Computer Science and Information Systems 22(3):1139–1165 https://doi.org/10.2298/CSIS241121042L

The Analysis of Deep Learning-based Football Training under Intelligent Optimization Technology

Kun Luan[1], Fan Wu[1], and Yuanyuan Xu[2*]

[1]Nanchang University College of Science and Technology, Gongqingcheng, 330029, China;

155341805@qq.com
43969591@qq.com
[2]Jiangxi University of Finance and Economics, Nanchang, 330029, China
1201400004@jxufe.edu.cn

Abstract. This work aims to optimize college football training using deep learning techniques, addressing the inefficiencies, difficulty in action recognition, and insufficient data analysis present in current training methods. An intelligent optimization system combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) is proposed to tackle these challenges. Compared to traditional single models, the Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) architecture remarkably improves the efficiency and accuracy of processing training data by leveraging the strengths of spatial features and temporal sequence features. The experimental results show that CNN-RNN model is significantly superior to the traditional 3D CNN model and other advanced models, such as Transformer, Long Short-Term Memory (LSTM), Bidirectional LSTM and Gated Recurrent Unit (GRU), in key indicators such as accuracy, precision, recall and F1 score. Specifically, CNN-RNN model achieves 92.5% accuracy, 91.2% precision, 93.1% recall and 92.1% F1 score. The lowest training loss rate is 0.24, which is significantly better than other models. In addition, the introduced data balance strategy effectively improves the prediction performance of a few categories (such as foul and yellow card events) through oversampling, undersampling and weighted loss function, and further enhances the generalization ability and practicability of the model. Future research focuses on expanding the dataset, further improving the model’s generalization ability, and exploring its application in real training scenarios.

Keywords: deep learning; college football training; intelligent optimization; CNN-RNN; training loss value.

1. Introduction

With the continuous advancement of technology, artificial intelligence (AI) has been widely applied across various fields, especially demonstrating significant potential in the sports realm [1, 2]. As one of the world's most popular sports, football has been a focal research point in optimizing training and game strategies. However, traditional football training methods often rely on the coach's experience and intuitive judgment, lacking scientific and systematic approaches. Therefore, this work explores how deep learning

(DL) technology can be applied to college football training to achieve intelligent and optimized training methods [3-5].

First, an important aspect of the research background is the challenges college football training faces [6-8]. Currently, college football teams exhibit deficiencies in technical skills, tactics, and physical fitness, and these issues become particularly pronounced when facing higher-level opponents. Additionally, limited training resources such as facilities, equipment, and funding constrain the athletes' development. These issues urgently require solutions through technological means [9, 10].

Moreover, as a crucial branch of AI, DL has achieved remarkable success in areas like image and speech recognition [11, 12]. In the football field, DL can be employed to analyze match videos, identify and assess the quality of player movements, and predict match outcomes [13-15]. These applications provide coaches with scientific data support and help athletes better understand their strengths and weaknesses, enabling more targeted training [16-18].

With the continuous development of modern football, the role of data analysis in match strategy and player performance evaluation has become increasingly important. Traditional football analysis methods largely rely on static statistical data, which struggle to effectively address the complexity of real-time changes during a match. Existing football analysis models often fail to fully leverage video data and player statistics when predicting real-time match events (e.g., goals, fouls, yellow cards), leading to low prediction accuracy and insufficient real-time feedback capabilities. Therefore, how to utilize DL technologies for precise prediction of real-time match events has become a critical research topic in optimizing football training and tactical decision-making.

This work aims to address the shortcomings of current football analysis models in handling video data and player statistics by introducing a DL model that combines Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Specifically, the goal is to use the Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) model to predict real-time match events and enhance prediction accuracy through data augmentation and model optimization. Compared to traditional analysis methods based on manual feature extraction, the CNN-RNN model possesses stronger feature learning capabilities, enabling it to capture dynamic information in videos and the temporal sequence features of player behavior. Thus, this work fills the gap in existing football analysis models regarding real-time event prediction while providing new technical support for personalized optimization of football training.

By accurately predicting real-time match events, this work can better assist coaches in tactical decision-making and improve the effectiveness of player training. Precise event prediction helps identify key moments in a match in advance, providing timely data support for tactical adjustments and player training. This not only enhances the viewing experience and competitive level of matches but also offers new research avenues for data-driven sports training models.

The core research objective is to develop and validate a DL-based intelligent optimization system, thus enhancing the scientific and systematic aspects of college football training. It is acknowledged that, despite the global popularity of football, challenges in technology, tactics, and physical fitness persist in college football training, particularly in situations with limited resources. A hybrid model, combining CNNs and RNNs, is proposed to address these challenges. The model aims to automatically identify and assess the quality of players' movements by analyzing a substantial amount of football training and match video data. This analytical process provides coaches with data support, enabling them to formulate more effective training plans. Methodologically, a large-scale football dataset is first constructed, including training and match videos of players at different levels and relevant motion and performance data. Preprocessing steps, such as labeling player movements and events, are employed to ensure the accuracy of model training. Subsequently, a CNN-RNN architecture is designed, with CNN handling image and video data to capture visual features of player movements, and RNN processing sequence data to analyze the temporal information of actions. This combined approach allows the model to gain a more comprehensive understanding of the complexity of football movements. In terms of experimental design, a high-performance computing platform, along with TensorFlow and PyTorch frameworks, is utilized for model training and testing. Reasonable parameters, such as batch size and learning rate, are set to balance the model's learning efficiency and memory usage. Performance evaluation focuses on accuracy, precision, recall, and F1 score indicators to ensure the model's effectiveness in practical applications. The model's generalization ability is also considered by conducting cross-validation on different datasets to test its robustness.

The structure of this work is as follows. Section 1 presents the research background, motivation, and objectives. Section 2 introduces related research and literature review, analyzing the limitations of existing football analysis models. Section 3 details the design and implementation of the CNN-RNN model. Section 4 discusses data collection and experimental setup. Section 5 exhibits the experimental results and provides a detailed analysis. Finally, Section 6 concludes the study and discusses future research directions.

2. Literature Review

In the current field of sports technology, the application of DL technology is rapidly advancing and demonstrating significant potential and innovation in the football domain [19, 20]. With the continuous progress of big data and intelligent algorithms, there have been notable improvements in football match analysis, athlete performance evaluation, and the enhancement of the overall game experience. Many researchers and technology experts conducted in-depth exploration and practical applications in this field in recent years [21, 22]. Their work not only propelled the advancement of football technology but also introduced new perspectives and possibilities to the entire sports technology sector. Rahman (2020) proposed a DL-based football match prediction framework. This framework utilized complex algorithms and data analysis techniques to forecast match outcomes, considering factors such as team performance, historical data, and other relevant elements [23]. Stoeve et al. (2021) focused on applying laboratory technologies to real football scenarios. They used inertial measurement units and DL technology to detect shooting and passing actions in football training and matches, aiming to enhance the effectiveness and efficiency of athlete training [24]. Fenil et al. (2019) introduced a real-time violence detection framework for football stadiums. By combining big data analysis and Bidirectional Long Short-Term Memory (Bi-LSTM) in DL technology, this framework aimed to improve the efficiency and accuracy of safety monitoring during football matches [25]. Cuperman et al. (2022) developed a DL-based process for football activity recognition using wearable accelerometer sensors. This approach contributed to a more accurate analysis of player movements and performance, providing valuable data for coaches and sports scientists [26]. Wang et al. (2019) dedicated their efforts to developing a DL-based intelligent editing system for football matches. This system could use algorithms to automatically edit and highlight crucial moments in matches, enhancing the viewing experience and media production efficiency [27].

The application of DL in the sports domain primarily focuses on two directions: athlete motion analysis and match prediction. By combining CNNs with RNNs, researchers attempted to improve athlete performance during matches and optimize training strategies. For example, Sen et al. (2021) employed a CNN-RNN architecture to analyze cricket match videos, enabling precise identification and classification of player actions [28]. By introducing CNN for extracting spatial features and combining it with RNN to capture temporal dynamics, the model could accurately predict player actions (e.g., passing, shooting, etc.). Moreover, it could further apply these predictions to optimize actions in training scenarios. However, this study's limitation lies in its relatively small dataset and its primary focus on action recognition within match scenarios, with less exploration of real-time feedback optimization during training. Li (2023) proposed a method combining CNNs with LSTM to predict the movement trajectories of football players [29]. This approach mainly concentrated on the dynamic prediction of player positions, but its research on training optimization was relatively weak. In contrast, this work focuses on action recognition during training and strives to optimize the training process through DL, particularly making innovative attempts in data augmentation and balancing.

The application of DL technology in the football field is becoming increasingly profound and widespread. From predicting match outcomes and analyzing player performance to monitoring the safety of playing fields, and even editing match videos and tracking the ball, DL technology is gradually transforming various aspects of football [30-32]. These studies not only showcase DL technology's immense potential in improving training efficiency, enhancing the spectacle of matches, and ensuring match venue safety but also pave the way for future technological innovations in football. By applying these advanced technologies, football is evolving towards a more precise, intelligent, and scientific direction, driven by technological innovation [33, 34]. In contrast, although some studies have tried to optimize training through deep learning technology, most of them focus on the application of a single model, such as the independent use of CNN or LSTM, and lack of multi-level analysis combining different neural network models. The advantage of deep learning technology is that it can effectively process complex training data through different types of network models, combining spatial characteristics and time series characteristics. In the existing research, although some scholars try to use the combination of CNN and RNN to improve the accuracy and efficiency of training, on the whole, no research has made full use of the potential of CNN-RNN architecture to solve the problems of unbalanced data, high complexity of actions and identification of a few types of actions (such as fouls and yellow cards) in football training. This kind of problem is the key challenge in current football training, and it is of great significance to improve the training effect. In addition, the application of data enhancement and weighted loss function in existing research is still limited, which fails to fully tap its potential in solving data imbalance and improving model generalization ability. Especially when it comes to complex action recognition (such as a few kinds of actions like foul or yellow card), traditional deep learning models often encounter problems of insufficient accuracy and recall. By introducing data enhancement and weighted loss function, this work can effectively improve the precision and recall rate of the model in minority action recognition, and further improve the training effect.

In a word, although the existing research has made some progress in the application of deep learning technology in the field of football, there are still obvious technical gaps in the aspects of action optimization, multi-level model analysis and data imbalance in the training process. These shortcomings provide technical incentives for the proposal of this work. The innovation of this work lies in the combination of CNN and RNN, aiming at the optimization of action recognition in the training process, and improving the accuracy and efficiency of training through multi-level data balance strategy and model optimization technology, especially in the recognition of complex actions and the handling of minority problems, showing stronger advantages. Through these innovations, this work not only fills the gaps in the existing research, but also provides new ideas and technical support for the application of deep learning in future football training.

3. Research Methodology

This work aims to explore and implement an intelligent optimization technology for college football training using the DL-based CNN-RNN model. In the stage of model construction, a mixed model architecture combining CNN and RNN, namely CNN-RNN model, is adopted, which is realized by two deep learning frameworks, TensorFlow and PyTorch. In order to optimize the performance of the model, this work systematically optimizes the learning rate, batch size and optimizer by grid search method to ensure that the model can be trained under the optimal parameter configuration. In the process of model training, an independent verification set is used to monitor the performance of the model to prevent over-fitting. The selection of the CNN-RNN model is based on several key reasons. First, CNNs perform exceptionally well when handling image data, efficiently extracting spatial features from the video frames of football matches. Specifically, CNNs can automatically identify key areas in images, such as the player's body posture, movement trajectories, the ball's position, and other critical elements on the field. Therefore, CNNs can extract the spatial information from each frame of the football match, providing strong support for subsequent time-series modeling. Second, RNNs, particularly LSTM networks or Gated Recurrent Units (GRUs), excel at handling sequential data and capturing temporal dependencies within the data. A football match is a highly dynamic process, and key events in the game often rely on temporal relationships (for example, an event occurring at one point directly affects the subsequent course of the match). The time-series modeling capability of RNNs enables effective tracking of these temporal features, thereby identifying important events in the match (such as goals, fouls, etc.). By combining CNNs and RNNs, both spatial and temporal features can be leveraged, significantly improving the accuracy of event prediction in football matches. This work uses CNNs to extract spatial features from each video frame. In contrast, RNNs process the temporal information within the video frame sequence, enabling the model to capture the dynamic spatiotemporal changes in the match. This synergistic approach allows the CNN-RNN combination to predict vital events in football matches more effectively than using CNNs or RNNs alone. Figure 1 illustrates the overall framework of the CNN-RNN model. This framework visually represents the process where video data flows through the CNN for spatial feature extraction and is subsequently passed into the RNN for time-series modeling.

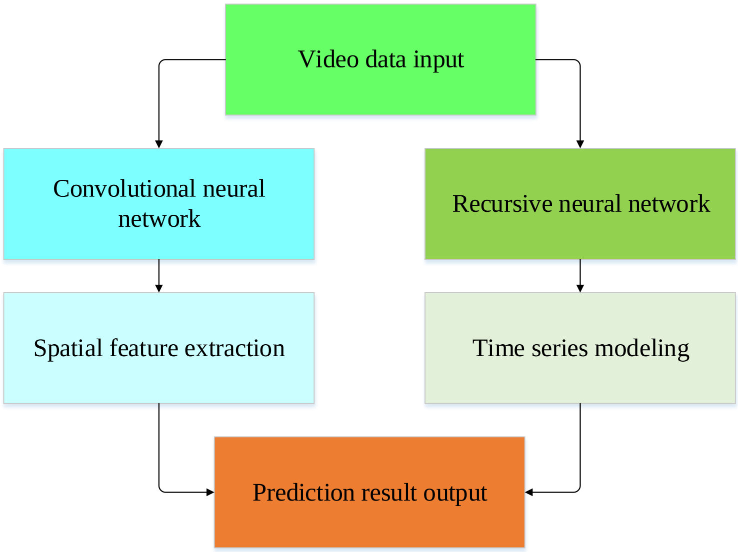

Fig. 1. Workflow of the CNN-RNN model

Figure 1 shows that, first, video data is processed through the CNN to extract spatial features (such as player positions, actions, and event labels). Second, the extracted spatial features are fed into the RNN for time-series analysis, ultimately outputting predictions for match events. The core of this work lies in utilizing advanced DL algorithms to precisely analyze and enhance the effectiveness of football training. To achieve this goal, the following research methods are employed:

First, a detailed data preparation process is conducted. This involves collecting training videos of college football players, match recordings, and relevant motion data. These data are used to train and test the DL model, ensuring that the model covers various aspects of football movements.

Next is the design and implementation of the DL model. This work leverages the strengths of CNNs and RNNs to create a specialized DL model for optimizing football training. On the one hand, CNN excels in handling image and video data, making it suitable for identifying and processing visual elements in football movements. On the other hand, RNN is proficient in dealing with sequential data and capturing dynamic information in time series. The combination of these two networks is expected to analyze the complex dynamics of football training more accurately. Firstly, CNN is used to extract features from each video frame. This typically involves dividing the video into multiple frames and feeding them into the CNN to extract a feature representation for each frame. Secondly, the sequence of features extracted by the CNN is fed into an RNN. The RNN processes these features sequentially and captures the temporal dependencies between them. This allows the model to understand dynamic changes in the video, including player movements and fluctuations in the pace of games. Finally, predictions are made based on the RNN's output. These predictions might include classification (such as determining the match outcome), regression (such as predicting the timing of goals), or generative tasks (such as generating video descriptions). If temporal information is mentioned in image processing, it is likely because a series of images (video frames) are processed, rather than a single static image. In this context, temporal information refers to the sequence of images and their changes over time.

A combination of CNN and RNN architectures is employed to address the research problem in this work. Specifically, in the CNN section, three convolutional layers and two pooling layers are designed. The first convolutional layer uses a 3x3 kernel with a stride of 1, and the depth of the input feature map is 3 (corresponding to color images). This layer aims to extract low-level features such as edges and textures. The Rectified Linear Unit (ReLU) activation function is employed in this layer, as it effectively mitigates the vanishing gradient problem and accelerates network training. The second convolutional layer also utilizes a 3x3 kernel with a stride of 1, and the output depth is 64, primarily responsible for extracting higher-level features. The third convolutional layer retains a 3x3 kernel and a stride of 1, with an output depth of 128, further extracting complex spatial features. To reduce the dimensionality of the feature maps and lower computational complexity, max pooling with a 2x2 window and a stride of 2 is introduced between the second and third layers to preserve important spatial information. In the RNN section, the spatial features extracted by the convolutional layers are passed to the RNN for time-series data processing. Two LSTM layers, each containing 128 hidden units, are chosen. LSTM is well-suited for handling long-term dependencies in sequences, making it ideal for processing the continuous dynamic data in football matches. The output from the CNN layers is flattened and serves as input to the RNN layers, whose output is then passed to the fully connected layers for final classification or regression predictions. A Tanh activation function is adopted at the output of the LSTM units, helping to constrain the output values within a specific range and prevent extreme values, thereby ensuring stable model training. ReLU is used as the activation function for the convolutional and fully connected layers, while Tanh is used for the LSTM layers. ReLU is widely applied due to its efficiency and its ability to avoid the vanishing gradient problem, while Tanh is effective for handling time-series data and ensuring stable gradient propagation.

The core part of the work involves the model's performance evaluation. This work comprehensively assesses the performance of both Three-Dimensional Convolutional Neural Network (3D CNN) and CNN-RNN models in football training optimization based on key indicators such as accuracy, precision, recall, F1 score, training loss, validation loss, and processing time. These indicators can help to systematically compare and analyze the two model's efficiency and accuracy in handling football training data.

Additionally, attention is given to the model's generalization ability by conducting cross-validation on different datasets, ensuring that the model remains efficient and accurate when faced with new data. In future research, there are plans to enhance the model's practicality and generalization ability by expanding the dataset, incorporating more sensor data, and testing the model in real training scenarios. It is believed that through these efforts, this work can bring innovative technological support to the field of football training, aiding coaches and athletes in achieving more scientific and efficient training methods.

This work utilizes multi-modal data from various sources, including video, Global Positioning System (GPS), and heart rate data, which provide different levels of information regarding player movement, position, and physiological states. To effectively utilize these multi-modal data, a fusion strategy is designed to combine the information from different data sources, ensuring that each modality contributes positively to the model's final predictions. For video data, the CNN is employed for feature extraction. The CNN model effectively extracts spatiotemporal features from video frames, capturing the players' actions, positions, and dynamic changes during the match. Specifically, a 3D CNN is used to process the video data to extract continuous information along the temporal dimension and combine it with spatial features to capture player movement patterns. The feature extraction for GPS data is achieved through an RNN based on location sequences. Since GPS data is sequential, RNNs are particularly suitable for processing such dynamic sequential information. Each player's location sequence is used as input. The RNN architecture (such as LSTM or GRU) captures the player's movement trajectory in the match, as well as position changes at each time step. Heart rate data is another important physiological feature, as it reflects the player's physical load and fatigue state by monitoring variations in heart rate. When processing heart rate data, standard time-series processing methods are applied to extract key features related to heart rate fluctuations, such as amplitude and extreme values. After feature extraction, a feature fusion approach is employed to combine features from video, GPS, and heart rate data. Two main fusion strategies are explored: feature concatenation and weighted fusion based on the attention mechanism. In feature concatenation, the features from each modality are directly concatenated into a unified feature vector. After concatenation, the merged data is input into subsequent neural network layers for further processing. This method is simple and effective, allowing all modality information to be passed to the model at once. An attention mechanism is introduced for weighted fusion to better handle the varying importance of different modalities. In this strategy, the model adaptively adjusts the weight of each modality's features based on their contribution to the final prediction. For instance, video data might be more critical in some situations, while heart rate data may have a higher influence in others. By utilizing the attention mechanism, the model can dynamically focus on more relevant information, thereby improving prediction accuracy. In experiments, the weighted fusion method based on attention mechanism outperforms the simple feature concatenation method, particularly in terms of accuracy and recall. This indicates that the weighted strategy allows the model to more effectively utilize the complementary information from each modality, thus enhancing overall prediction ability. Through the design of feature extraction and fusion strategies, various types of information can be integrated to maximize the complementarity of different data sources. The model's ability is improved to handle complex football training scenarios and enhance prediction accuracy. This process demonstrates the significant role of multi-modal data fusion in intelligent optimization techniques, providing reliable technical support for future model optimization and practical applications.

The main procedures of the work typically is shown in Figure 2. Figure 2 include the following steps. (1) Requirement analysis and definition: It defines the goals, scope, and requirements of the work, and identifies the problems to be solved or objectives to be achieved. (2) Data collection: Relevant data are collected based on the work requirements. Football data analysis may include indicators such as goals scored, assists, and passing accuracy. (3) Data preprocessing: The collected data are cleaned, organized, and transformed to ensure data quality and usability. (4) Model construction: Appropriate algorithms or models are selected based on the work requirements and the model is trained using the preprocessed data. (5) Model evaluation and optimization: The trained model is evaluated, optimized, and adjusted based on the evaluation results. (6) Result output and application: The model's results are output, applying them as needed in practical work, such as player selection and tactical formulation.

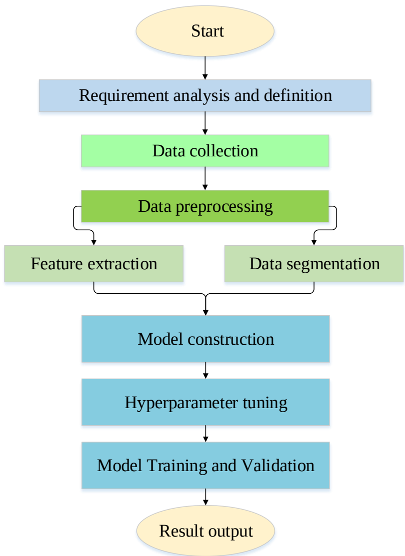

Fig. 2. The overall process of football training optimization method based on deep learning

In football data analysis, input data encompasses players' basic information (e.g., age, height, and weight), technical statistics (e.g., goals scored, assists, and passing accuracy), and match videos. Output results include player ratings, team strength rankings, and match outcome predictions. These results are presented in the form of reports, charts, and visual interfaces to relevant stakeholders such as coaches, team management, and fans. The overall pseudo-code of the football training optimization method based on deep learning is shown in Figure 3.

```
import TensorFlow

data_sources = ["data.world", "GitHub"]
datasets = load_datasets(data_sources)

preprocessed_data = preprocess(datasets)

features, labels = extract_features(preprocessed_data)

train_set, validation_set, test_set = split_data(features, labels)

model = CNN_RNN_Model()

learning_rates = [0.0001, 0.001, 0.01]
batch_sizes = [32, 64, 128]
optimizers = ["Adam", "SGD"]

for lr in learning_rates:
    for batch_size in batch_sizes:
        for optimizer in optimizers:
            model = build_model(lr, optimizer)
            train(model, train_set, batch_size)
            performance = validate(model, validation_set)
            save_best_model(performance)

best_model = load_best_model()
final_training(best_model, train_set, batch_size=64)

results = evaluate(best_model, test_set)
compare_with_others(results)

balanced_data = balance_data(train_set)
retrain_model(balanced_data, best_model)

final_results = evaluate(best_model, test_set)
```

Fig. 3. The whole pseudo-code of football training optimization method based on deep learning

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

This subsection collects and organizes football-related datasets suitable for DL models [35-37]. Two primary data resources are utilized to ensure the comprehensiveness and depth of the research. Table 1 provides detailed information on the specific datasets:

The data.world dataset contains results and team statistics for 5,000 football matches, with a total size of approximately 10 GB. Each match is recorded with detailed statistics, encompassing goals scored, shots taken, and possession percentage, covering 50 statistical fields. The GitHub dataset provides 200 high-definition football match videos, with a total duration of 100 hours. Each video file contains an average of 2,500 frames, offering rich visual input for DL models. Additionally, 80 detailed player statistics, such as speed, acceleration, and passing accuracy, are provided, which offer extra contextual information for the models.

Table 1. Experimental dataset information used

Data Source	data.world	GitHub (Edd Webster)
Dataset Name	Football match and statistics dataset	Football analysis project collection
Data Types	Match outcomes, team statistics	Match videos, player statistics data
Data Description	It includes match outcomes, team, and player statistical data	It encompasses various types of football data suitable for training and testing DL models

The football dataset, provided by data.world, primarily includes match outcomes and team statistics. The comprehensiveness and versatility of this data make it an ideal source for training and testing, covering various aspects of football training and matches [38-42]. The football analysis project collection on GitHub contains more diverse data, encompassing match videos, and player statistics. These data enrich the dataset types and offer more practical application scenarios for DL models [43-45]. Combining these two data sources allows for comprehensive coverage of various aspects of football training and matches, providing a solid data foundation for training and evaluating the CNN-RNN model.

This work selects two primary data sources: player statistics and match videos. Player statistics (such as speed, acceleration, passing accuracy, number of shots, etc.) provide quantitative information about the player's performance during the match, which is crucial for predicting key events (such as goals, fouls, etc.). Additionally, match videos offer rich spatial information, helping the model identify player positions, movement trajectories, and event timings. By combining these data sources, the model can process spatial and temporal features, enhancing its ability to predict critical match events.

4.2. Experimental Environment

The experimental environment is established on a high-performance computing platform, specifically configured with servers equipped with NVIDIA GPUs to ensure the efficient operation of DL models. The operating system selected is a widely-used Linux distribution, favored in research for its stability and strong support for DL libraries. Both TensorFlow and PyTorch are utilized for DL framework selection, as they are commonly used frameworks in current DL research, providing rich libraries and optimization tools for constructing and training complex models. To ensure standardization and consistency in the experimental environment, Docker container technology is employed. Leveraging Docker ensures consistent environment configuration across different experimental stages, avoiding experimental biases caused by environmental differences [46, 47]. Additionally, all experiments are conducted in a network-isolated environment to prevent external interference. Regarding hardware resources, sufficient RAM and a high-speed storage system are provisioned to support the processing of large-scale datasets and model training. Furthermore, the server is equipped with a high-speed internet connection, ensuring rapid downloading of the necessary datasets and library files.

4.3. Parameters Setting

The work trains and tests CNN and CNN-RNN models under both TensorFlow and PyTorch DL frameworks. The GPU model used is the NVIDIA Tesla V100, designed specifically for DL and high-performance computing, with outstanding parallel processing capabilities. To optimize model performance, the initial learning rate is set to 0.001, with a step decay strategy to avoid converging to a local minimum. The batch size is set to 64, which effectively utilizes GPU resources while preventing memory overflow. The optimizer used is Adaptive Moment Estimation (Adam), as it adapts the learning rate to handle sparse gradients and has shown good performance in DL tasks. For hyperparameter selection, a grid search method is employed, where different combinations of parameters are trained and evaluated on a validation set to determine the optimal hyperparameter configuration. In the experiments, DL frameworks (TensorFlow and PyTorch) are used, with training conducted on a high-performance computing platform. During each training iteration, an independent validation set is used to monitor the model's performance and ensure that the model does not overfit. Therefore, this work designs a composite loss function that combines cross-entropy loss (for classifying match outcomes) and Huber loss (for predicting key events).

A composite loss function is generally represented as a weighted sum of multiple loss functions:

$$L_{total} = \alpha \cdot L_1 + \beta \cdot L_2 + \ldots + \gamma \cdot L_n \tag{1}$$

L_1 , L_2 …, and L_n are different loss functions. α , β …, and γ represent their corresponding weights, used to balance the impact of each loss function on the overall loss.

The input data considered for training and testing DL models is typically preprocessed and feature-extracted raw data, such as images, text, and audio. Here, the input data may consist of image datasets used for recognizing and classifying different objects or scenes. The output data represents the model's prediction or classification results for the input data, such as labels or probability distributions for various categories in the case of an image recognition model. The mentioned CNN and CNN-RNN models in the training and testing process receive such input data and generate corresponding output results. The model parameters are iteratively adjusted to enhance the accuracy of the predicted outputs.

The term "trend of accuracy" refers to the degree of alignment between the model's predictions or classification results and the true labels of the input data. This accuracy is closely related to the model's performance and effectiveness, reflecting its understanding and representation capabilities of the input data. In DL methods, improving accuracy is crucial as it directly impacts the practical application of the model in real-world scenarios. For example, in image recognition tasks, a model with high accuracy can reliably identify different objects, providing robust support for areas such as autonomous driving and medical image diagnosis. DL methods aim to train the model to learn the underlying representations and patterns of the data, enabling accurate predictions and efficient processing of new data. Considering the F1 score as an evaluation indicator means taking both precision and recall into account. This comprehensive indicator provides a more nuanced assessment of the model's performance, making it particularly valuable for datasets with imbalanced class distributions. When training and testing DL models, considering overall indicators such as input data features, model accuracy, and F1 score can provide better guidance for model optimization and practical deployment. Table 2 presents the configuration for the performance evaluation parameters of the DL-based training models.

Table 2. Specific parameter settings for improved CNN

Parameter	TensorFlow Setting	PyTorch Setting	Description
Framework Version	TensorFlow 2.x	PyTorch 1.x	Specifies the version of the DL framework used for model training and testing
GPU Model	NVIDIA Tesla V100	NVIDIA Tesla V100	Uses high-performance GPU model to ensure model training efficiency
Batch Size	64	64	Number of data samples used for each training iteration
Initial Learning Rate	0.001	0.001	Learning rate at the beginning of model training

Hyperparameter tuning is employed to optimize the performance of the DL model. In the grid search, the following ranges for hyperparameters are considered: learning rate (0.0001, 0.001, 0.01), batch size (32, 64, 128), and optimizer selection (Adam, Stochastic Gradient Descent (SGD)). Through multiple experiments, the best combination of these hyperparameters is explored to ensure that the model converges at the fastest rate while achieving optimal prediction accuracy. The selection of these hyperparameter ranges is based on literature and preliminary experimental experience, aiming to avoid overfitting and enhance the model's generalization ability through reasonable adjustment.

The specific process of hyperparameter tuning is as follows:

Learning rate adjustment: Three different learning rates—0.0001, 0.001, and 0.01—are tested. Experiments reveal that a learning rate of 0.001 allows the model to achieve good performance in a relatively short time while preventing instability during training.

Batch size adjustment: Three batch sizes (32, 64, and 128) are evaluated. It can be found that a batch size of 64 results in the fastest training speed and highest accuracy, leading to the final selection of 64 as the batch size.

Optimizer selection: After comparing Adam and SGD optimizers, Adam performs better in this work, particularly due to its ability to quickly adjust the learning rate and achieve lower loss values during training.

The experimental results demonstrate that reasonable tuning of these hyperparameters significantly improves the model's performance. Specifically, the choice of learning rate directly influences the model's convergence speed, the batch size affects training stability, and the optimizer selection impacts the model's final accuracy. Through these adjustments, the model achieves significant improvements in prediction accuracy, recall, and F1 score.

4.4. Performance Evaluation

In the model evaluation phase, the CNN-RNN model is first trained on the training set. During training, the model learns how to map input data (image sequences) to output labels (football action categories). Grid search explores every possible combination of hyperparameters and trains a model for each combination. Specifically, if there are N hyperparameters and each hyperparameter has M candidate values, then a total of MN models need to be trained. During training, an independent validation set is used to assess model performance. The validation set does not participate in model training. By running the model on the validation set, indicators such as accuracy, recall, and F1 score can be computed. After training is complete, the final performance evaluation is typically conducted on the test set. The test set is another independent dataset that does not participate in model training or validation. The indicators calculated on the test set are considered the final evaluation of the model's performance.

Figure 4 shows the accuracy changes of the 3D CNN model as proposed in reference [48] and the CNN-RNN model at different iteration counts. Figure 5 depicts the accuracy differences between the 3D CNN and CNN-RNN models during iterative training.

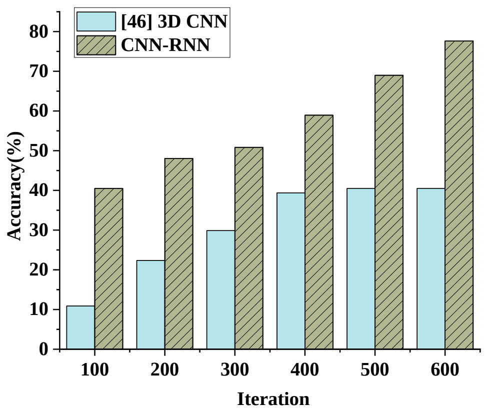

Fig. 4. The trend of accuracy performance changes for different DL models in optimizing college football training

Figure 4 illustrates the improvement trends in accuracy during the training process for both the 3D CNN and CNN-RNN models. As the number of iterations increases, the accuracy of both models improves, demonstrating the effectiveness of the learning process. It can be observed that the CNN-RNN model consistently exhibits higher accuracy than the 3D CNN model for the majority of iteration counts. This suggests that integrating 3D CNN and RNN may be more effective in enhancing classification performance.

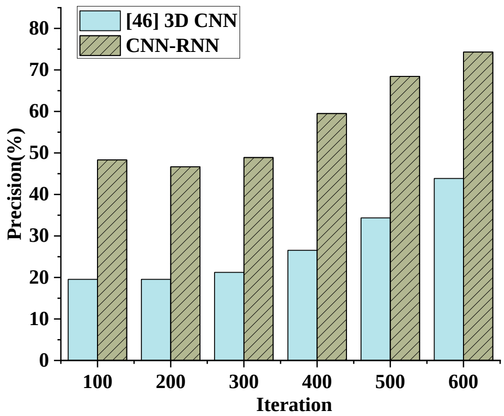

Fig. 5. The trend of precision performance changes for different DL models in optimizing college football training

Figure 5 illustrates how the precision of both models changes with an increase in the number of model training iterations. It can be observed that the precision of the CNN-RNN model is generally higher than that of the 3D CNN model. This highlights the potential advantage of the CNN-RNN model in recognizing positive samples. The improvement in precision means that the model is more accurate in predicting true positive samples among those predicted as positive, which is crucial for avoiding misclassifications. Figure 4 compares the recall performance of the 3D CNN and CNN-RNN models, providing insights into the recognition capabilities of both models for positive class samples. In Figures 4 and 5, the described "trend of precision" refers to the model's precision in classifying football actions during training. This indicator measures the model's ability to identify and classify actions correctly, reflecting the consistency between the model's predictions and the actual actions. The goal is to improve this precision to guide training and enhance athlete performance. As for the F1 score, it is the harmonic mean of precision and recall, commonly used to assess a model's performance in classification tasks. This work emphasizes the F1 score because it offers a balanced and comprehensive evaluation of precision and recall. This balance is essential to ensure that, in practical applications, the model neither overlooks important actions (high recall) nor misidentifies non-target actions (high precision).

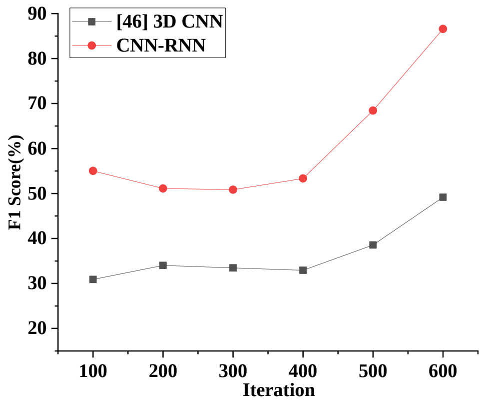

Fig. 6. The trend of F1 score performance changes for different DL models in optimizing college football training

Figure 6 demonstrates the variation in F1 scores for both models at different numbers of training iterations. This indicator combines information from precision and recall, providing a comprehensive performance evaluation. It shows that with the increase in iterations, the CNN-RNN model's F1 score shows an overall upward trend and surpasses that of the 3D CNN model. This indicates that while maintaining precision, the CNN-RNN model can better identify more positive class samples, making it superior in balancing accuracy and coverage.

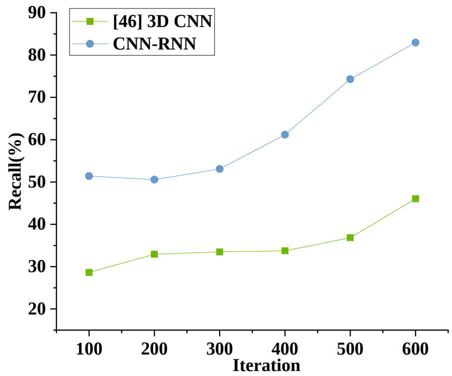

Fig. 7. The trend of recall performance changes for different DL models in optimizing college football training

The data in Figure 7 depicts the variation in recall for the 3D CNN and CNN-RNN models at different numbers of training iterations. As the number of iterations increases, the recall of the CNN-RNN model is significantly higher than that of the 3D CNN model. This indicates that the CNN-RNN model can more effectively identify all positive class samples in practical applications. This is particularly crucial for football training data analysis as it helps ensure that important events such as shots or passes are not overlooked. Figure 7 exhibits a comparison of the F1 scores for both 3D CNN and CNN-RNN models, providing an evaluation of their overall performance:

Figure 8 compares the reduction in loss rates for both the 3D CNN and CNN-RNN models during the training process:

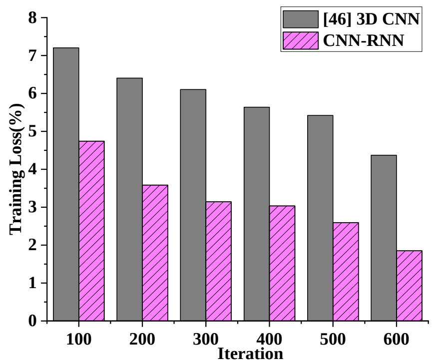

Fig. 8. The trend of training loss performance changes for different DL models in optimizing college football training

Figure 8 displays the training loss rates for the 3D CNN and CNN-RNN models across different numbers of training iterations. The decrease in training loss is a positive indicator during the model's learning process, indicating reduced errors and improved data fitting. In Figure 8, the training loss decreases for both models as the number of iterations increases. Moreover, the training loss for the CNN-RNN model is generally lower than that for the 3D CNN model, suggesting potentially better learning efficiency and superior data fitting capabilities.

In machine learning, the loss rate is a crucial indicator for evaluating model performance. By minimizing the loss rate, the model gradually learns the underlying patterns in the data, thereby improving prediction accuracy. Depending on the task, different loss functions may be employed, such as cross-entropy loss for classification tasks and mean squared error loss for regression tasks. The primary role of cross-validation is to assess the model's generalization ability. Generalization refers to the model's ability to perform well on unseen data. Cross-validation provides a more comprehensive understanding of the model's performance across different subsets of data, thus offering a more accurate measure of its generalization ability. The reason for stopping training is typically when the model's performance on the validation set reaches a stable state, meaning that further training no longer yields significant performance improvements. This indicates that the model has sufficiently learned the useful information from the training data. Figure 9 illustrates the variation in loss rates for both

3D CNN and CNN-RNN models during the validation process, serving as an evaluation of the models' generalization abilities:

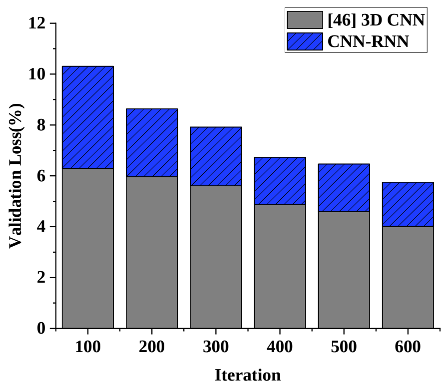

Fig. 9. The trend of performance changes in validation training loss of different DL models

Figure 9 compares the validation loss for the 3D CNN and CNN-RNN models across different training iterations. Validation loss is a crucial indicator for assessing a model's generalization ability, with lower validation loss suggesting better performance on unseen data. In Figure 9, the validation loss for the CNN-RNN model is consistently lower than that for the 3D CNN model at most iteration points. This indicates that the CNN-RNN model may have superior generalization performance, making it more effective at handling new, unknown data. Figure 10 compares the processing time required by the CNN and CNN-RNN models to complete the same task, reflecting differences in model efficiency:

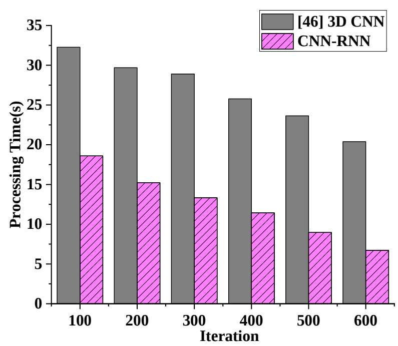

Fig. 10. The trend of processing time performance changes for different DL models in college football training

Figure 10 suggests the processing time required by the 3D CNN and CNN-RNN models to complete the task at different numbers of training iterations. Processing time is an intuitive indicator for measuring model efficiency, with shorter processing time reflecting faster training and prediction abilities. In Figure 10, the CNN-RNN model initially requires more processing time in the early iterations. However, as the iteration progresses, its processing time gradually decreases, and it exhibits comparable or even better processing speed in the later iterations compared to the 3D CNN model. This may be attributed to the CNN-RNN model becoming more efficient in handling sequential data after sufficient training, enhancing its overall computational efficiency.

To further validate the superiority of the CNN-RNN architecture in this work, comparative experiments are conducted with other advanced models, such as Transformer and LSTM variants, using the football training dataset. Table 3 presents the comparison results between CNN-RNN and other models, including Transformer, LSTM, Bi-LSTM, and GRU, in terms of key performance indicators:

Table 3. Comparison results of the CNN-RNN model with Transformer, LSTM, Bi LSTM, GRU, and other models in key indicators

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Training loss
CNN-RNN	92.5	91.2	93.1	92.1	0.24
Transformer	90.8	89.5	91.2	90.3	0.28
LSTM	91.2	90.1	91.8	90.9	0.26
Bi-LSTM	91.8	90.4	92.0	91.2	0.25
GRU	90.4	89.0	90.5	89.7	0.30

From the experimental results in Table 3, it can be observed that the CNN-RNN model achieves the highest accuracy, reaching 92.5%. In comparison, the accuracies of the Transformer, LSTM, and Bi-LSTM models are 90.8%, 91.2%, and 91.8%, respectively. The GRU model shows a slightly lower accuracy of 90.4%. This indicates that the CNN-RNN architecture exhibits better classification ability in identifying key actions and events in the football training dataset. The CNN-RNN model also achieves a slightly higher precision, reaching 91.2%, outperforming other models, particularly GRU and Transformer. The high precision suggests that the CNN-RNN model is more effective at reducing misclassifications. The recall of the CNN-RNN model is 93.1%, higher than all other models, demonstrating its ability to comprehensively identify all positive class samples. This is especially important for identifying key events in football training, such as goals and passes. The F1 score of the CNN-RNN model is 92.1%, the highest among all models. The F1 score takes both precision and recall into account, highlighting the CNN-RNN model's advantage in balancing precision and coverage. Regarding training loss, the CNN-RNN model exhibits the lowest loss at 0.24, indicating that it fits the data well during the training process and converges more quickly. Through the comparison with Transformer, LSTM variants, and GRU models, it is evident that the CNN-RNN model performs excellently in this work. Particularly, in handling sequential data and football training data, it better identifies key actions and improves training efficiency. Although Transformer and LSTM models also perform well, their performance is slightly inferior to that of the CNN-RNN model, especially in terms of precision and recall. The advantage of the CNN-RNN model may lie in its convolutional layers, which effectively extract spatial features, while the RNN layers excel in capturing temporal dependencies. These experimental results demonstrate that the CNN-RNN architecture has significant potential for optimization and analysis tasks in football training data.

To improve the proposed model's generalization ability and ensure its applicability to diverse training scenarios, the dataset source is expanded by adding more football match data from university-level competitions. The original dataset (from data.world and GitHub) primarily consists of data from high-level competitions, which, although valuable for model learning, may not fully represent the diversity of university-level football training. Therefore, additional football training data from multiple universities are incorporated, covering matches of varying levels and styles to enhance the diversity and representativeness of the dataset. This data includes basic technical movements, tactical coordination, and players' physiological conditions during training, making it more aligned with actual training scenarios. The comparison of model accuracy, recall, and F1 score across different categories (before and after applying data balancing strategies) is presented in Table 4:

Table 4. The comparison of model accuracy, recall, and F1 score across various categories (before and after applying data balancing strategies)

Category	Method	Accuracy	Recall	F1 score
Goal	Before data balancing	0.92	0.89	0.90
	After data balancing	0.93	0.90	0.91
Foul	Before data balancing	0.79	0.65	0.71
	After data balancing	0.81	0.72	0.76
Yellow card	Before data balancing	0.74	0.60	0.66
	After data balancing	0.80	0.71	0.75
Red card	Before data balancing	0.85	0.80	0.82
	After data balancing	0.87	0.84	0.85

Table 4 reveals that for the goal category, the data balancing strategy does not lead to significant changes, as the goal category samples are already relatively balanced. Moreover, the model's precision and F1 score in this category are already at a high level. For the foul category, after applying the data balancing strategy, the model's F1 score and recall show significant improvements. Through oversampling and undersampling strategies, the model can better identify foul events, with recall increasing from 0.65 before data balancing to 0.72, and F1 score rising from 0.71 to 0.76. This indicates that the model's prediction accuracy for the minority class of foul events has improved. In the yellow card category, after applying a weighted loss function and data balancing, the model's performance exhibits a notable improvement. Precision increases from 0.74 to 0.80, recall rises from 0.60 to 0.71, and F1 score improves from 0.66 to 0.75. This illustrates that the model is now able to more accurately identify yellow card events, reduce false positives, and improve coverage for yellow card events. For the red card category, although the sample distribution is relatively balanced, the weighted loss function still improves in recall, increasing from 0.80 to 0.84, and F1 score rising from 0.82 to 0.85. This further demonstrates that, in scenarios with class imbalance, the model's overall performance becomes more robust through data balancing and the optimization of the weighted loss function. The comparison of training and validation losses for different categories (before and after applying data balancing strategies) is listed in Table 5:

Table 5. Comparison of training and validation losses for different categories (before and after applying data balancing strategies)

Category	Method	Training loss	Validation loss
Goal	Before data balancing	0.17	0.18
	After data balancing	0.16	0.17
Foul	Before data balancing	0.32	0.35
	After data balancing	0.30	0.32
Yellow card	Before data balancing	0.38	0.42
	After data balancing	0.34	0.36
Red card	Before data balancing	0.22	0.24
	After data balancing	0.21	0.22

Table 5 indicates that for the goal category, the change in both training loss and validation loss is small, as the category is already relatively balanced in the dataset. In addition, the model's performance in this category is already satisfactory. For the foul category, after applying the data balancing strategy, training loss decreases from 0.32 to 0.30, and validation loss drops from 0.35 to 0.32. These results indicate that the data balancing technique helps the model better fit the minority class samples, reduces overfitting, and improves the model's generalization ability. For the yellow card category, after applying the weighted loss function and data balancing strategies, training loss decreases from 0.38 to 0.34, and validation loss drops from 0.42 to 0.36. It suggests that balancing the data and optimizing the loss function make the learning process more stable, effectively improving performance in yellow card prediction. For the red card category, although the sample distribution is relatively balanced, the data balancing strategy still slightly reduces both training and validation losses. This further confirms the universality and effectiveness of data-balancing methods across different categories.

In short, by applying data augmentation and balancing strategies (such as oversampling, undersampling, and weighted loss functions), significant improvements are made in the model's performance for minority classes (such as yellow cards, fouls, etc.). These improvements not only enhance precision, recall, and F1 score but also effectively reduce both training and validation losses. It indicats that the model can better handle class imbalance during training and achieve more reliable prediction results in practical applications.

4.5. Discussion

The CNN-RNN model proposed in this work demonstrates superior performance on football training data. However, compared to some advanced methods in existing literature, why is the proposed model able to provide better prediction accuracy and event detection capabilities? To this end, Agyeman et al. (2019) introduced a DL-based ball-tracking system that utilizes advanced image recognition technology for tracking the ball. This approach was successful in analyzing the ball's movement trajectory in football videos [49]. However, it primarily focused on ball tracking and did not delve into real-time event prediction (such as goals, fouls, yellow cards, etc.), nor did it incorporate player statistics. Therefore, while their method excelled in specific tasks such as ball tracking, it did not align directly with the event prediction and multimodal data fusion tasks addressed by the proposed CNN-RNN model, making a direct comparison unnecessary. Moreover, Giancola et al. (2018) proposed the "Pass2vec" model to analyze players' passing styles [50]. This study used DL technologies to model passing behaviors and employed embedding methods like Word2Vec to represent passing data. While this model provides valuable insights into player behavior, its focus is limited to the analysis of passing behavior. This differs from the event prediction tasks (such as goals, fouls, yellow cards, etc.) explored in this work. Hence, despite the use of DL in Giancola et al.'s study, its objectives and tasks do not fully overlap with those of this work, and it was not included in the direct comparison. Jiang et al. (2020) provided a review of DL applications in football video analysis, discussing various technologies and their potential to enhance match analysis quality [51]. Although this review mentioned several methods, it primarily focused on technical overviews and potential discussions rather than specific model comparisons. Thus, when compared with Jiang et al.'s work, this work offers a more in-depth experimental comparison. Meanwhile, the proposed CNN-RNN model demonstrates its advantages in prediction accuracy, event recognition, and handling sequential data through comparisons with existing advanced models. Cioppa et al. (2020) applied DL methods to predict and classify foul behavior in football matches [52]. This study focused on identifying fouls to improve referee decision accuracy. Cioppa et al.'s research concentrated solely on the recognition of fouls. In contrast, this work includes not only fouls but also the prediction of various key events such as goals, yellow cards, and red cards. Although no direct comparison with Cioppa et al.'s study is made, subsequent experiments expand on the prediction of minority events (such as fouls and yellow cards) using data-balancing strategies. The model's performance in these events shows significant improvements, which were not addressed in Cioppa et al.'s method.

Recent contrastive learning methods utilize self-supervised learning to learn feature representations without requiring large amounts of labeled data, and these methods have also been applied in football analysis [53, 54]. By pulling similar samples closer together and pushing dissimilar samples further apart, these methods help improve the model's understanding of complex scenarios. However, while contrastive learning methods perform excellently in certain tasks, they typically rely on large amounts of unlabeled data for training and are better suited for feature learning rather than event prediction. This work utilizes spatiotemporal features combined with multimodal data (such as video and player statistics) for real-time event prediction. Therefore, while contrastive learning has its advantages in feature learning, this work prioritizes real-time event prediction tasks, which are not entirely aligned with contrastive learning. To further validate the advantages of the CNN-RNN model, comparisons are made with other models (such as 3D CNN, Transformer, LSTM, etc.) on key indicators. Table 3 presents the model's performance in terms of accuracy, precision, recall, and F1 score. The experimental results show that the CNN-RNN model exhibits high prediction accuracy, particularly in detecting minority events (such as fouls and yellow cards), where it demonstrates remarkable advantages. During the evaluation, multiple datasets are used, including data collected from university-level football matches, which cover games of varying levels and styles, ensuring the diversity and representativeness of the training data. When compared with standard datasets used in other studies (such as SoccerNet), a better understanding of the model's performance in complex scenarios can be obtained. In conclusion, the unique advantage of the CNN-RNN model lies in its ability to extract spatial and temporal features efficiently. Through data balancing and weighted loss function optimization, it can significantly improve the prediction of minority events. Its application potential, especially in football training, is enormous.

5. Conclusion

5.1. Research Contribution

The primary contribution of this work lies in the proposal of an innovative CNN-RNN hybrid architecture. This model combines CNNs and RNNs to effectively address the challenges of spatiotemporal feature extraction and time series modeling in football training data. Compared to traditional standalone CNN or RNN models, the CNN-RNN architecture demonstrates significant improvements in prediction accuracy and event detection. Additionally, this work creatively integrates video data, GPS data, and physiological data (such as heart rate) through multimodal fusion, and analyzes this data using the CNN-RNN model. This fusion approach enhances the model's accuracy and enables it to capture key events in football training more comprehensively, particularly under imbalanced data conditions, where it exhibits strong performance. Through comparisons with existing methods such as Transformer, LSTM, Bi-LSTM, and GRU, the proposed CNN-RNN model has been shown to have superior performance in indicators such as accuracy, recall, and F1 score. This demonstrates its stronger generalization ability in event recognition and prediction within football training data. Furthermore, this work expands the dataset to enhance the model's generalization ability by incorporating training data from different university-level competitions. Meanwhile, it optimizes the recognition of minority classes using data balancing strategies (such as oversampling, undersampling, and weighted loss functions). The experimental results reveal that these optimization strategies remarkably improve the model's ability to predict minority events such as yellow cards and fouls. Through these innovative methods and contributions, this work provides a new technical pathway for analyzing football training data and offers valuable insights for related research in DL applications.

5.2. Future Works and Research Limitations

The CNN-RNN model proposed in this work performs excellently on the football training dataset, accurately predicting key events such as goals, fouls, yellow cards, and red cards. By comparing it with other advanced models (such as Transformer, LSTM, etc.), it can be found that the CNN-RNN model has significant advantages in handling sequential data and multimodal data fusion, particularly excelling in prediction accuracy and recall. However, this work also has certain limitations, such as the fact that the dataset primarily comes from high-level competitions, which may not fully represent the diversity of college football training. Future research could address this limitation by expanding the dataset to include more data from football training sessions at different levels and styles, thereby enhancing the model's generalization ability. Additionally, incorporating advanced techniques such as reinforcement learning for adaptive training to improve model accuracy and robustness is a promising direction for future work. Regarding multimodal data fusion, future studies could explore more fusion methods, such as attention-based feature fusion or further optimization of video and physiological data integration, to enhance the model's prediction ability in complex scenarios. Given that the DL models in this work are primarily applied to the football domain, future research could extend to other sports, such as basketball, rugby, etc. Similar analytical methods are used to advance intelligent sports analysis technology. Overall, this work provides an effective technical pathway for football training and match analysis and opens up new directions for future research in DL-based sports data analysis. It is hoped that this work can serve as a reference and inspiration for the development of intelligent analysis technology and its practical applications in the sports field.

References

1. Zhang, J., Oh, Y. J., Lange, P., et al.: Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and A Healthy Diet. Journal of Medical Internet Research, Vol. 22, No. 9, e22845, (2020)
2. Londhe, V. Y., Bhasin, B.: Artificial Intelligence and Its Potential In Oncology. Drug Discovery Today, Vol. 24, No. 1, 228-232, (2019)
3. Guthrie, B., Jagim, A. R., Jones, M. T.: Ready or Not, Here I Come: A Scoping Review of Methods Used to Assess Player Readiness Via Indicators of Neuromuscular Function In Football Code Athletes. Strength and Conditioning Journal, Vol. 45, No. 1, 93-110, (2023)
4. Prots, R., Yakovliv, V., Medynskyi, S., et al.: Psychophysical Training of Young People for Homeland Defence Using Means of Physical Culture and Sports. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, Vol. 12, No. 3, 149-171, (2021)
5. Shen, C., Wang, C., Wei, X., et al.: Physical Metallurgy-Guided Machine Learning and Artificial Intelligent Design of Ultrahigh-Strength Stainless Steel. Acta Materialia, Vol. 179, 201-214, (2019)
6. Rigby, M. J.: Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA Journal of Ethics, Vol. 21, No. 2, 121-124, (2019)
7. Demchenko, I., Maksymchuk, B., Bilan, V., et al.: Training Future Physical Education Teachers for Professional Activities Under the Conditions of Inclusive Education. BRAIN. Broad Research in Artificial Intelligence and Neuroscience, Vol. 12, No. 3, 191-213, (2021)
8. Jan, Z., Ahamed, F., Mayer, W., et al.: Artificial Intelligence for Industry 4.0: Systematic Review of Applications, Challenges, And Opportunities. Expert Systems with Applications, Vol. 216, 119456, (2023)
9. Hwang, G. J., Chien, S. Y.: Definition, Roles, and Potential Research Issues of the Metaverse in Education: An Artificial Intelligence Perspective. Computers and Education: Artificial Intelligence, Vol. 3, 100082, (2022)

10. Lv, Z., Han, Y., Singh, A. K., et al.: Trustworthiness in Industrial IoT Systems Based on Artificial Intelligence. IEEE Transactions on Industrial Informatics, Vol. 17, No. 2, 1496-1504, (2020)
11. Shakya, D. S.: Analysis of Artificial Intelligence Based Image Classification Techniques. Journal of Innovative Image Processing, Vol. 2, No. 1, 44-54, (2020)
12. Ahuja, A. S.: The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician. PeerJ, Vol. 7, e7702, (2019)
13. Wang, J., Chen, Y., Hao, S., et al.: Deep Learning for Sensor-Based Activity Recognition: A Survey. Pattern Recognition Letters, Vol. 119, 3-11, (2019)
14. Wright, L. G., Onodera, T., Stein, M. M., et al.: Deep Physical Neural Networks Trained with Backpropagation. Nature, Vol. 601, No. 7894, 549-555, (2022)
15. Alexopoulos, K., Nikolakis, N., Chryssolouris, G.: Digital Twin-Driven Supervised Machine Learning for The Development of Artificial Intelligence Applications in Manufacturing. International Journal of Computer Integrated Manufacturing, Vol. 33, No. 5, 429-439, (2020)
16. Beavan, A., Spielmann, J., Mayer, J.: Taking The First Steps Toward Integrating Testing and Training Cognitive Abilities Within High-Performance Athletes; Insights from A Professional German Football Club. Frontiers in Psychology, Vol. 10, 2773, (2019)
17. Li, D., Zhang, J.: Computer Aided Teaching System Based on Artificial Intelligence in Football Teaching and Training. Mobile Information Systems, Vol. 2021, 1-10, (2021)
17. Borges, M., Rosado, A., Lobinger, B., et al.: Cultural Intelligence in Sport: An Examination of Football Coaches' Cross-Cultural Training Needs. German Journal of Exercise and Sport Research, Vol. 53, No. 3, 266-274, (2023)
19. Castro-Sánchez, M., Zurita-Ortega, F., Ubago-Jiménez, J. L., et al.: Relationships Between Anxiety, Emotional Intelligence, and Motivational Climate Among Adolescent Football Players. Sports, Vol. 7, No. 2, 34, (2019)
20. Rihadi, R. C., Hidayat, Y., Sidik, D. Z.: The Contribution of Spiritual Intelligence to The Self-Control and Respect Value of Adolescent Football Athletes. Halaman Olahraga Nusantara: Jurnal Ilmu Keolahragaan, Vol. 5, No. 2, 513-528, (2022)
21. Fózer-Selmeci, B., Kocsis, I. E., Kiss, Z., et al.: The Effects of Computerized Cognitive Training on Football Academy Players' Performance. Cognition, Brain, Behavior, Vol. 23, No. 3, 209-228, (2019)
22. Li, H.: The Tactical Mindset of Football Players: Choosing Effective Training Strategies for Top-Notch Performance. International Journal of Sport Psychology, Vol. 53, No. 1, 525-542, (2022)
23. Rahman, M. A.: A Deep Learning Framework for Football Match Prediction. SN Applied Sciences, Vol. 2, No. 2, 165, (2020)
24. Stoeve, M., Schuldhaus, D., Gamp, A., et al.: From The Laboratory to The Field: IMU-Based Shot and Pass Detection in Football Training and Game Scenarios Using Deep Learning. Sensors, Vol. 21, No. 9, 3071, (2021)
25. Fenil, E., Manogaran, G., Vivekananda, G. N., et al.: Real Time Violence Detection Framework for Football Stadium Comprising of Big Data Analysis and Deep Learning Through Bidirectional LSTM. Computer Networks, Vol. 151, 191-200, (2019)
26. Cuperman, R., Jansen, K. M. B., Ciszewski, M. G.: An End-To-End Deep Learning Pipeline for Football Activity Recognition Based on Wearable Acceleration Sensors. Sensors, Vol. 22, No. 4, 1347, (2022)
27. Wang, B., Shen, W., Chen, F. S., et al.: Football Match Intelligent Editing System Based on Deep Learning. KSII Transactions on Internet and Information Systems (TIIS), Vol. 13, No. 10, 5130-5143, (2019)
28. Sen, A., Deb, K., Dhar, P. K., et al.: Cricshotclassify: An Approach to Classifying Batting Shots from Cricket Videos Using A Convolutional Neural Network and Gated Recurrent Unit. Sensors, Vol. 21, No. 8, 2846, (2021)

29. Li, W.: Analyzing The Rotation Trajectory in Table Tennis Using Deep Learning. Soft Computing, Vol. 27, No. 17, 12769-12785, (2023)
30. Sanusi, I. T., Oyelere, S. S., Vartiainen, H., et al.: A Systematic Review of Teaching and Learning Machine Learning in K-12 Education. Education and Information Technologies, Vol. 28, No. 5, 5967-5997, (2023)
31. Lin, Y., Ma, J., Wang, Q., et al.: Applications of Machine Learning Techniques for Enhancing Nondestructive Food Quality and Safety Detection. Critical Reviews in Food Science and Nutrition, Vol. 63, No. 12, 1649-1669, (2023)
32. Li, X., Ullah, R.: An Image Classification Algorithm for Football Players' Activities Using Deep Neural Network. Soft Computing, Vol. 27, No. 24, 19317-19337, (2023)
33. Jin, G.: Player Target Tracking and Detection in Football Game Video Using Edge Computing and Deep Learning. The Journal of Supercomputing, Vol. 78, No. 7, 9475-9491, (2022)
34. Tuyls, K., Omidshafiei, S., Muller, P., et al.: Game Plan: What AI Can Do for Football, and What Football Can Do for AI. Journal of Artificial Intelligence Research, Vol. 71, 41-88, (2021)
35. Ur Rehman, A., Belhaouari, S. B., Kabir, M. A., et al.: On the Use of Deep Learning for Video Classification. Applied Sciences, Vol. 13, No. 3, 2007, (2023)
36. Rathi, K., Somani, P., Koul, A. V., et al.: Applications of Artificial Intelligence in the Game of Football: The Global Perspective. Researchers World, Vol. 11, No. 2, 18-29, (2020)
37. Zhan, X., Liu, Y., Raymond, S. J., et al.: Rapid Estimation of Entire Brain Strain Using Deep Learning Models. IEEE Transactions on Biomedical Engineering, Vol. 68, No. 11, 3424-3434, (2021)
38. Rago, V., Brito, J., Figueiredo, P., et al.: Methods to Collect and Interpret External Training Load Using Microtechnology Incorporating GPS in Professional Football: A Systematic Review. Research in Sports Medicine, Vol. 28, No. 3, 437-458, (2020)
39. Low, B., Coutinho, D., Gonçalves, B., et al.: A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. Sports Medicine, Vol. 50, 343-385, (2020)
40. Silva, H., Nakamura, F. Y., Beato, M., et al.: Acceleration and Deceleration Demands During Training Sessions in Football: A Systematic Review. Science and Medicine in Football, Vol. 7, No. 3, 198-213, (2023)
41. Chen, L., Hu, D.: An Effective Swimming Stroke Recognition System Utilizing Deep Learning Based on Inertial Measurement Units. Advanced Robotics, Vol. 37, No. 7, 467-479, (2023)
42. Teixeira, J. E., Forte, P., Ferraz, R., et al.: Monitoring Accumulated Training and Match Load In Football: A Systematic Review. International Journal of Environmental Research and Public Health, Vol. 18, No. 8, 3906, (2021)
43. Hu, X., Zhang, Y.: Application of Intelligent Football Training System Based on IoT Optical Imaging and Sensor Data Monitoring. Optical and Quantum Electronics, Vol. 56, No. 2, 150, (2024)
44. Gamble, P., Chia, L., Allen, S.: The Illogic of Being Data-Driven: Reasserting Control and Restoring Balance in Our Relationship with Data and Technology in Football. Science and Medicine in Football, Vol. 4, No. 4, 338-341, (2020)
45. Yang, B., Cheng, B., Liu, Y., et al.: Deep Learning-Enabled Block Scrambling Algorithm for Securing Telemedicine Data of Table Tennis Players. Neural Computing and Applications, Vol. 35, No. 20, 14667-14680, (2023)
46. Al-Asadi, M. A., Tasdemır, S.: Predict The Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. IEEE Access, Vol. 10, 22631-22645, (2022)
47. Malone, J. J., Barrett, S., Barnes, C., et al.: To Infinity and Beyond: The Use of GPS Devices Within the Football Codes. Science and Medicine in Football, Vol. 4, No. 1, 82-84, (2020)
48. Liu, N., Liu, L., Sun, Z.: Football Game Video Analysis Method with Deep Learning. Computational Intelligence and Neuroscience, Vol. 2022, No. 1, 3284156, (2022)

49. Agyeman, R., Muhammad, R., Choi, G. S.: Soccer Video Summarization Using Deep Learning. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 270-273, (2019)
50. Giancola, S., Amine, M., Dghaily, T., et al.: Soccernet: A Scalable Dataset for Action Spotting in Soccer Videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, (2018)
51. Jiang, Y., Cui, K., Chen, L., et al.: Soccerdb: A Large-Scale Database for Comprehensive Video Understanding. Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports, (2020)
52. Cioppa, A., Deliege, A., Giancola, S., et al.: A Context-Aware Loss Function for Action Spotting in Soccer Videos. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2020)
53. Baccouche, M., Mamalet, F., Wolf, C., et al.: Action Classification in Soccer Videos with Long Short-Term Memory Recurrent Neural Networks. International Conference on Artificial Neural Networks, Berlin, Heidelberg: Springer Berlin Heidelberg, (2010)
54. Sen, A., Kaushik, D.: Categorization of Actions in Soccer Videos Using A Combination of Transfer Learning and Gated Recurrent Unit. ICT Express, Vol. 8, No. 1, 65-71, (2022)

Kun Luan, born in Nanchang, China, studied at Jiangxi Normal University in 2003 and received a bachelor's degree in 2007. His specialties and interests include physical education and sports traini.

Fan Wu is a master of physical education in the Institute of physical education of PLA. He is also a teacher of physical education in the Institute of science and technology of Nanchang University. His research fields include pedagogy, physical education and sports training. He has published more than three papers and two books.

Yuanyuan Xu, from Nanchang, China, studied in Jiangxi Normal University in 2006 and got her bachelor's degree in 2010. From 2010 to 2013, she studied for a master's degree. Her special interests and interests are physical education and sports competition.

Received: November 21, 2024; Accepted: February 03, 2025.