SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

*Molecular& Cellular Biomechanics* 2025, 22(2), 797.
https://doi.org/10.62617/mcb797

Article

# Biomechanical research on the construction and optimization of youth basketball training system based on the integration of sports and education

**Zechun Hu[1,*], Zhengfeng Huang[2]**

[1] City University of Hefei, Hefei 238000, China
[2] Wuhu Institute of Technology, Wuhu 241300, China
*** Corresponding author:** Zechun Hu, hzc@cuhf.edu.cn

**CITATION**

Hu Z, Huang Z. Biomechanical research on the construction and optimization of youth basketball training system based on the integration of sports and education. Molecular & Cellular Biomechanics. 2025; 22(2): 797. https://doi.org/10.62617/mcb797

**ARTICLE INFO**

Received: 14 November 2024
Accepted: 20 November 2024
Available online: 17 January 2025



**Abstract:** The development and improvement of a youth basketball training program founded on the fusion of education and sports is investigated in this study. Athlete performance and academic advancement must be balanced in light of the growing need for comprehensive youth development. Biomechanical factors play a significant role in both sports performance and injury prevention, making it essential to integrate them into the training program design. To increase the effectiveness and design of training programs, the suggested model makes use of the Tabu Search Optimized Intelligent Random Forest (TSO-IRF) algorithm. The TSO-IRF identifies important physical, technical, and cognitive elements affecting basketball play by combining search-based optimization with machine learning (ML) approaches from a biomechanical perspective. It focuses on elements such as joint forces, muscle activation patterns, and movement kinematics, which are fundamental in determining an athlete's performance and injury risk. The research gathers information on youth basketball training programs, with a specific emphasis on biomechanical aspects. This includes information on players' body mechanics during different basketball movements, like jumps, shots, and passes. By integrating this data, the study ensures that the goals of educational development and sports training are aligned, while also considering the biomechanical requirements of the athletes. TSO-IRF is used to evaluate these multidimensional features and provides individualized training suggestions in line with the performance and educational objectives of both sports. Experimental results indicate that the TSO-ERF model can perform better than traditional methods, providing higher prediction recall (94.26%), accuracy (97.81%) and precision (97.21%) in development metrics for players. Additionally, the model shows improved adaptability across various skill levels as it can adjust training recommendations based on an athlete's unique biomechanical characteristics. The proposed youth basketball training system optimizes loads in training, reduces risks of injury, and develops young athletes over the long term. It facilitates athletic success but fosters cognitive and emotional development so that the fields of sport and education may converge. Future work involves the application of this model in other sports disciplines and algorithm refinement to take care of larger datasets that would help deliver real-time performance feedback.

**Keywords:** sports and education; biomechanics; youth basketball training system; Tabu Search Optimized Intelligent Random Forest (TSO-IRF)

## 1. Introduction

Considering the educational system as the primary training field, "the integration of sports into education" is a remarkable effort. Its ultimate objective is to support human development comprehensively through competitive sports and contribute to the national [1]. Basketball is a special type of interpretation and implementation that is identifiable by the cheerful aspect of physical activity [2]. Basketball has outstanding academic value because of its numerous positive impacts on human personality. Basketball's value-formative features explain its enormous appeal in the educational surrounding, including the desire of instructors in physical education to encourage it at the highest level of school sports [3]. Participating in physical education and sports can help young people improve their personal and social skills. Physical education is increasing in popularity as a means of educating youth about the challenges of daily living [4].

However, young people's brain and musculoskeletal systems are indeed developing. Teenagers' and adolescents' control and coordination of stretch and shorten cycles are significantly impacted by differences in neural conduction speed, control ability, and muscle structure [5]. Intensive youth sports programs can have both positive and negative outcomes, given their popularity and the possibility that most participants fail to succeed in their basketball training. Understanding the overall development implications for youth. Athletes' involvement in these initiatives is essential to guaranteeing that they foster healthy development [6]. Athletes have varied rates of physical growth and performance during this phase, leading to variations in timing, intensity, and pace. Youth male basketball athletes with advanced maturity typically outperform their late-maturing colleagues in static endurance, power, sprinting, agility, jump, and shooting abilities [7].

Using sports analytics as a performance tool during training might increase the interest and motivation of young athletes who are driven to develop [8]. There are instructional components, such as academic tutoring and life skills coaching, which again can be specific to student-athletes in terms of learning requirements [9]. Tutoring in core subjects such as math, language arts and science helps athletes better meet academic requirements while maintaining training schedules. Life skills such as time management, teamwork, leadership and resilience should also be incorporated into training. This will promote personal growth and character development. For example, some preparation in time management will help young athletes deal with setbacks on the court, but also in pursuing academic opportunities to develop other useful life skills [10]. The study addressing the aim of this study is to improve youth basketball training program that combines education and sport as well as uses the TSO-IRF method to optimize training, improve performance, and cognitive and emotional growth.

Contribution of this study

- It integrates data from a number of youth basketball training programs, in particular physical, technical, and cognitive attributes that drive basketball performance.
- Pre-processing of the data further involves Z-score normalization to bring all data into one comparable scale. This removes any form of bias in the dataset that may be emanating from a difference in unit or magnitude.
- While feature extraction involves the use of Principal Component Analysis. It decreases the dimensionality of the data as it identifies the most relevant features, and principal components, providing the maximum variance in data.
- The study increases the effectiveness and design of training programs using TSO-IRF, a hybrid approach integrating search-based optimization techniques

with machine learning, which will be applied for the evaluation of multidimensional features.

The rest of the study is the following parts: Part 2 describes the related article on young player basketball training, Part 3 provides a methodology to clarify a proposed method process, the performance of the evaluated result and discussion are shown in Part 4, finally, Part 5 expresses the overall research conclusion.

## 2. Related work

To evaluate juvenile basketball training performance using an evaluation model based on the Particle Swarm Optimization (PSO) method was the purpose of Ouyang and Wu [11]. The technique created a performance index system using the Analytic Hierarchy Process (AHP) to determine index weights, and then optimizes the model using PSO. Results reveal that the strategy improved evaluation accuracy while decreased evaluation time, exhibiting beneficial training performance insights.

Zhong [12] to improve basketball training by precisely recognizing motions using Basketball Spatiotemporal Action Recognition Network (BSTARNet), an AI-based action recognition technique. An encoder-decoder model was improved with Darknet by using Convolutional Long Short-Term Memory (ConvLSTM) for spatiotemporal extraction of data and Attention Long Short-Term Memory (AttLSTM) for concentrated attention on action zones. Experiments reveal that BSTARNet recognizes basketball actions with 89.5% mAP and 95.4% accuracy.

Xu-Hong et al. [13] used 3D convolutional neural network architecture to enhance basketball technical action recognition. The approach tackled the complexity of diverse basketball tactics by processing two different resolution image inputs from a basketball action dataset. The framework's ability to identify activities in basketball footage was demonstrated by experimental findings, and improved sports video analysis accuracy.

The aim of Khobdeh et al. [14] was to improve basketball action recognition for Human-Computer Interaction by providing officials and players with helpful information. In order to overcome obstacles like complicated backdrops and illuminated, it used YOLO (You Only Look Once) for player identification in conjunction with a deep fuzzy LSTM network for action classification. When the suggested model was tested on the SpaceJam and Basketball-51 datasets, it outperformed all baseline models in terms of accuracy.

Lower extremity muscle strains (LEMSs) in National Basketball Association (NBA) players were examined by Lu et al. [15], which also evaluated machine learning algorithms for injury prediction. With 736 LEMS incident data, models such as XGBoost outperformed logistic regression with good prediction accuracy (AUC 0.840). XGBoost was the best-performing model, and important factors were age, playing statistics, and injury history.

Javadpour et al. [16] sought to maximize offensive play in Division 1 women's basketball by using deep learning to predict the best play given a certain set of game parameters. The information came from basketball games played by a private university in a top-25 league. Using deep learning techniques, gaming scenarios were analyzed to find patterns that increased the likelihood of a successful shot. To improve decision-making and team performance during close games, the model was trained using raw game data.

By utilizing random forest, support vector machine, and multi-linear regression algorithm to evaluate players' skill levels and track their progress, Yao and Li [17] sought to create an effective evaluation system for juvenile sports instruction. The technique entails building a system model that assesses the impact of training, monitors the development of physical fitness, and gives coaches measurable criteria to customize training regimens. The system's effectiveness in helping trainers assessed players' skills and make better training selections was demonstrated by the results.

By combining 3D Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks, Wang et al. [18] attempted to improve basketball action detection by capturing spatiotemporal aspects of basketball actions. Adaptive rates of learning and normalization were used to optimize the model after basketball actions from publically accessible datasets have been preprocessed. With an accuracy improvement of 15.1% beyond frame difference-based techniques and 12.4% over optical flow-based techniques, the suggested method surpassed conventional approaches. Its average accuracy was 93.1%, indicating great durability under a variety of scenarios.

By utilizing big data and intelligent systems to optimize the ecological environment of basketball courses, Xue [19] aimed to improve basketball instruction in higher education. The process entails created a model to extract student learning data for individualized basketball instruction, constructing an intelligent motion capture device, and transferring data to a server. Comparative trials were used to assess the efficacy of the system, which showed an 87% identification rate. According to the findings, using big data in basketball instruction may greatly increase students' physical fitness and promote sustainable growth plans for collegiate athletics.

With an emphasis on the ecological context of basketball education, Xuexiang [20] intended to investigate the creation and administration of collegiate basketball talent training utilizing data mining (DM) technology. To enhance the training model, the technique entails gleaning valuable insights from sports talent data. The accuracy of sports talent information mining has improved by 23.85% when compared to traditional approaches, according to the results, indicating that basketball education has to change to satisfy social demands.

Wang and Du [21] were to use the Internet of Things (IoT) and machine learning technologies to improve sports education and training. To increase prediction accuracy, it refines hidden layer mapping and parameter optimization in the Extreme Learning Machine (ELM). The ELM analyzed and forecasted the outcomes of sports training using the continuous data gathering enabled by IoT technology. The system efficiently optimized the process of sport education and training, as demonstrated by experimental findings.

Yang [22] used neural networks to create a customized physical training index system for basketball players. Tests of agility, strength, and endurance were combined to provide an objective framework for assessing athletes' physical preparation. Using a neural network model, the technique examined data from 100 elite basketball players and predicted a performance boost of 6.68%. It emphasized how well machine learning works in sports science studies.

Xu and Tang [23] aimed to avoid injury in sports and enhanced the efficiency of shooting during basketball training by considering machine learning-based path planning for an intelligent robot. One could find a strategy on that to analyze basketball trajectories of motion, create a model of shooting motion, and introduced an improved Q-Learning algorithm of robot path planning. Results have shown that the path resulting from the improved algorithm was smoother and takes less time to build an optimal path. The approach avoided collisions effectively and enhanced safety during basketball training.

Problem statement

The integration of sports and education in youth basketball training systems is a new area of focus, as this meets the emerging need for all-round development of youths, balancing athletic performances with academic and cognitive development. Traditional training models often stress the physical skills and technical developments of young athletes, with limited regard for the large cognitive and emotional component that underpins enduring success both in sport and academically. The last gap created the need for training systems that, while enhancing athletic performance, would also nurture the intellectual and emotional maturity of the young athlete. Even with such potential benefits, integrating the two domains yet lacks personalized and data-driven approaches in optimizing the training loads and injury risks and aligning performance with educational objectives. The challenge at present would be to construct training programs that could be adaptive to great differences in skill levels by effectively balancing the goals of physical fitness, mental development, and academic progress. This problem is addressed by the paper, which designs and improves sports and education integration based on the Tabu Search Optimized Intelligent Random Forest (TSO-IRF) algorithm in the field of a basketball training program for youth. Key factors will be identified in terms of physical, technical, and cognitive influences on player performance. Optimize the training load to suggest personalized training in developing better adaptability among skill levels to promote long-term athlete development.

## 3. Methodology

The approach integrates education and sports to improve the training process for young basketball players. First, the data is collected from youth basketball training dataset available in Kaggle. Players in the Kaggle dataset were selected based on certain criteria, including age, skill levels, and their performance. Next, the gathered data is preprocessed by applying Z-score normalization to unify the features. Followed by this, feature extraction is done through Principal Component Analysis (PCA) to reduce the number of dimensions while retaining crucial information. At the core of the methodology is a TSO-IRF algorithm that integrates search-based optimization techniques with machine learning to identify key physical, technical, and cognitive factors affecting basketball performance. **Figure 1** presents an entire flowchart that elaborates the methodology applied, which includes data collection to the determination of key performance factors.

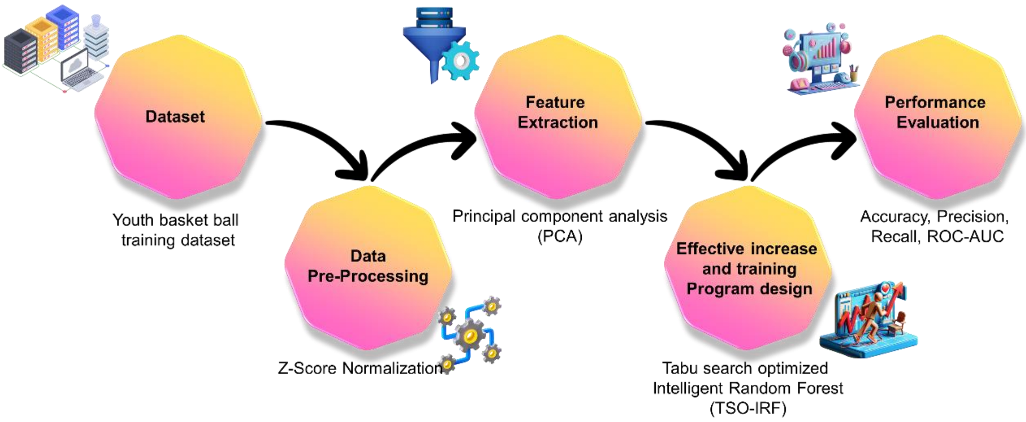


**Figure 1.** Basic concept of study proposed framework.

### 3.1. Dataset

The Youth Basketball Training data set on Kaggle [24] provides performance information related to young basketball players in terms of shooting accuracy, dribbling speed, and physical endurance. The dataset is divided into 1000 rows, each representing an individual adolescent athlete, and 25 columns that contain demographic, physical, technical, cognitive, educational, as well as training-related characteristics. The sample was gathered based on the participation of players in structured youth basketball training programs. Data collection was through the tracking of player performance over a selected period while performing drills and exercises. The dataset comprises various features, enabling the in-depth examination of technical, physical, and cognitive factors that affect basketball performance. This data is pre-labeled and sorted for analysis.

### 3.2. Data pre-processing

This data is standardized using the *Z*-score normalization to transform these features into a common scale with a zero mean and unit standard deviation. It makes the input features comparable to each other by avoiding biases due to different units or ranges of feature values, hence allowing for more accuracy. One statistical method used to deal with the problem of outliers is *Z*-score normalization. This technique transforms the values of the chosen feature by using its standard deviation and mean. Specifically, the change is carried out using Equation (1) as follows:

$$z' = \frac{z - \mu}{\sigma} \tag{1}$$

where $z'$ denotes the standardized value, $z$ is the actual value, $\mu$ is the feature's mean, and $\sigma$ is the standard deviation. *Z*-score normalization maps values under the mean to negative numerals, values over the mean to positive numerals, and values equivalent to the mean to zero. This approach guarantees that the information is normalized and that outliers are efficiently handled.

### 3.3. Feature extraction

Feature extraction uses Principal Component Analysis (PCA) in this study to decrease the dimensionality of the collected data. PCA emphasizes the main variables that exist behind basketball training by converting the original correlated variables into a smaller set of uncorrelated components. This reduction allows for better and faster evaluation and optimization during training while maintaining the main details in the context to provide accurate predictions. Extracting features and decreasing dimensionality are two well-established applications of principal component analysis (PCA). Representing the d-dimensional data in a lower-dimensional space is our goal while using PCA. This will lessen the complexity of space and time as well as the degrees of freedom. By eliminating the category label, every observation in a data set of $l$ occurrences is statistically n-dimensional. Let $w_1, w_2, \ldots, w_l \in \mathfrak{q}^n$. The subsequent processes involved in PCA computation. Determine the mean vector μ in m dimensions by the following Equation (2).

$$\mu = \frac{1}{l}\sum_{i=1}^{l} w_j \tag{2}$$

For the observed data, calculate the calculated matrix of covariance $T$ by following the Equation (3).

$$T = \frac{1}{l}\sum_{l=1}^{l}(w_j - \mu)(w_j - \mu)^s \tag{3}$$

Determine the latent values and matching latent vectors of $T$, here $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$. The $l$ primary elements were determined from the $l$ initial variables by following the Equation (4).

$$\begin{gathered} z_1 = b_{11}w_1 + b_{12}w_2 + \cdots + b_{1k}w_l \\ z_2 = b_{21}w_1 + b_{22}w_2 + \cdots + b_{2k}w_l \\ z_l = b_{l1}w_1 + b_{l2}w_2 + \cdots + b_{kk}w_l \end{gathered} \tag{4}$$

It is important that $z_l$'s is unrelated. The initial variation in the data set is explained to the greatest extent feasible by $z_1$, the remainder of the variance is explained to the greatest extent possible by $z_2$, etc. The most useful data sets often have a few numbers of bigger latent values that predominate over the rest that is Equation (5).

$$\gamma_l = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{\lambda_1 + \lambda_2 + \cdots + \lambda_n + \cdots + \lambda_l} \geq 80\% \tag{5}$$

Where the proportion kept in the data representation is denoted by $\gamma_l$. The primary components that can account for at least 80% of the total variance should be kept when employing PCA for feature extraction.

### 3.4. Tabu Search Optimized Intelligent Random Forest (TSO-IRF)

The Tabu Search Optimized Intelligent Random Forest (TSO-IRF) algorithm, as it effectively utilizes search-based optimization and machine learning, is ideally appropriate to the complexity, non-linearity, and multidimensional nature of youth basketball training data in this study. Tabu Search optimization serves to identify critical features and helps explain the tuning of the model by exploring a broader solution space, whereas the Random Forest algorithm is remarkably good in examining these features for predicting and personalizing training strategies. With this hybrid approach, the model can precisely assess critical physical, technical, and cognitive variables that affect performance, thereby increasing training efficiency, injury risk mitigation and athletic pathway success. With its proven adaptability, the TSO-IRF is ideal for multiple levels of play and can support real-time, individualized feedback necessary to balance athlete and academic development. Algorithm 1 procedure of Tabu Search Optimized Intelligent Random Forest (TSO-IRF).

**Algorithm 1** Process of TSO-IRF

```
Initialization
Initialize Tabu List (TL) to store visited solutions
Initialize Random Forest (IRF) with default parameters
Define stopping criteria (max iterations or convergence threshold)
Define neighborhood search space (M)
Define aspiration criteria function T(l)
Set initial solution w_0 (a random selection of features)
Define Tabu Search parameters
α = 1/2 [1+sin[jθ/M_neigh ]]  Coefficient for neighbor solution variation
Set parameters for T(l) using sigmoid function
Tabu Search Process
for iteration in range(1, max_iterations):
   Step 1: Generate neighborhood solutions around the current solution w
    neighbors = generate_neighbors(w, M)
     Step 2: Evaluate the quality of the neighbors using Random Forest
    best_neighbor = None
    best_accuracy = −∞
    For a neighbor in neighbors:
        accuracy = evaluate_rf(neighbor, IRF)  Evaluate RF model on current feature set
        if accuracy > best_accuracy:
            best_accuracy = accuracy
            best_neighbor = neighbor
     Step 3: Check if the best neighbor violates Tabu List constraints
    if best_neighbor is in TL:
         Apply aspiration criteria if necessary
        if aspiration_criteria(best_neighbor, iteration):
            w = best_neighbor   Accept the neighbor despite being in the Tabu List
    Else:
        Accept the best non-tabu neighbor
        w = best_neighbor
Step 4: Update the Tabu List (add the current solution and remove the oldest)
    update_tabu_list(TL, w)
Step 5: Check stopping criteria (convergence or max iterations reached)
    if stopping_criteria(w, iteration):
        break
Final Solution
optimized_features = w   Features selected by the optimized solution
final_model = train_rf(optimized_features) Train Random Forest with the selected
features
```

### 3.4.1. Intelligent random forest

The Intelligent Random Forest is used to analyze and assess the multidimensional features established to affect youth performance in basketball.

These will be processed by the IRF algorithm to predict the true development of the athlete's condition. It enables personalized recommendations for training, thus achieving a balance between physical, technical, and cognitive aspects and adapting to different levels of skills in the best training outcomes. Numerous studies demonstrate that the random forest approach improves classification accuracy, tolerates aberrations and noise, and is resistant to overfitting. The node's splitting Equations (6 and 7) display the data gain and Gini index achieved by dividing the specimen set $C$ based on characteristics $b$.

$$Gain(C,b) = Ent(C) - \sum_{u=1}^{U} \frac{|C^u|}{|C|} Ent(C^u) \tag{6}$$

$$Gini(C,b) = \sum_{u=1}^{U} \frac{|C^u|}{|C|} Gini(C^u) \tag{7}$$

The symbol $C^u$ denotes that the $u$ branching node includes all samples in $C$ with a value of $b^u$ for the parameter $b$.

$$Ent(C) = -\sum_{l=1}^{|z|} o_l \log_2 o_l \tag{8}$$

$$Gini(C) = \sum_{l=1}^{|z|} \sum_{l' \neq l} olol' = 1 - \sum_{l=1}^{|z|} ol^2 \tag{9}$$

The goal of node dividing is to increase the integrity of the data after separation (Equations 8 and 9). This may be achieved by a mix of node-splitting formulas and adaptive methods. The variable selection procedure is as follows:

$$G = \underset{\alpha, \beta \in R}{\min} \; E(C,b) = \alpha Gini(C,b) - \beta Gain(C,b) \tag{10}$$

$\alpha, \beta$ denote the value of the attribute dividing strength coefficient, Equation (10). Meanwhile, $G$ has a low value. The dynamic parameter selection procedure is used to get the best combination of variables. The experiment evaluates performance using classification error and accuracy rates. The categorization error rate of the specimen $C$ is expressed as Equation (11).

$$F(e;C) = \frac{1}{n} \sum_{j=1}^{n} II(e(w_j) \neq z_j) \tag{11}$$

The precision percentage is calculated using Equation (12).

$$acc(e;C) \frac{1}{n} \sum_{j=1}^{n} II\big(e(w_j) = z_j\big) = 1 - f(e;C) \tag{12}$$

The next part compares and verifies the outcomes of the experiment. **Figure 2** illustrates the intelligent random forest architecture.

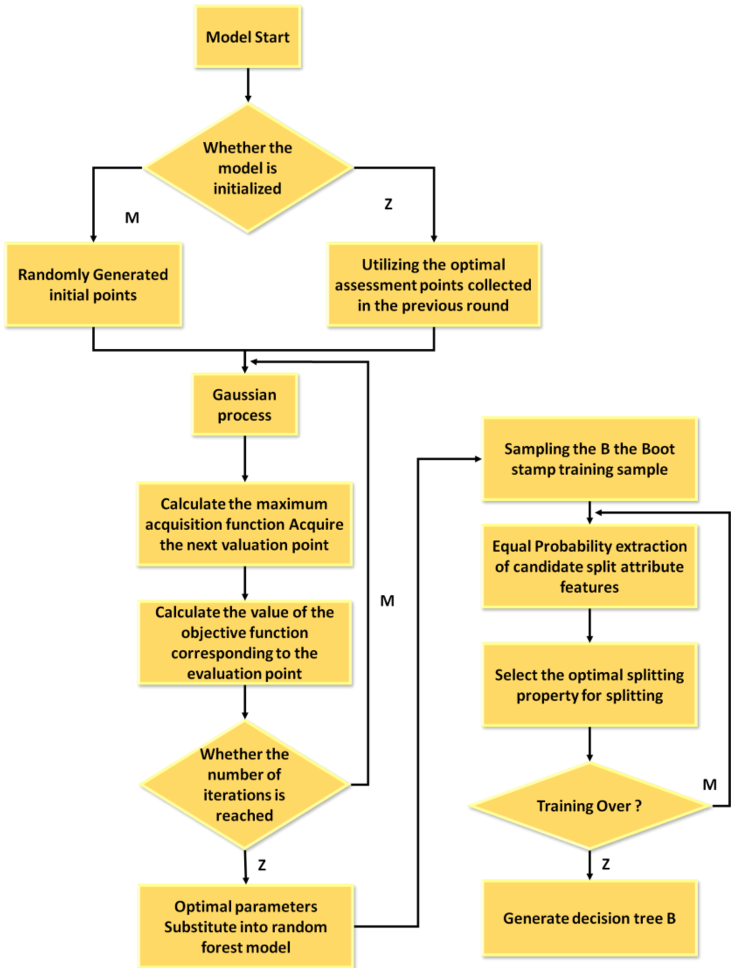


**Figure 2.** Intelligent random forest architecture.

### 3.4.2. Tabu search optimization

The TSO method was implemented in this study that allowed an efficient exploration of youth basketball training features and relevance identification, including physical, technical, and cognitive aspects. This has optimized the choice of relevant features so that the search mechanism would operate in a larger solution space to enhance the accuracy of performance predictions while also allowing trainers to have sport-specific, personalized, and effective plans for each athlete.

*Generations of neighbors and searching for neighborhoods*

To function optimize $g(w)$ worldwide from all of the potential outcomes $w \in W$ in the space $W$, a structure near the space of solutions and the initial solution must be specified. The search continues to adjust the present answer to generate a list of potential answer in the region of solution space. The amount of answers in close proximity of the result area, $M(w_j)$, and the total amount of repetitions, $l$, determine the solution number traversed by TS throughout the course of the search. The operation is evaluated for $M(w_j)$ solutions in every repetition. Around the whole solution space, the best move is selected. The following iteration of the search looks for a solution close to the relocation that was approved. Therefore, the TS uses a historical record of the search to provide a collection of workable options.

*List of tabu*

The data through the previously visited answer is stored in the tabu list in TS. The list, which changes continuously while the search continues on, features the biggest recent moves. The tabu list's contents help direct the change from the current response to the next one. Every time the search procedure is carried out, the tabu list is updated. Additionally, by avoiding re-visiting recent neighbors that are shown, the tabu list saves computing time.

When neighbor solutions are being generated, a coefficient called $\alpha$ is employed to regulate the difference between new neighboring answers and the existing ones. The process of creating new neighbors involves multiplying the change from the present position by α. The form of the coefficient $\alpha$ is a sine function, Equation (13).

$$\alpha = \frac{1}{2}\left[1 + sin\left[\frac{j\theta}{M_{neigh}}\right]\right] \tag{13}$$

Here, $\theta$ is a parameter that regulates the period of oscillation of $\alpha$, $M_{neigh}$ is the entire amount of neighborhood answers produced at every iterations, and $j$ was the neighbor's index.

*Aspiration criteria*

Sometimes moves leading to unvisited solutions are prevented by the TS criteria. A requirement known as the aspiration criteria has the power to overrule a move's tabu status. In some circumstances, the ambition criteria may invalidate the tabu property to prevent some missing solutions throughout the search and preserve a suitable balance between intensification and diversification. Aspiration criteria are created using the sigmoid function provided by the Equation (14).

$$T(l) = \frac{1}{1 + f^{-\sigma}(l - l_{center} \times N)} \tag{14}$$

The tuning parameters $l_{center}$, $l$, $\sigma$, and $N$ serve as a representation of the greatest number of repetition, another tunable variable, and the current repetition number. The range of $l_{center}$ value in numbers is 0.30 to 0.70, while the spectrum of σ is 5 to 10/M. For every cycle, a uniform distribution generates a random number $O$ ranging from zero to one. If $O$ is bigger than$T(l)$, the tabu characteristic is activated, and the greatest non-tabu neighbors are chosen as an initial beginning point. If $O$ is either smaller or closer to $T(l)$, the aspirations criteria disregard the tabu characteristic.

*Criterion of stopping*

When the ideal situation is reached, a stopping condition must be used to end the search. The halting criterion might be a select frequency of repetitions or a threshold for solution convergence. The maximum time terminate and terminate-on convergence requirements are also employed as search stoppers.

$$\left|\frac{f_l(w) - g_{l-\Gamma}(w)}{g_{l-\Gamma}(w)}\right| < \delta \tag{15}$$

where $\delta$ is the proportion of the alteration in the significance of the objective operation, $\Gamma = \eta M$ when $\eta$ was the percentage of the highest number of repetitions $(N)$ by which the objective function alteration is compared, Equation (15). The stopping criteria suggest that if the development across $\Gamma$ generation is less than a threshold ($\delta$), additional repetitions may be futile and the search should be stopped. **Figure 3** illustrates the flow chart of the TS algorithm.

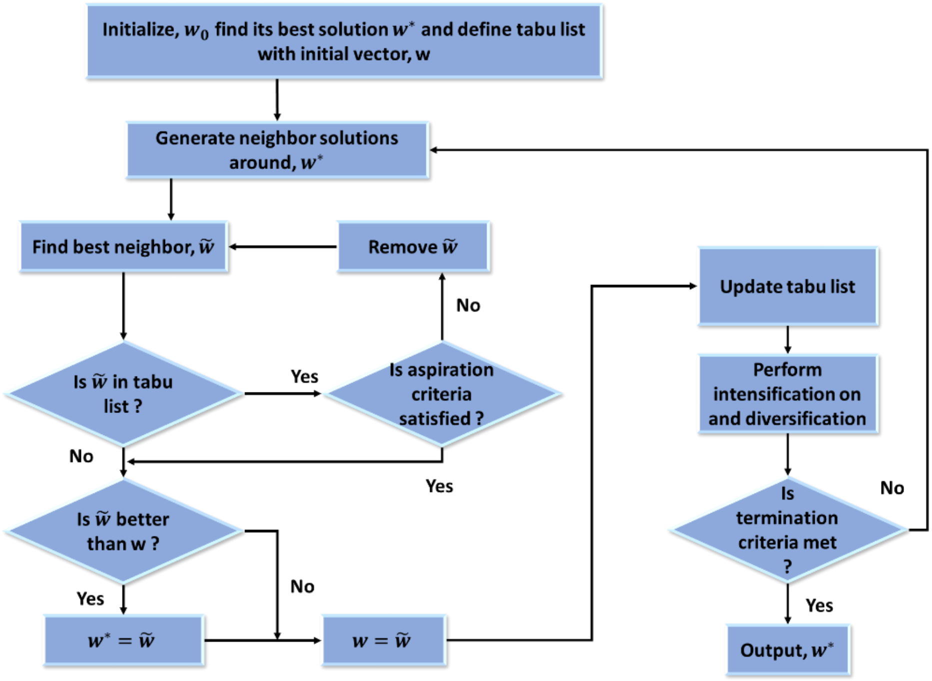


**Figure 3.** Flow chart of the TS algorithm.

## 4. Result and discussion

The Python programming language was utilized in combination with an Intel® Core i9 CPU-powered Windows 10 operating system with 16.00 GB of RAM. **Figure 4** illustrates how the TSO-IRF has to adapt and tailor training for such diverse needs of player development. The TSO-IRF model optimizes the key metrics that are tailored from one player to another according to one's needs at different levels of skills. At the beginner level, recommendations for training involve developing rudimentary physical fitness and cognitive understanding accompanied by technical skills. In the case of higher ranks, the model develops specified strengths in areas of technical and cognitive skills and physical fitness to continue enhancing general performance and minimize injury.

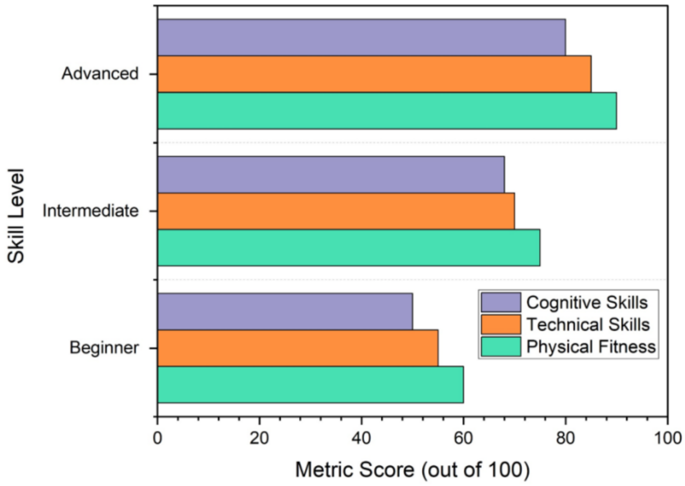


**Figure 4.** TSO-IRF model optimizes the key metrics by skill level.

Accuracy and loss

Accuracy and loss for the proposed TSO-IRF model increases the values of accuracy while reducing the values of loss with time, which improves the predictions with each model training, as shown in **Figure 5.** Training accuracy grows before stabilizing, while testing accuracy may be low because of overgeneralization. **Figure 5** shows loss will be steady downward, showing better performance by the model. The model should be able to generalize well to new data without overfitting if the training and testing accuracy match and the loss decreases.

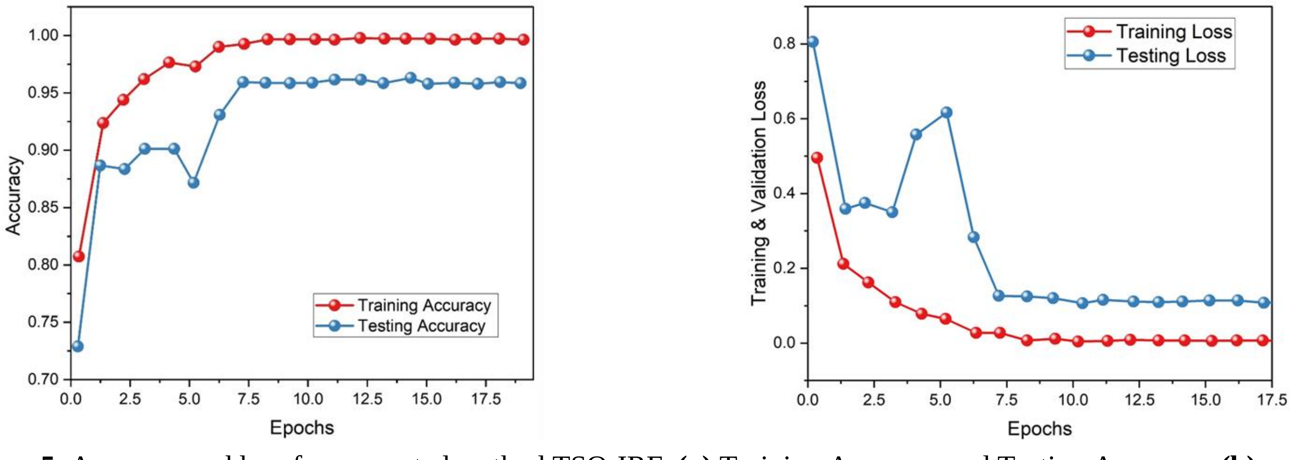


**Figure 5.** Accuracy and loss for suggested method TSO-IRF. **(a)** Training Accuracy and Testing Accuracy; **(b)** Training Loss and Testing Loss.

Receiver Operating Characteristic—Under the Curve (ROC-AUC)

The ROC-AUC measures the ability of the proposed TSO-IRF model to distinguish between successful and unsuccessful player development in youth basketball training, as shown in **Figure 6.** Good predictability of player progression is indicated by high values close to 1 for AUC; the corresponding True Positive Rate (TPR) would be high, while the False Positive Rate (FPR) would be very low; it would not be problematic to distinguish between those players who are improving and those who are not (at least according to physical, technical, and cognitive ability). A value closer to 0.5 would correspond to the almost random-guessing performance of the model. A higher ROC-AUC value, for example, suggests an even better model in terms of personalization while being effective in recommending training towards better outcomes for the player contrarily. Lower AUC shows potential aspects that could be improved in the model or feature optimization.

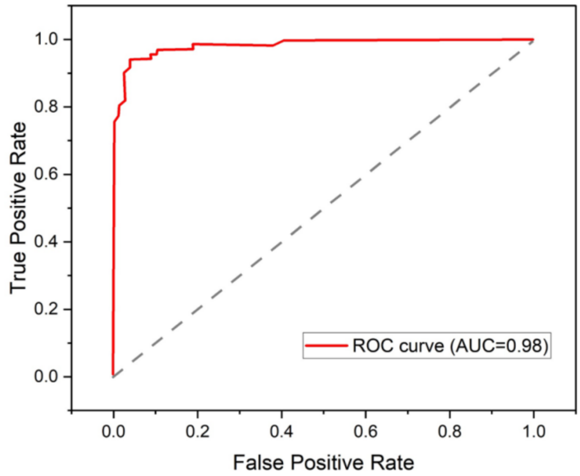


**Figure 6.** ROC-AUC curve for proposed TSO-IRF approach.

The young basketball training program uses TSO-IRF performance evaluation in the study to integrate sports and education. The comparative assessment uses several parameters, including recall, accuracy, and precision. Deep neural network (DNN) [25], Gated residual network (GRN) [25], and decision tree (DT) [25] are the conventional comparison techniques.

Accuracy

Accuracy is defined as the ratio of correct predictions made to all predictions made in the form of both true positives and true negatives. Accuracy indicates the overall correctness of the TSO-IRF model regarding its prediction related to whether the player's development at physical, technical, and cognitive levels would stay correct and on track. **Figure 7** shows accuracy values. If the value of accuracy for the same traditional techniques such as GRN, DNN, and DT is less than TSO-IRF, it means that the TSO-IRF model is better at handling more complex multidimensional data and making accurate predictions regarding the development of players.

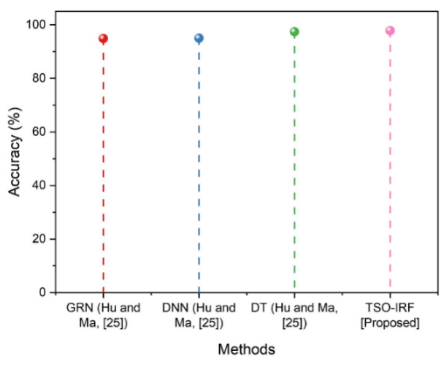


**Figure 7.** Comparison accuracy values of various traditional models vs. proposed method.

Precision

Precision is the number of correctly classified positive cases out of the total number of cases or positives that the classifier expects to be positive. A false positive could be classifying a player as improving when, in fact, they are not. **Figure 8** demonstrates the precision values. When this model classifies a player as improving, there's the likelihood that such classification is correct when the precision is high. If TSO-IRF gives higher precision than other methods like GR, DNN, and DT, it would mean that the model is more likely to be a good predictor in terms of positive comments concerning the development of players and minimizes false positives.

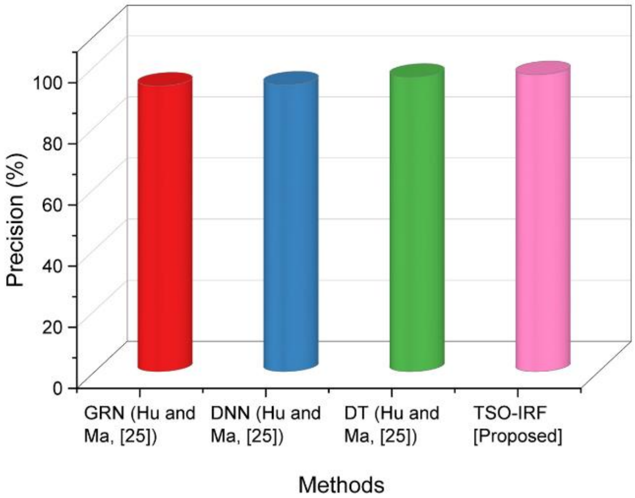


**Figure 8.** Comparison precision values of various traditional models vs. proposed method.

Recall

Recall (or Sensitivity) is the number of correct true positives out of all actual positives, i.e., the percentage of correctly identified positive instances out of all actual positives. A player who is improving should not be categorized as a false negative. High recall would mean the model correctly identifies most of the players who are improving even though it also predicts some false positives. **Figure 9** describes the recall values. If recall is better for TSO-IRF than for other methods, including GR, DNN, and DT, then the model is more targeted at improving players and therefore provides more utility in an early intervention context where players may not have developed as well as the analyst would have anticipated. **Table 1** illustrates the comparison performance of the proposed method with traditional methods.

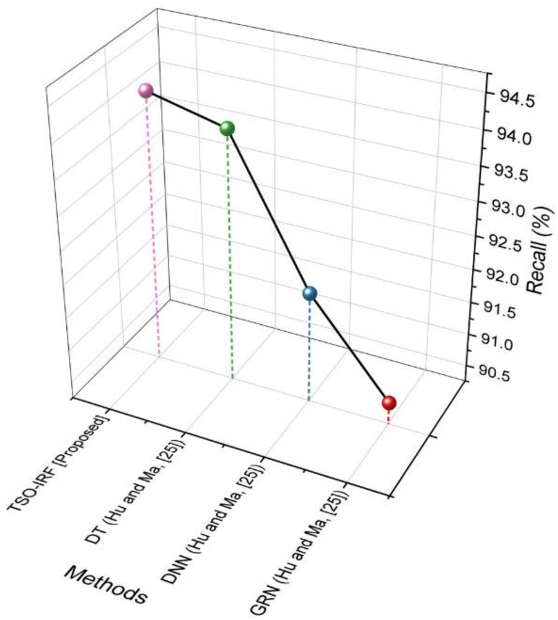


**Figure 9.** Comparison recall values of various traditional models vs. proposed method.

**Table 1.** Model performance comparison based on proposed method with traditional models.

| Methods | Accuracy | Precision | Recall |
|---|---|---|---|
| GRN (Hu and Ma [25]) | 94.84 | 93.51 | 90.58 |
| DNN (Hu and Ma [25]) | 94.97 | 93.85 | 91.92 |
| DT (Hu and Ma [25]) | 97.37 | 96.37 | 93.99 |
| TSO-IRF [proposed] | 97.81 | 97.21 | 94.26 |

## Discussion

The youth basketball training system integrates the sports and educational aspects of development towards overall development. A drawback with the traditional methods, including Deep Neural Network (DNN), Gated Residual Network (GRN), and Decision Tree (DT), is that they might not operate efficiently in dealing with the complexity and multi-dimensionality of player development data, both in terms of sports and education. DNN [25] has proved to be powerful but tends to suffer from overfitting, especially in limited data or a poorly preprocessed dataset. It also happens that high computational resources were required for training DNNs, which might not prove to be effective in cases where real-time feedback is required in youth basketball training. DNNs require a significant amount of training time, which is not suitable for systems that need rapid adjustments during live training sessions. Although GRNs [25] were implemented to handle complex temporal data, they probably do not handle multi-faceted factors appropriately, like physical, technical, and cognitive skills, in the absence of additional optimization techniques. The applicability of GRNs might be restricted in cases of lack of noisy data, which often result in incorrect models and influence player development decisions. They are also computationally intensive and require more sophisticated tuning to avoid underfitting or overfitting. DT [25] tends to overfit the training data, especially if there are many variables or in cases of more complex relations among features. This drawback in DT models can prevent them from generating generalized insights, which are crucial for optimizing training outcomes across a range of player profiles. Moreover, they are not capable of managing intricate inter-feature interaction capabilities, which may be very important for holistic player development, especially in sports and education.

To overcome such limitations, the proposed hybrid model combines the strengths of different techniques to reach an optimum solution, as exemplified by the TSO-IRF model. The hybrid model will have the ability to analyze the player more comprehensively at different dimensions, such as physical, technical, and cognitive, while simultaneously overcoming problems related to the great computational requirements of classical methods. The TSO-IRF model is so structured that it accommodates large datasets efficiently without consuming much computational power, thus providing quick real-time feedback for youth basketball training. It has a feedback loop that updates the training strategies through continuous data input and provides personalized advice to the players based on their changing needs. This proposed study has the advantage of integrating education and sports through the TSO-IRF model, aiming to maximize training effectiveness by identifying crucial physical, technical, and cognitive factors towards personalized advice based on data analysis. The method delivers more accuracy in prediction, skill level adaptation, reduced chances of injury, and long-term athlete development both in athletics and cognitive development.

## 5. Conclusion

Young player's education and sports development should be balanced with their physical achievement, based on this study. The study established an integrated structure for youth development that provides for both cognitive and physical development by integrating sports training programs with educational objectives. The TSO-IRF approach is developed in the research using search-based optimization and ML. Thus, it is possible to determine the critical technical, cognitive, and physical factors that influence basketball performance, improve training procedures, and provide specific suggestions for player development. It used Z-score normalization for data preparation to ensure that the data is consistent and appropriate for data analysis. Furthermore, PCA was employed for feature extraction, which reduces the multidimensional training data while retaining important data. The experimental findings demonstrate that the TSO-IRF model surpasses conventional techniques in terms of forecast precision (97.21%), recall (94.24%), and accuracy (97.81%). It provides better measures to evaluate player development, as well as enhanced flexibility across various skill levels, hence improving the training program's effectiveness.

## 6. Limitation and future score

In integrated sports and education for youth basketball training, resource shortages, a lack of specialized coaches, and finding a balance between academic work and athletic commitments are some of the major challenges. Students engage themselves with different levels of engagement and performance. Future development needs to consider data-driven approaches, such as machine learning for personalized training, improvement of the current education of coaches, and designing flexible schedules to fit better with both academic and athletic objectives. The studies may also delve into scalable models that incorporate technologies for monitoring progress and performance optimization.

**Author contributions:** Conceptualization, ZH (Zechun Hu) and ZH (Zhengfeng Huang); methodology, ZH (Zechun Hu) and ZH (Zhengfeng Huang); writing—original draft preparation, ZH (Zechun Hu) and ZH (Zhengfeng Huang); writing—review and editing, ZH (Zechun Hu) and ZH (Zhengfeng Huang). All authors have read and agreed to the published version of the manuscript.

**Funding:** Project in Humanities and Social Sciences of Anhui Province，Research on current situation and countermeasures of the Youth Three-player Basketball in Anhui Province under the background of the integration of sports and education. (NO: 2022AH052462).

**Ethics approval:** Not applicable.

**Conflict of interest:** The authors declare no conflict of interests.

## References

1. Wu, D. and Du, H., 2021. Sports Science Contributes to the Development of the “Integration of Sports and Education” in China. Journal of Innovation and Social Science Research ISSN, 2591, p.6890. 10.53469/jissr.2021.08(10).22
2. Gottlieb, R., Shalom, A. and Calleja-Gonzalez, J., 2021. Physiology of basketball–field tests. Review Article. Journal of human kinetics, 77(1), pp.159–167. https://doi.org/10.2478/hukin-2021-0018
3. Gorgan, C.M., Oancea, B.M. and Ciocan, C.V., 2024. Using Basketball Game as an Educational Instrument for Children’s Motor Qualities Development. Revista Romaneasca pentru Educatie Multidimensionala, 16(2), pp.376–394.
4. Opstoel, K., Chapelle, L., Prins, F.J., De Meester, A., Haerens, L., Van Tartwijk, J. and De Martelaer, K., 2020. Personal and social development in physical education and sports: A review study. European Physical Education Review, 26(4), pp.797–813. https://doi.org/10.1177/1356336X19882054
5. Zhou, J.Y., Wang, X., Hao, L., Ran, X.W. and Wei, W., 2024. Meta-analysis of the effect of plyometric training on the athletic performance of youth basketball players. Frontiers in Physiology, 15, p.1427291.
6. Thompson, F., Rongen, F., Cowburn, I. and Till, K., 2022. The impacts of sports schools on holistic athlete development: a mixed methods systematic review. Sports medicine, 52(8), pp.1879–1917. https://doi.org/10.1007/s40279-022-01664-5
7. Guimarães, E., Baxter-Jones, A.D., Williams, A.M., Tavares, F., Janeira, M.A. and Maia, J., 2021. The role of growth, maturation and sporting environment on the development of performance and technical and tactical skills in youth basketball players: The INEX study. Journal of Sports Sciences, 39(9), pp.979–991. https://doi.org/10.1080/02640414.2020.1853334
8. Marshall, B., Loya, A., Drazan, J., Prato, A., Conley, N., Thomopoulos, S. and E. Reuther, K., 2021. Developing a STEM+M identity in underrepresented minority youth through biomechanics and sports-based education. Journal of Biomechanical Engineering, 143(4), p.041009. https://doi.org/10.1115/1.4047548
9. Kirby, L. and Amason, P., 2020. Academic success: Perceptions of student-athletes, learning specialists, and academic advisors. Journal of Higher Education Athletics & Innovation, 1(7), pp.33–60.
10. Chartier, M.C., Falcão, W.R. and Trottier, C., 2021. Student-athletes’ development of life skills in a university sports setting. International Journal of Coaching Science, 15(2), pp.3–30.
11. Ouyang, B. and Wu, R., 2022. Evaluation Model of Youth Basketball Training Performance Based on PSO Algorithm. Wireless Communications and Mobile Computing, 2022(1), p.1830318.

12. Zhong, S., 2022. Application of Artificial Intelligence and Big Data Technology in Basketball Sports Training. Wireless Communications & Mobile Computing (Online), 2022.
13. Xu-Hong, M., Hong-Ying, S. and Wei-Hong, S., 2022. Analysis of Basketball Technical Movements Based on Human-Computer Interaction with Deep Learning. Computational Intelligence and Neuroscience: CIN, 2022.
14. Khobdeh, S.B., Yamaghani, M.R. and Sareshkeh, S.K., 2024. Basketball action recognition is based on the combination of YOLO and a deep fuzzy LSTM network. The Journal of Supercomputing, 80(3), pp.3528–3553.
15. Lu, Y., Pareek, A., Lavoie-Gagne, O.Z., Forlenza, E.M., Patel, B.H., Reinholz, A.K., Forsythe, B. and Camp, C.L., 2022. Machine learning for predicting lower extremity muscle strain in National Basketball Association athletes. Orthopaedic Journal of Sports Medicine, 10(7), p.23259671221111742.
16. Javadpour, L., Blakeslee, J., Khazaeli, M. and Schroeder, P., 2022. Optimizing the best play in basketball using deep learning. Journal of Sports Analytics, 8(1), pp.1–7. http://dx.doi.org/10.3233/JSA-200524
17. Yao, J. and Li, Y., 2022. Youth sports special skills' training and evaluation system based on machine learning. Mobile Information Systems, 2022(1), p.6082280.
18. Wang, J., Zuo, L. and Cordente Martínez, C., 2024. Basketball technique action recognition using 3D convolutional neural networks. Scientific Reports, 14(1), p.13156.
19. Xue, X., 2024. The Construction Industry and Design of Basketball Teaching System Based on Digital Art and Intelligent Model.
20. Xuexiang, Y., 2024. A NEW MODE OF TRAINING SPORTS BASKETBALL PROFESSIONALS BASED ON THE IMPACT OF ECOLOGICAL ENVIRONMENT. Revista multidisciplinar de las Ciencias del Deporte, 24(96).
21. Wang, C. and Du, C., 2022. Optimization of physical education and training system based on machine learning and Internet of Things. Neural Computing and Applications, pp.1–16.
22. Yang, X., 2024. Construction of measurement index system of basketball players' specific physical fitness training based on AI intelligence and neural network. Molecular & Cellular Biomechanics, 21(1), pp.250–250.
23. Xu, T. and Tang, L., 2021. Adoption of machine learning algorithm-based intelligent basketball training robot in athlete injury prevention. Frontiers in Neurorobotics, 14, p.620378.
24. The dataset available online: https://www.kaggle.com/datasets/ziya07/youth-basketball-training/data
25. Hu, N. and Ma, B., 2024. Evaluation and optimization of basketball tactics training effect in physical education: Application research using Decision Tree (DT) algorithm. Molecular & Cellular Biomechanics, 21(2), pp.368–368.