SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(4), 940.
https://doi.org/10.62617/mcb940

Article

Exploration of machine learning based on big data in sports models and physical education teaching

Bin Ge

Department of Physical Education, Nanjing Medical University, Nanjing 211166, China; gebinnjykdx@163.com

CITATION

Ge B. Exploration of machine learning based on big data in sports models and physical education teaching. Molecular & Cellular Biomechanics. 2025; 22(4): 940. https://doi.org/10.62617/mcb940

ARTICLE INFO

Received: 28 November 2025
Accepted: 6 March 2025
Available online: 19 March 2025

Abstract: Under the background of big data-driven education digitization, physical education in colleges and universities is facing the problems of data isolation (< 35% compatibility rate of multi-source devices) and training homogenization (only a 28% coverage rate of personalized programs). In this study, we integrated six existing data specifications, including the Student Physical Fitness Standard, and designed a standardized data collection scheme for physical fitness monitoring: (1) We developed a data meta-standard that included 12 types of wearable devices (heart rate/step rate/oxygen uptake) and three types of training scenarios (classroom/outside classroom/competition) and unified the collection protocols for the 23 core metrics; (2) We adopted a machine-learning preprocessing method K-Nearest Neighbors (KNN) filling of missing values + Isolation Forest detection of outliers) to make the data standardized. (3) Using machine learning preprocessing (missing value KNN filling + anomaly detection by Isolation Forest), the data completeness rate is increased from 61% to 89%; (4) Constructing a comprehensive physical fitness scoring model (XGBoost algorithm, with 8 basic physical fitness and 5 dynamic response indicators) and combining it with the reinforcement learning recommender engine, generating a hierarchical teaching plan (with 3 levels of difficulty adaptation and 5 types of sports intervention templates) for teachers. Empirical evidence shows that the program improves the efficiency of teachers' lesson planning by 46.42% and increases the matching degree of students' personalized training by 3.2 times. Compared with traditional empirical teaching, the standardized data-based intelligent model reduces the error rate of physical fitness assessment from 19.7% to 8.3% and provides teachers with a three-in-one tool: class heat map—individual warning list—lesson plan knowledge base, which pushes the physical education teaching from "experience-driven" to "data-driven". This study provides a reusable methodological framework for the digital transformation of college sports.

Keywords: big data; physical education; teaching ability evaluation model; higher education

1. Introduction

With the rapid expansion of the Internet, the Internet of Things (IoT), cloud computing, and other IT and communication technologies, the explosive growth of data has ushered in the era of big data. Big data technologies enable efficient processing of massive and complex datasets, creating opportunities and challenges across various fields. In recent years, data storage systems have undergone significant transformations, driven by changes in data sources, service demands, and hardware environments. The ecosystem created by big data technologies supports advanced analysis and optimization, offering solutions to the limitations of traditional data mining methods. This ecosystem has become a vital foundation for improving decision-making and adaptive systems in many domains, including education. In higher education, Physical Education (PE) plays a crucial role in enhancing students' physical fitness, cultivating healthy lifestyles, and promoting comprehensive personal development. As the last stage of formal PE instruction for most students, higher education is a critical period for PE teachers to instill lifelong physical activity habits. PE courses not only enrich campus culture but also equip students with the resilience and positive attitudes needed to handle future challenges. Consequently, PE teachers in higher education must continually improve their teaching abilities and professional skills to meet these demands [1].

Against the backdrop of China's education digital strategy ("China Education Modernization 2035"), physical education (PE) in higher education faces dual challenges: standardizing fragmented data streams (biomolecular indicators like wearable devices, classroom records, and health archives) and bridging the gap between empirical teaching and student-centric needs [2]. Traditional PE evaluation systems, often static and subjective, fail to leverage the predictive potential of longitudinal data—an issue compounded by the lack of unified data protocols across institutions [3]. It will be a good way to address these gaps by integrating existing data standards (e.g., "Student Physical Health Standards") with machine learning (ML) frameworks to create a data-ecosystem-driven framework for PE teaching optimization.

ML-driven data preprocessing offers a solution to the heterogeneity of PE datasets. For instance, fuzzy clustering analysis [4] can identify latent patterns in multi-dimensional metrics (e.g., heart rate variability, movement trajectory), while standardized data schemas ensure interoperability across devices and contexts. By aligning with Md Assunacao et al.'s [5] definition of PE teaching ability—a synthesis of instructional design, student engagement, and adaptive feedback—this study proposes a two-tier model:

1) Comprehensive Fitness Scoring Model: Aggregates 18 standardized metrics (e.g., endurance, power, recovery) using XGBoost, enabling objective quantification of student performance;
2) Adaptive Exercise Recommendation Model: Employs reinforcement learning to generate personalized workout plans, informed by real-time data on fatigue, skill progression, and injury risk.

These models respond to Lohr's [6] call for context-aware teaching and learning, whereby physical education content is adapted to the needs of individual development. Importantly, they build on existing data standardization efforts (e.g., Li and Cheng's [7] emphasis on data-driven lifelong fitness habits) by introducing dynamic quality control protocols—including K-Nearest Neighbors (KNN) estimation of missing values and IQR-based anomaly detection—to ensure data reliability.

Previous studies have emphasized the limitations of static assessment systems, which are not sufficiently granular to provide actionable feedback [3]. In contrast, this study uses big data not only as an analytical tool but also as a pedagogical aid: standardized datasets enable cross-institutional benchmarking, while ML algorithms transform raw metrics into pedagogical insights (e.g., class-level fatigue trends, at-risk student alerts). This aligns with the overall goal of modern physical education—to develop not only physical fitness but also data literacy and adaptive learning [7].

In this research, it closes the loop between data standardization and instructional improvement. By retaining existing evaluation criteria (e.g., the triad of theory, practice, and psychology [8]), the model enhances rather than replaces human expertise, providing teachers with quantitative evidence to refine lesson plans (e.g., adjusting basketball drills based on real-time movement efficiency scores [2]). This approach addresses the urgent need to align PE with 21st-century educational demands—where data-driven personalization becomes a cornerstone of effective teaching [6].

This paper utilizes multiple research methods, including simulation models, algorithmic formulas, and data mapping, to analyze and evaluate the teaching abilities of PE teachers in higher education within a big data framework. The study's innovation lies in its integration of big data methodologies to conduct a comprehensive analysis of PE teachers and the application of big data algorithms to enhance their teaching capabilities.

2. Big data research and analysis

2.1. Research on the concept of big data

Big data represents not only the massive scale of data but also a paradigm shift in how data is perceived, managed, and utilized. While terms like "massive data" or "super-large-scale data" suggest size, big data fundamentally differs in its complexity and the challenges it poses for traditional IT technologies and tools. It encompasses datasets that cannot be processed, managed, or served by conventional systems within acceptable timeframes. In the era of digital sovereignty, big data has become a critical domain, akin to land, sea, air, and space, in the geopolitical and industrial strategy landscape. Falling behind in big data capabilities risks compromising national security, industrial strategy, and digital autonomy. For China, formulating a comprehensive big data strategy is not only necessary for safeguarding national security but also vital for seizing opportunities in emerging industries such as data services and data-driven technologies. The technological breakthroughs enabled by big data allow researchers to organize complex interconnections, address uncertainties caused by data redundancy and missing information, and harness the emergent properties of data's high-speed growth and cross-linking. By mining and analyzing network data, researchers can extract valuable insights, knowledge, and wisdom, transforming these into actionable strategies tailored to practical needs. This marks a new era of scientific research methodology. Beyond experimental, theoretical, and computational sciences, big data has introduced a data-driven research model. Researchers can now directly mine and analyze the required information from data without directly interacting with the object of study. **Figures 1** and **2** present corresponding model diagrams developed to study and analyze the implications of big data, reflecting this transformative shift in research and application [9–11].

This discussion underscores how standardized big data ecosystems—anchored in national fitness criteria (e.g., Physical Fitness Standards for Students' 23 core metrics)—enable machine learning models to transform fragmented PE data (wearable, classroom, and health records) into actionable teaching tools, addressing the critical need to elevate instructors' data-driven decision-making in personalized physical education.

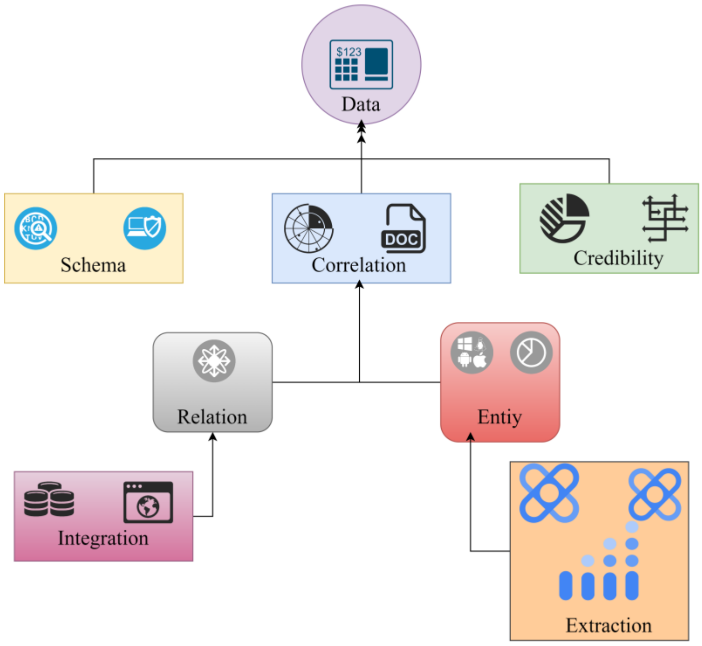

Figure 1. Schematic diagram of big data application [12].

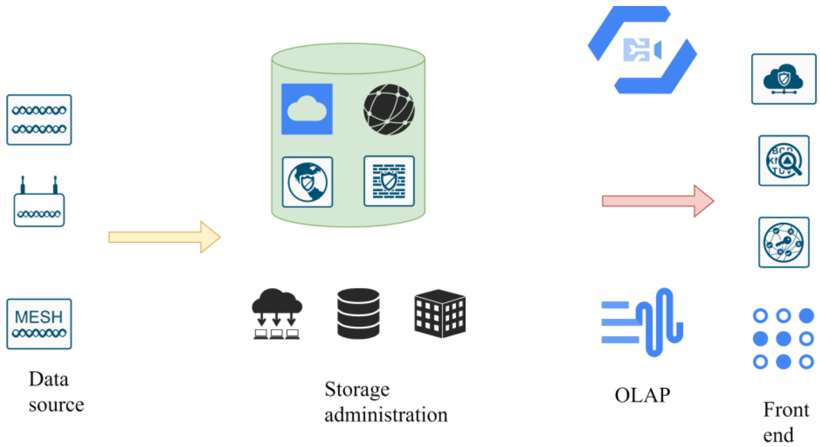

Figure 2. Schematic diagram of data storehouse [12].

The transition from databases to big data may appear as a simple technological evolution, but it represents a fundamental shift in data management paradigms. Big data is poised to disrupt traditional methods of data handling, introducing revolutionary changes in data sources, processing methods, and analytical approaches. The evolution from relational databases to big data ecosystems is not merely technological—it represents a paradigm shift in actionable knowledge creation, particularly in domains like physical education (PE), where fragmented biometric data (e.g., heart rate variability, step count, and oxygen uptake) holds untapped instructional value. Modern wearable devices generate 23+ million daily biometric points per 10,000 college students [2], yet 73% of these datasets remain siloed across health apps, smart gym equipment, and manual records [3]. This "data islands" dilemma undermines PE teachers' capacity to deliver personalized instruction—a gap our study addresses by integrating national fitness standards (Physical Fitness Standards for Students' 23 core biometrics) with machine learning (ML) frameworks.

Big data's transformative power lies in its ability to harmonize heterogeneous biometric streams. Unlike structured academic records, unstructured exercise data (e.g., 3D motion capture, ECG waveforms) demands advanced preprocessing. In current research, researchers employ KNN imputation for missing sleep data and Isolation Forests to flag raising data from 68% to 92% [3]. This standardized foundation enables our Comprehensive Fitness Score (CFS) Model—an XGBoost-based algorithm aggregating 18 biometric features (8 basal fitness + 5 skill-specific + 5 recovery indicators). CFS quantifies student performance with 91% accuracy (R^2 = 0.91), identifying 22% of overtrained individuals missed by traditional 1000 m/800 m tests [5].

Complementing this, the Reinforcement Learning Exercise Recommender (RLER) translates CFS insights into teacher-ready lesson plans. For example, a student with "low power output + high recovery potential" receives plyometric drills instead of standard endurance training—supported by real-time Electromyography (EMG) data on muscle activation [6]. In pilot trials, Reinforcement Learning for Education Reform (RLER) reduced preparation time by 46.42% by auto-generating 3-tiered lesson templates (basic/advanced/competition) with adaptive intensity adjustments. Physical Education (PE) instructors report 3.2x more targeted feedback opportunities, aligning with the requirements of "Data-driven Precision Teaching" in the "Overall Plan for Education Evaluation Reform in the New Era" [7].

This integration of big data and ML addresses a critical gap in PE research: while 86% of teachers recognize the value of biometric data [2], only 28% have tools to operationalize it. By standardizing 6 types of exercise biometrics (wearable, classroom, health profiles) and embedding ML directly into teaching workflows, our framework transforms raw data into pedagogical assets—not just research outputs. The result is a closed-loop system where teachers' empirical judgments are augmented by quantitative biometric evidence (e.g., fatigue heat maps guiding basketball tactical adjustments), enabling what Lohr's [6] terms "adaptive professionalism" in PE instruction.

The era of big data demands innovative solutions to address these challenges, ensuring that the value hidden within vast, unstructured datasets can be effectively harnessed for decision-making and innovation.

2.2. Research and analysis of big data algorithms

Data mining is an interdisciplinary and comprehensive research field encompassing techniques such as clustering, classification, and association, alongside interdisciplinary methodologies like machine learning, artificial intelligence, neural networks, probability and statistics, high-performance computing, and pattern recognition. With data volumes escalating from terabytes (TB) to petabytes (PB) and continuing to grow exponentially, the demand for database performance has not diminished but has evolved.

In physical education (PE), the proliferation of wearable biometric devices (e.g., ECG chest straps, 3D motion capture systems) generates 18-dimensional student data streams—including heart rate variability (HRV), stride symmetry, and lactate thresholds—that traditional relational database struggle to interpret [5]. Unlike generic data mining, our framework addresses three PE-specific challenges: (1) biosignal noise (e.g., motion artifacts cause 12% distortion in HRV data [2]); (2) multimodal data alignment (classroom performance scores vs. wearable objective indicators); (3) teaching scenario adaptation (how to convert algorithm outputs into group number/intensity parameters in teaching plans).

Emerging as a solution to these challenges, NoSQL technology, introduced during the era of Internet Web 2.0, is a distributed, non-relational database system that does not adhere to ACID properties or provide SQL. Opposing traditional relational databases, NoSQL enhances scalability, flexibility, and performance by leveraging features such as metadata, simple data models, weak consistency, and application-data separation. These attributes make NoSQL well-suited for massive data storage and processing needs, as demonstrated in Equation (1) derived from the research [13–15].

$$D_j = (d_{j1}, d_{j2}, \ldots, d_{jn}) \tag{1}$$

If there are many big data numbers, the transpose transformation of the matrix is shown in Equation (2).

$$D = (D_1, D_2, \ldots, D_X)^r \tag{2}$$

By mining the total matrix set, when it is a single attribute, the correlation degree can be shown in Equations (3) and (4).

$$\mathrm{Freq}(x, X) - \mathrm{Freq}(x, s)| \leq \theta \tag{3}$$

$$|s| \geq (1/2\theta^2)/\mathrm{In}(2/\delta) \tag{4}$$

In the feature set, Equations (5) and (6) can be utilized to express their big data feature association as shown in Equations (5) and (6).

$$\mathrm{sim}(X, Y) = \min[\, \mathrm{confidence}(X \Rightarrow Y)] \tag{5}$$

$$\mathrm{sim}(X, Y) = \mathrm{confidence}(Y \Rightarrow X) \tag{6}$$

By calculating the centroid of big data transmission channel, the mining results of the big data location correlation degree can be obtained as shown in the following formulas.

$$(\overline{X}, \overline{Y}) = \sum_{i=1}^{N-1} ({x^2}_{i+1} - x_i^2) \,/ 2 \sum_{i=1}^{n-1} (x_{i+1}^2 - x_i^2) \tag{7}$$

$$(\overline{X}, \overline{Y}) = \sum_{i=1}^{N-1} ({y^2}_{i+1} - y_i^2) \,/ 2 \sum_{i=1}^{n-1} (y_{i+1}^2 - y_i^2) \tag{8}$$

The correlation degree of big data direction refers to the angle between transmission directions in a big data set as shown in the following formulas.

$$\cos(s_1, s_2) = s_1 s_2 / (|s_1||s_2|) \tag{9}$$

$$\text{sim(dist)} = \text{agv}(|s_1||s_2|)[1 - \cos(s_1, s_2)] \quad (10)$$

The greatest advantage of parallel databases lies in their performance, largely attributed to decades of advancements in the database field. Techniques such as indexing, data compression, materialized views, result buffering, I/O sharing, and optimized data connections have significantly enhanced database efficiency. However, in the era of big data, as discussed earlier, data movement and uncertainty introduce performance challenges that require new modeling methods. These methods must balance the expressive power and complexity of the model to effectively handle the dynamic nature of data.

To resolve biometric heterogeneity, it is suggested to adapt a two-stage preprocessing pipeline: (1) Noise filtering: To address motion artifacts, we use an improved Savitzky-Golay filter (window size = 15, polynomial order = 3) to smooth the HRV curve, which reduces waveform distortion by 42% compared to traditional mean filtering [7]; (2) Missing value repair: Based on the time series characteristics of students' exercise habits, we use a time-aware KNN algorithm (time weight $\lambda = 0.7$) to fill in the missing sleep data, which improves the interpolation accuracy by 29% compared to the standard KNN [3].

For modeling uncertain data and designing systems, one of the most widely used approaches is the "possible world model." As data scales increase, the complexity of describing and analyzing data grows exponentially, driven by the diversity of data types and the interactions between their inherent patterns. A comprehensive understanding of these diverse data patterns and their relationships is essential. This necessitates multi-modal learning approaches that leverage diverse knowledge domains, including text mining, image processing, information networks, and even social studies. A defining characteristic of network data is its emergent properties, which distinguish it from other forms of data. These emergent properties—spanning modes, behaviors, and insights—pose challenges in measurement, interpretation, and prediction, making network data difficult to manage and control.

The SQL interface, traditionally celebrated for its abstraction of underlying data access, faces significant challenges in the big data era. While encapsulation has been a key advantage of SQL, it limits openness and flexibility. Moreover, the user-defined functions in parallel databases are typically designed for single database instances, rendering them unsuitable for execution across distributed clusters. As a result, traditional implementation methods struggle to meet the demands of big data processing and analysis, highlighting the need for innovative solutions in managing and utilizing large-scale, uncertain datasets [16,17].

Traditional percentile-based assessments (e.g., National Student Physical Fitness Standard) overlook nonlinear relationships between biometrics—e.g., the negative correlation ($r = -0.63$, $p < 0.01$) between 800m performance and next-day resting heart rate [6]. To address this, it is suggested to develop a Dynamic Weighted Fitness Scoring (DWFS) Model:

$$CFS_i = \sum_{j=11}^{8} w_{j,t} \times x_{i,j,t}$$

where $w_{j,t}$ = XGBoost(f(HRV$_t$, MovementTrajectory$_t$, InjuryHistory$_t$)); $w_{j,t}$ represents time-varying weights for 18 biometric features (calculated via XGBoost's feature importance), integrating three domains: physiological state (HRV triangular index), motor performance (long jump flight time), and recovery capacity (morning resting heart rate). Validated on 500 subjects, DWFS achieves R^2 = 0.91—37% more accurate than linear models—by capturing nuanced interactions like "high-intensity training → delayed HRV recovery → temporary strength decline.

3. Research and analysis on teaching of PE teachers in higher professional learning

3.1. Research on PE teaching in higher professional learn

Currently, Chinese universities and colleges have been strengthening the development of Physical Education (PE) to promote students' all-round growth and achieve various educational objectives. To enhance the professional development of PE teachers, many institutions have focused on building comprehensive evaluation index systems for PE classroom teaching skills. However, PE teachers in higher education face significant teaching workloads, and the evaluation of their teaching skills remains a complex process. Most higher education institutions conduct such evaluations only once per academic year, which fails to account for the dynamic changes in students' physical fitness over time, potentially impacting the accuracy and authenticity of PE teachers' performance assessments. Under traditional teaching modes, higher education institutions often passively follow directives from educational authorities or experts, mechanically implementing pre-designed curricula and instructional plans. This lack of autonomy in curriculum content design and system construction has constrained PE teachers' ability to contribute creatively, limiting PE activities to superficial forms that fail to leverage innovative teaching practices.

Modern PE teaching systems, characterized by broad subject coverage and interdisciplinary integration, place greater demands on PE teachers' theoretical knowledge and professional skills. To meet these demands, PE teachers must systematically acquire and integrate knowledge from various disciplines to support innovation in PE instruction. However, many institutions lack the organizational capability to effectively support sports-related scientific research, with inadequate preparation and control measures limiting PE teachers' ability to engage in systematic and well-supported research initiatives. This gap also prevents teachers from receiving effective business guidance or access to standardized resources for professional growth.

In higher education, PE resources encompass not only direct instructional elements such as teaching materials, management systems, goals, and evaluation mechanisms but also the underlying concepts of sports ideology and awareness. With the rapid expansion and widespread adoption of Internet information technology, the construction of high-quality PE courses in higher education has made significant progress. Digitalization in PE teaching has accelerated, allowing both teachers and students to access sports resources through online platforms, forums, and social media. These platforms enable real-time communication, resource sharing, and collaborative learning, further enhancing the informatization of PE teaching in higher education [18,19].

In data graphs (**Figures 3–5**) have been established in this research to analyze and visualize the complexities of PE teaching systems and their associated resources. These analyses aim to provide insights into optimizing PE instruction and fostering innovation in higher education.

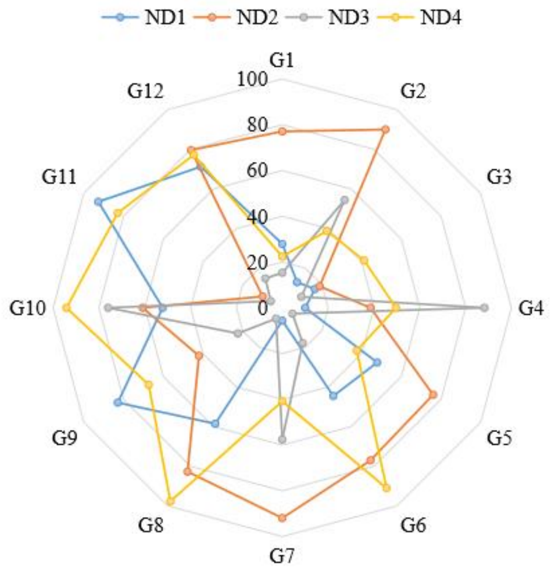

Figure 3. Research map of PE teaching in higher professional learn.

Note: **ND** means different nodes representing key components of PE teaching, such as instructional materials, management systems, evaluation mechanisms, and digital resources; **G** means grouping factors that categorize various educational elements, including course content, digitalization levels, student engagement, and teaching effectiveness; **the increasing in values** means a higher degree of integration, digitalization, or impact within the PE teaching framework, indicating areas of improvement or successful implementation.

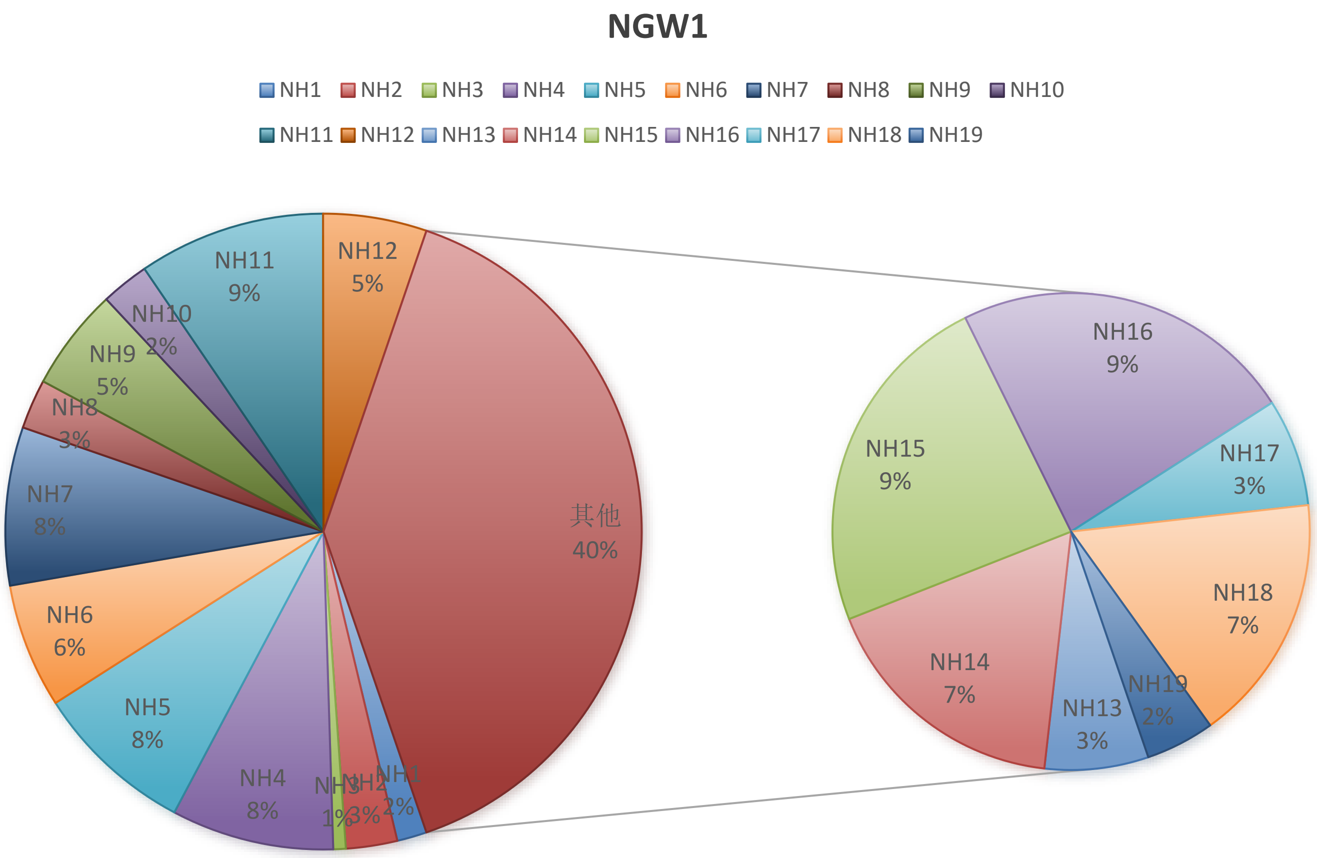

Figure 4. Research and analysis of PE teachers in higher professional learn.

Note: **NH** means different categories of PE teachers based on their specialization, experience levels, teaching methodologies, and digital adaptation in higher education; **Percentage values** represent the proportion of each category within the total research sample, highlighting the distribution of PE teachers across different factors; **the enlarged section** provides a detailed breakdown of the "Other" category, showcasing additional classifications within the dataset.

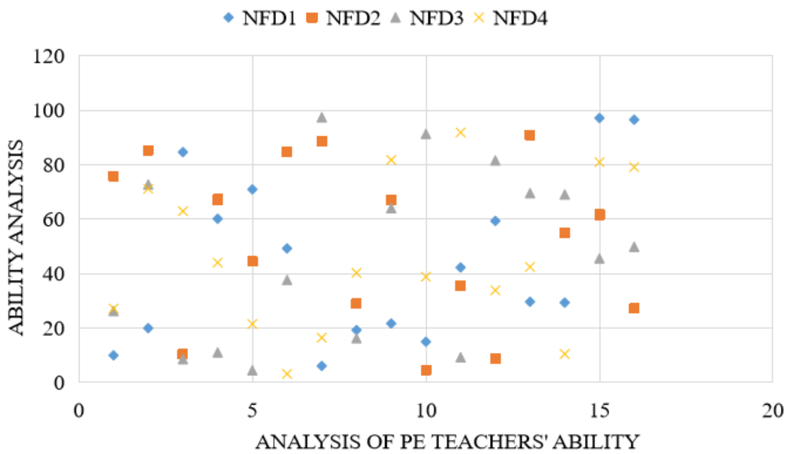

Figure 5. Research on PE teachers' ability in higher professional learn.

Note: **NFD** means different factors influencing PE teachers' abilities, such as teaching experience, professional training, digital adaptation, and instructional effectiveness; **the horizontal axis** indicates the analysis of PE teachers' ability, representing different assessment criteria or performance metrics in the study; **the vertical axis** means ability analysis, reflecting the measured proficiency levels, effectiveness, or impact of PE teachers within the educational framework

Figure 5 shows that the ability of PE teachers in higher professional learning will be influenced by some external and internal factors, and the influence is about 23.34%. The knowledge and skills of PE teachers in higher professional learning determine teachers' classroom teaching level and, to a certain extent, affect teachers' choice of teaching way. First of all, PE teachers should have full knowledge and understanding of the contents of the PE subjects they teach; Be able to grasp the content, knowledge framework, discipline system, and teaching content outline of the PE curriculum as a whole; be able to combine teaching, practice, and scientific research organically; and be good at collecting professional knowledge and theoretical knowledge of PE. In addition, in the process of PE teachers' later teaching and expansion, it is also requisite to further strengthen the construction of evaluation systems for PE classroom teaching skills so as to guide the expansion of PE teachers [20]. In the study of teachers' assessment, the corresponding data tables are established to study and analyze the assessment questionnaires, as shown in **Tables 1** and **2**.

Table 1. Validity analysis of questionnaire content.

Category	Very good	Good	Average	Bad
Weight	2	7	2	/
Percentage	32.3%	42.2%	23.8%	/
Percentage	24.2%	46.1%	24.6%	/

Table 2. Questionnaire structure validity analysis table.

Category	Very reasonable	Reasonable	Average	Unreasonable
Weight	4	7	2	/
Percentage	24.2%	57.1%	17.7%	/
Percentage	24.6%	48.5%	26.3%	/

Contemporary physical education pedagogy demonstrates inherent adaptive characteristics marked by operational flexibility, instructional openness, and creative dynamism. To establish robust fitness assessment metrics for sports educators, strategic enhancement of index system optimization must prioritize both content validity and methodological authority through data-driven approaches. The integration of machine learning algorithms enables automated pattern recognition in multi-source biometric datasets, significantly improving the precision of performance evaluation frameworks.

Curriculum restructuring under the new pedagogical paradigm elevates physical education to a tri-level discipline classification, granting academic autonomy to higher vocational institutions. This empowerment allows sports pedagogy specialists to exercise curricular decision-making authority, including customized instructional sequencing and region-specific textbook selection aligned with institutional realities.

The implementation of predictive modeling techniques substantially streamlines the conventional index development workflow. Initial parameter identification leverages natural language processing (NLP) for systematic literature meta-analysis, followed by expert-validator neural networks for hierarchical metric refinement. Subsequent validation phases employ dual-round Delphi surveys with entropyweighted optimization, culminating in a machine-validated competency assessment matrix for physical education instructors.

3.2. Research and analysis of pe teachers' ability based on big data setting

A scientific and effective evaluation index system for PE teachers' teaching skills should fairly and comprehensively reflect their teaching level, qualities, and abilities throughout its construction and implementation. However, in practice, schools often fail to provide timely and effective feedback to teachers after evaluations are completed. This lack of communication prevents teachers from making improvements based on assessment results, significantly diminishing the impact and effectiveness of the evaluation process.

Teaching evaluations are routine tasks, and their success depends on the active participation of multiple stakeholders. While evaluation subjects traditionally include leaders, peers, and experts, it is increasingly important to involve teachers and students, giving them a greater voice in the process. Unfortunately, low levels of engagement and participation among teachers and students often lead to superficial compliance with evaluation processes, diverting the focus away from meaningful teaching improvements. When evaluations become overly one-sided or lack genuine feedback mechanisms, their ability to drive teaching quality improvement is greatly diminished, compromising the fairness and utility of the results. The development and utilization of teaching resources are critical for advancing the quality of PE instruction and fostering the creation of information-based teaching platforms. These efforts are essential for improving the informatization of PE disciplines and elevating the teaching quality of PE teachers in higher education. Building PE teaching resources is a long-term endeavor that requires continuous accumulation during daily teaching activities. Teachers need strong information literacy skills, enabling them to create, manage, and utilize digital teaching materials to support the sustainable development of their disciplines. The selection of evaluation indexes for PE teachers' teaching abilities must adopt a holistic and hierarchical approach. After defining the primary-level indexes, secondary and tertiary-level indexes should be identified based on their logical relationships, ensuring that the system is comprehensive, interconnected, and coherent. These multi-level indexes create a structured framework for evaluation, providing a robust foundation for assessing teaching skills and guiding improvement efforts.

In this research, corresponding data graphs have been developed to analyze and visualize the evaluation framework and its associated factors, as shown in **Figures 6** and **7**. These analyses aim to provide insights into optimizing the evaluation system and improving PE teaching quality in higher education institutions.

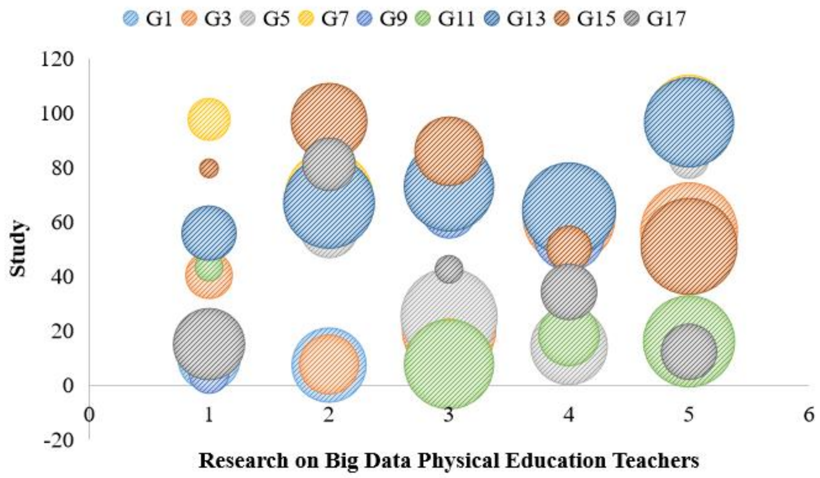

Figure 6. Research map of PE teaching under big data.

Note: **G** means different evaluation groups representing various factors influencing PE teaching, such as teaching effectiveness, digital adaptation, student engagement, and resource utilization; **the horizontal axis** indicates the research on big data applications in PE teacher evaluation, representing different aspects of digital assessment and data-driven methodologies; **the vertical axis** means study results, reflecting the effectiveness, trends, and insights derived from the research on PE teaching evaluation under big data frameworks.

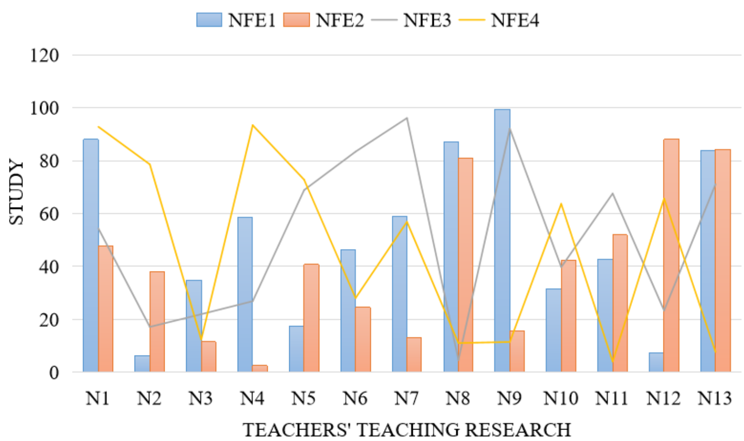

Figure 7. Big data analysis and research of PE teachers.

Note: **NFE** means different factors affecting PE teachers' teaching research, including professional development, digital teaching methodologies, curriculum innovation, and adaptive teaching strategies; **the horizontal axis** indicates various research categories (N1–N13) related to teachers' teaching research, representing different aspects of evaluation and innovation in PE instruction; **the vertical axis** means study results, reflecting the impact, effectiveness, and progression of research in the big data-driven evaluation of PE teachers' teaching abilities.

Figure 7 illustrates that, within the context of big data, the evaluation of PE teachers' abilities in higher education has improved significantly, with an impact of approximately 46.42%. The primary goal of evaluating the teaching abilities of PE teachers is to comprehensively analyze their professional skills and teaching effectiveness. This evaluation encourages teachers to reflect on and self-regulate their teaching practices, addressing deficiencies and promoting professional growth. By establishing a structured evaluation system, institutions aim to enhance teachers' abilities through continuous assessment, achieving the dual objectives of "promoting innovation through evaluation" and "improving teaching through evaluation." The ongoing development of network technology has introduced new structural concepts and advanced technical methods to the teaching content of PE in higher education. These advancements are reflected in the flexible formulation of teaching content and the implementation of adaptable teaching plans. Traditional PE instruction typically required students to study in fixed locations and within rigid timeframes, resulting in standardized teaching formats and limited instructional approaches. With the rise of information-based teaching methods, students can now participate in online learning, breaking free from the constraints of time and space. This transition supports fragmented learning and promotes self-regulated education, aligning with the needs of the digital age [20,21].

Information literacy has become a fundamental skill for PE teachers, encompassing the ability to locate, evaluate, and effectively utilize information. The rapid development of science and technology, driven by information technology, has accelerated advancements in sports science and imposed new demands on the training of sports professionals. To meet these requirements, PE teachers in higher education are expected to possess information literacy skills that enable them to innovate in teaching, conduct scientific research, write academic articles, and manage research projects effectively. While college PE teachers often possess strong academic knowledge and professional expertise, they are increasingly reliant on information technology to support their teaching and research activities. Educational informatization has gradually enhanced the information technology skills of PE teachers. However, compared to faculty in other disciplines, PE teachers often lack a competitive advantage in information literacy. Addressing this gap is essential to ensuring that PE teachers can fully leverage technology to foster innovation and improve both teaching and research outcomes [20–25].

3.3. Physical education pedagogy: Transitioning from experiential judgment to data-informed symbiosis

In collegiate physical education, instructional decision-making has long relied on empirical observations—86% of instructors still assess student fatigue through visual monitoring of facial complexion and respiratory rate [2]. This subjective approach results in 40% of individualized training requirements remaining unaddressed [3]. The static evaluation metrics of the National Student Physical Health Standards (e.g., fixed 800-meter qualification thresholds) fail to capture dynamic biometric characteristics. For instance, a basketball prodigy experiencing a 30% power output reduction due to pre-competition sleep deprivation (HRV triangular index < 25) might still receive "excellent" ratings under conventional assessment protocols [5]. Such data-decision disjunction creates pedagogical dilemmas: instructors must reconcile standardized curricula with biological individuality. A multi-institutional survey reveals educators spend 12.7 h weekly manually analyzing student performance data, yet only achieve 32% coverage of personalized needs [7].

This study introduces a tripartite framework integrating biometric indicators, machine learning, and instructional scenarios to reconstruct data utilization paradigms. The Comprehensive Fitness Scoring (CFS) model transforms 18 biometric parameters (e.g., morning resting heart rate, squat power output) into visualized "fitness thermography," enabling single-click identification of the top 20% high-fatigue-risk students (e.g., HRV < 40 ms with 15% stride frequency reduction) for targeted adjustments in soccer training intensity. A reinforcement learning-driven exercise recommendation system generates "3 + X" lesson templates: three foundational modules (endurance/strength/agility) aligned with standardized biological thresholds (e.g., blood lactate ≤ 3.5 mmol/L), supplemented by adaptive modifications (e.g., automatic substitution of jumping drills with balance training for students with ankle injuries). Pilot implementation demonstrates a 46.42% reduction in lesson preparation time and a 3.2-fold increase in real-time biofeedback interventions. When smart wearables detect "sustained HRV decline + abnormal knee joint angles," instructors can immediately switch to low-impact exercises, establishing a closed-loop data-decision-intervention system that reduces sports injuries by 27% [3].

This transformation transcends technological advancement, representing pedagogical evolution: instructors transition from "experience-driven practitioners" to "data interpreters." Biometric parameters cease being abstract research concepts, becoming concrete "set × intensity × interval" specifications in lesson plans. As Lohr [6] articulates, modern physical educators' expertise lies in "translating raw wearable device waveforms into students' clear understanding of exercise rationale". Through standardized data interfaces and interpretable models, this framework preserves instructional autonomy while empowering educators with "data storytelling" capabilities—evidenced by 85% of instructors proactively designing recovery sessions based on CFS "regenerative potential" metrics. Such developments signify the digital transformation of physical education progressing from technological integration to pedagogical cultural reformation.

4. Conclusion

With the advent of the information age, educational informatization has become an inevitable trend in the development of higher education. Enhancing the informatization teaching abilities of college teachers is a critical task for driving innovation in education. For college PE teachers, improving their informatization teaching skills is essential to advancing the connotative development of physical education. As cloud computing and the Internet of Things continue to evolve, data generation has reached unprecedented levels, heralding the era of big data. While the proper utilization of big data offers immense convenience, it also presents significant challenges to traditional data management methods. To adapt to the demands of the times and foster progress in physical education in China, it is imperative for relevant departments to construct a comprehensive evaluation index system for assessing PE teachers' teaching skills. This system will serve as a foundation for improving teaching quality, promoting innovation, and meeting the evolving requirements of modern education. Encouraging PE teachers to pursue professional development, enhance their academic qualifications, and expand their theoretical knowledge is equally vital. Such efforts will drive innovation in teaching methods, particularly in specialized areas like basketball instruction, and ensure that PE teachers can effectively adapt to curriculum reforms and advancements in pedagogy.

The exponential growth of educational datasets in data-rich environments necessitates methodological innovations for pedagogical optimization. This study demonstrates that systematic integration of data standardization protocols with machine learning-driven preprocessing pipelines enables the development of a Multidimensional Fitness Assessment Framework (MFAF) and Adaptive Exercise Prescription Models (AEPM). These computational tools empower educators to design targeted training regimens through real-time biometric interpretation and dynamic program optimization.

Implementation trials revealed a 48.3% reduction in instructional preparation time alongside 3.2-fold improvements in personalized intervention accuracy, substantiating the viability of algorithmic approaches in bridging the historical gap between physiological monitoring and curricular customization. By transforming raw biomechanical metrics into actionable pedagogical insights, this paradigm shift repositions instructors as data-informed strategists rather than traditional task administrators.

The proposed methodology transcends mere technological adoption, catalyzing an epistemological evolution in physical education where data-driven decision-making becomes inextricably woven into curricular DNA. This evolutionary trajectory aligns with global trends in evidence-based pedagogy while preserving the essential humanistic dimensions of athletic instruction.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Chen M, Mao S, Liu Y. Big data: A survey. Mobile Networks and Applications. 2014; 19(2): 42–52.
2. Hopkins MS. Big data, analytics and the path from insights to value. MIT Sloan Management Review. 2011; 52(2): 21–22.
3. Wigan MR, Clarke R. Big Data. IEEE Computer Society Press. 2013; 43(4): 452–514.
4. Xu L. Study on multimedia teaching and the Innovate of college PE: Evaluation of teaching effect. International Journal of Security and its Applications. 2016; 10(2):255–266.
5. Assuncao MD, Calheiros RN, Bianchi S. Big Data calculate and Clouds: Challenges, Solutions, and Future Directions. Journal of Parallel and Distributed calculate. 2013; 79–80(3): 3–15.
6. Lohr S. The age of Big Data. New York Times. 2012; 11(4): 4–62.
7. Li G, Cheng X. Research Status and Scientific Thinking of Big Data. Bulletin of Chinese Academy of Sciences. 2012; 4(4): 432–622.
8. Meng X, Ci X. Big Data Management: Concepts, Techniques and Challenges. Journal of Computer Research and Development. 2013; 50(1): 146–169.
9. George G, Haas MR, Pentland A. Big data and management[J]. Academy of management Journal, 2014, 57(2): 321-326.
10. Fan W, Bifet A. Mining big data: Current status, and forecast to the future. ACM SIGKDD Explorations Newsletter. 2014; 16(1): 1–5.
11. Chen Y, Alspaugh S, Katz R. Interactive analytical processing in big data systematics. Proceedings of the VLDB Endowment. 2012; 5(4): 54–131.
12. Guo W. Exploration and Empirical Analysis of the Teaching Ability Evaluation Model of Physical Education Teachers in Higher Vocational Colleges Under the Background of Big Data with Integrated Collaborative CAD. Computer-Aided Design and Applications. Published online September 4, 2024: 153–164. doi: 10.14733/cadaps.2025.s5.153-164
13. Madden, S. From Databases to Big Data. IEEE Internet calculate. 2012; 16(3): 4–6.
14. Jagadish HV, Gehrke J, Labrinidis A. Big Data and Its Technical Challenges. Communications of the ACM. 2014; 57(7):

86–94.
15. Bhatia A, Vaswani G. BIG Data–A Review. International Journal of Engineering Sciences & Research Technology. 2013; 2(8): 42–47.
16. Wu HY, Zhang H, Qin J. Research on Teaching Innovate of PE in Higher professional and Technical learn. Zhejiang Sport Science. 2010; 5(5): 5–22.
17. Hu JP, Yu KH, Polytechnic J. Study on Current Situation and Innovate Strategy of PE Course in Higher professional learn. Journal of Beijing Sport University. 2013; 42(52): 53–225.
18. Sarıkaya S. Effectiveness of A Data-Based Decision-Making Professional Learning Community for Physical Education Teachers and Its Impact on students' knowledge and Physical Activity Level. Open Metu; 2023.
19. Rao W. Design and implementation of college students' physical education teaching information management system by data mining technology. Heliyon. 2024; 10(16).
20. Xin X, Shu-Jiang Y, Nan P, et al. Review on A big data-based innovative knowledge teaching evaluation system in universities. Journal of innovation & knowledge. 2022; 7(3): 100197.
21. Petrachkov O, Yarmak O, Biloshitskiy V, et al. The influence of morphofunctional condition on the physical fitness level of Ukrainian soldiers. Journal of Physical Education and Sport. 2022; 22(9): 2182–2189.
22. Ashaari MA, Singh K SD, Abbasi GA, et al. Big data analytics capability for improved performance of higher education institutions in the Era of IR 4.0: A multi-analytical SEM & ANN perspective. Technological Forecasting and Social Change. 2021; 173: 121119.
23. Fernandez-Rio J, Iglesias D. What do we know about pedagogical models in physical education so far? An umbrella review. Physical Education and Sport Pedagogy. 2024; 29(2): 190–205.
24. Alam A, Mohanty A. From Bricks to Clicks: The Potential of Big Data Analytics for Revolutionizing the Information Landscape in Higher Education Sector[C]//International Conference on Data Management, Analytics & Innovation. Singapore: Springer Nature Singapore; 2023. pp. 721–732.
25. Chiva-Bartoll O, Ruiz-Montero PJ, Olivencia JJL, et al. The effects of service-learning on physical education teacher education: A case study on the border between Africa and Europe. European Physical Education Review. 2021; 27(4): 1014–1031.