

Analysis of Physical Education Teaching and Prediction of Students' Sports Performance Based on Artificial Intelligence Algorithms

Zhe Zhang^{1*}, Xiaoyuan Huang²

¹*Teaching and Research Department for Sports, Beijing University Of Civil Engineering and Architecture, Beijing, China.*

²*Electrical and Information Engineering Department, Beijing University Of Civil Engineering and Architecture, Beijing, China.*

**Corresponding Author. Email: zhangzhe@bucea.edu.cn*

Abstract: Taking students' sports performance as the research object, this study launched a multidimensional exploration through systematic data analysis and intelligent algorithm modeling. First, descriptive statistics were used to characterize the data of sports performance-related indicators, calculate the mean, median and extreme value distribution of each parameter, and reveal the overall characteristics of the data set. On this basis, the correlation analysis technique was used to quantitatively assess the influence weights of each indicator on physical education performance, and it was found that physical education performance scores, course participation rates and course grades constituted the core influence dimensions, and their correlation strengths were significantly higher than those of other observed indicators. For the construction of the prediction model, the Sparrow Search Algorithm (SSA) was innovatively introduced to optimize the parameters of the random forest model, and the SSA-RF hybrid model was established to realize the synergistic optimization of feature selection and hyper-parameter tuning. The experimental results show that the model achieves 99.75% and 96% prediction accuracies in the training set and test set, respectively, and its confusion matrix shows a balanced recognition ability for all types of samples. The study further explores the application value of AI technology in the field of physical education, and confirms that the intelligent algorithm can not only improve the prediction accuracy of sports performance, but also provide data support for the development of personalized training programs and curriculum content optimization. By constructing a whole-process framework of “data analysis-model optimization-education application”, this study provides an operable methodological system for the implementation of intelligent physical education, and its research results are of great practical significance for promoting the digital transformation of physical education teaching and establishing an accurate evaluation mechanism based on big data.

Keywords: Student athletic performance, Descriptive statistics, Correlation analysis, Random forests, artificial intelligence.

1. Introduction

Traditional physical education has long faced challenges such as insufficient individualized instruction and limited teaching efficiency. Teachers rely on experience to observe students' movements, which makes it difficult to capture detailed differences in real time, and when facing large classes, targeted feedback lags behind, resulting in inefficient correction of technical movements [1]. In addition, the differences in students' physical fitness base and sports habits make it difficult for uniform training programs to meet individual needs. With the breakthrough of artificial intelligence technology, especially the integration of computer vision, machine learning and wearable devices, physical education is ushering in intelligent transformation [2]. Through high-precision sensors and video capture technology, AI is able to collect real-time motion data, analyze student performance with algorithmic models, and provide a scientific basis for dynamic and precise teaching.

Artificial intelligence algorithms analyze students' movement postures frame by frame through computer vision technology to accurately identify movement norms. Based on machine learning algorithms, AI is able to integrate students' historical sports data, construct personalized ability portraits, and generate training programs that are adapted to their goals [3]. For example, smart wearable devices continuously monitor heart rate, step frequency and other data, and the algorithm dynamically adjusts the training intensity to avoid excessive fatigue or ineffective training. In addition, AI can also predict the risk of sports injuries, such as through gait analysis to identify abnormal knee joint loading in running, and recommend adjusting posture or increasing rehabilitation training in advance to protect students' sports safety. AI integrates students' school sports data through a cloud-based platform to form a visualized health profile, helping parents understand their children's physical changes and progress trajectories [4]. In this paper, we firstly conduct descriptive statistical analysis of students' sports performance, analyzing the mean, median, maximum and minimum of each index; secondly, we conduct correlation analysis of the data to analyze the key indicators affecting students' sports performance, and finally we use the sparrow search algorithm to optimize random forests to predict students' sports performance, and then we analyze the significance of AI in the students' sports performance as well as in the the role of physical education curriculum reform.

2. Sources of data sets

The dataset contains a comprehensive set of indicators designed to assess student performance in physical education. It includes a variety of quantitative and qualitative measures that reflect students' levels of physical fitness, classroom participation, and overall performance outcomes [5]. The dataset contains several indicators including age, gender, sex, grade level, strength, scores on endurance capacity based on physical fitness tests, scores on flexibility levels, scores on speed and agility, body mass index, scores on physical skills related to sports and physical activity, motivation to participate in physical education classes, percentage of participation in physical education classes, scores on motivation to participate in physical education, composite scores on overall performance in sports , indicators of improvement in sports performance over time, performance in physical education classes, average number of hours of physical activity outside of school, and the predictor variable was the students' overall sports performance, which was categorized into three bands of 1, 2, and 3, i.e., low, intermediate, and high levels of performance. Selected partial data sets are shown in Table 1.

Table 1: Selected data sets.

Age	Gender	Grade Level	Final Grade	Previous Semester PE Grade	Performance
16	1	1	1	1	1
17	2	2	2	2	2
14	2	3	3	1	2

Table 1: (continued).

16	1	3	1	3	2
16	2	3	2	1	2
17	1	2	3	1	1
14	3	1	2	4	2
14	2	1	2	2	2
16	3	2	1	1	1
15	1	1	1	3	3
16	3	2	3	1	2

The data set was analyzed with descriptive statistics for the maximum, minimum, mean, standard deviation and median of each variable as shown in Table 2.

Table 2: Descriptive Statistical Analysis.

Variable Name	Sample Size	Max	Min	Average	Standard deviation	Median
Age	500	17	14	15.592	1.151	16
Strength_Score	500	76.324	21.037	51.04	9.778	51.21
Endurance_Score	500	88.317	14.944	50.497	12.318	50.218
Flexibility_Score	500	97.066	4.707	50.269	14.464	50.013
Speed_Agility_Score	500	81.099	23.643	50.316	9.801	50.335
BMI	500	35	15	21.987	3.804	21.797
Health_Fitness_Knowledge_Score	500	100	45.706	75.202	10.564	75.396
Skills_Score	500	100	27.414	69.16	14.461	69.244
Attendance_Rate	500	99.987	75.098	87.4	7.343	87.304
Overall_PE_Performance_Score	500	100	41.67	69.281	9.984	69.261
Improvement_Rate	500	10	0	4.972	2.029	4.976
Hours_Physical_Activity_Per_Week	500	9.989	2	6.08	2.219	6.041

The sample consisted of physical performance data of adolescents aged 14-17 years (mean 15.59 years), showing significant polarization in physical performance indicators for endurance (standard deviation 12.32) and flexibility (extremes 4.71-97.07), with balanced performances for strength and speed-agility (means around 50-51); overall healthy BMI (mean 21.99) but with individuals at risk of overweight/obesity (up to 35); high health knowledge (mean 75.2) and skill level (mean 69.16) but large variations in the latter (standard deviation 14.46), good attendance (mean 87.4%), but room for improvement in overall physical education performance (mean 69.28) and rate of improvement (mean 4.97/10).

In addition, we used correlation analysis to explore the relationship between students' athletic performance and each indicator, using correlation coefficients to quantify and plotting the correlation rankings based on the magnitude of the correlation coefficients, which are shown in Figure 1.

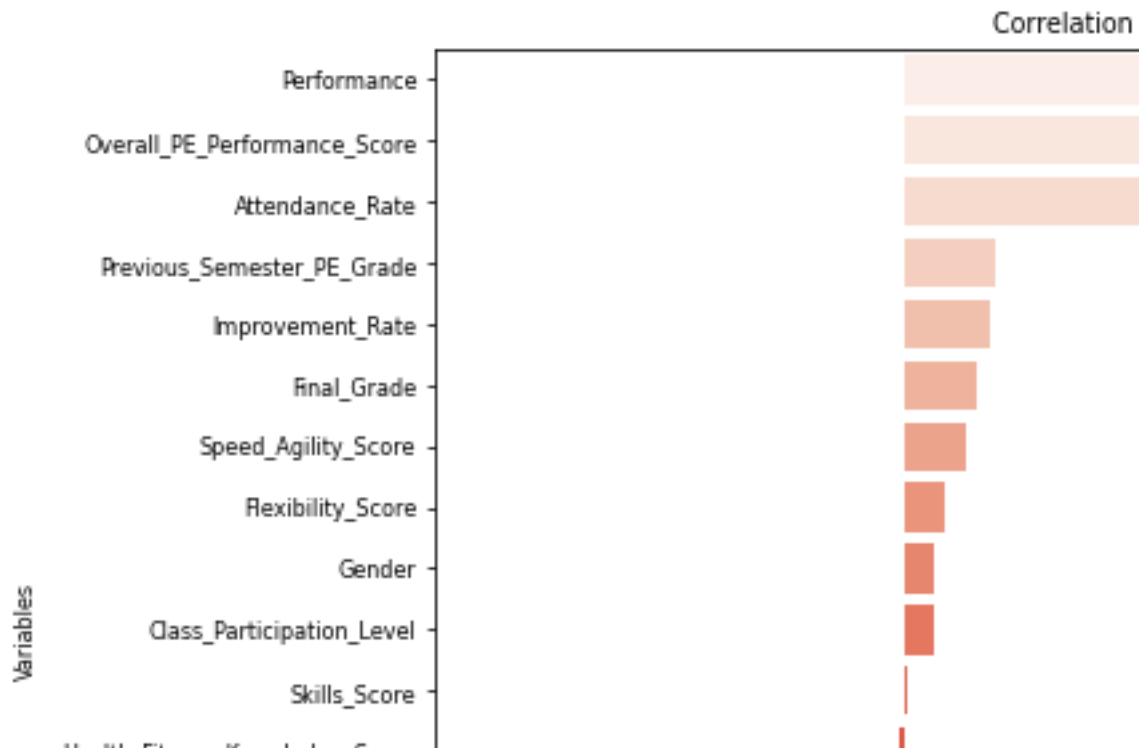


Figure 1: Ranking of correlation coefficients

As shown by the correlation ranking, the indicator most positively correlated with the physical education performance rating is the physical education performance score, followed by participation in physical education courses, and scores in physical education courses.

3. Method

3.1. Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is a meta-heuristic optimization algorithm inspired by the foraging and anti-predation behaviors of sparrow populations, and its core principle simulates the collaborative mechanism of “finder-follower” and the vigilance behaviors of sparrows when encountering dangers in the foraging process, and searches for the optimal solution by balancing the global exploration and the local exploitation capabilities [6].

3.2. Random Forest

Random forest is a machine learning algorithm based on the idea of integrated learning, which improves model performance by constructing multiple decision trees and aggregating their predictions [7]. Its core principle contains two key mechanisms: bagging and random feature selection. First, the algorithm utilizes Bootstrap sampling technique to randomly select multiple sample subsets from the original dataset with putback to provide differentiated training data for each decision tree, and this data perturbation effectively reduces the variance of the model. Second, in the node splitting process of each tree, the algorithm is not evaluating all the features, but randomly selecting a subset of features as the candidate splitting features, and this randomness of the feature space further enhances the diversity of the base learner [8].

3.3. Random Forest Optimized by Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) optimizes the key hyperparameters (e.g., number of decision trees, maximum depth, and node splitting thresholds) of the random forest by simulating the exploration and exploitation mechanisms in the foraging behavior of sparrow groups [9]. The algorithm makes use of the information sharing and dynamic following strategy among sparrow individuals to efficiently search for parameter combinations within the global range, avoiding the problem that traditional methods are prone to falling into local optimums, and at the same time balancing the exploration and exploitation by adaptively adjusting the search range to improve the prediction accuracy and generalization ability of the random forest model, and in particular, optimizing the synergistic effect of feature selection and parameter configuration when dealing with high-dimensional data [10].

4. Result

We use the sparrow search algorithm to optimize the random forest model for predicting students' sports performance. In the experiment, the population size of the sparrow search algorithm was set to 30, the maximum number of iterations was 50, the proportion of discoverers was 20%, the proportion of vigilantes was 10%, and the safety threshold was 0.8. The fitness function was cross-validated using a 5-fold fold, and the randomized seed was fixed at 42 to ensure reproducibility. For dataset division, 70% of the data is randomly selected as the training set and the remaining 30% of the data is used as the test set.

Firstly, the confusion matrices of the training and test sets are output, the confusion matrix of the training set is shown in Fig. 2 and the confusion matrix of the test set is shown in Fig. 3.

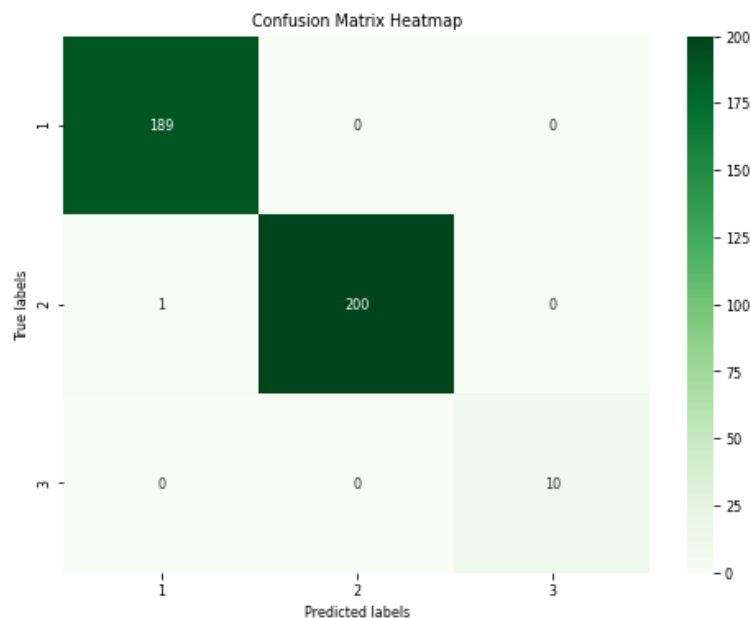


Figure 2: The confusion matrix of the training set.

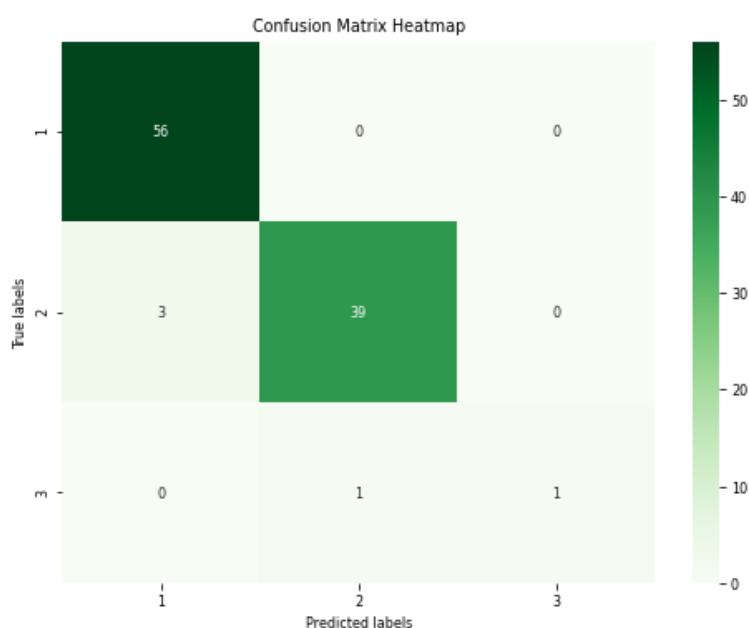


Figure 3: The confusion matrix of the test set

5. Implications of Artificial Intelligence for the Study of Student Sports Performance

The rapid development of Artificial Intelligence (AI) technology provides a new perspective and method for studying students' sports performance. Traditional sports teaching mostly relies on subjective observation and empirical judgment, making it difficult to accurately quantify students' athletic ability, technical details, and physical fitness status. Through wearable devices, computer vision technology, sensors and other tools, AI is able to collect multi-dimensional data of students in sports in real time, such as heart rate, movement trajectory, muscle force pattern, reaction speed, etc., and analyze these data through machine learning algorithms, revealing the potential patterns of individual sports performance. AI can identify the postural deviation of students' shooting, running or swimming through motion capture technology, providing instant feedback to optimize the technical movements; through the long term, AI can identify the postural deviation of students' shooting, running or swimming, and provide instant feedback to optimize the technical movements. feedback to optimize technical movements; through long-term tracking data, AI can also assess the trend of students' physical progress and predict the risk of sports injuries to help develop personalized training plans. In addition, the application of AI in team sports can analyze students' collaborative ability and tactical execution effect, providing coaches with scientific training adjustment suggestions. This data-driven analysis not only improves the objectivity and accuracy of sports research, but also provides technical support for tapping students' athletic potential and realizing precision teaching, promoting the transformation of sports science from empiricism to data intelligence.

6. The significance of predicting students' sports performance for physical education curriculum reform

The ability to predict students' sports performance based on AI technology has far-reaching practical value for physical education curriculum reform. Traditional physical education curriculum design often adopts a “one-size-fits-all” model, ignoring the differences in physical quality, skill base and interest direction of individual students, resulting in uneven teaching results. Through the AI prediction model, teachers can identify students' athletic potential and shortcomings in advance, and

then design a tiered, dynamic curriculum program. For students with weak physical ability but strong coordination, the curriculum can focus on skillful programs (such as gymnastics and ball games); for students with outstanding explosive power but insufficient endurance, cardiorespiratory fitness can be improved through stage-by-stage intensive training. This kind of personalized teaching can not only improve students' participation and self-confidence, but also avoid frustration or sports injuries caused by uniform training. In addition, the predicted results can help optimize the course evaluation system, shifting from a single outcome assessment (e.g., running timings) to a process-based assessment that focuses on the students' progress, effort, and room for technical improvement, thus establishing a more scientific incentive mechanism. In the long run, AI-driven predictive capabilities can also support educational equity by helping physical education teachers focus on disadvantaged student groups, design adaptive activity programs, and ensure that every student has access to growth opportunities in the physical education curriculum. This kind of data-based curriculum reform not only fits the educational concept of “teaching students according to their abilities”, but also promotes the transformation of school sports from skill transfer to health promotion and lifelong sports habit cultivation, laying the foundation for comprehensively improving the physical fitness and sports literacy of young people.

7. Conclusion

In this study, a multidimensional study of students' sports performance is conducted through a systematic analysis method, and a complete research framework from basic characterization to intelligent prediction is constructed. In the data analysis dimension, descriptive statistics are used to quantitatively portray the sports performance indicators, analyze the concentration trend, the degree of dispersion and the distribution characteristics, and cover the core parameters such as mean, median, and extreme value, so as to present the complete statistical characteristics of the data set. In the correlation dimension, the core factors affecting sports performance are identified through the construction of Pearson's correlation coefficient matrix, in which the standardized sports performance scores, course participation frequency and course assessment results constitute the key echelons of influence, which provides a scientific basis for the selection of variables in the subsequent modeling.

In the prediction model construction session, the SSA-RF fusion prediction model was constructed by innovatively combining bio-heuristic algorithms with integrated learning, and adopting the Sparrow Search Algorithm (SSA) to globally search for optimization in the hyperparameter space of Random Forest (RF). Experimental data show that the model achieves 99.75% fitting accuracy in the training set and 96% accuracy in the test set by confusion matrix validation, and its generalization performance is significantly better than that of traditional machine learning models. Notably, the results of the model feature importance ranking are highly consistent with the pre-correlation analysis, forming a double validation at the research methodology level and confirming the scientific nature of the research design.

This study confirms the potential application of intelligent algorithms in the field of sports science, and its innovative value is reflected in the construction of a closed loop of “feature analysis-model optimization-educational application” at the methodological level, and the development of a transferable sports performance prediction tool at the practical level. More importantly, the study reveals the cross-fertilization path between educational technology and athletic training, and provides a theoretical framework and technical solution for the establishment of an intelligent sports education ecosystem based on artificial intelligence. This not only marks a new stage of intelligent physical education evaluation, but also reshapes the value cognition of “sports education for human beings” from the level of technology philosophy - to realize the comprehensive development of human beings through technological empowerment, which is the ultimate demand of the reform of education modernization.

References

- [1] Varillas-Delgado, David, et al. "Genetics and sports performance: the present and future in the identification of talent for sports based on DNA testing." *European journal of applied physiology* 122.8 (2022): 1811-1830.
- [2] Lochbaum, Marc, et al. "Revisiting the self-confidence and sport performance relationship: a systematic review with meta-analysis." *International Journal of Environmental Research and Public Health* 19.11 (2022): 6381.
- [3] Melin, Anna K., et al. "Direct and indirect impact of low energy availability on sports performance." *Scandinavian journal of medicine & science in sports* 34.1 (2024): e14327.
- [4] De Fazio, Roberto, et al. "Wearable sensors and smart devices to monitor rehabilitation parameters and sports performance: an overview." *Sensors* 23.4 (2023): 1856.
- [5] Richlan, Fabio, et al. "Virtual training, real effects: A narrative review on sports performance enhancement through interventions in virtual reality." *Frontiers in Psychology* 14 (2023): 1240790.
- [6] Hassabo, Ahmed G., et al. "Applications of nanotechnology in the creation of smart sportswear for enhanced sports performance: Efficiency and comfort." *Journal of Textiles, Coloration and Polymer Science* 20.1 (2023): 11-28.
- [7] Biz, Carlo, et al. "Is kinesio taping effective for sport performance and ankle function of athletes with chronic ankle instability (CAI)? A systematic review and meta-analysis." *Medicina* 58.5 (2022): 620.
- [8] Beebe, Kelzie E., et al. "Collegiate athletes' perceptions of the connection between mental health and sport performance: A qualitative investigation." *Journal of Clinical Sport Psychology* 18.3 (2023): 357-373.
- [9] Afonso, José, et al. "Why sports should embrace bilateral asymmetry: a narrative review." *Symmetry* 14.10 (2022): 1993.
- [10] Lemelin, Emilie, et al. "Autonomy support in sport contexts: The role of parents and coaches in the promotion of athlete well-being and performance." *Sport, Exercise, and Performance Psychology* 11.3 (2022): 305.