# Personalized Training-Oriented Basketball Teaching Data Analysis and Intelligent Decision Support System

**Yu Lei***

School of Physical Education, Huazhong University of Science and Technology, Wuhan 430074, China

**Keywords:** Basketball teaching; Personalized training; Data analysis; Decision support system; Machine learning; Explainable AI; Wearable sensors

Citation: Lei Y.: Personalized Training-Oriented Basketball Teaching Data Analysis and Intelligent Decision Support System. Data Intelligence, 2026, Vol. 8, Art. No.: 20260102. DOI: https://doi.org/10.3724/2096-7004.di.2026.0102

## ABSTRACT

The rapid integration of artificial intelligence, the Internet of Things, and big-data analytics has transformed how basketball is taught, trained, and evaluated. However, conventional basketball instruction is still dominated by experience-based, one-size-fits-all programs that struggle to address heterogeneous athlete profiles, dynamic loads, and individual injury risks. To bridge this gap, this paper proposes a personalized-training-oriented data analysis and Intelligent Decision Support System (IDSS) for basketball teaching. The system fuses multimodal data—Inertial Measurement Unit (IMU) signals, video-based pose estimation, technical-tactical statistics, physiological metrics, and subjective wellness questionnaires—within a four-layer architecture comprising data acquisition, feature engineering, intelligent decision, and visualization. A hybrid recommendation engine that couples Improved Neural Collaborative Filtering with content-based athlete profiling is introduced for personalized training prescription, and an XGBoost-SHAP model is employed for interpretable performance prediction and lower-extremity injury risk estimation. A workload-balanced optimization formulation is presented to schedule training sessions under fatigue, skill, and time constraints. Empirical evaluation on a curated dataset of 120 collegiate basketball players over 18 weeks demonstrates that the proposed system improves shooting accuracy by 4.35%, reduces non-contact injury incidence by 28.6%, and yields a recommendation precision of 0.873—outperforming conventional baselines. The findings indicate that data-driven, explainable IDSS frameworks can serve as a credible companion to coaches and physical-education teachers in modern basketball pedagogy.

*Corresponding author: Yu Lei (E-mail: Ray9889@163.com, ORCID: 0009-0006-0265-4800).

© 2026 Chinese Academy of Sciences. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

## 1. INTRODUCTION

Basketball is one of the most widely practiced and televised sports in the world. In educational contexts—from secondary school physical-education (PE) classes to collegiate athletic programs—basketball serves both as a vehicle for general fitness and as a competitive discipline requiring sophisticated technical, tactical, physical, and psychological development [1-2]. Yet, traditional basketball teaching remains largely uniform: instructors deliver standardized drills calibrated to the average learner, and progression decisions hinge on the subjective judgment of a single coach who must simultaneously observe many players. This conventional model exhibits three systemic limitations.

First, learner heterogeneity is poorly accommodated. Adolescents and young adults differ markedly in anthropometric attributes, biomechanical efficiency, motor-learning rate, prior experience, and psychological readiness [3-4]. A drill that elicits productive overload in one student may induce undue fatigue or even injury in another. Second, training-load monitoring is often coarse—based on session duration or rating of perceived exertion (RPE)—which limits early detection of maladaptive responses such as autonomic-recovery suppression or kinematic asymmetries that precede injury [5-6]. Third, decision-making is generally retrospective: coaches review past games or videos to plan the next training cycle, leaving little room for in-session adaptation.

Recent advances in sports analytics offer a path forward. High-frequency tracking technologies—IMU-based wearables, optical multi-camera systems, and ball-in-shoe sensors—now generate rich, time-resolved descriptions of every dribble, jump, sprint, and shot [7-10]. In parallel, machine-learning (ML) and deep-learning (DL) models have demonstrated accuracy beyond 90% in tasks such as action recognition [7, 11], shot-outcome prediction [12-13], game forecasting [13-14], and injury-risk assessment [12, 15]. Decision support systems (DSSs) tailored to elite sport are also maturing: structured frameworks such as those proposed by Schelling and Robertson translate complex time-series data into actionable insights [16-17].

Although these technologies have proliferated, their concerted application to basketball teaching—as opposed to elite competition—remains nascent. Two specific gaps motivate the present work. (i) Most reported systems target a single function rather than orchestrating an end-to-end loop from data ingestion to individualized prescription [16-17]. (ii) Black-box predictive models fail to satisfy the explanatory needs of teachers and learners; without transparency, recommendations are unlikely to be trusted or followed [13, 18-19].

This paper aims to address these gaps. The principal contributions are as follows:

- We design a four-layer Intelligent Decision Support System (IDSS) tailored to basketball teaching, integrating multi-source data acquisition, ML-based analysis, optimization-driven decision, and human-centered visualization.
- We propose a hybrid personalized-training prescription engine that combines Improved Neural Collaborative Filtering (I-NCF) with content-based filtering on athlete profile vectors, mitigating the cold-start problem common in sport recommender systems [20-21].

- We adopt an XGBoost-SHAP scheme for interpretable performance and injury-risk prediction, producing global feature attributions for coaches and instance-level explanations for individual learners [13-14, 19].
- We formulate the weekly training-plan generation as a constrained optimization problem balancing skill development, recovery, and exposure to game-like scenarios.
- We empirically evaluate the system on 120 collegiate players over 18 weeks, showing significant improvements in shooting accuracy, non-contact injury rate, and adherence relative to conventional baselines.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 details the system architecture. Section 4 describes the multidimensional data analysis pipeline. Section 5 introduces the personalized training model. Section 6 elaborates the intelligent decision engine. Section 7 reports experiments and case studies. Section 8 discusses limitations.

## 2. RELATED WORK

### *2.1 Sports Analytics and Machine Learning in Basketball*

The application of ML to basketball began with shot-classification studies on the seminal SportVU optical-tracking dataset, where recurrent neural networks learned to predict three-point success solely from positional sequences and outperformed feature-rich static classifiers [12]. More recent work has extended these techniques across the basketball analytics landscape: gradient-boosted ensembles such as XGBoost have been combined with SHapley Additive exPlanations (SHAP) to predict NBA game outcomes with over 90% post-game accuracy and to surface the metrics—three-point percentage, offensive rebounds, assists—that most influence wins [13]. Stacked ensemble models that blend Naïve Bayes, AdaBoost, MLP, KNN, XGBoost, Decision Tree, and Logistic Regression learners have further improved interpretability [14]. Beyond game prediction, deep models such as LSTM and Transformer architectures have been compared on multi-season NBA datasets, with Transformers attaining the highest discriminative AUC and LSTMs offering better calibration on imbalanced data.

### *2.2 Personalized Training and Recommender Systems*

Personalization in sport draws heavily on recommender-system techniques developed for e-commerce and education. Felfernig and colleagues survey the field of sports recommender systems, identifying training-plan recommendation, equipment selection, and route discovery as the principal use cases, and pointing out that data sparsity and contextual relevance are the central challenges [21]. Constraint-based and case-based recommenders dominate the marathon-training literature, but for team sports collaborative filtering with deep representations has proven more effective. The Improved Neural Collaborative Filtering (I-NCF) algorithm of Tao integrates residual connections and attention to handle high-dimensional sparse data, and has reported state-of-the-art accuracy on football training-plan datasets [20]. In basketball, AI-driven personalized programs that fuse anthropometric, technical, and psychological inputs have been

shown to improve individual development outcomes more than conventional programs in randomized studies of Chinese youth athletes [22-23].

### *2.3 Injury Risk Modeling and Workload Monitoring*

Lower-extremity non-contact injuries account for nearly half of training injuries in basketball [15], creating a strong incentive to detect maladaptive responses early. Cohan and colleagues introduce METIC, a deep-learning model that predicts NBA injuries from longitudinal player data and significantly outperforms classical baselines on imbalanced data [24]. Lu and colleagues use ML to predict lower-extremity muscle strain in NBA athletes and identify recent back injury as the principal risk factor [12]. Machine-learning systematic reviews report that tree-based ensembles, support vector machines, and artificial neural networks achieve AUCs in the 0.7–0.87 range across diverse sports, with feature interpretation remaining a critical research need [6]. Wearable IMUs anchor most workload-monitoring pipelines: the validity of inertial wearables for measuring acceleration and deceleration has been verified against three-dimensional motion-capture references in indoor team sport [9], and recent female-basketball studies have correlated IMU-derived physical demands with technical statistics such as steals and turnovers [10].

### *2.4 Decision Support Systems in Sport*

Decision support frameworks structure raw data into actionable workflows. Schelling and Robertson articulate a five-stage DSS development pipeline for high-performance sport that has been adopted in basketball 'Most Demanding Scenario' analyses [16-17]. Fuzzy-logic DSSs have been advocated to handle the linguistic uncertainty inherent in fatigue and fitness assessment [3, 25], while big-data analytics frameworks for sports performance optimization emphasize visualization and role-specific interaction paradigms for coaches, physiotherapists, and athletes [26]. Despite this progress, the literature reveals comparatively few systems that explicitly target the teaching context—where pedagogical priorities (skill acquisition, motivation, safety) differ from those of pure performance maximization [2, 4, 27].

### *2.5 Computer-Vision and Pose Estimation*

Vision-based action recognition complements sensor-based pipelines, especially in classroom environments where wearables are impractical for every student. Multi-stream networks that integrate RGB, temporal, and contextual streams have been shown to substantially improve basketball pose-estimation robustness in occlusion-heavy scenes [18]. Pang and Zhang fuse computer vision and sports-biomechanics principles to characterize player actions with millimeter-level precision, providing kinematic features such as joint angle, segment velocity, and centre-of-mass trajectory [27]. These representations form the visual counterpart of the kinematic features extracted from wearables.

In summary, the literature provides strong but fragmented building blocks. The system described below integrates these components into a unified pedagogical loop.

## 3. SYSTEM ARCHITECTURE

### *3.1 Overall Framework*

The proposed Intelligent Decision Support System adopts a four-layer architecture:

- Data Acquisition Layer—Heterogeneous sensors (IMUs, wide-angle cameras, smart basketballs, heart-rate chest straps), match statistics, and electronic questionnaires feed raw signals into a Kafka-based message bus.
- Data Analysis Layer—A processing pipeline performs synchronization, denoising, segmentation, feature extraction, and multimodal fusion, producing the Athlete Feature Vector used by all downstream tasks.
- Intelligent Decision Layer—Three coupled engines (Personalization, Prediction, Optimization) generate prescriptions; each is paired with an explainability module that surfaces the rationale.
- Visualization & Interaction Layer—Web dashboards and mobile apps deliver role-specific views to coaches, students, and physical-education teachers, while a feedback channel returns adherence and outcome data.

### *3.2 Data Acquisition Layer*

The data-acquisition layer concurrently captures six categories of signals (see Table 1). IMU vests sample tri-axial acceleration, gyroscope, and magnetometer data at 100 Hz, capturing jumps, accelerations, decelerations, and changes-of-direction with concurrent validity acceptable for indoor team sport [9]. Optical capture uses two ceiling-mounted cameras at 60 fps for multi-view pose estimation following the multi-stream network of Sun et al. [18]. A smart basketball senses spin and impact, while heart-rate (HR) chest straps provide R-R intervals for downstream heart-rate-variability (HRV) analysis [5]. Match statistics—points, assists, rebounds, turnovers, steals, blocks, true shooting percentage—are imported automatically through a CSV API. Finally, daily wellness questionnaires (sleep quality, soreness, stress, mood) and weekly RPE values are entered through a mobile interface.

**Table 1.** Data sources and sampling characteristics.

| Data Type | Modality | Sampling Rate | Latency |
|---|---|---|---|
| IMU acceleration/gyroscope | Time-series | 100 Hz | ≤30 ms |
| Optical pose estimation | Image/keypoints | 60 fps | ≤80 ms |
| Smart-basketball | Time-series | 200 Hz | ≤40 ms |
| Heart-rate (R-R intervals) | Time-series | 1 Hz | ≤500 ms |
| Match statistics | Tabular | per match | manual |
| Wellness questionnaires | Categorical | 1/day | manual |

### 3.3 Data Analysis Layer

Within this layer, the synchronization service aligns multi-rate streams to a common timeline using a hardware trigger broadcast at the start of each session. Denoising employs the complementary filter for IMU orientation estimation, which has been shown to yield lower RMSE than low-pass filtering when validated against motion-capture ground truth [9]. Feature engineering extracts time-domain, frequency-domain, and biomechanical descriptors (mean acceleration magnitude, jerk, stride asymmetry, knee-valgus angle), aggregating them at multiple temporal scales (drill, session, week). Pose-based action recognition follows the the XGBoost-based action recognition approach of Acikmese et al., who report 96.9% overall accuracy on basketball actions using a single IMU vest and a 250-sample window with 75% overlap [7]. The fused Athlete Feature Vector $x_i \in \mathbb{R}^d$ (with $d = 87$ in our deployment) becomes the input to the decision layer.

### 3.4 Intelligent Decision Layer

The decision layer hosts three engines:

- Personalization Engine: hybrid recommender that prescribes drills.
- Performance Prediction Engine: XGBoost models of skill acquisition and injury risk.
- Optimization Engine: constrained scheduler that assembles prescribed drills into weekly plans.

All three engines share a common explainability module that uses SHAP to translate model outputs into natural-language statements rendered in the visualization layer.

### 3.5 Visualization & Interaction Layer

Different stakeholders see different views of the system, an interaction paradigm validated by recent big-data sports DSS literature [26]. Coaches view squad-level dashboards summarizing readiness, training adherence, and risk zones. Students see their personal trajectory, recommended drills, and progress badges. Teachers in PE classes obtain compact reports compatible with curriculum-evaluation rubrics. Feedback loops capture adherence, perceived effort, and qualitative comments, which are streamed back to the data-analysis layer for online model updating.

## 4. MULTIDIMENSIONAL DATA ANALYSIS PIPELINE

### 4.1 Athlete Feature Vector

Let $x_i = [p_i; t_i; l_i; m_i; w_i]$ denote the feature vector for athlete $i$, where $p_i$ holds anthropometric and demographic features (height, weight, age, position, dominant hand), $t_i$ the technical features (shooting percentage, dribbling tempo, defensive stance index), $l_i$ the load features (PlayerLoad, jump count, total acceleration, weekly acute-chronic workload ratio), $m_i$ the medical/physiological features (resting HR, LnRMSSD, prior injuries), and $w_i$ the wellness features (sleep, soreness, RPE). A z-score normalization is applied per feature using the cohort's distribution.

### *4.2 Workload Computation*

PlayerLoad PL is computed from tri-axial acceleration as Eq. (1) and Eq. (2):

$$PL_t = \sqrt{(a_{x,t} - a_{x,t-1})^2 + (a_{y,t} - a_{y,t-1})^2 + (a_{z,t} - a_{z,t-1})^2} \tag{1}$$

$$PL_{session} = \sum_{t_1=1}^{T} PL_t \tag{2}$$

while the acute-chronic workload ratio (ACWR), a widely used injury-risk indicator [6], is computed as Eq. (3):

$$\text{ACWR} = \frac{PL_{7d}}{PL_{28d}} \tag{3}$$

ACWR values outside the 0.8–1.3 range are flagged as elevated risk, consistent with recent female-basketball IMU literature [10].

### *4.3 Action Segmentation and Recognition*

Continuous IMU streams are segmented with a sliding window of 2.5 s and 75% overlap, then classified into seven basketball-specific actions: standing, walking, running, jumping, dribbling, set-shot, and jump-shot. Following Acikmese et al. [7], we benchmarked KNN, Decision Tree, Random Forest, AdaBoost, and XGBoost; XGBoost achieved the best F1 score on our cohort (0.952), confirming literature findings.

### *4.4 Pose-based Skill Quality Index*

For each shot or layup attempt captured on video, the multi-stream pose-estimation network of Sun et al. [18] returns 17 anatomical key-points. Skill quality is summarized by a Shooting Form Index (SFI) defined in Eq. (4):

$$SFI = w_1 \cdot s_{elbow} + w_2 \cdot s_{knee} + w_3 \cdot s_{wrist} + w_4 \cdot s_{release} \tag{4}$$

where each $s^*$ is a Gaussian similarity to a position-specific gold template derived from elite-player libraries, and the weights $w_k$ are estimated by Bayesian optimization to maximize correlation with shooting accuracy. Larger SFI values indicate forms more consistent with high-percentage shooters.

### *4.5 Multimodal Fusion*

To reconcile redundancy and complementarity across modalities, we apply a late-fusion strategy: each modality is processed by a modality-specific feature extractor, the resulting embeddings are concatenated, and an attention-weighted projection produces the fused representation. This approach has shown superior generalization to early fusion in basketball pose-estimation studies [18].

## 5. PERSONALIZED TRAINING MODEL

### 5.1 Problem Formulation

Let $U = \{1, \ldots, N\}$ be the set of athletes, $D = \{1, \ldots, M\}$ the catalogue of drills, and $r_{ij} \in \mathbb{R}$ an implicit feedback rating that combines completion rate, performance gain, and self-reported enjoyment for athlete $i$ and drill $j$. The personalization task is to predict missing entries of the matrix $R = [r_{ij}]$ and recommend the top-k drills that maximize learning while respecting fatigue and skill constraints.

### 5.2 Hybrid Recommendation

Pure collaborative filtering is undermined by data sparsity and the cold-start problem ubiquitous in sport [21]. We therefore combine I-NCF [20] with a content-based filter that exploits the Athlete Feature Vector $x_i$ and a Drill Feature Vector $y_j$ (capturing intensity, target skill, equipment).

The I-NCF component embeds athletes and drills into latent vectors $u_i,\ v_j \in \mathbb{R}^{32}$ and predicts ratings via a residual MLP with multi-head attention, as shown in Eq. (5):

$$r_{ij}^{NCF} = f_\theta(u_i \oplus v_j \oplus u_i \odot v_j) \tag{5}$$

where $\oplus$ denotes concatenation and $\odot$ element-wise product.

The content-based component computes cosine similarity between athlete and drill profiles in a shared metric space learned by a Siamese network, as given in Eq. (6):

$$r_{ij}^{CBF} = cos(\phi(x_i),\ \psi(y_j)) \tag{6}$$

The hybrid score blends both signals as in Eq. (7):

$$r_{ij} = \alpha \cdot r_{ij}^{NCF} + (1-\alpha) \cdot r_{ij}^{CBF} \tag{7}$$

with $\alpha$ tuned on a validation set; in our deployment $\alpha = 0.6$. Training minimizes a weighted binary-cross-entropy loss on observed and uniformly sampled negative pairs.

### 5.3 Drill Catalogue

The drill catalogue contains 312 items grouped into seven pedagogical categories (ball-handling, shooting, passing, defense, conditioning, tactics, recovery). Each drill is annotated with metabolic intensity, neuromuscular load, target skill, equipment, and minimum prerequisite level. Table 2 summarizes the catalogue.

**Table 2.** Drill catalogue summary.

| Category | # Drills | Median Duration (min) | Mean Intensity (1–10) | Equipment |
|---|---|---|---|---|
| Ball-handling | 58 | 8 | 5.2 | Ball |
| Shooting | 76 | 12 | 4.8 | Ball + Hoop |
| Passing | 41 | 10 | 4.5 | Ball |
| Defense | 35 | 9 | 6.6 | Cones |
| Conditioning | 44 | 14 | 7.9 | None |
| Tactics | 38 | 18 | 5.5 | Court |
| Recovery | 20 | 15 | 2.1 | Mat/Foam Roller |

### *5.4 Constrained Plan Generation*

Recommended drills must be assembled into a weekly plan $\Pi = \{\pi_1, \ldots, \pi_7\}$, where $\pi^d$ is the set of scheduled drills of drills on day d. Define a binary variable $z_j,^d \in \{0, 1\}$ equal to 1 if drill j is scheduled on day d. The optimization problem is given in Eq. (8):

$$max_z \sum_{d_1=1}^{7} \sum_{j_1=1}^{M} r_{ij} \cdot z_{j,d} \tag{8}$$

subject to the daily time budget constraint in Eq. (9):

$$\sum_j \tau_j \cdot z_{j,d} \le T_d, \quad \forall d \tag{9}$$

the daily load ceiling constraint in Eq. (10):

$$\sum_j \rho_j \cdot z_{j,d} \le L_d^{max}, \quad \forall d \tag{10}$$

and the ACWR safety-band constraint in Eq. (11) [6]:

$$0.8 \le \frac{\sum_d L_d}{L_i^{chronic}} \le 1.3 \tag{11}$$

where $\tau_j$ is the duration, $\rho_j$ the load contribution, $T^d$ the daily time budget, and $L^{d_{max}}$ the daily load ceiling. The problem is solved with a branch-and-bound mixed-integer solver; for typical parameter values it returns an optimal weekly plan in under 0.5 s on commodity hardware.

## 6. INTELLIGENT DECISION SUPPORT ENGINE

### *6.1 Performance Prediction with XGBoost*

Predicting short-term performance—for instance, will the player's free-throw percentage improve next week?—provides actionable feedback to coaches and learners. We train an XGBoost regressor on the

Athlete Feature Vector to predict the next-week change in three KPIs: free-throw percentage ΔFT%, three-point percentage Δ3P%, and assist-to-turnover ratio ΔA/TO. XGBoost is selected because it handles tabular data efficiently, accommodates missing values, and has shown the best accuracy among NBA outcome predictors [13-14]. Hyperparameters are tuned by Bayesian optimization (number of trees, max depth, learning rate, subsample, column subsample), aligning with prior reports in sport analytics [13].

### *6.2 Injury Risk Modeling*

We model the probability of a non-contact lower-extremity injury within the next 7 days as a binary classification task, drawing on the multimodal-fusion approach proposed by Tan et al. for youth basketball [15]. Class imbalance is addressed with SMOTE oversampling and focal-loss re-weighting. The classifier outputs a calibrated probability $p_i \in [0, 1]$; athletes with $p_i \geq 0.6$ are placed on a precautionary protocol that substitutes a recovery drill for the riskiest scheduled drill.

### *6.3 Explainability with SHAP*

For each prediction we compute SHAP values $\phi_k$ that allocate the model's output across input features $x_k$, satisfying Eq. (12):

$$y_i = \phi_0 + \sum_{k_1=1}^{d} \phi_{k,i} \tag{12}$$

Global SHAP plots reveal that jump count, prior ankle injury, ACWR, and LnRMSSD exhibit the largest mean absolute attributions for injury risk, consistent with the variables identified by Migliaccio et al. in their semi-professional cohort [5]. Instance-level explanations are converted into natural-language messages—for example, 'Risk increased by 12% because your jump count exceeded the weekly average by 23%; recovery work added on Day 4.'

### *6.4 Decision Rule Composition*

The three engines are coupled through a small set of business rules implemented in a rule engine. Examples include:

- IF $p_i \geq 0.6$ AND drill $j$ involves >50 jumps THEN substitute drill $j$ with a low-impact alternative of the same skill category.
- IF ΔFT% is predicted to decline by ≥2% THEN add a 15-minute form-shooting drill on at least two days of the upcoming week.
- IF wellness sleep score < 4/10 for two consecutive days THEN cap session intensity at ≤ 5/10.

Such rules act as guardrails, preventing the optimizer from overriding pedagogical or medical commonsense.

## 7. EXPERIMENTAL RESULTS AND ANALYSIS

### 7.1 Dataset and Experimental Protocol

We evaluated the system on a cohort of 120 collegiate basketball players (mean age 19.4 ± 1.6 years; 64% male) drawn from three universities over an 18-week competitive semester. Participants were randomly assigned to either the Treatment group ($n$ = 60), which used the proposed IDSS, or the Control group ($n$ = 60), which followed conventional coach-prescribed training. All participants wore IMU vests and HR chest straps during every training session. Match statistics were obtained from the league's official scoring system, while shooting accuracy was measured by standardized weekly tests. Ethical approval was obtained from the participating institutions; all participants provided informed consent.

### 7.2 Evaluation Metrics

We measured (i) shooting accuracy (free-throw and three-point percentages); (ii) lower-extremity non-contact injury incidence per 1000 hours of exposure; (iii) drill recommendation precision (Precision@5, Recall@5, NDCG@5) using a held-out split of athlete-drill interactions; (iv) injury-risk classifier performance (AUC, F1, sensitivity, specificity); and (v) qualitative adherence and satisfaction surveys.

### 7.3 Recommendation Performance

We compared the hybrid recommender against three baselines: User-based Collaborative Filtering (UCF), Matrix Factorization (MF), and pure I-NCF. Table 3 summarizes results.

**Table 3.** Recommendation performance (mean over 5-fold cross-validation).

| Method | Precision@5 | Recall@5 | NDCG@5 | Cold-start MAE |
|---|---|---|---|---|
| UCF | 0.612 | 0.487 | 0.604 | 1.42 |
| MF (BPR) | 0.704 | 0.561 | 0.683 | 1.18 |
| I-NCF | 0.821 | 0.674 | 0.781 | 0.93 |
| **Hybrid (Ours)** | **0.873** | **0.728** | **0.836** | **0.71** |

The hybrid model improves Precision@5 by 6.3% over I-NCF and reduces cold-start MAE by 23.7%, confirming that the content-based component effectively compensates for sparse interaction histories.

### 7.4 Performance and Injury Prediction

Evaluated on the held-out test set with the natural class distribution (approximately 5% positive rate), the classifier achieved AUC = 0.842, sensitivity = 0.74, and an F1 score of 0.29 for the minority (injury) class—a characteristically low value reflecting the class imbalance inherent in injury prediction tasks.

SMOTE oversampling was applied exclusively to the training fold in each cross-validation split to prevent data leakage. This is within the strong-performance band [6, 15]. Top-10 features contributing to the injury risk model can be seen in Figure 1.

| Rank | Feature | Mean \|SHAP\| | Relative Importance |
|---|---|---|---|
| 1 | Jump count (week) | 0.231 |  |
| 2 | ACWR | 0.187 |  |
| 3 | Prior ankle injury | 0.162 |  |
| 4 | LnRMSSD trend | 0.143 |  |
| 5 | Sleep score (3-day) | 0.120 |  |
| 6 | Knee-valgus index | 0.107 |  |
| 7 | Soreness (3-day) | 0.094 |  |
| 8 | Position (Centre) | 0.082 |  |
| 9 | Stride asymmetry | 0.073 |  |
| 10 | Age | 0.051 |  |

**Figure 1.** Top-10 features contributing to the injury risk model (mean absolute SHAP value).

### *7.5 Pedagogical Outcomes*

Across the 18-week intervention, the Treatment group outperformed the Control group on all pedagogical metrics (Table 4).

**Table 4.** Pedagogical outcomes after 18 weeks (mean ± SD).

| Outcome | Control | Treatment | Δ | *p* |
|---|---|---|---|---|
| Free-throw % gain | +3.6 ± 4.1% | +9.7 ± 4.8% | +6.1% | <0.001 |
| Three-point % gain | +1.8 ± 3.2% | +4.4 ± 3.9% | +2.6% | 0.002 |
| Shooting Form Index | 0.71 ± 0.10 | 0.83 ± 0.08 | +0.12 | <0.001 |
| Non-contact injury rate (per 1000 h) | 6.3 | 4.5 | −28.6% | 0.041 |
| Adherence to plan | 71% | 88% | +17% | <0.001 |
| Satisfaction (1–7 Likert) | 4.6 ± 1.1 | 5.9 ± 0.9 | +1.3 | <0.001 |

The 4.35% relative improvement in shooting performance (averaged across free-throw and three-point gains) and 28.6% reduction in non-contact injury incidence are practically meaningful. Adherence improvement is consistent with the engagement boost reported in technologically enhanced PE programs [2, 4].

The trajectory exhibits a near-monotonic increase from 60% to 83%, with the steepest gains in weeks 5–11 when form-shooting drills were most heavily prescribed.

### *7.6 Ablation Study*

To quantify the contribution of individual modules, we ablated each in turn (Table 5). Removing the content-based component degrades cold-start performance most strongly; removing the SHAP explainer degrades adherence most strongly, supporting the literature claim that explainability fosters trust and compliance [19].

**Table 5.** Ablation study.

| Configuration | Precision@5 | Adherence | Injury Rate (per 1000 h) |
|---|---|---|---|
| Full system | 0.873 | 88% | 4.5 |
| – Content-based filter | 0.821 | 84% | 4.7 |
| – SHAP explanations | 0.870 | 76% | 5.0 |
| – Constrained optimizer | 0.864 | 79% | 5.6 |
| – Injury rule engine | 0.872 | 86% | 6.2 |

### *7.7 Case Study*

Athlete A12 (point guard, 19 y, 1.83 m, 79 kg) entered the trial with FT% = 60%, 3P% = 28%, and a history of a mild ankle sprain six months prior. The system's profile encoded the prior injury and produced a plan front-loaded with form-shooting drills (12 sessions in weeks 1–4) and ankle-stability work, while limiting jump counts during weeks of high academic load (mid-term examinations). By week 12, FT% had risen to 78%, while ACWR remained within 0.9–1.2.

## 8. DISCUSSION

The results validate the hypothesis that an explainable, multimodal IDSS can elevate basketball teaching beyond the limits of conventional one-size-fits-all instruction. The system improves shooting accuracy and reduces injuries while increasing learner adherence—a triad rarely achieved simultaneously in sports pedagogy. Three lessons stand out.

**First, multimodal fusion is decisive.** No single modality—IMU, video, statistics, or wellness—is sufficient on its own. IMU data capture instantaneous kinematics but miss long-horizon technical drift; video reveals form but adds little about internal load; questionnaires reveal subjective state but are noisy and infrequent. Late fusion of complementary signals produces a richer Athlete Feature Vector and consequently better predictions [18, 26-27].

**Second, explainability is operationally important, not merely academic.** The ablation study shows that removing SHAP explanations degrades adherence more than it degrades raw model accuracy. Athletes

follow recommendations they understand; coaches accept tools whose reasoning they can audit. This finding echoes the broader scoping review of XAI in sports science [19].

**Third, personalized coaching is augmented, not replaced, by automation.** The IDSS recommends drills and surfaces risks, but final scheduling decisions remain with the coach. This human-in-the-loop design avoids the 'automation complacency' failure mode and respects the educator's pedagogical authority [16-17].

Several limitations should be acknowledged. The cohort, while diverse across institutions, is relatively small ($n$ = 120) and confined to collegiate-level athletes; generalization to youth basketball or elite professionals requires further study, ideally building on the multidimensional youth-basketball protocol of Liu et al. [23]. The drill catalogue, although extensive, was authored at participating universities and may embed local pedagogical biases. Additionally, our experiments used a single competitive semester; longitudinal effects across multiple seasons remain unexplored. Finally, the system depends on continuous wearable use, which raises practical issues of cost, comfort, and data privacy that schools must address.

Future work will (i) scale the cohort to thousands of learners across age groups, (ii) integrate large-language-model interfaces so that coaches can interrogate the system in natural language, (iii) extend the optimization layer to multi-team scheduling with travel and academic constraints [16], and (iv) investigate federated learning approaches that allow knowledge sharing across schools without exchanging raw athlete data.

## REFERENCES

[1] J. Wang, C. Li, and X. Zhou, "Decoding the court: Insights into basketball training and performance optimization through time-motion analysis," *Educ. Inf. Technol.*, vol. 30, pp. 1–30, 2024.

[2] L. Wang, J. Li, X. Chen, and Q. Liu, "Application of digital-intelligent technologies in physical education: A systematic review," *Front. Public Health*, 2025.

[3] R. M. Yousef and W. M. F. Ahmad, "Real-time athlete fatigue monitoring using fuzzy decision support systems," *Int. J. Comput. Intell. Syst.,* vol. 18, 2025.

[4] A. N. Abady, M. Zaki, and R. Suherman, "Technological innovations and pedagogical advancements in basketball skill learning: A systematic review of high school physical education," *Inspiree: Indones. Sport Innov. Rev.,* vol. 6, no. 1, 2025.

[5] G. M. Migliaccio, J. Padulo, and L. Russo, "An explainable machine learning approach to explain the effects of training and match load on ultra-short-term heart rate variability in semi-professional basketball players," *Sports,* vol. 13, no. 11, p. 392, 2025.

[6] H. Van Eetvelde, L. D. Mendonça, C. Ley, R. Seil, and T. Tischer, "Machine learning methods in sport injury prediction and prevention: A systematic review," *J. Exp. Orthop.,* vol. 8, no. 1, p. 27, 2021.

[7] Y. Acikmese, S. Gunay, and B. Yilmaz, "Action recognition in basketball with inertial measurement unit-supported vest," *Sensors,* vol. 25, no. 2, p. 563, 2025.

[8] A. Hoelzemann, J. F. Romero, M. Bock, K. Van Laerhoven, and L. Lu, "Hang-Time HAR: A benchmark dataset for basketball activity recognition using wrist-worn inertial sensors," *Sensors,* vol. 23, no. 13, p. 5879, 2023.

[9] L. S. Luteberget, B. R. Holme, and M. Spencer, "Player monitoring in indoor team sports: Concurrent validity of inertial measurement units to quantify average and peak acceleration values," *Front. Physiol.,* vol. 9, p. 1325, 2018.

[10] S. Bergeles, P. Vlachakis, and N. Stavropoulos, "Relationship between physical demands and player performance in professional female basketball players using inertial movement units," *Sensors,* vol. 24, no. 19, p. 6373, 2024.

[11] J. Liu, Y. Wang, and S. Liu, "Artificial intelligence technology in basketball training action recognition," *Front. Neurosci.,* vol. 16, 2022.

[12] Y. Lu, A. Pareek, O. Z. Lavoie-Gagne, et al., "Machine learning for predicting lower extremity muscle strain in National Basketball Association athletes," *Orthop. J. Sports Med.,* vol. 10, no. 7, p. 23259671221111742, 2022.

[13] X. Ouyang, Y. Zhang, S. Liu, and Y. Wu, "Integration of machine learning XGBoost and SHAP models for NBA game outcome prediction and quantitative analysis methodology," *PLoS ONE,* vol. 19, no. 7, p. e0307478, 2024.

[14] M. Y. Hung, R. Q. Ramos, and A. R. C. Donna, "Stacked ensemble model for NBA game outcome prediction analysis," *Sci. Rep.,* vol. 15, 2025.

[15] J. Tan, S. Lin, Z. Wang, and X. Wu, "The impact of sport-specific physical fitness change patterns on lower limb non-contact injury risk in youth female basketball players: A pilot study based on field testing and machine learning," *Front. Physiol.,* vol. 14, 2023.

[16] X. Schelling and S. Robertson, "A development framework for decision support systems in high-performance sport," *Int. J. Comput. Sci. Sport,* vol. 19, no. 1, pp. 1–23, 2020.

[17] X. Schelling and S. Robertson, "Decision support system applications for scheduling in professional team sport: The team's perspective," *Front. Sports Act. Living,* vol. 3, p. 678489, 2021.

[18] K. Sun, Y. Liu, and J. Cao, "Learning spatio-temporal context for basketball action pose estimation with a multi-stream network," *Sci. Rep.,* vol. 15, 2025.

[19] J. Schmid, M. Müller, and L. Donath, "A scoping review of explainable artificial intelligence in sports science," *Discov. Artif. Intell.,* vol. 5, 2025.

[20] Y. Tao, "Improved neural collaborative filtering algorithm for personalized training plan recommendations," *J. Circuits Syst. Comput.,* 2026. DOI: 10.1142/S0218126626500374.

[21] A. Felfernig, M. Stettinger, M. Atas, R. Samer, J. Nerlich, S. Scholz, J. Tiihonen, and M. Pilsl, "Sports recommender systems: Overview and research directions," *J. Intell. Inf. Syst.,* 2024.

[22] B. Li and W. He, "Research the role of artificial intelligence in developing personalized training plans to maximize the potential of basketball players," *J. Comput. Methods Sci. Eng.,* vol. 25, pp. 4833–4848, 2025.

[23] Y. Liu et al., "AI and big data personalized training protocol for Chinese youth basketball," *Sci. Rep., vol.* 15, 2025.

[24] A. Cohan, J. Schuster, and J. Fernandez, "A deep learning approach to injury forecasting in NBA basketball," *J. Sports Anal.,* vol. 7, no. 4, pp. 277–289, 2021.

[25] N. Pichainarongk and P. Chaiyapinunt, "Fuzzy decision support systems to improve the effectiveness of training programs in the field of sports fitness," *Int. J. Comput. Intell. Syst.,* vol. 17, 2024.

[26] M. T. Belova, "Big data analytics framework for decision-making in sports performance optimization," *Data,* vol. 10, no. 7, p. 116, 2025.

[27] H. Pang and D. Zhang, "Actions of basketball players integrating computer vision and sports biomechanics," *J. Comput. Methods Sci. Eng.,* vol. 24, pp. 395–409. 2025.