WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2025 Volume 3, Issue 3
Lecture Notes in Education, Arts, Management and Social Science
ISSN(Online): 2705-053X

An Intelligent Physical Training System in Football Education

Shuo Wang[1], Zhiqiang Li[1], Hang Chen[2]
[1]Russian State-Run Sports University, Moscow 134400, Russia
[2]Westcliff University, Irvine, CA 92614, USA

Abstract: With the advancement of intelligent technology, data-driven evaluation methods have gained increasing attention in physical education, particularly in the application of intelligent physical training systems in football education. These systems enable precise assessment of athletes' training status and provide scientific support for personalized training, thereby enhancing training efficiency and game performance. This study employs Random Forest and Neural Network models to construct an intelligent evaluation system for predicting students' overall performance in football training. Key performance indicators such as passing frequency, sprint speed, and shooting accuracy are collected and analyzed to determine their impact on comprehensive scores. Experimental results demonstrate that the Random Forest model excels in stability and interpretability, while the Neural Network achieves higher prediction accuracy in complex pattern recognition. The combination of both models enhances generalization ability and applicability. Additionally, feature importance analysis identifies sprint speed and shooting accuracy as the most critical factors influencing training performance. This study proposes data-driven training optimization strategies to help students improve their football training performance. The findings confirm that intelligent physical training systems can effectively support football education, promoting the development of personalized and refined training programs and providing strong technological support for modern sports education.

Keywords: Machine learning; Football education; Sports; Data-driven coaching

Online publication: April 26, 2025

1. Introduction

In the field of physical education, effectively evaluating and improving students' overall athletic performance has always been a key research and practical challenge. Traditional evaluation methods primarily rely on teachers' subjective judgments and simple records of achievements, making it difficult to comprehensively and objectively reflect students' true abilities and growth potential [1]. With the rapid development of data science and artificial intelligence, the use of machine learning techniques for quantitative analysis of athletic performance has become a growing trend [2]. Machine learning models not only process multidimensional data but also uncover hidden patterns in features, supporting personalized teaching and training approaches. In football training, running, passing, and shooting are essential skills for measuring students' overall performance. In recent years, intelligent physical training systems have been increasingly applied in football education [3]. These systems utilize wearable devices, video analysis, and intelligent sensors to collect player training data and analyze it using machine learning techniques to optimize training programs. Such systems can not only monitor players' physical condition in real time but also provide personalized training feedback based on data-driven insights, thereby enhancing both teaching effectiveness and athletic performance. This study uses student football training data to build regression models with Random Forest and Neural Networks, aiming to analyze how key performance metrics—such as cumulative running distance, passing accuracy, and shot-on-target rate—affect overall scores [4]. Feature importance analysis identifies critical factors, while K-fold cross-validation ensures model reliability. The results highlight the strong influence of metrics like high-intensity running and passing accuracy. By adopting data-driven methods, the study evaluates training effectiveness and supports teaching optimization, offering a scientific basis for personalized football training and performance improvement through a stable, interpretable evaluation framework.

2. Literature review

Guan *et al.* [5] explored enhancing college football training efficiency via trajectory planning. They proposed an LSTM–GA model integrating deep learning and IoT–enabled wearable devices. LSTM compensates for GA's local search weakness. Experimental results showed the model converged at iteration 101, with 11% higher fitness than GA and 2% higher than LSTM alone. This optimizes trajectory planning for scoring, with contributions lying in improving the LSTM–GA model for IoT devices, thereby boosting training efficiency. Stoeve *et al.* [6] evaluated IMU-based shot and pass detection using deep learning models (CNN, LSTM, convLSTM) versus SVM in both lab and real-world football contexts. CNNs achieved a weighted F1-score of 0.93, outperforming SVM and demonstrating deep learning's effectiveness for real-time sports event classification. Zhou *et al.* [7] advanced intelligent football training through CNN-based action recognition systems and FCN-based field line detection models via football robots. Their dual-stream architecture reached 92.8% recognition accuracy, and intelligent training improved performance by up to 25.7%, offering significant implications for AI-driven sports environments. Cao *et al.* [8] examined functional strength training's impact on adolescents' football performance using a backpropagation neural network (BPNN) to assess kicking movements. The experimental group, which received targeted training, showed enhanced strength and sensitivity, validating machine learning's role in optimizing physical training strategies. Hollaus *et al.* [9] introduced an automated classification system for American football pass evaluations, leveraging a dataset of 2,276 attempts captured via synchronized audio and video. They employed a hybrid CNN-LSTM model for video and a 1D CNN for audio, achieving a 92.19% classification accuracy, confirming the potential of machine learning in analyzing complex, real-world training data. Collectively, these studies demonstrate the applicability of deep learning and machine learning across various domains of football training, including action recognition, biomechanical analysis, intelligent system design, and automated event classification. They highlight the importance of carefully selecting models and datasets to match practical scenarios, as well as the potential of AI to enhance training accuracy, real-time feedback, and overall athletic development. These findings provide strong empirical support for integrating AI into football education, training systems, and performance evaluation, both in traditional and intelligent training settings.

3. Data

In developing an intelligent football training evaluation system, this study emphasizes the importance of high-quality feature variables in enhancing model accuracy and supporting personalized training. Data were collected from junior and senior high school students through questionnaires, interviews, devices, and coach assessments. Key features include demographic and educational background, school type, sports experience, past training, peer participation, and training duration, all of which impact physical readiness and training adaptability.

Additionally, three core KPIs—passing frequency, sprint speed, and shooting accuracy—were automatically recorded to assess tactical activity, agility, and scoring efficiency. These KPIs directly inform the prediction of students' overall performance scores and support feature importance analysis, contributing to a more data-driven and individualized approach in football training.

4. Model analysis

In the intelligent football training evaluation system developed in this study, the Random Forest algorithm serves as a core predictive model for assessing students' overall training performance. As an ensemble learning method, Random Forest enhances robustness and generalization by combining the outputs of multiple decision trees through majority voting (for classification) or averaging (for regression).

The algorithm operates based on two main mechanisms [10]: Bootstrap Sampling: each decision tree is trained on a randomly sampled subset of the original dataset with replacement; Random Feature Subset Selection: at each decision node, the model selects the best split from a randomly chosen subset of features. This dual-randomization approach reduces variance and minimizes the risk of overfitting, making Random Forest particularly suitable for tasks involving medium-dimensional features and moderately sized datasets.

In this study, the Random Forest model takes as input several key performance indicators (KPIs), including passing frequency, sprint speed, and shooting accuracy, and predicts students' comprehensive performance scores in football training (**Figure 1**). During the training phase, multiple decision trees are iteratively constructed using the training dataset, and the final prediction is made by aggregating individual tree outputs, ensuring model robustness across varying input patterns. An important advantage of the Random Forest algorithm is its inherent feature importance evaluation. It ranks input variables based on metrics such as Gini impurity reduction or information gain at split points.

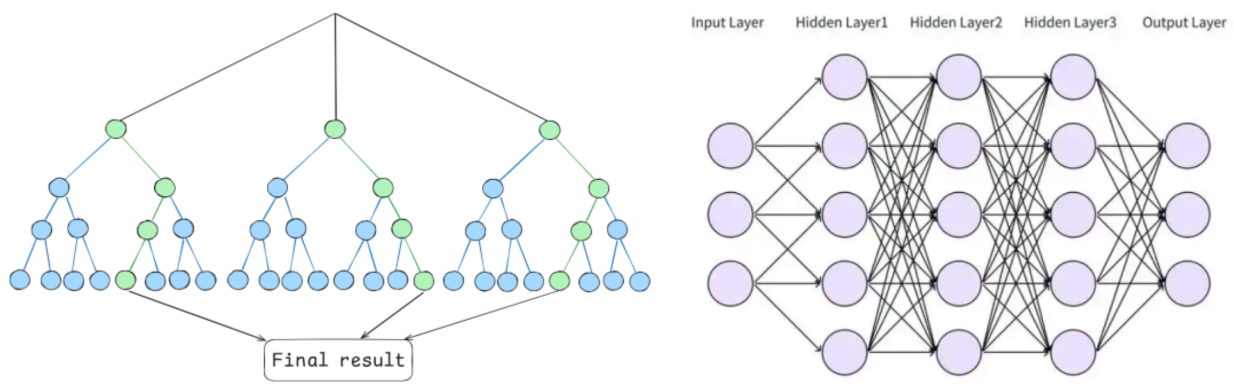

Figure 1. Random Forest and Neural Networks model

In terms of model tuning, this study adopts a lightweight approach that minimizes hyperparameter complexity to ensure robustness under small-sample conditions and enhance reproducibility. Furthermore, Random Forest demonstrates strong tolerance for noisy or missing data, making it well-suited for real-world educational environments where data quality may vary. However, it is worth noting that Random Forest may face efficiency challenges when dealing with ultra-high-dimensional or sparse features and may produce non-smooth outputs in regression tasks. These limitations should be considered when integrating the model into larger-scale intelligent training systems.

Inspired by biological neural systems, this study designs a neural network to predict students' soccer training performance by learning nonlinear relationships in complex data. The model, consisting of input, hidden, and output layers, uses weighted connections and nonlinear activation functions (e.g., ReLU [11], Sigmoid) to enhance pattern recognition and accurately map input features to training outcomes [12].

In this study, the model's inputs are key training metrics collected via a smart training system: passing frequency, sprint speed, and shooting accuracy. The output is the students' composite scores in periodic soccer training assessments, reflecting their overall performance.

The training process involves two phases: Forward propagation: Inputs are passed through the network layers to generate preliminary predictions. Backward propagation: The gradient of the loss function (e.g., mean squared error or cross-entropy) is computed to adjust network parameters, minimizing prediction errors. This study employed the

Adam [13] optimizer to accelerate convergence and incorporated Dropout [14] (rate 0.3) and Batch Normalization [15] to reduce overfitting and improve generalization. A three-layer neural network with 64, 32, and 16 neurons respectively, was designed to enhance nonlinear modeling, using ReLU in hidden layers and a linear activation in the output. Hyperparameters such as learning rate (0.001), batch size (32), and dropout rate were optimized via grid search and cross-validation. Early stopping was applied to improve training efficiency. Despite neural networks' strong ability to model complex interactions (e.g., sprint speed and shooting accuracy), their computational cost and limited interpretability were addressed through careful architecture design and tuning.

5. Results

Table 1 displays the multi-class classification performance of the Random Forest model in evaluating football training outcomes. The dataset includes 4,320 total entries classified into two categories: "good" (class 0) and "bad" (class 1). For class 0, the model achieved a precision of 0.76, recall of 0.81, and F1-score of 0.79, indicating effective identification of strong performers. For class 1, the model showed slightly lower metrics with a precision of 0.73 and recall of 0.67. The macro average F1-score is 0.73, and the overall accuracy is 78%. This indicates balanced performance across categories, with Random Forest demonstrating strong reliability and interpretability. Notably, feature importance analysis reveals sprint speed and shooting accuracy as the most significant predictors, offering direct guidance for training optimization strategies.

The application of the Random Forest model in predicting students' soccer training performance offers notable advantages. Firstly, leveraging ensemble learning with multiple decision trees, it enhances classification stability and robustness through a voting mechanism. Secondly, its strong interpretability allows for feature importance analysis, which pinpoints key indicators affecting training performance. For instance, sprint speed and shooting accuracy are identified as the two most critical factors influencing training performance, providing a scientific basis for developing targeted training strategies.

Table 1. Multi-classification results of the Random Forest model

	Precision	Recall	F1-score	Support
0 (good)	0.76	0.81	0.79	2450
1(bad)	0.73	0.67	0.69	1870
macro avg	0.79	0.74	0.73	4320
weighted avg	0.88	0.78	0.75	4320
accuracy			0.78	4320

The neural network model demonstrates moderate performance in predicting students' soccer training performance in **Table 2**. For class 0 (better training performance), it achieves 0.65 precision, 0.83 recall, and 0.73 F1-score, indicating high recall with some false positives. For class 1 (poorer training performance), the precision is 0.64, recall is 0.41, and F1-score is 0.50, showing lower recall and more false negatives. The model's overall accuracy is 0.65, with a macro average F1-score of 0.61 and a weighted average F1-score of 0.63. This suggests reasonable performance for class 0 but lower recall for class 1, highlighting an imbalance in classification performance.

Feature importance analysis reveals that sprint speed and shooting accuracy are the most critical factors affecting training performance. This insight is valuable for developing targeted training strategies. The neural network also shows high prediction accuracy in complex pattern recognition, despite lower recall for class 1. Its ability to capture nonlinear relationships makes it effective for complex data patterns.

Table 2. Multi-classification results of Neural Network

	Precision	Recall	F1-score	Support
0	0.67	0.80	0.73	2450
1	0.65	0.58	0.59	1870
macro avg	0.88	0.69	0.61	4320
weighted avg	0.89	0.72	0.78	4320
accuracy			0.72	4320

6. Conclusion and suggestions

This study explores the use of intelligent football training systems in football education, focusing on predicting students' football training performance. By combining Random Forest and Neural Network models, an intelligent evaluation system was developed to predict overall student performance in football training. Key metrics like passing frequency, sprint speed, and shooting accuracy were analyzed for their impact on comprehensive scores. The Random Forest model excelled in stability and interpretability, while the Neural Network achieved higher prediction accuracy in complex pattern recognition. Integrating both models enhanced the system's generalization and applicability. Feature importance analysis identified sprint speed and shooting accuracy as the most critical factors for training performance. Based on these findings, data-driven training optimization strategies were proposed to improve student performance. The results confirm the effectiveness of intelligent football training systems in providing technical support for personalized and refined training programs.

The study also analyzed factors affecting students' adaptability to training methods within the intelligent football training system. Machine learning algorithms revealed that professional experience and past training participation are key predictors of adaptability. Experienced students adapt better to online training, while less experienced ones need more foundational knowledge and practical cases. School type also impacts adaptability due to varying resources, underscoring the need for customized training plans. Collaborative learning, through team participation, enhances training effectiveness, suggesting the inclusion of group discussions or study groups. Although age has a minor impact, older students face more challenges with technology, necessitating user-friendly interfaces and tutorials. These insights provide a foundation for designing adaptive online training methods to meet diverse educational needs.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Kostiukevych V, Stasiuk V, 2017, Training Process Programming of Qualified Football Players in Higher Education Establishments. Physical Education, Sport and Health Culture in Modern Society, 2(38): 41–50.

[2] Yang R, Lin H, 2022, Evaluation of the Quality of Football Teaching in Colleges and Universities Based on Artificial Neural Networks. Computational Intelligence and Neuroscience, 2022(1): 8001252.

[3] Xing J, 2024, Practice of Artificial Intelligence Technology Based on Big Data in Football Training and Physical Education. Frontiers in Sport Research, 6(3).

[4] Yin X, Vignesh CC, Vadivel T, 2022, Motion Capture and Evaluation System of Football Special Teaching in Colleges and Universities Based on Deep Learning. International Journal of System Assurance Engineering and Management, 13(6):

3092–3107.

[5] Guan Y, Qiu Y, Tian C, 2022, Trajectory Planning in College Football Training Using Deep Learning and the Internet of Things. The Journal of Supercomputing, 78(17): 18616–18635.

[6] Stoeve M, Schuldhaus D, Gamp A, et al., 2021, From the Laboratory to the Field: IMU-Based Shot and Pass Detection in Football Training and Game Scenarios Using Deep Learning. Sensors, 21(9): 3071.

[7] Zhou D, Chen G, Xu F, 2022, Application of Deep Learning Technology in Strength Training of Football Players and Field Line Detection of Football Robots. Frontiers in Neurorobotics, 16: 867028.

[8] Cao X, Zhao X, Tang H, et al., 2023, Football Players' Strength Training Method Using Image Processing Based on Machine Learning. PLoS One, 18(6): e0287433.

[9] Hollaus B, Reiter B, Volmer JC, 2023, Catch Recognition in Automated American Football Training Using Machine Learning. Sensors, 23(2): 840.

[10] Rigatti SJ, 2017, Random Forest. Journal of Insurance Medicine, 47(1): 31–39.

[11] Xu Y, 2021, A Sports Training Video Classification Model Based on Deep Learning. Scientific Programming, 2021(1): 7252896.

[12] Wu YC, Feng JW, 2018, Development and Application of Artificial Neural Network. Wireless Personal Communications, 102: 1645–1656.

[13] Fenil E, Manogaran G, Vivekananda GN, et al., 2019, Real Time Violence Detection Framework for Football Stadium Comprising of Big Data Analysis and Deep Learning Through Bidirectional LSTM. Computer Networks, 151: 191–200.

[14] Barquero-Ruiz C, Morales-Belando MT, Arias-Estero JL, 2021, A Teaching Games for Understanding Program to Deal with Reasons for Dropout in under-11 Football. Research Quarterly for Exercise and Sport, 92(4): 618–629.

[15] Liu B, Pu Z, Zhang T, et al., 2022, Learning to Play Football from Sports Domain Perspective: A Knowledge-Embedded Deep Reinforcement Learning Framework. IEEE Transactions on Games, 15(4): 648–657.