ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Implementation Outcomes and Evaluation Modeling of Personalized Exercise Programs in University Physical Education

Hongmei Gao[1,*]

[1] Department of Physical Education, Hetao College, Bayannur 015000, Inner Mongolia Autonomous Region, China

SUMMARY: *This study examined the implementation outcomes of personalized exercise programs in university physical education and developed an evaluation model for identifying effective individual responses. A 16-week quasi-experimental design was used with 586 valid undergraduate participants from compulsory physical education classes. The intervention group received individualized exercise programs generated from baseline fitness tests, exercise preference, self-efficacy, injury-risk screening and weekly workload records, while the control group followed conventional class-based instruction. The evaluation framework combined standardized fitness improvement, moderate-to-vigorous physical activity, adherence, safety and preference matching. LightGBM was used to predict whether students reached the predefined response threshold, and SHAP was applied to explain variable contribution. After 16 weeks, the intervention group increased its physical fitness composite score by 8.49 points, compared with 3.57 points in the control group. Weekly MVPA increased by 57.4 min in the intervention group and by 24.6 min in the control group. The intervention group also showed higher adherence, lower overload rate and fewer discomfort reports. The response prediction model achieved an AUC of 0.842, an F1 score of 0.786 and an RMSE of 3.18 for score-gain prediction. Ablation analysis indicated that removing workload records caused the largest performance decline. The results suggest that personalized exercise programs can improve university physical education when baseline diagnosis, process workload control and interpretable evaluation are integrated within the same teaching cycle.*

KEYWORDS: *university physical education; personalized exercise program; exercise prescription; implementation outcome; evaluation model*

1 Introduction

Public physical education classes in colleges and universities are taught to significantly different populations. Students in the same teaching class do not have the same cardiorespiratory endurance, muscular strength, flexibility, body composition, exercise experience, and injury history. Uniform programs, intensity, and assessment standards facilitate scheduling and management, but make it difficult to deal with the dispersion of students' true starting points. Low-fitness students are more prone to avoidance, fatigue accumulation, and movement deformation in running and jumping and resistance tasks; students who already have a sports foundation often lack effective stimulation due to insufficient classroom load. Physical education that only reads results from end-of-term physical tests makes it difficult for teachers to determine whether students are moving enough, loading appropriately, and identifying risks

---

*gaohongmei2109@163.com

https://doi.org/10.65102/is2026767

Published: 30 April, 2026

in a timely manner over the course of the semester.

Physical activity guidelines provide basic health goals for college physical education programs, and WHO guidelines emphasize the need for a certain level of frequency, intensity, and duration of physical activity in order to develop clear health benefits [1], and the ACSM Guidelines for Exercise Testing and Prescribing further standardize exercise prescription into actionable parameters of frequency, intensity, duration, type, total amount, and progression [2]. When these principles are introduced into college physical education classes, the question no longer remains "schedule more exercise" but rather the need to translate student fitness shortcomings, tolerable loads, exercise preferences, and extracurricular performance conditions into manageable instructional programs. Exercise prescription research has also pointed out that long-term training programs must be continuously monitored and reprogrammed based on individual responses, and that training logs, load records, and individual needs are important bases for prescription adjustments [3].

The already-built researches of college sports give an experience-based foundation for personal motion plans. Movement formulation guidance can enhance movement quality, body shape, heart-lung endurance, and partial psychological health indices in university students. Randomized tests of depression symptoms among college students have indicated that individualised aerobic and resistance workout plans have good possibility of being carried out over a 12-week time frame and raise physical activity degrees and participation in intervention. Digital health interference means have also been utilized to strengthen physical activity among college students, and a systematical review indicated that such tools had a more consistent promotion for step numbers, but differences in impacts on moderate to high strength physical activity and static behavior still existed. Another systematical summary of physical movement interference items for college students has pointed out that campus interference items can enhance total physical movement, MVPA, and step numbers, but the interference effects were affected by differences in sample scale, execution time length, interference form, and behavior skills.

Curriculum formats are also being adapted for teaching physical education in higher education. a 16-week randomized trial of basketball instruction showed that a blended learning model improved lung capacity, flexibility, strength, and speed metrics [8]. Prescriptive health education combined with functional movement screening has been shown to improve movement quality and quality of life related metrics in college student populations [9]. Meanwhile, wearable devices and mobile recordings are beginning to enter the campus sports scene. Wearable exercise devices have been used to explain college students' exercise adherence behaviors, with roles related to exercise motivation and social support [10]. IoT wearable research geared toward college physical education classes has further shown that student acceptance of devices affects athletic performance improvement, and that technology access by itself is not a substitute for instructional design [11].

Artificial intelligence and machine learning provide new assessment tools for personalized exercise programs. applications of AI in sports programs have covered automatic feedback, posture recognition, data acquisition, training assistance and predictive analytics [12]. Studies on machine learning exercise prescription based on fitness evaluation and BMI have shown that LightGBM, CNN and attention mechanisms can be used for individualized prescription generation and interpretation, but most of such studies have focused on fitness classification, BMI identification, or a single health goal [13]. Higher education physical education classes are more complex to evaluate, focusing on both physical test scores and whether students consistently perform in and out of the classroom, complete tasks within safe load intervals, and whether teachers are able to make pedagogical adjustments based on model outputs.

Three gaps remain in existing research. First, the logic of generating personalized prescriptions is not sufficiently connected to the classroom implementation component. Some studies illustrate that prescriptions are effective without clearly accounting for how baseline body measurements, exercise preferences, injury risk, and out-of-classroom implementation conditions enter into the prescription parameters. Second, effectiveness evaluations are often compressed into pre- and post-physical testing differences. Higher education physical education classes need to simultaneously account for improvements in fitness, changes in MVPA, adherence, overload reporting, and pain and discomfort, and a single outcome does not reflect the quality of instructional delivery. Third, digital evaluation emphasizes collection over interpretation. Records of heart rate, step counts, RPE, punch cards, and teacher adjustments tend to exist in a fragmented manner, lacking comprehensive models that can identify responders, explain variable contributions, and provide feedback on instructional adjustments.

Accordingly, this paper focuses on personalized exercise programs in college physical education, with core questions including: how to generate actionable exercise prescriptions based on student profiles; whether personalized programs can improve fitness, activity level, adherence, and safety in a 16-week course; and whether evaluation models can identify responders and explain key influences. The contribution of this paper consists of three points: first, the construction of a student portrait consisting of baseline body measurements, preferences, self-efficacy, injury risk, and process load records; second, the establishment of a comprehensive assessment index that includes fitness improvement, MVPA, adherence, safety, and preference matching; and third, the use of LightGBM and SHAP to form a feedback mechanism for response prediction and interpretation, which provides a deployable solution for individualized moderation in college physical education classes. deployment program.

2 Methods

2.1 Study Design, Participants and Variable Construction

This research adopted a 16-week non-true experimental design, the subjects come from a required public physical education curriculum in an ordinary university. At the beginning step, 624 undergraduate students were brought into the research, and after we eliminated the samples that had missing baseline physical tests, missed more than 4 weeks of classes in the intervention period, had less than 70% of effective wearable record days, and had missing items in the final test, a valid sample that contained 586 students was got. Among these people, 294 were in the intervention group and 292 were in the control group. The intervention group obtained one individual movement plan which was based on normal body education courses, and the control group used unified class teaching with unified after-class movement suggestions. Two groups all finished two in-class body movement classes each week, and the intervention group finished one more outside-class exercise assignment which was made according to student situation files.

Data sources included four categories. The first category was baseline physical measurement data, containing BMI, spirometry body mass index, 50 m run, standing long jump, seated forward bend, 1000 m for boys or 800 m for girls performance, pull-ups for boys or sit-ups for girls performance. The second category was behavioral process data, which included weekly average MVPA, step count, class attendance, extracurricular exercise clocking, heart rate interval attainment ratio, and session-RPE load. The third category was subjective data, containing exercise self-efficacy, exercise pleasure, fatigue scores, and pain and discomfort reports. The fourth category was instructional implementation data containing teacher adjustment records, number of prescription revisions, and average weekly time spent on administration per student.

The questionnaires and behavioral indicators used established instruments or common collection rules. The self-reported physical activity section referred to the IPAQ's activity categorization and time recording methods [14], and the wearable data cleaning referred to the principles of valid days, wearing time and missing treatment for accelerometer collection and processing [15]. Training load adopted session-RPE idea to combine subjective exertion level and single activity duration to avoid evaluating training stimulus by step count only [16]. Exercise pleasure uses the PACES short version of the idea into the preference matching module, so that exercise program selection does not rely solely on teacher experience [17]. Baseline characteristics and variable structure, as shown in Table 1.

Table 1: Baseline Characteristics and Variable Structure

Variable	Intervention Group (n=294)	Control Group (n=292)	p-value	Data Role
Age / year	19.42 ± 1.08	19.39 ± 1.11	0.741	Covariate
Male / %	48.6	49.3	0.866	Covariate
BMI / kg·m^{-2}	22.91 ± 3.12	22.84 ± 3.05	0.784	Baseline Feature
Physical Fitness Composite Score / point	68.42 ± 8.73	68.61 ± 8.69	0.792	Primary Outcome
Weekly MVPA / min	112.6 ± 47.8	113.9 ± 46.5	0.738	Process Outcome
Exercise Self-Efficacy / point	3.18 ± 0.74	3.15 ± 0.71	0.617	Psychological Feature
Exercise Enjoyment / point	3.24 ± 0.68	3.21 ± 0.70	0.601	Preference Feature
Prior Sports Injury / %	13.9	14.4	0.861	Safety Feature
Valid Wearable-Record Days / %	91.3 ± 6.8	90.9 ± 7.1	0.486	Data-Quality Feature

The body measurement indicators were first assimilated. 50 m run, 800 m or 1000 m time category indicators were reversed in the direction of the lower the better performance, while the other indicators retained the original direction. Different quantitative indicators into the model before the use of standardization processing, as shown in equation (1).

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{1}$$

where x_{ij} denotes the raw or assimilated value of Student i on Indicator j, $\mu_j$denotes the sample mean of Indicator j, σ_j denotes the sample standard deviation of Indicator j, and z_{ij} denotes the standardized indicator value. This treatment allowed the physical, behavioral, subjective, and instructional implementation variables to enter into the unified modeling universe. The student portrait is comprised of physical measures, preferences, self-efficacy, injury risk, and extracurricular implementation conditions, as shown in Figure 1.

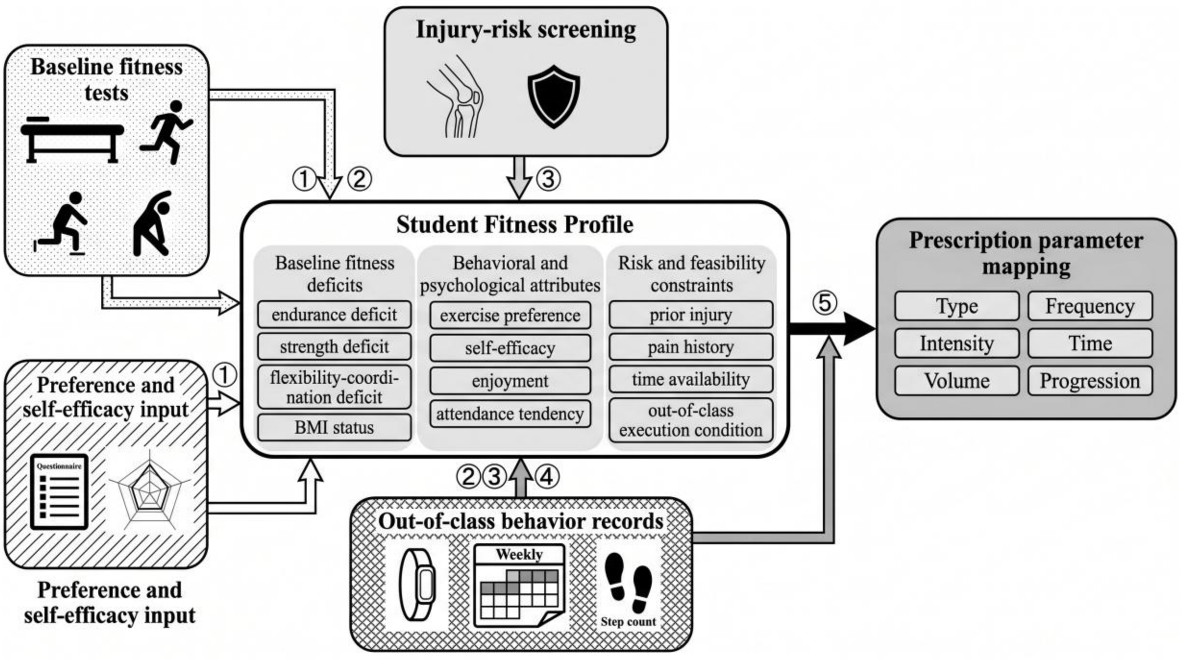

Figure 1: Student Profile Construction and Prescription Mapping Mechanism.

2.2 Personalized Exercise Program Design

Personalized exercise programs were generated by dominant portrait. Each student was categorized into four dominant portraits based on baseline body measurements of shortness, BMI, injury history, exercise preference, and self-efficacy: endurance deficient, strength deficient, flexibility coordination deficient, and overweight with low compliance risk. Students can have multiple weaknesses, but prescriptions are generated with only one dominant portrait and two secondary weaknesses to avoid dispersing prescriptions to the point where they are difficult to implement. Prescriptions are looked back at bi-weekly and loads are adjusted based on MVPA completion rates, RPE intervals, heart rate reserve attainment percentage, pain reports, and attendance.

The intervention group completed two in-class physical activity and three out-of-class exercise tasks per week. The target intensity was 50%-75% heart rate reserve or RPE 11-15 as the primary zone. If task completion exceeded 85% for two consecutive weeks and pain scores were less than 3, the total amount of training was increased by 5-10% for the next cycle. If there were two consecutive reports of overload or a pain score of 4 or more, the intensity was reduced, the duration shortened, or low-impact exercise replaced. The control group received uniform classroom content and uniform out-of-class exercise recommendations, and teachers did not base prescription callbacks on individual profiles. Individualized exercise prescription rules based on student profiles are shown in Table 2.

Table 2: Personalized Exercise Prescription Rules by Student Profile

Student Profile	Main Weakness	In-Class Task	Out-of-Class Task	Intensity Target	Progression Rule	Safety Threshold
Endurance Deficit	800/1000 m and MVPA Insufficiency	Interval jogging, aerobic games	Brisk walking, jogging, cycling	55%–75% HRR; RPE 12–15	+5%–8% duration every 2 weeks	Pain ≥4 or RPE >16 twice
Strength Deficit	Pull-up/Sit-up and Standing Long Jump	Bodyweight resistance, core training	Squat, push-up regression, elastic band	RPE 11–15	+1 set or +2 repetitions every 2 weeks	Joint pain or technique collapse
Flexibility-Coordination Deficit	Sit-and-Reach and Movement Control	Mobility drills, balance, low-speed skill tasks	Stretching, mobility circuit	RPE 9–13	+10% range or task complexity	Dizziness, sharp pain
Overweight and Low-Adherence Risk	BMI and Low Self-Efficacy	Low-impact aerobic circuit	Walking, cycling, short frequent bouts	50%–65% HRR; RPE 10–13	+5 min per session after 2 stable weeks	Knee/Achilles discomfort or fatigue accumulation

Prescription matching considers not only physical fitness shortcomings, but also students' willingness to perform. If endurance-deficient students prefer ball games, the extracurricular task can be adjusted from continuous jogging to small-field soccer, badminton mobile practice or cycling; if shoulder discomfort exists in strength-deficient students, elastic band rowing, kneeling push-ups and core stabilization can be used instead of pull-up training. The teacher's end only displays the dominant portrait, the current week's tasks, abnormal reminders and adjustment suggestions, reducing the interference of irrelevant data on teaching decisions. The personalized exercise regimen was implemented using a two-week callback mechanism to keep load progression, compliance and safety thresholds in the same closed loop, as shown in Figure 2.

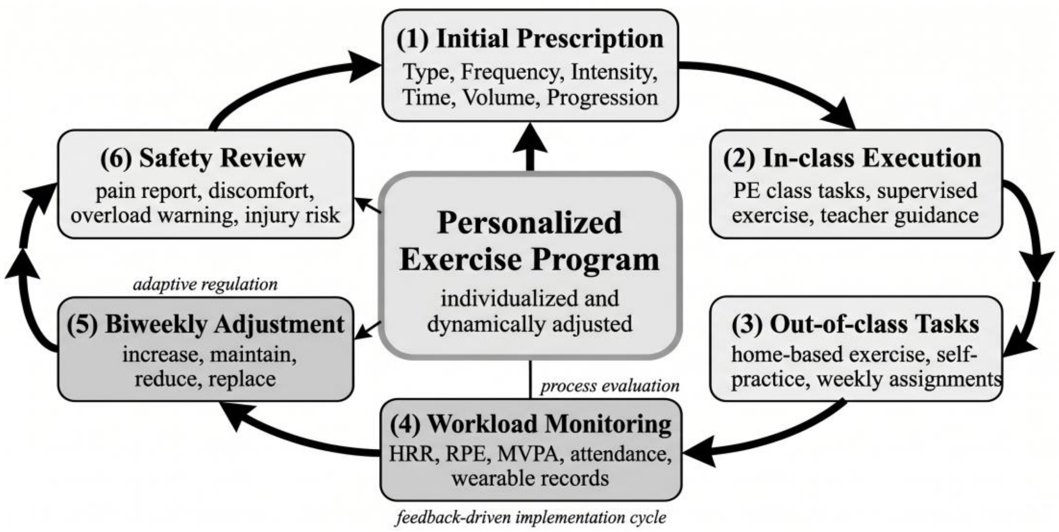

Figure 2: Closed-loop Implementation Mechanism of Personalized Exercise Programs.

The process load fitness score is used to determine whether the student is consistently in the target training interval, as shown in equation (2).

$$L_i(t) = 0.40h_i(t) + 0.25r_i(t) + 0.20m_i(t) + 0.10a_i(t) - 0.15p_i(t) \tag{2}$$

where $h_i(t)$ denotes the proportion of students in i who achieved the target heart rate reserve interval in week t, $r_i(t)$ denotes the proportion of RPE in the target range, $m_i(t)$ denotes the proportion of MVPA completions, $a_i(t)$ denotes the proportion of attendance and clock-in completions, $p_i(t)$ denotes the standardized score for reports of pain or discomfort, and $L_i(t)$ denotes the weekly load fit score. This score combines effective stimulation, quality of implementation, and risk constraints as the basis for prescription callback.

2.3 Evaluation Model and Statistical Protocol

The evaluation model is composed of two parts, the Comprehensive Effectiveness Index and the Reaction Forecasting Model. The Combined Effective Index is utilized to assess whether the intervention can bring an overall advantage, and the Response Forecasting Model is utilized to determine which students have the higher possibility to reach effective response threshold values. The input contents of the model include baseline physical measurement values, student-centered feature descriptions, prescription parameter values, weekly load matching scores, MVPA variation quantities, compliance rates, pain feedback materials, preference matching degrees, and the quantities of teacher adjustment operations. The main result was the change of 16-week combined body ability score, and secondary results included the change of weekly average MVPA, exercise keeping rate, change of exercise self-confidence, overburden report, and occurrence of ache and uncomfort. The composite effect index is given definition as what equation (3) shows.

$$E_i = \omega_1 s_{i1} + \omega_2 s_{i2} + \omega_3 s_{i3} + \omega_4 s_{i4} + \omega_5 s_{i5} \tag{3}$$

where s_{i1} denotes the physical fitness improvement score, s_{i2} denotes the MVPA improvement score, s_{i3} denotes the adherence score, s_{i4} denotes the security score, s_{i5} denotes the preference matching score, ω_1 to ω_5 denote the weights of the five categories of indicators, and E_i denotes the composite effect index for the ith student. The weights were set to 0.382, 0.247, 0.136, 0.128, and 0.107, respectively, indicating that improvement in physical fitness is still the main outcome, but the quality of implementation and security also enter the evaluation caliber.

Responders are defined to be students that satisfy two conditions at the same time: an improvement which is not less than 5 points in the 16-week combined physical fitness score and not less than 80% compliance with the exercise task. The predictive model construction work was carried out by utilizing LightGBM, and it was then compared with logistic regression, random forests, as well as XGBoost. LightGBM is fit for the table form teaching data, and its gradient boosting structure has high efficiency in the middle-sample, many-variable, and non-linear relation situation cases [18]. Model interpretation uses SHAP to identify the marginal contribution of different variables to the response probability [19]. Model optimization is done using joint loss function as shown in equation (4).

$$\mathcal{L} = \lambda \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + (1-\lambda) \frac{1}{n} \sum_{i=1}^{n} [-r_i \log(\hat{r}_i) - (1-r_i) \log(1-\hat{r}_i)] \tag{4}$$

where y_i denotes the actual change in the composite physical fitness score of student i, $\hat{y}_i$ denotes the predicted change value, r_i denotes the actual response state, $\hat{r}_i$ denotes the predicted response probability, n denotes the sample size, λ denotes the balance coefficient of the regression task and the classification task, and $\mathcal{L}$ denotes the joint loss. In this paper, λ is set to 0.55 so that the model emphasizes both improved magnitude prediction and responder identification. Comprehensive evaluation, response prediction, and explanatory feedback are put into the same model framework, as shown in Figure 3.

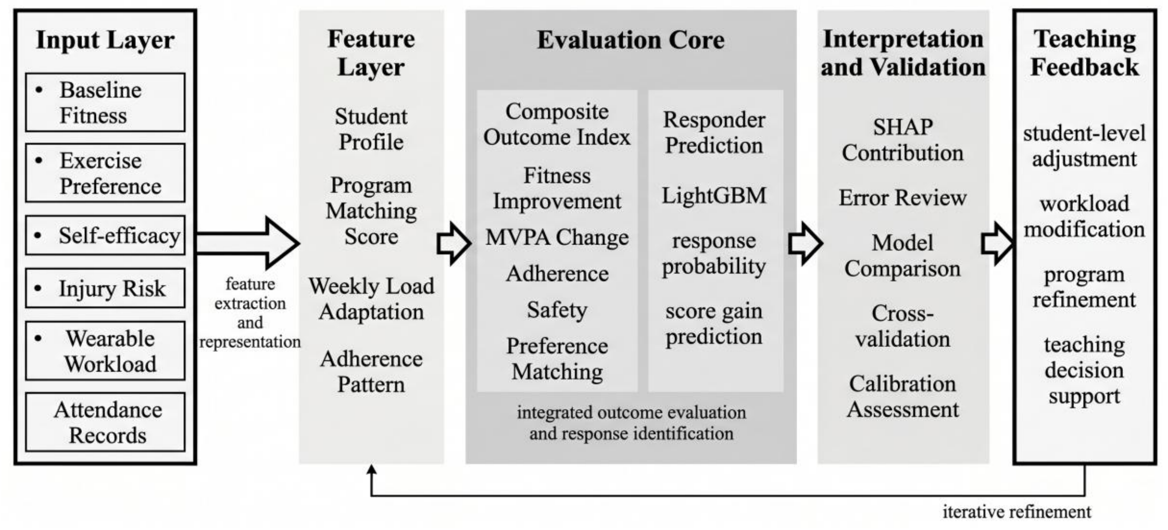

Figure 3: Evaluation Modeling Framework and Validation Protocol.

Statistical analyses were performed using baseline-corrected between-group comparisons. Means and standard deviations were reported for continuous variables and proportions for categorical variables. Between-group differences were analyzed using analysis of covariance to control for baseline values, gender, and school origin; week-by-week trends were handled using linear mixed models; and multiple comparisons were controlled for false discovery rates using the Benjamini-Hochberg method [20]. Model training was performed using a 7:3 training test division, with 50% discounted cross-validation within the training set. Performance metrics included AUC, F1, RMSE, calibration slope, and inference time per 100 students.

3 Results and Discussion

3.1 Overall Implementation Outcomes

This section answers the question of whether the personalized exercise program resulted in overall gains over the 16 weeks of physical education. The evaluation went beyond reading final physical test scores and examined physical activity levels, compliance, overload reporting, and pain and discomfort simultaneously to avoid attributing effects simply to exercise additions.Changes in key outcome indicators after 16 weeks are shown in Table 3.

Table 3: Changes in Main Outcomes after 16 Weeks

Outcome	Intervention Baseline	Intervention Week 16	Control Baseline	Control Week 16	Between-Group Difference	Effect Size
Physical Fitness Composite Score / Point	68.42 ± 8.73	76.91 ± 8.05	68.61 ± 8.69	72.18 ± 8.44	4.92	0.63
Weekly MVPA / Min	112.6 ± 47.8	170.0 ± 52.4	113.9 ± 46.5	138.5 ± 49.7	32.8	0.58
Exercise Adherence / %	68.9 ± 14.5	84.7 ± 10.6	69.4 ± 14.1	71.2 ± 13.8	14.0	0.71
Self-Efficacy / Point	3.18 ± 0.74	3.79 ± 0.66	3.15 ± 0.71	3.36 ± 0.69	0.40	0.52
Exercise Enjoyment / Point	3.24 ± 0.68	3.82 ± 0.61	3.21 ± 0.70	3.43 ± 0.67	0.36	0.49
Overload Report / %	8.7	5.8	8.5	9.9	-4.1	—
Pain Reports / Per 1000 Sessions	3.6	2.7	3.5	4.1	-1.4	—

Through the passage of 16 weeks, the physical fitness comprehensive score was raised from 68.42 ± 8.73 to 76.91 ± 8.05 in the intervention group, having an average increase of 8.49 points, and from 68.61 ± 8.69 to 72.18 ± 8.44 in the control group, having an average increase of 3.57 points. After we do the baseline correction work, the difference of the change quantity between the two groups is 4.92 points, and the effect size it has is 0.63. The average weekly MVPA has obtained an increase of 57.4 minutes in the intervention group, and 24.6 minutes in the control group, hence there exists a difference between the two groups of 32.8 minutes. In respect of adherence situation, the intervention group has its rate rise from 68.9% to 84.7%, and the control group has its rate rise from 69.4% to 71.2%. From the perspective of safety, the proportion of overload reports decreased to 5.8% in the intervention group, and increased to 9.9% in the control group. The number of pain discomfort reports was 2.7 times per 1000 exercise tasks in the intervention group, and 4.1 times per 1000 exercise tasks in the control group. The changes of weekly average MVPA are presented in Figure 4.

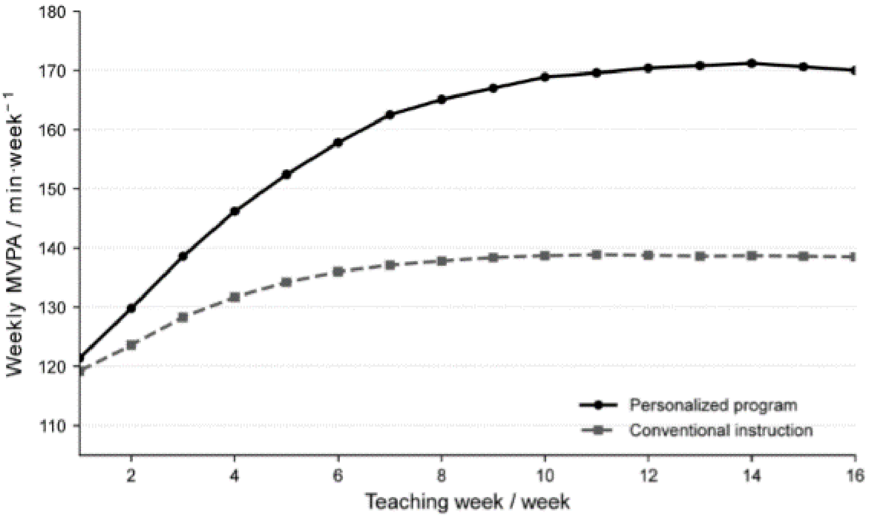

Figure 4: Weekly MVPA Trajectories during the 16-week Intervention.

In Figure 4, the intervention group increased from 121.4 min to 146.2 min in weeks 1-4, continued to rise to 168.9 min in weeks 5-10, and stayed around 170 min after week 11. The control group also showed an increase in the first 4 weeks, but basically stayed above and below 138 min after week 8. This trend suggests that the benefits of personalized programs occur primarily in the mid- to late-week period, when students complete the first round of adaptation and prescription callbacks begin to match their actual load tolerance. Digital health research has suggested that technology reminders alone do not promote MVPA consistently, and the results in this paper suggest that MVPA gains are more likely to be sustained when technology records are combined with teacher prescription adjustments. Changes in adherence rates are shown in Figure 5.

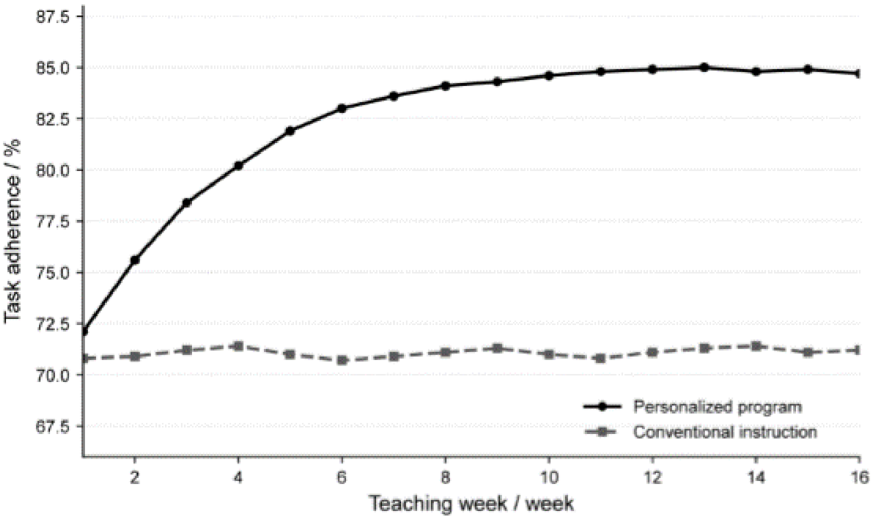

Figure 5: Weekly Adherence Trajectories during the 16-week Intervention.

In Figure 5, the adherence rate in the intervention group increased from 72.1% at week 1 to 84.7% at week 16, and stabilized above 83% after week 6. The control group remained between 70% and 72% for a long time. This result is consistent with the difference in adherence rates in Table 3 and explains the source of the MVPA difference. Exercise self-efficacy research indicates that physical activity can influence students' sustained engagement through mental toughness and sense of competence [21]; in this study, the intervention group's self-efficacy improved by 0.61 points, suggesting that students' willingness to engage and stability of execution improved in tandem as they continued to accomplish their goals in an affordable task. To further illustrate which process variables drive the results, this paper constructs three-dimensional response surfaces for load fitness scores, adherence rates, and incremental fitness composite scores, as shown in Figure 6.

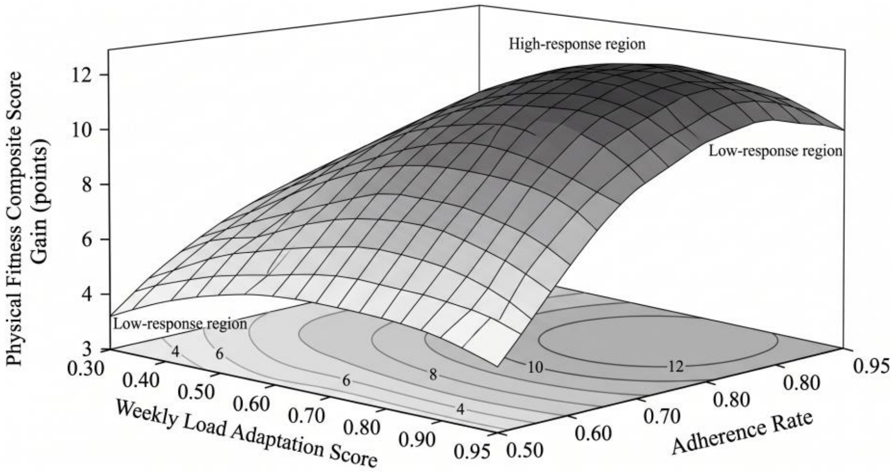

Figure 6: Three-dimensional Response Surface of Implementation Outcomes.

In Figure 6, the highest incremental body mass composite score, about 9-12 points, was found when the weekly load fit score was at 0.64-0.80 and the adherence rate was at 0.80-0.90. Regions with low load-fitting and low adherence tend to have body mass increments below 6 points. This result suggests that simply increasing activity does not equate to higher gains; students need to consistently complete tasks within the appropriate intensity zone. The mHealth exercise intervention for overweight college students also showed a dose relationship between physical activity and body fat change, but the relationship needs to be interpreted in the context of individual status and performance conditions [22].

3.2 Heterogeneous Effects across Student Profiles

The overall mean indicates that the intervention is effective, but college physical education is more concerned with differences in gains across student portraits. This section develops analyses by dominant portraits, outcome indicators, and load response relationships to determine whether individualized programs actually have a corresponding effect on student shortcomings. Research on exercise interventions for college students has shown that there are differences in the effects of different exercise types on mental and physical outcomes, and program design needs to avoid putting all students into the same training logic [23]. The heterogeneity results are shown in Figure 7.

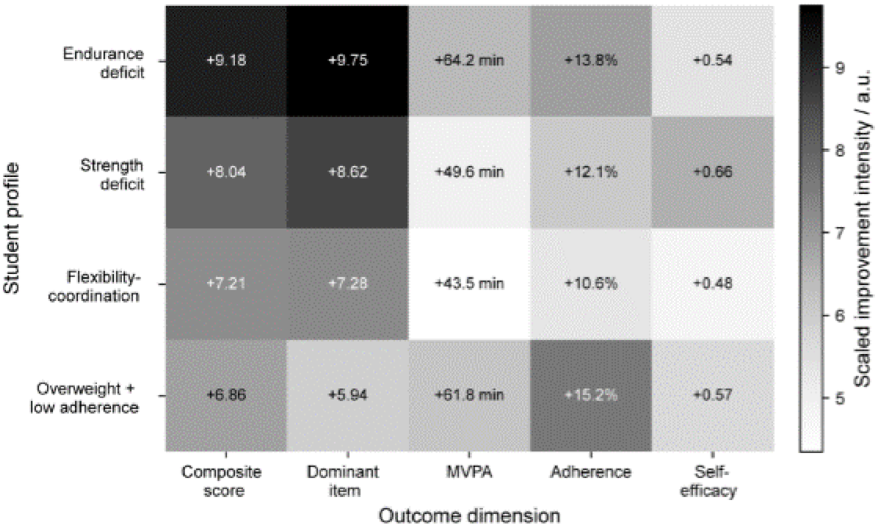

Figure 7: Heterogeneous Effects by Student Fitness Profile.

In Figure 7, the endurance-deficient students showed a 9.18-point increase in the composite fitness score, a 9.75-point increase in the dominant weakness score, and a 64.2-min increase in MVPA; the strength-deficient students showed an 8.62-point increase in the dominant weakness score and a 0.66-point increase in self-efficacy; the flexibility-coordination-deficient students showed a 7.28-point increase in the dominant weakness score; and the overweight students with a low risk of compliance showed a 6.86-point increase in the composite fitness score, but a 15.2-point increase in compliance rate and a 61.8-min increase in MVPA. score, but a 15.2 percentage point increase in adherence rate and a 61.8 min increase in MVPA.This result suggests that the benefits of the personalized program are not uniformly distributed, but rather are concentrated in the changeable metrics across the portraits. Figure 8 further demonstrates the relationship between weekly load fit scores and incremental fitness composite scores.

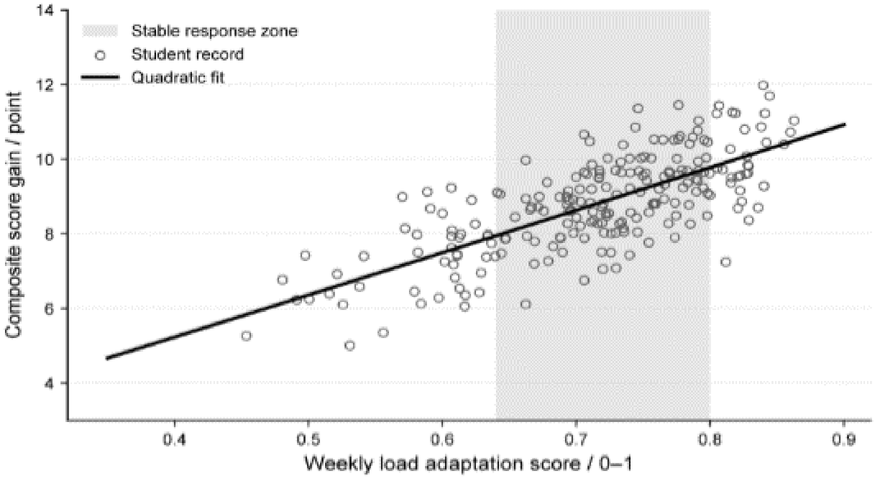

Figure 8: Load-response Relationship between Adaptation Score and Fitness Gain.

In Figure 8, the scatter distribution shows that for load fit scores below 0.50, body mass gains were mostly concentrated in the 4-7 range; after the scores moved into the 0.64-0.80 range, the gains shifted upward to the 8-12 range; and after the scores moved above 0.84, gains did not continue to increase and some students experienced a drop in gains. This result explains why safety thresholds need to be entered into the prescription model. If increments are added only to the physical test shortfall, some low fitness students may experience high RPE, pain reports, and task interruptions, ultimately counteracting the training stimulus. A systematic review of social media and text message-based physical activity interventions also showed that whether the intervention was effective was influenced by the form of interaction, executive feedback, and individual motivation [24], suggesting that the mechanisms of behavioral maintenance are just as important as the parameters of exercise prescription.

In terms of instructional delivery, the heterogeneous results require teachers to avoid grouping students only by program or total score. Two students with similar composite physical fitness scores may correspond to different weaknesses, with one short in 800 m or 1000 m and the other short in upper body strength or core strength. If the same running add-on program is used, the former may improve endurance and the latter may have decreased compliance. The model in this paper incorporates dominant portraits, auxiliary weaknesses, exercise preferences, and risk thresholds simultaneously into prescription matching to bring stratified instruction closer to students' real task demands.

The results of gender stratification showed that boys' physical fitness composite score improved by 8.73 points and girls' improved by 8.26 points, with small differences between groups. The differences were mainly in the improvement indicators: slightly higher improvement in strength-related items for boys and slightly higher improvement in flexibility and self-efficacy for girls.

This result suggests that gender can be entered into the model as a covariate, but should not be used as the sole basis for grouping. Mobile resistance training research has pointed out that resistance prescriptions often face problems with movement quality, progression, and underreporting in digital platforms [25], and that resistance tasks in college physical education classes are more likely to require on-site corrections by the instructor in conjunction with simplified design of the task outside of the classroom.

3.3 Model Performance, Ablation and Deployment Implications

Once implementation effects are confirmed, evaluating the model also requires determining which students will respond and which data modules are most critical. This section compares the predictive performance of different models and examines the models' reliance on process records, preference matching, injury risk, and pictorial clustering through ablation analysis. Research on technology-enhanced physical education states that digital tools should be integrated with learning engagement, feedback, and long-term activity habit formation, and that pursuing device access alone is difficult to steadily improve instructional outcomes [26]. The predictive and ablative performance of the model was assessed, as shown in Table 4.

Table 4: Changes in Main Outcomes after 16 Weeks

Outcome	Intervention Baseline	Intervention Week 16	Control Baseline	Control Week 16	Between-Group Difference	Effect Size
Physical Fitness Composite Score / Point	68.42 ± 8.73	76.91 ± 8.05	68.61 ± 8.69	72.18 ± 8.44	4.92	0.63
Weekly MVPA / Min	112.6 ± 47.8	170.0 ± 52.4	113.9 ± 46.5	138.5 ± 49.7	32.8	0.58
Exercise Adherence / %	68.9 ± 14.5	84.7 ± 10.6	69.4 ± 14.1	71.2 ± 13.8	14.0	0.71
Self-Efficacy / Point	3.18 ± 0.74	3.79 ± 0.66	3.15 ± 0.71	3.36 ± 0.69	0.40	0.52
Exercise Enjoyment / Point	3.24 ± 0.68	3.82 ± 0.61	3.21 ± 0.70	3.43 ± 0.67	0.36	0.49
Overload Report / %	8.7	5.8	8.5	9.9	-4.1	—
Pain Reports / Per 1000 Sessions	3.6	2.7	3.5	4.1	-1.4	—

Table 4 shows that LightGBM achieves AUC 0.842 and F1 0.786 in respondent identification, and RMSE 3.18 in predicting the change of composite score of physical fitness, which is better than logistic regression, random forest, and XGBoost. the calibration slope is 0.94, which indicates that the prediction probability is closer to the actual response rate. The inference time is 0.06 s/100 students, which can meet the needs of batch evaluation in general physical education classes. After deleting the process load record, the AUC decreased by 0.049 and the RMSE increased by 0.69, which is the ablation setting with the largest decrease in performance. The AUC decreased by 0.027 after deleting preference matching and 0.041 after deleting the dominant portrait, indicating that the model performance mainly comes from the combined effect of process load, portrait structure and willingness to perform.The ROC curves are shown in Fig. 9.

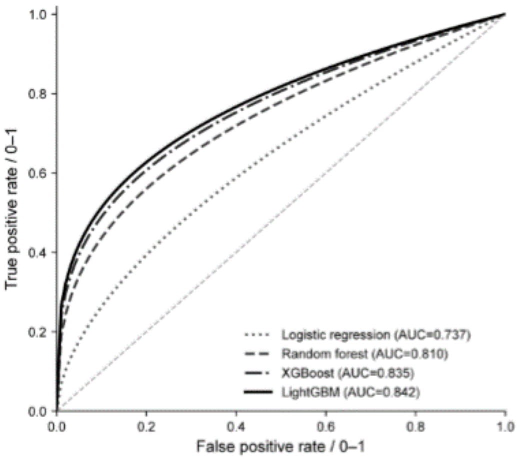

Figure 9: ROC Comparison of Responder Prediction Models.

In Figure 9, the LightGBM curve is overall higher than the other models and still maintains a high true positive rate in the low false positive rate interval. The logistic regression curve has a more limited improvement, indicating that responder identification is not a simple linear relationship. Random Forest and XGBoost perform close to each other, but calibration and inference are slightly less efficient than LightGBM.For college physical education, the models need to not only be accurate, but also be able to be quickly recalled by the faculty end during the semester. A review of machine learning in adaptive physical activity also points out that the value of individualized sport models depends on monitoring, implementation paths, and explanatory mechanisms, not just predicting scores [27]. The feature contributions are shown in Figure 10.

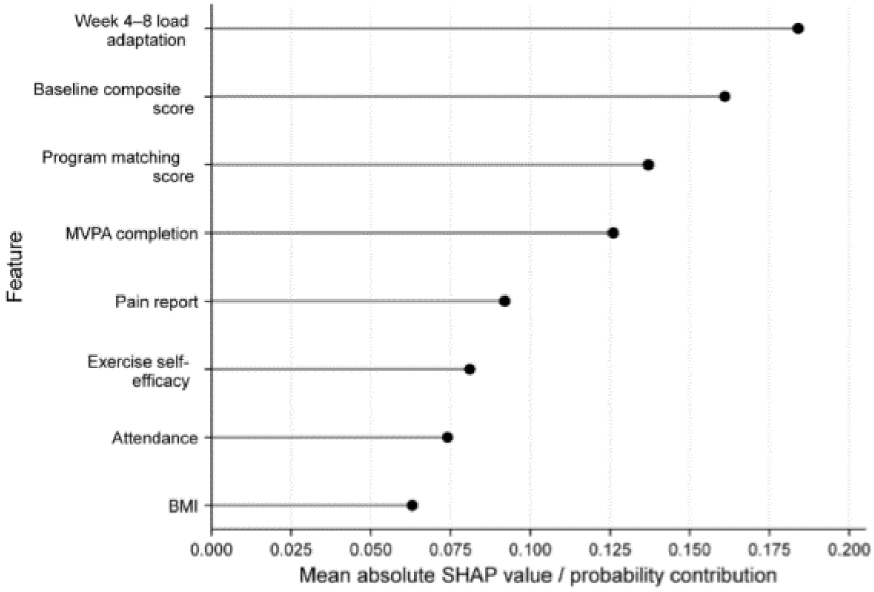

Figure 10: Feature Contribution of the Evaluation Model.

In Figure 10, the average load fitness score for weeks 4-8 has the highest mean absolute SHAP value which is 0.184, following is the baseline physical fitness composite score of 0.161, prescription match score of 0.137, MVPA completion rate of 0.126, pain report of 0.092, self-efficacy of 0.081, attendance of 0.074, and BMI of 0.063. This ordering indicates that the model relies the most greatly not on one single baseline physical examination, but on whether students are constantly in the suitable loading zone in the middle of the semester. The 4th to 8th weeks lie at the crossing place of the intervention adjustment and load increasing parts, hence students who keep high fitness marks in this time period have more stable follow-up fitness promotion.

Error analysis showed that the RMSE of the model was 2.74 in the low fitness group, 3.11 in the medium fitness group, and 3.86 in the high fitness group.The high fitness group had a larger error, mainly because there were more extracurricular independent training in this group, and it was difficult for the school curriculum records to cover their complete exercise input. Another source of error comes from wearable wearing habits. Despite the exclusion of samples with less than 70% valid records in this paper, some students still had missing heart rates before and after ball games, strength training, or bathing. The third type of error comes from subjective data, where preference, self-efficacy and pain reports are influenced by students' expressive habits and may underestimate true fatigue or discomfort.

On the aspect of deployment, the model is able to undertake three kinds of tasks. Before class, teachers are able to make an initial arrangement according to student materials, thus decreasing the burden that comes from grouping students completely depending on past experience. In the teaching process, the model can find out those students whose load fitness scores are lower than 0.45 and whose compliance rates are lower than 65% for two continuous weeks, therefore it reminds the teacher to decrease the strength or change the kind of exercise. After the class is finished, the model is able to combine fitness progress enhancement, implementation procedure, and safe incidents into course quality feedback. The time of management which is consumed at the teacher's side showed that the average time which is spent on file building and prescription making for each student was 7.8 min in week 1, and the average weekly adjusting time fell to 2.4 min after week 4. For an average physical education class that has about 40 students, this work load can be further compressed through the use of batch templates, exception prompts, and automatic scoring.

4 Conclusion

This paper focuses on the implementation effect and assessment model of personalized exercise program in college physical education, and constructs a teaching evaluation framework that includes student portrait, prescription matching, process load record, comprehensive effect index and response prediction.The caliber of the 16-week data shows that the personalized program outperforms the regular teaching in terms of composite physical fitness scores, MVPA, adherence rate, self-efficacy, and exercise enjoyment, and at the same time reduces the overload reporting and pain discomfort incidence.

(1) At the level of object organization, this paper incorporates baseline body measurements, exercise preferences, self-efficacy, injury risk, and out-of-class performance records into student portraits, shifting physical education class groupings from single-item or total score tiering to prescription matching with co-constraints of dominant shortcomings and performance conditions.

(2) At the methodology and outcome level, the composite effect index was able to simultaneously evaluate fitness improvement, activity level, adherence, safety and preference matching. the LightGBM response prediction model achieved AUC 0.842, F1 0.786 and RMSE 3.18, and the ablation results showed that process load record was a key module affecting model performance.

(3) This paper still suffers from the limitations of short data period, limited sample range and insufficient records of autonomous movements outside the school. Subsequent studies should expand the cross-school sample, add post-semester follow-up, and introduce a more stable interface for extracurricular exercise records to test the model's suitability in different colleges and universities, different sports, and different teaching resource conditions.

About the Author

Hongmei Gao was born in Bayannur, Inner Mongolia, China in 1979. She received her bachelor's degree from Inner Mongolia Minzu University and holds a master's degree from Xi'an Physical Education Institute. She currently works in the School of Physical Education at Hetao College. Her research interests include physical education teaching, sports training, and sports and health.

References

[1] World Health Organization. (2020). WHO guidelines on physical activity and sedentary

behaviour. Geneva: World Health Organization.

[2] American College of Sports Medicine. (2025). ACSM’s guidelines for exercise testing and prescription (12th ed.). Philadelphia: Wolters Kluwer.

[3] Wackerhage, H., & Schoenfeld, B. J. (2021). Personalized, evidence-informed training plans and exercise prescriptions for performance, fitness and health. Sports Medicine, 51(9), 1805-1813.

[4] Zhong, X. L., Sheng, D. L., Cheng, T. Z., et al. (2023). Effect of exercise prescription teaching on exercise quality and mental health status of college students. World Journal of Psychiatry, 13(5), 191-202.

[5] Zhao, Y., Wang, W., Wang, M., et al. (2023). Personalized individual-based exercise prescriptions are effective in treating depressive symptoms of college students during the COVID-19: A randomized controlled trial in China. Frontiers in Psychiatry, 13, 1015725.

[6] Bi, S., Yuan, J., Wang, Y., et al. (2024). Effectiveness of digital health interventions in promoting physical activity among college students: Systematic review and meta-analysis. Journal of Medical Internet Research, 26, e51714.

[7] Yuan, F., Peng, S., Khairani, A. Z., et al. (2024). A systematic review and meta-analysis of the efficacy of physical activity interventions among university students. Sustainability, 16(4), 1369.

[8] Wang, C., Yuan, Y., & Ji, X. (2024). The effects of a blended learning model on the physical fitness of Chinese university students: A cluster randomized controlled trial in basketball education. BMC Public Health, 24, 2451.

[9] Li, B., Su, Y., Zhu, L., et al. (2025). Prescription-based health education integrating functional movement screening: A controlled trial in Chinese university students. Frontiers in Public Health, 13, 1649030.

[10] Han, Y., Zhang, J., Liu, H., et al. (2025). The impact of wearable sports equipment on college students’ physical exercise persistence: A mediated model of physical exercise motivation moderated by social support. Frontiers in Sports and Active Living, 7, 1691032.

[11] Xu, Y., Li, J., Wang, H., et al. (2024). From wearables to performance: How acceptance of IoT devices influences physical education results in college students. Scientific Reports, 14, 24146.

[12] Canzone, A. C. A., Belmonte, G. B., Patti, A. P., et al. (2025). The multiple uses of artificial intelligence in exercise programs: A narrative review. Frontiers in Public Health, 13, 1510801.

[13] Mo, M., Li, B., Yang, Y., et al. (2026). A machine learning framework for personalized exercise prescription based on BMI and physical fitness assessment. Scientific Reports.

[14] Craig, C. L., Marshall, A. L., Sjöström, M., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. Medicine & Science in Sports &

Exercise, 35(8), 1381-1395.

[15] Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., et al. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. Sports Medicine, 47(9), 1821-1845.

[16] Haddad, M., Stylianides, G., Djaoui, L., et al. (2017). Session-RPE method for training load monitoring: Validity, ecological usefulness, and influencing factors. Frontiers in Neuroscience, 11, 612.

[17] Chen, C., Weyland, S., Fritsch, J., et al. (2021). A short version of the Physical Activity Enjoyment Scale: Development and psychometric properties. International Journal of Environmental Research and Public Health, 18(21), 11035.

[18] Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (Vol. 30, pp. 3146-3154).

[19] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765-4774).

[20] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289-300.

[21] Qiu, W., Wang, X., Cui, H., et al. (2025). The impact of physical exercise on college students’ physical self-efficacy: The mediating role of psychological resilience. Behavioral Sciences, 15(4), 541.

[22] Zhang, Q., Li, Z., Jiang, L., et al. (2025). Effects of an exercise intervention based on mHealth technology on the physical health of male university students with overweight and obesity: Randomized controlled trial. Journal of Medical Internet Research, 27, e69451.

[23] Xiao, Y., Shi, C., Zhang, X., et al. (2025). Effectiveness of different exercise interventions on depressive symptoms among college students: A network meta-analysis. BMC Public Health, 25, 1845.

[24] Buja, A., Lo Bue, R., Mariotti, F., et al. (2024). Promotion of physical activity among university students with social media or text messaging: A systematic review. Inquiry, 61, 00469580241248131.

[25] Cox, E. R., Beacroft, S., Jansson, A. K., et al. (2025). Effects of mHealth interventions to prescribe resistance training: A systematic review and meta-analysis of randomized controlled trials. International Journal of Behavioral Nutrition and Physical Activity, 22, 127.

[26] Martín-Rodríguez, A., & Madrigal-Cerezo, R. (2025). Technology-enhanced pedagogy in physical education: Bridging engagement, learning, and lifelong activity. Education Sciences, 15(4), 409.

[27] Greco, G., Petrelli, A., Poli, L., et al. (2026). Machine learning in adapted physical activity: Clinical applications, monitoring, and implementation pathways for personalized exercise in chronic conditions: A narrative review. Journal of Functional Morphology and Kinesiology, 11(1), 106.