nodes can be represented as a graph $G=(V,E)$, where $V=\{v_1,\dots,v_n\}$ denotes the set of N nodes and E denotes the set of edges. E contains two sets of edges, and the first set of edges is the intra-frame node connections, denoted as $E_s=\{(v_{ti},v_{tj})\mid i,j=1,\dots,V;t=1,\dots,T\}$. The intra-frame node relationships of this graph can be represented by the adjacency matrix $A\in R^{N\times N}$, with the matrix element v_{ij} taken as either 0 or 1, indicating whether the node v_i is connected to v_j. The second set of edges denotes the correlation of the nodes between frames, denoted as $E_T=\{(v_{ti},v_{(t+1)i})\mid i=1,\dots,V;t=1,\dots,T\}$. The whole skeleton sequence can be represented as $X\in R^{C_{in}\times T_{in}\times N_{in}}$, where C_{in},T_{in}, and N_{in} stand for the node coordinate dimensions, the number of skeleton frames, and the number of nodes, respectively.

II. A. 2) Spatio-temporal graphical convolutional networks

A spatio-temporal graph convolution block consists of a spatial graph convolution with a temporal convolution [15]. The graph convolution operation for the t th frame can be expressed as:

$$f_{out}(v_{ti})=\sum_{v_{tj}\in N(v_{ti})}\frac{1}{Z_{ti}(v_{tj})}f_{in}(v_{tj})\cdot w(l_{ti}(v_{ij})) \tag{1}$$

where v_{ti} denotes the i th node of the t th frame, and f_{in} and f_{out} represent the input as well as the output feature maps, respectively. The $N(v_{ti})$ denotes the set of neighbor nodes of v_{ti}, the weight function $w(\cdot)$ is used to assign weights to the inputs, and the mapping function $l_{ti}(\cdot)$ is used to assign different labels to all nodes within the set $N(v_{ti})$. In ST-GCN, there exist three mapping functions, which are unique labeling, distance function, and spatial configuration. Among them, the spatial configuration is the best, which classifies the neighboring nodes into three categories, the first one is the node itself, the second one is the neighboring nodes whose distance from the central node is closer than the distance from the root node, and the third one is the neighboring nodes whose distance from the root node is closer than the distance from the central node.

The regular term $Z_{ti}(v_{tj})$ is used to balance the distribution of different classes of nodes.

For practical application of ST-GCN, Eq. (1) is transformed into:

$$f_{out}=\sum_{k=1}^{K_v}W_k(f_{in}\cdot A_k)\odot M_k \tag{2}$$

where K_v denotes the number of subsets of neighboring nodes, and under the spatial configuration, K_v is 3. $A_k=\Lambda_k^{-\frac{1}{2}}\bar{A}_k\Lambda_k^{-\frac{1}{2}}$, where $\overline{A_k}$ denotes the adjacency matrix of the neighbor nodes of the k th class. The $\Lambda_k^{ii}=\sum_j(\bar{A}_k^{ij})+\alpha$ is the normalized diagonal matrix, and α is set to zero. 001 preventing empty lines. The learnable parameters W_k and M_k are used for the convolution operation with adjusting the weights of the edges and $\odot$ for the dot product operation, respectively.

Temporal convolution is generally a 1-dimensional convolution with a convolution kernel size of $K_t\times 1$, where K_t is the temporal sensing field, which helps to aggregate the node information of neighboring frames within the sensing field.

II. A. 3) Differential Convolutional Networks

AMD-GCN introduces differential information into the temporal convolution operation, which further enriches the input features by taking into account the interframe node motion information.

Ordinary convolution directly aggregates the features in the nearby region and operates on pixel p_0 as follows:

$$y(p_0)=\sum_{p_n\in R}w(p_n)\cdot x(p_0+p_n) \tag{3}$$

where p_0 is the center pixel position, p_n is the bias from the center pixel position, $w(p_n)$ is the weight corresponding to this bias, and R is the set of biases of the pixels in the nearby region. The differential convolutional network aggregates the differential information of the nearby region and its operation can be expressed as:

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot(x(p_0+p_n)-x(p_0)) \tag{4}$$

Conventional convolution and differential convolution can be related by the following equation:

$$\begin{aligned} y(p_0)&=\theta\cdot\sum_{p_n\in R} w(p_n)\cdot(x(p_0+p_n)-x(p_0))+(1-\theta)\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n) \\ &=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n)+\theta\cdot(-x(p_0))\cdot\sum_{p_n\in R} w(p_n) \end{aligned} \tag{5}$$

where the hyperparameter $\theta\in[0,1]$ is used to adjust the weights of the conventional convolution and differential convolution, Eq. (5) degenerates to Eq. (3) when θ is 0, and Eq. (5) degenerates to Eq. (4) when θ is 1.

II. B.Adaptive spatial graph convolution module

The graph convolution operation based on the fixed adjacency matrix is shown in Equation (2), where the topology of the graph is determined by the adjacency matrix A_k and the mask M_k, where whether two nodes are connected or not is determined by A_k, and the strength of the connection is controlled by M_k. In order to utilize both the fixed adjacency matrix and the adaptive adjacency matrix, Eq. (2) can be modified as:

$$f_{out}=\alpha_1\cdot\left(\sum_{k}^{K_v} W_k(f_{in}A_k)\odot M_k\right)+\alpha_2\cdot\Phi f_{in}B \tag{6}$$

Two types of adjacency matrices, A_k and B, are included here. The spatial map convolution results based on them are controlled by two learnable parameters α_1 and α_2. $A_k\in R^{V\times V}$ denotes the adjacency matrix corresponding to the set of neighboring nodes k, which characterizes the connectivity structure of the human skeleton itself. $B\in R^{N\times C\times V\times V}$ is the adaptive adjacency matrix, where N,c,v represent the Batchsize size, the number of channels and the number of nodes in the frame, respectively. With this adjacency matrix, a specific node topology can be provided for different channels of each sample, and thus the flexibility of action modeling can be greatly enhanced.

In order to construct the correlation between nodes, AMD-GCN first projects the node features into the same feature space, and then measures the similarity of the node features using dot product. Firstly, for the input feature map $f_{in}\in R^{N\times C_{in}\times T\times V}$, it is mapped to $f_e\in R^N\times C_{out}\times T\times V$. Subsequently a 3×1 convolution is used to aggregate the local temporal features and the output is $f_c\in R^N\times C_{out}\times T_{new}\times V$. Here the shapes of f_c and f_e are not the same because the convolution operation was performed without padding to avoid introducing irrelevant information. The T and V dimensions of f_c are then exchanged to obtain the new feature $f_t\in R^N\times C_{out}\times V\times T_{new}$. Subsequently f_c and f_t are matrix multiplied and passed through the softmax layer to obtain the similarity matrix $B\in R^{N\times C_{out}\times V\times V}$, where B^{ij} denotes the nodes v_i and v_j have similarity. This similarity matrix is matrix multiplied with the features f_e to obtain the output $f_o\in R^{N\times C_{out}\times T\times V}$. The process can be expressed as:

$$f_o=B\otimes f_e=softmax(f_t\otimes f_c)\otimes f_e \tag{7}$$

where $\otimes$ denotes matrix multiplication.

II. C.Attention-enhanced multi-scale temporal difference convolution module

The motion patterns of different actions are often very different, and the traditional single time-series convolution is difficult to be modeled effectively. Therefore, AMD-GCN proposes multi-scale temporal convolution to obtain multi-scale temporal information.

In order to more fully model the motion information, AMD-GCN combines differential convolution with time series convolution to obtain time series differential convolution. Different from the traditional differential convolution, this section uses a non-shared convolution kernel to go increase its flexibility. That is, for the i th node x_t^i of the t th frame, the timing differential convolution can be expressed as:

$$y(x_i^i)=\sum_{j=1}^{K} w_{i1}\cdot x^i_{\left\lfloor t-\frac{K}{2}+j\right\rfloor}+\sum_{j=1}^{K} w_{i2}\cdot\left(x^i_{\left\lfloor t-\frac{K}{2}+j\right\rfloor}-x_i^i\right) \tag{8}$$

The formula combines ordinary temporal convolution with temporal differential convolution. $y(x_t^i)$ represents the output of the temporal difference convolution, the first term on the right side of the equal sign is the ordinary temporal convolution and the second term represents the temporal difference convolution, and their weights are adjusted using w_{i1} and w_{i2}, respectively. Here K is the time-sensitive field size of the time-series convolution.

II. D.Multi-stream data fusion

In order to more fully utilize the skeleton sequence information, this study also considers the skeleton, node motion, and bone movement information. Each skeleton can be represented as a vector of source nodes pointing to a target node. For the source node $v_i = (x_i, y_i, z_i)$ and the target node $v_j = (x_j, y_j, z_j)$, the corresponding skeleton vectors can be represented as $b_{ij} = (x_j - x_i, y_j - y_i, z_j - z_i)$. Since the skeleton graph is a graph that does not contain a ring, assigning bones to the target nodes will result in the number of joints being greater than the number of bones. Therefore here the bone vectors corresponding to the joint points that are not assigned bones are set to all 0 vectors. The node motion information is the vector pointing from the current node to the node at the corresponding position in the next frame, for the node $v_t = (x_t, y_t, z_t)$ in the t th frame and the corresponding node $v_{t+1} = (x_{t+1}, y_{t+1}, z_{t+1})$, and the node motion vector can be expressed as $jm_{v_t v_{t+1}} = (x_{t+1} - x_t, y_{t+1} - y_t, z_{t+1} - z_t)$. Similarly, the skeletal motion vector is the displacement vector $bm_{vt}v_{t+1} = (x_{t+1}^b - x_t^b, y_{t+1}^b - y_t^b, z_{t+1}^b - z_{t+1}^b)$ of the adjacent frames of bones at the same position, where the corresponding t frame skeleton is denoted as $b_t = (x_t^b, y_t^b, z_t^b)$, and the skeleton of the $t+1$ th frame is denoted as $b_t + 1 = (x_{t+1}^b, y_{t+1}^b, z_{t+1}^b)$.

III. Skill assessment model based on CNN and decision fusion

III. A. PC-CNN network model

In this paper, the PC-CNN network model is constructed as shown in Fig. 1. This network is a spatial pyramid pooling layer SPP8 added to the convolutional layer C7 and the fully connected layer FC9 to realize that the length of the output vectors passing through the SPP8 layer is fixed and uniform, and the traditional stochastic gradient descent (SGD) is replaced by the COCOB optimization algorithm which does not need to set the learning rate during the training process of the network in order to achieve faster convergence [16].The PC-CNN network model in C1 layer is a convolutional layer containing 32 filters with convolutional kernel size of 3×3 and step size of 1. The input is a gait contour image.S2 layer is a pooling layer, which does maximum pooling in a 2×2 window with a step size of 2.C3 layer is a convolutional layer containing 32 filters with convolutional kernel size of 3×3 and step size of 1.C4 layer is a convolutional layer containing 32 filters with convolutional kernel size of 3×3 and step size of 1.S5 layer is a pooling layer which does maximum pooling in a 2×2 window and step size of 1.S6 layer is a pooling layer which does maximum pooling in a 2×2 window and step size of 2. The S5 layer is a pooling layer that does maximum pooling within a 2 × 2 window with a step size of 2. The C6 layer is a convolutional layer containing 64 filters with convolutional kernel size of 3 × 3 with a step size of 1. The C7 layer is a convolutional layer containing 64 filters with convolutional kernel size of 3 × 3 with a step size of 1. The SPP8 layer is a spatial pyramid pooling layer, and the extraction of the feature blocks consists of three scales: 1 × 1, 2 × 2, and 4 × 4 , each feature block is maximally pooled (i.e., maximum pooling is done within a window of 1×1, 2×2, and 4×4) with a step size of 1. The FC9 layer is a fully connected layer, the input to the fully connected layer is the result of the spatial pyramid pooling layer, and the number of nodes in the output of the fully connected layer is 64.

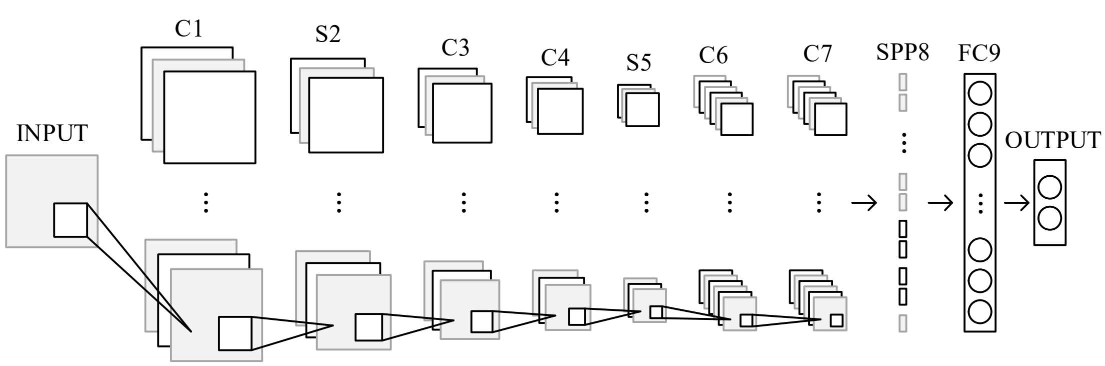

Figure 1: PC-CNN network structure

In this paper, the AMD-GCN model constructed in the previous paper is used to monitor the joint motion and eye tracking of soccer players as a factor for skill assessment. Then PC-CNN network is used to train and recognize the limb contour images of the athletes. The athlete's limb contour images were recognized by PC-CNN combining SPP and COCOB optimization algorithms, and the recognition accuracy of the PC-CNN network model was used as the modal data input of the limb contour images for the decision fusion model.

III. B. Skill evaluation modeling

III. B. 1) Multimodal data fusion models

Multimodal feature fusion refers to the fusion of features from different modalities [17] to improve the performance and accuracy of the model, which is mainly divided into two stages: feature extraction and feature fusion. Multi-channel feature fusion is widely used in the field of computer vision, and in the past, most models were fused at the decision layer, which can lead to the following problems, for a certain sample, whose category label is forehand striking, when the prediction result of the bone branch and the joint point branch are both backhand striking, and the prediction result of the RGB branch is a right-footed shot, the fusion at the decision layer takes into account the confidence of the multi-branching and it is easy to incorrectly determine the prediction result as having a left-footed shot. The fusion of the decision layer considering the confidence of multiple branches can easily misjudge the prediction result as having a left-footed shot. Even if the RGB branch can predict correctly after adding multimodal information, the fusion based on the model prediction results will weaken the decision power of the RGB branch. The same can be generalized to the case when one branch predicts correctly with a lower confidence level and the other two branches predict incorrectly with a higher confidence level, the decision level fusion may reduce the robustness of the model in this case. A schematic of feature fusion classification is shown in Figure 2.

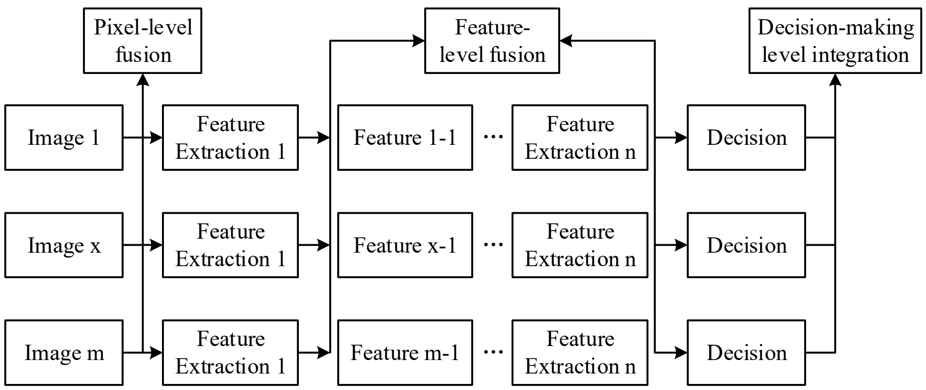

Figure 2: Schematic diagram of feature fusion classification

The specific process of the multimodal fusion layer is shown in Fig. 3, where the final output features of multiple models are fused at the feature level, and the fused multimodal features are mapped into a subspace for similarity matching between multimodal features, and the feature distances of multiple modalities are considered comprehensively for obtaining a more balanced similarity ranking.

Specifically, after obtaining the joint point and bone prediction results, channel-level splicing is performed for the skeletal modalities from the two sources, followed by dimensionality transformation via a convolutional layer with a convolutional kernel size of 1×1 while preserving the nonlinear modeling capability of the network, and then a global average pooling layer to obtain the channel-level feature values representing the importance of the features at the channel level. This set of weights is used to interact with the features of the RGB modality to obtain the skeleton related RGB features, which are then entered into the convolutional layer for the Reshape operation and spliced with the bi-stream features of the skeleton modality at the channel level to obtain the final feature output.

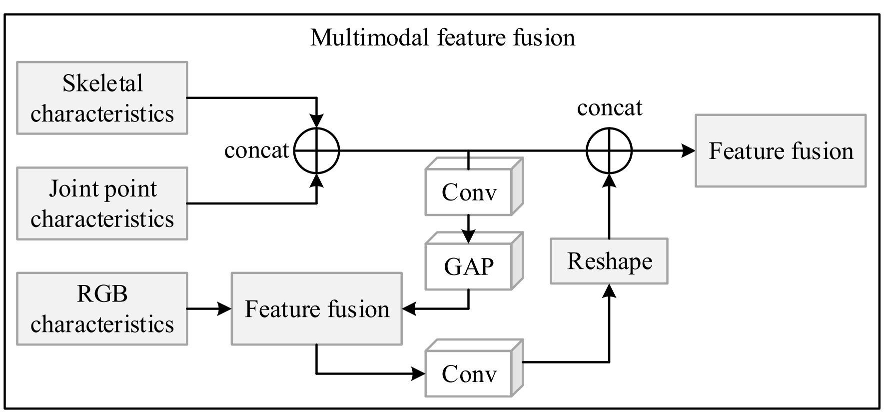

Figure 3: Multimodal feature fusion

III. B. 2) Data analysis

Joint angle data were low-pass filtered at 6 Hz to remove noise.GRF data were low-pass filtered at 20 Hz and amplitudes were normalized by body weight to reduce the effect of measurement differences between subjects. Due to the inconsistency in the length of the sample data due to the differences between the recorded gait cycles, the two modal data, joint angle and GRF, were time normalized to 101 points (representing the motion period from 0% to 100%), respectively, and then, the time normalized data from the different channels were concatenated to form a sample vector of the respective modal data. To reduce redundant information and extract effective gait features, PCA was used for feature extraction of joint angles with GRF. The number of principal components retained was determined by the 90% variance contribution.

The features extracted from both modal data, joint angle and GRF, are used as inputs to the classification algorithm. In this paper, Simple Bayes (NB), k Nearest Neighbor Algorithm (KNN), Support Vector Machine (SVM), and Random Forest (RF) are used as candidate classifiers, and the best classifier for the two modalities is selected by comparing the recognition performance. In the fusion stage, the two selected optimal classifiers and the output probabilities of Softmax are used to train the decision fusion model. Usually, decision fusion models can be built in two ways: classification-based fusion algorithms and rule-based fusion algorithms. The former utilizes data-based classification techniques to fuse different modal information, while the latter is based on predefined computational rules. In this paper, five fusion algorithms, namely neural network (NN), KNN, NB, SVM and RF, are used to build the fusion model. The input of the fusion model is the output probability of different modal information, and the integration of different modal information enhances the complementarity of motion information.

IV. Skills assessment and performance analysis

IV. A. Skills assessment performance

In order to verify the performance of the PC-CNN evaluation model in this paper, this paper collects the athlete's movements, eye tracking, EEG signals and other data during the soccer game. The athlete's joint movements are extracted and the number of occurrences of each movement is counted. Eye tracking metrics are calculated at full screen range, three different interest domains respectively. The EEG signals were decomposed into five different frequency bands, and the time and frequency domain features of each band were calculated separately. Among them, the three different interest domains are the results of the division of eye tracking data by each element, which are soccer ball trajectory, field of play, goal position, shooting distance, passing distance and defense line adjustment, so each interest domain needs to calculate the corresponding six eye tracking metrics separately. The overall features obtained from the raw data of the three modalities by feature extraction were used for subsequent data analysis and evaluation model construction. Five different machine learning algorithms, namely Bayesian Network (Bayes), Random Forest (RF), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), were used to compare the assessment models with the PC-CNN skill evaluation model in this paper.

Statistically, a total of 180,000 soccer player movement data, 360,000 eye tracking data, and 1.48 million EEG signal data were obtained in this study. The three modal data were feature extracted to form a feature matrix. Meanwhile, the original skill levels were divided into low, medium and high levels by their value ranges in equal amounts, which were used as skill level labels for subsequent data analysis and assessment model construction. In this study, the assessment accuracies of six machine learning algorithm models (Bayes, RF, LDA, SVM, KNN, PC-CNN) on the validation set are compared, and the results are shown in Table 1. It can be seen that the PC-CNN model in this paper has the best overall performance, not only the overall skill assessment accuracy rate reaches 94.3%, but also its skill assessment accuracy rate for low, medium and high levels is over 90%.

Table 1: Evaluation accuracy of six machine learning algorithms in verification set

Model	Accuracy in low level	Accuracy in medium level	Accuracy in high level	Overall evaluation accuracy
Bayes	75.4%	83.5%	74.8%	77.9%
RF	76.3%	80.8%	83.3%	80.1%
LDA	75.7%	78.8%	89.9%	81.5%
SVM	79.2%	87.6%	84.9%	83.9%
KNN	78.3%	87.2%	82.6%	82.7%
PC-CNN	90.4%	92.5%	100%	94.3%

In order to understand the influence of each modal data on the evaluation results of the six machine learning algorithm models, this study compares the accuracy of different modal data and their mean values in the six machine learning algorithm models, and the results are shown in Table 2. It can be seen that joint movements have the greatest influence on the evaluation results of the six machine learning algorithm models (the highest average value is 78.17%), followed by eye tracking (the second highest average value is 63.02%), and EEG signals have the least influence (the lowest average value is 55.47%). It can be seen that there is a stronger link between joint movements and skill level compared to eye tracking and EEG signals. However, the mean value of the multimodal data was greater than that of any of the unimodal data, thus verifying the superiority of the multimodal fusion method for skill level assessment.

Table 2: The effect of each modal data on evaluation results of six machine learning algorithms

Modal	Bayes	RF	LDA	SVM	KNN	PC-CNN	Average
Joint action	67.97%	73.55%	73.47%	82.45%	84.77%	86.83%	78.17%
Eye tracking	45.25%	66.33%	65.83%	65.34%	66.72%	68.64%	63.02%
Brain electrical signal	37.86%	61.27%	50.14%	55.92%	62.46%	65.16%	55.47%
Multimodal data	75.43%	85.68%	85.75%	87.67%	87.96%	89.47%	85.33%

IV. B. Performance analysis

Since soccer is a typical team sport, the assessment of athletes should ultimately come down to the analysis of the overall performance of the team. Based on the fact that the victory or defeat of a team is determined by the number of goals scored in the game, at the same time, the level of goal-scoring technology can be directly manifested from the characteristics of goals scored, and the two are inseparable. Therefore, this paper will be A, B, C three teams shooting technology indicators through the “shooting mode, shooting area 2 aspects of research.

IV. B. 1) Analysis of shot patterns

The shooting performance and shooting effect of the three teams A, B, and C are shown in Figure 4 and Figure 5. As can be seen from Figures 4 and 5, the shooting methods of the three teams A, B and C are as follows: (1) Grab the point and shoot directly, Team A shoots 14 times and scored 3 times, with a success rate of 21.4%. Team B had 17 shots on goal and scored 5 goals, with a success rate of 29.4%. Team C had 10 shots on goal and scored 5 goals, with a success rate of 50%. (2) Receiving the ball and adjusting the shot, Team A shot 7 times and scored 4 times, with a success rate of 57.1%. Team B had 13 shots on goal and scored 3 times, with a success rate of 23.1%. Team C had 7 shots on goal and scored 0 goals, with a success rate of 0%. (3) Set-piece shots, Team A shot 4 times, scored 2 times, and the success rate was 50%. Team B had 3 shots on goal and scored 0 goals, with a success rate of 0%. Team C had 6 shots on goal and scored 0 goals, with a success rate of 0%. (4) Penalty kick shots, Team A had 8 shots on goal and scored 6 times, with a success rate of 75%. Team B had 6 shots on goal and scored 4 times, with a success rate of 66.7%. Team C had 3 shots on goal and scored 2 goals, with a success rate of 66.7%. (5) Supplementary shots, Team A shot 6 times, scored 3 times, and the success rate was 50%. Team B had 0 shots on goal and scored 0 goals, with a success rate of 0%. Team C had 0 shots on goal and scored 0 goals, with a 0% success rate. (6) Header shots, Team A shot 4 times, 0 goals, 0% success rate. Team B had 6 shots on goal and 1 goal, with a success rate of 16.7%. Team C had 5 shots on goal and scored 1 goal, with a success rate of 20%. (7) Sudden shooting, teams A, B, and C all had 0 shots on goal, 0 goals were scored, and the success rate was 0%. (8) Long shots, Team A had 13 shots on goal and scored 0 goals, with a success rate of 0%. Team B had 7 shots on goal and scored 0 goals, with a success rate of 0%. Team C had 10 shots on goal and scored 0 goals with a 0% success rate.

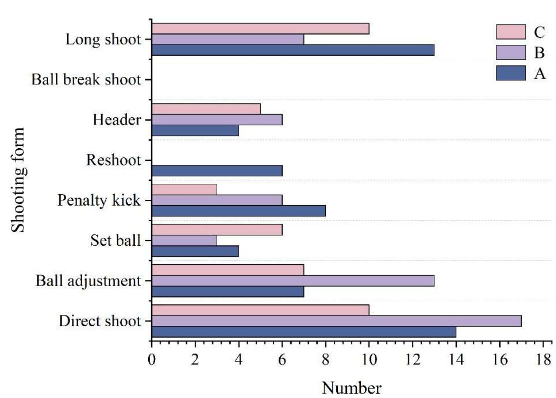

Figure 4: Shooting form

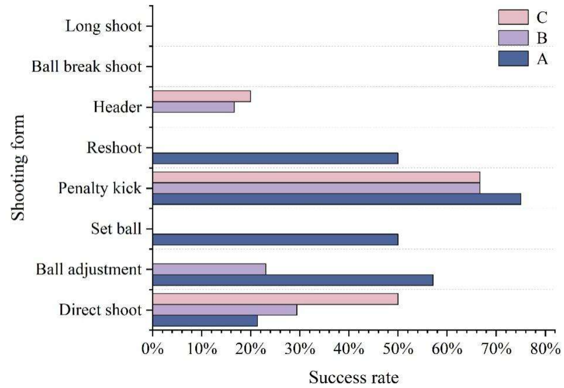

Figure 5: Shooting effect regional analysis of shotguns

The shooting area and shooting effect of the three teams A, B, and C are shown in Figure 6 and Figure 7. As can be seen from Figures 6 and 7, the shooting areas of teams A, B, and C are as follows: (1) Zone 1, team A has 8 shots on goal and 6 goals, with a shot success rate of 75%. Team B had 11 shots on goal and scored 3 goals, with a shot success rate of 27.3%. The C team had 8 shots on goal and scored 3 times, with a shot success rate of 37.5%. (2) In Zone 2, Team A has 4 shots on goal, 0 goals, and a shot success rate of 0%. Team B had 7 shots on goal and scored 2 goals, with a shot success rate of 28.6%. Team C had 3 shots on goal and 0 goals and a shot success rate of 0%. (3) In Zone 3, Team A had 15 shots on goal and scored 7 goals, with a shot success rate of 46.7%. Team B had 22 shots on goal and scored 7 goals, with a shot success rate of 31.8%. The C team had 14 shots on goal and scored 2 goals, with a shot success rate of 14.3%. (4) Zone 4, Team A has 20 shots on goal and 0 goals, with a shot success rate of 0%. Team B had 8 shots on goal and scored 3 times, with a shot success rate of 37.5%. Team C had 12 shots on goal and scored 3 goals, with a shot success rate of 25%. (5) In Zone 5, Team A has 5 shots, Team B has 2 shots, and Team C has 5 shots, but the success rate of shots on goal is 0%.

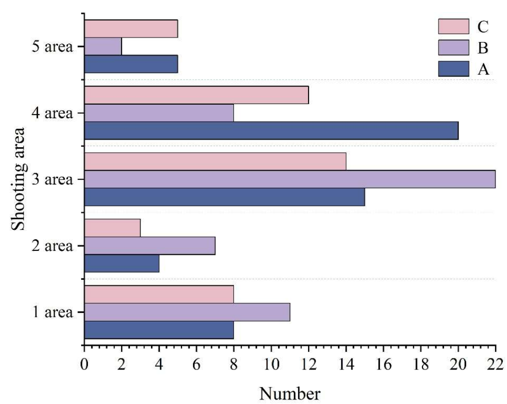

Figure 6: Shooting area

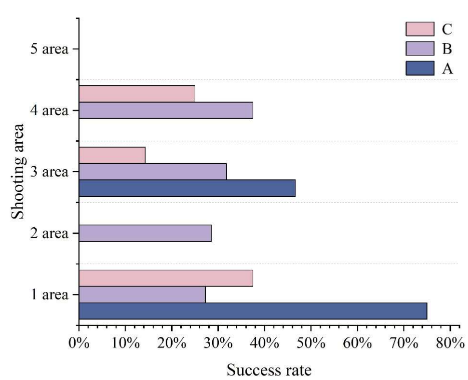

Figure 7: Shooting effect

V. Conclusion

The article firstly recognizes the athlete's movement by constructing AMD-GCN model, and secondly, CNN is fused with decision making to construct PC-CNN skill assessment model by high multimodal data fusion. The model was used to assess and analyze the skills and performance of soccer players.

(1) In the skill assessment of soccer players at low, middle and high levels, the accuracy of the PC-CNN model in this paper for low, middle and high levels of skill assessment is 90.4%, 92.5% and 100% respectively, and the overall skill assessment accuracy reaches 94.3%, exceeding the other comparative models. Among all the multimodal data, joint movements had the greatest influence on the assessment model (78.17%), followed by eye tracking (63.02%), and the least influence was the EEG signal (55.47%).

(2) The performance of the soccer teams in terms of shooting style and shooting area was analyzed.Teams A, B, and C had the highest success rate of penalty kick shooting technique with 75%, 66.7%, and 66.7%, respectively.Teams A and C had the highest success rate of shooting in Zone 1 with 75% and 37.5%, respectively.Team B had the highest success rate of shooting in Zone 4 with 37.5%.

References

[1] Kokstejn, J., & Musalek, M. (2019). The relationship between fundamental motor skills and game specific skills in elite young soccer players. Journal of Physical Education and Sport, 19, 249-254.

[2] Bojkowski, Ł., Kalinowski, P., Śliwowski, R., & Tomczak, M. (2022). The importance of selected coordination motor skills for an individual football player's effectiveness in a game. International journal of environmental research and public health, 19(2), 728.

[3] Preljević, A., Špirtović, O., Ahmić, D., Toskić, L., & Zećirović, A. (2020). The relationship between specific motor skills and performance success in football players. Facta Universitatis, Series: Physical Education and Sport, (1), 197-206.

[4] Nusri, A., Prima, A., Ardi, N. F., Ockta, Y., Setiawan, Y., Orhan, B. E., ... & Karaçam, A. (2024). Design of basic football skills test instrument for university students. Retos: nuevas tendencias en educación física, deporte y recreación, (59), 649-657.

[5] Jukic, I., Prnjak, K., Zoellner, A., Tufano, J. J., Sekulic, D., & Salaj, S. (2019). The importance of fundamental motor skills in identifying differences in performance levels of U10 soccer players. Sports, 7(7), 178.

[6] McIntosh, S., Kovalchik, S., & Robertson, S. (2019). Comparing subjective and objective evaluations of player performance in Australian Rules football. PLoS one, 14(8), e0220901.

[7] Wright, C., Carling, C., Lawlor, C., & Collins, D. (2016). Elite football player engagement with performance analysis. International Journal of Performance Analysis in Sport, 16(3), 1007-1032.

[8] Qader, M. A., Zaidan, B. B., Zaidan, A. A., Ali, S. K., Kamaluddin, M. A., & Radzi, W. B. (2017). A methodology for football players selection problem based on multi-measurements criteria analysis. Measurement, 111, 38-50.

[9] Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., ... & Lemmink, K. A. P. M. (2021). Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. European Journal of Sport Science, 21(4), 481-496.

[10] Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Transactions on Intelligent Systems and Technology (TIST), 10(5), 1-27.

[11] Yang, T., Yuan, G., & Yan, J. (2021). Health analysis of footballer using big data and deep learning. Scientific Programming, 2021(1), 9608147.

[12] Yang, H., Zhang, S., Zhang, J., & Wang, C. (2024). Evaluating the performance of athletes in various sports using data mining and big data analytics. Soft Computing, 28(4), 2875-2890.

[13] Herberger, T. A., & Litke, C. (2021). The impact of big data and sports analytics on professional football: A systematic literature review. Digitalization, digital transformation and sustainability in the global economy: risks and opportunities, 147-171.

[14] Yu, H., & Li, J. (2024). Developing a Reliable Hybrid Machine Learning Model for Objective Soccer Player Valuation. International Journal of Advanced Computer Science & Applications, 15(6).

[15] Jiaqi Wang,Yuqing Fan,Xi Pan,Jun Sun & Limao Zhang. (2025). Multi-source information fusion for dynamic safety risk prediction of aerial building machine using spatial–temporal multi-graph convolution network. Advanced Engineering Informatics,65(PB),103261-103261.

[16] Sri Vinitha V. & Renuka D. Karthika. (2024). Diverse ensemble classifier driven Email spam classification using multiple word embedding's with COCOB optimizer. Journal of Intelligent & Fuzzy Systems,46(1),2941-2954.

[17] Jinghan Cai,Wenzheng Liu,Tonghai Liu,Fanzhen Wang,Zhihan Li,Xue Wang & Hua Li. (2025). APO-CViT: A Non-Destructive Estrus Detection Method for Breeding Pigs Based on Multimodal Feature Fusion. Animals,15(7),1067-1067.