



OPEN ACCESS

EDITED BY
Chin-Ling Chen,
Chaoyang University of Technology, Taiwan

REVIEWED BY
Yi Gu,
Jiangnan University, China
Liu Ling-Chun,
National Chung Hsing University, Taiwan

*CORRESPONDENCE
Cui Cui
✉ cuicui201425@163.com

RECEIVED 17 April 2025
ACCEPTED 30 June 2025
PUBLISHED 25 July 2025

CITATION
Cui C and Yin J (2025) Analysis of the exercise
intention-behavior gap among college
students using explainable machine learning.
Front. Public Health 13:1613553.
doi: 10.3389/fpubh.2025.1613553

COPYRIGHT
© 2025 Cui and Yin. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Analysis of the exercise intention-behavior gap among college students using explainable machine learning

Cui Cui^{1,2*} and Jixin Yin¹

¹Department of Sports, Huanghe Jiaotong University, Jiaozuo, Henan, China, ²Gangneung Wonju University, Gangneung, Gangwon Province, Republic of Korea

Introduction: The physical fitness of college students is a growing global public health concern. A critical challenge in improving student fitness is addressing the intention-behavior gap—the disconnect between students' intentions to engage in physical activity and their actual behavior.

Methods: This study utilized survey data from TikTok-using college students, incorporating variables such as gender, academic grade, health belief perceptions, and planned behavior perceptions. Multiple machine learning models were developed to predict the presence of the intention-behavior gap. The performance of these models was evaluated, and SHapley Additive exPlanations (SHAP) was applied to the best-performing model to interpret feature importance.

Results: Among the models tested, SHAP analysis revealed that perceived barriers were the most influential factor contributing to the intention-behavior gap. Furthermore, the results indicated that male students in higher academic grades, with fewer perceived barriers and stronger subjective norms regarding physical activity, were significantly less likely to exhibit this gap.

Discussion: These findings suggest that university health promotion strategies should focus on reducing perceived barriers, cultivating a supportive campus environment for physical activity, and optimizing the allocation of physical education resources. Such measures may effectively support the transformation of students' physical activity intentions into consistent, health-promoting behaviors.

KEYWORDS

physical activity promotion, college student, intention-behavior gap, explainable machine learning, feature engineering

1 Introduction

College students represent the backbone of a nation's future development, and their physical health is closely linked to the quality and sustainability of national human resources (1–3). A physically fit student population contributes not only to individual well-being but also to the long-term economic and social vitality of a country. However, a growing body of research has recently highlighted a continuous decline in their physical fitness, with various studies reporting alarming trends in reduced cardiovascular endurance, muscular strength, and flexibility among this population (4, 5). Among the various interventions proposed, physical exercise is widely recognized by scholars as an effective measure. Physical exercise improves one's lifestyle as well as promotes both physical and mental health in the long term, which plays a crucial role in building healthy habits that extend into adulthood (6, 7).

Although governments and universities around the world have implemented various policies to promote physical activity among college students, their actual participation rates remain relatively low (8). Moreover, the frequency, intensity, and consistency of students' physical exercise often do not meet the recommended health guidelines (9). This phenomenon not only hinders the physical and mental well-being of college students but also, to some extent, restricts the development of physical education in higher education institutions. Therefore, it is essential to effectively foster their initiative and persistence in physical exercise.

Currently, a key challenge in promoting college students' active participation in physical exercise lies in the significant gap between their intentions and actual behaviors. This discrepancy is known as the intention-behavior gap (10, 11). Even when individuals have a strong willingness to exercise, various influencing factors often hinder them from translating that intention into actual physical activity. The intention-behavior gap has been confirmed to be a common phenomenon in health intervention practices. For example, Wee et al. (12) explored how reducing the intention-behavior gap promotes an increase in physical exercise behavior. Wang et al. (13) indicated that the intention-behavior gap can influence sustained physical activity.

Many scholars have conducted examinations on mediating or moderating variables between exercise intention and behavior by drawing on the foundational frameworks of classic theories such as the Health Belief Model and the Theory of Planned Behavior (14, 15). However, the factors contributing to the intention-behavior gap in exercise are diverse, and existing research lacks a comprehensive analysis from a multi-variable, holistic perspective. Recently, the continuous advancement of artificial intelligence technologies has equipped scientific research with a diverse array of efficient and powerful analytical tools (16, 17). SHapley Additive exPlanation (SHAP)-based interpretable machine learning offers the capability to simultaneously process multiple feature variables while identifying those that exert the most significant influence on outcomes (18–20). Thus, this approach is particularly well-suited for holistically analyzing the predictive effects of various feature variables on the intention-behavior gap in college students' physical exercise.

Based on the above analysis and discussion, this study systematically incorporates multiple feature variables—health belief and planned behavior—to construct a computational model for predicting the intention-behavior gap in college students' physical exercise. Specifically, this study utilizes questionnaire data collected from TikTok to identify the optimal predictive model among Logistic Regression (LR) (21), Support Vector Machine (SVM) (22), and Random Forest (RF) (23). Subsequently, the SHAP method is employed to rank feature importance within the optimal model. From a multivariate and holistic perspective, a systematic analysis is conducted on the intention-behavior gap in college students' physical exercise, which aims to provide evidence-based recommendations for promoting physical activity and enhancing the physical and mental well-being of university students. The main contributions of our work can be listed as follows:

1. We systematically identify and quantify key psychological and social factors that contribute to the exercise intention-behavior gap among college students by leveraging a large-scale dataset

collected from TikTok, which provides empirical evidence in a novel digital social context.

2. We develop an interpretable machine learning framework that not only predicts the intention-behavior gap but also explains the underlying decision mechanisms, advancing the application of explainable AI in health behavior research.
3. We propose a decision tree-based pathway analysis that reveals actionable intervention targets. This offers practical guidance for designing personalized university sports programs and policies aimed at effectively narrowing the exercise intention-behavior gap.

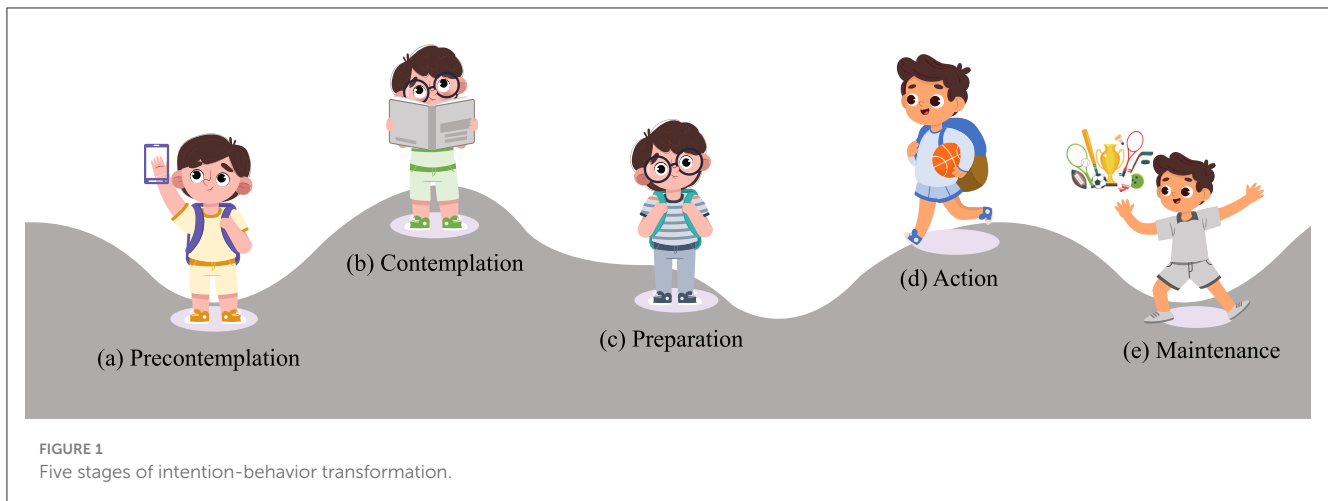
The remainder of this paper is organized as follows: Section 2 presents theoretical analysis of exercise intention-behavior gap in college students. Section 3 describes preliminaries about machine learning evaluation metrics and SHAP explanation. Section 4 presents the data sources and preprocessing methods, and introduces a comprehensive framework based on interpretable machine learning. Section 5 conducts a systematic analysis of the exercise intention-behavior gap among college students based on explainable machine learning methods. Section 6 summarizes the entire study and, based on the research findings, provides corresponding recommendations for addressing the exercise intention-behavior gap among college students. Section 7 discusses the limitations of this study and the directions for future improvements.

2 Theoretical analysis of exercise intention-behavior gap in college students

Overall, the research on the intention-behavior gap continues and expands upon traditional intention-behavior studies, still guided by the classical intention-behavior theoretical paradigm. In social psychology and health promotion, which extend from psychological theory, intention-behavior research contains two basic assumptions (24, 25). First, behavior is driven by behavioral intention. Second, most human behaviors are goal-oriented and exhibit rationality. In the context of university students' exercise behavior, this theoretical assumption reveals the gap between intention and actual behavior (26). University students' physical exercise behaviors generally go through a process of development from none to some and from weak to strong. Specifically, the formation of exercise behavior can be divided into five stages, as shown in Figure 1. When an individual's behavior remains at the first three stages and fails to transition into actual action or maintenance stages, the intention-behavior gap occurs. This is characterized by the presence of an intention to exercise but a lack of sustained actual behavior.

2.1 The perspective of health belief model

The health belief model, as a widely recognized theoretical framework for understanding individual health-related behaviors, posits that the adoption of such behaviors is largely influenced by an individual's subjective perceptions of health threats and their evaluations of coping strategies (27). The model comprises



several core dimensions, including perceived benefits, perceived barriers, perceived severity, perceived susceptibility, and self-efficacy (28). In the context of physical activity, individuals who believe that exercise can effectively prevent illness and promote health, and who feel confident in their ability to overcome exercise-related barriers, are more likely to initiate and sustain regular physical activity. College students represent a population at a pivotal stage of physical, psychological, and social development. This period is particularly sensitive for the establishment of long-term health behavior habits. While this group possesses unique advantages, such as strong autonomy and a high learning capacity, that support the adoption of healthy behaviors, they also encounter numerous challenges. These include academic pressures, career-related anxiety, adjustment to independent living, and the restructuring of social networks, all of which can undermine their exercise intentions and behaviors. Consequently, such factors may hinder the consistency and long-term stability of their health-related practices.

2.2 The perspective of theory of planned behavior

The theory of planned behavior is regarded as one of the most concise and predictive behavioral causal models in the fields of health psychology and health communication research, demonstrating considerable theoretical vitality over the subsequent decades (29). The theory suggests that an individual's intention to perform a behavior is the key proximal determinant of the actual behavior (30). Within the framework of the theory of planned behavior, the key variables include the individual's attitude toward a specific behavior, subjective norms, and the perceived ease or difficulty of performing the behavior. Thus, Based on the theory of planned behavior, it can be assumed that the more positive college students' attitudes toward physical activity are, the stronger the subjective norms, and the greater the perceived behavioral control, the stronger their intention to engage in physical activity, and the higher the likelihood of actually participating in such activities. However, despite some students expressing strong intentions to

exercise, they ultimately fail to translate these intentions into actual behavior due to factors such as poor time management, lack of willpower, or insufficient external support.

3 Preliminaries

3.1 Machine learning evaluation metrics

Machine learning models have been widely applied in various health-related research fields with the continuous development of computer science. For example, Yan et al. (31) developed a lung cancer diagnostic model using SVM. Roitman et al. (32) employed machine learning to effectively distinguish between healthy and hoarse voices. Khalid et al. (33) reliably estimated the vascular age of healthy individuals by RF. Based on the above applications, the predictive performance of machine learning models is typically evaluated using metrics such as Accuracy, Precision, Recall, and F1-Score. The corresponding calculation formulas are as follows:

- Accuracy: Measures the overall proportion of correct predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Where *TP* (True Positive), *TN* (True Negative), *FP* (False Positive), and *FN* (False Negative) represent the classification outcomes.

- Precision: Indicates the proportion of correctly predicted positive cases among all predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

- Recall: Represents the proportion of actual positive cases that are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

- F1-Score: The harmonic mean of Precision and Recall, providing a balanced measure especially useful for imbalanced

TABLE 1 Explanation of SHAP formula symbols.

Symbol	Description
SHAP(x_i)	Contribution of feature x_i to the prediction for a specific instance.
N	Set of all features, including x_i .
S	Subset of features, excluding x_i .
$f(S)$	Model's prediction when using the subset of features S .
$ S $	Number of features in subset S .
$ N $	Number of features in N .

datasets.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

3.2 SHAP explanation

The SHAP values can be used to explain the importance of different features in machine learning (34). According to (35), the calculation of SHAP values can be expressed as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

Explanation of the SHAP formula symbols as given in Table 1. In the SHAP method, each feature is assigned a SHAP value. The average SHAP value of a specific feature across all samples is considered as the contribution of that feature to the prediction outcome. Therefore, by introducing SHAP, the predictive logic of the exercise intention-behavior gap among college students can be explained.

4 Method

4.1 Data acquisition

In recent years, TikTok has garnered significant attention due to its large user base and high user engagement, particularly among college students. Therefore, we conducted an online questionnaire survey via the TikTok platform to collect data from university students aged 18 and above. To ensure data quality and relevance, participants were required to be current full-time undergraduate students and to complete all questionnaire items. A total of 2,000 responses were collected, of which 1,866 were deemed valid after excluding incomplete or inconsistent entries. The sample included students from different provinces across China, covering a variety of disciplines and university types, which enhances the representativeness of the data. Among the 1,866 valid responses, a small number of missing values are identified in individual variables. These missing values are handled using mean imputation to ensure data completeness and minimize potential bias.

4.2 Data preprocessing for exercise intention-behavior gap

Based on research by other scholars on the intention-behavior gap, the exercise gap in college students can be defined as the failure to meet the recommended physical activity standards despite having a strong intention to exercise or having a strong intention to exercise while meeting the recommended physical activity standards. In machine learning tasks, the intention-behavior gap can be treated as a binary classification label in supervised learning. According to the World Health Organization's "Physical Activity Guidelines," adults over the age of 18 should engage in at least 150 minutes of moderate-intensity physical activity per week. The research results show that among 1,866 college students with strong exercise intentions, approximately 56.27% exhibited the intention-behavior gap, while 43.73% did not. This phenomenon indicates that, despite most college students having strong intentions to exercise, there is still a significant gap in actual behavior. Specifically, more than half of the students, despite having strong exercise intentions, fail to meet the physical activity standards recommended by the World Health Organization.

4.3 Basic feature processing

To ensure a comprehensive analysis of the sample characteristics, we collected two basic demographic variables of college students: grade level and gender, which were used as feature variables in the machine learning models. According to the survey sample from TikTok, the results show significant differences in the exercise intention-behavior gap across different grade levels. Specifically, the largest number of students experiencing the gap are first-year students. This phenomenon can be attributed to the fact that first-year students are still adjusting to the new university environment. In contrast, senior students may experience a smaller intention-behavior gap due to greater autonomy, better time management, and higher levels of self-regulation developed through years of academic and social adjustment. They are also more familiar with available campus resources for physical activity, which can facilitate translating intention into action. Additionally, the proportion of females in the exercise intention-behavior gap is significantly higher than that of males. This disparity may stem from the fact that males generally have higher levels of interest in sports, competitive motivation, and exercise self-efficacy, making them more likely to actively engage in various forms of physical exercise.

4.4 Feature processing for health belief

Health belief refers to the psychological foundation upon which individuals base their engagement in health-promoting behaviors. This encompasses their cognitive and evaluative processes regarding health status, health risks, the effectiveness of health behaviors, and the feasibility of adopting such behaviors. Health beliefs among college students significantly influence their attitudes and decisions toward health behaviors. For instance,

TABLE 2 Health belief perceptual features.

Feature	Sub-feature
Perceived benefit (PBe)	1. Exercise is beneficial for health
	2. Exercise can help prevent diseases
Perceived barrier (PBa)	1. No suitable exercise facilities nearby
	2. Unable to find a suitable exercise method
Perceived severity (PSe)	1. Lack of exercise will lead to weight gain
	2. Lack of exercise will lead to low energy
Self-efficacy (SEf)	1. Exercise can be done every day
	2. Overcome the difficulties encountered during exercise
Perceived susceptibility (PSu)	1. The daily amount of exercise meets the standard
	2. Not engaging in proper exercise will accelerate aging

TABLE 3 Planned behavior perceptual features.

Feature	Sub-feature
Affective attitude (AAt)	1. A moderate level of exercise is beneficial for me
	2. A moderate level of exercise is enjoyable for me
	3. A moderate level of exercise is easy for me
Subjective norm (SNo)	1. My parents believe that exercise appropriately every day is necessary
	2. My doctor believes that I must exercise appropriately every day
	3. My partner believes that I must exercise appropriately every day
	4. I want to exercise according to my parents' wishes
	5. I want to exercise according to my doctor's advice
	6. I want to exercise according to my partner's wishes
Perceived behavioral control (PBC)	1. I can control my daily exercise level
	2. Proper exercise entirely depends on my willingness

students with a higher awareness of health risks are more likely to prioritize health-promoting behaviors, thereby increasing their likelihood of engaging in proactive health actions. Moreover, self-efficacy—the belief in one’s ability to successfully perform a specific health behavior—plays a crucial role within the context of health beliefs. The stronger a student’s self-efficacy, the more likely they are to initiate and sustain health behaviors. Consequently, the perceived dimensions of health beliefs can be categorized into five key areas: perceived benefits, perceived barriers, perceived severity, self-efficacy, and perceived susceptibility. The specific options for the variables in each of these dimensions are outlined in [Table 2](#).

TABLE 4 Result of the validation factor analysis.

Feature	Sub-feature	AVE	CR
PBe	PBe1	0.525	0.677
	PBe2		
PBa	PBa1	0.530	0.685
	PBa2		
PSe	PSe1	0.533	0.694
	PSe2		
SEf	SEf1	0.538	0.732
	SEf2		
PSu	PSu1	0.522	0.672
	PSu2		
AAt	AAt1	0.642	0.819
	AAt2		
	AAt3		
SNo	SNo1	0.511	0.862
	SNo2		
	SNo3		
	SNo4		
	SNo5		
	SNo6		
PBC	PBC1	0.586	0.771
	PBC2		

4.5 Feature processing for planned behavior

According to the Theory of Planned Behavior, an individual’s rational behavior is driven by a series of perceived conceptual variables, including affective attitude, perceived subjective norms, and perceived behavioral control. Affective attitude reflects the individual’s positive or negative evaluation of a certain behavior, often manifested through emotional responses. Perceived subjective norms refer to the social pressure or expectations from others that individuals perceive when performing a behavior, which can influence whether they choose to engage in a particular behavior. Perceived behavioral control refers to the individual’s cognition about whether they have sufficient resources, abilities, or opportunities to perform a behavior, which directly influences their behavioral intention and actual behavior. The specific options for setting the characteristics of the variables in the planned behavior perception are shown in [Table 3](#).

4.6 Analysis of the reliability and validity of perceived features

To assess whether the characteristics derived from the questionnaire settings align with the expected structure,

TABLE 5 Results of the discriminant validity test, where 3*** represents 0.003*** and 0*** denotes 0.000***.

	PBe	PBa	PSe	SEf	PSu
PBe	0.717				
PBa	−0.062(3***)	0.809			
PSe	0.354(0***)	0.237(0***)	0.728		
SEf	0.263(0***)	0.091(0***)	0.366(0***)	0.517	
PSu	0.144(0***)	0.224(0***)	0.329(0***)	0.554(0***)	0.519
	AAt	SNo	PBC		
AAt	0.807				
SNo	0.172(0***)	0.685			
PBC	0.299(0***)	0.469(0***)	0.531		

confirmatory factor analysis is conducted on the health belief perception and planned behavior perception characteristics, with the results presented in Table 4. According to the criteria (36) that the composite reliability (CR) of sub-factors within the same characteristic should exceed 0.7, or the average variance extracted (AVE) should be greater than 0.5, the extraction quality of the measurement indicators for each characteristic is deemed satisfactory. This indicates that the data structure of the characteristics aligns with the design expectations of the questionnaire. Furthermore, discriminant validity tests are performed to examine the correlations among the characteristics. The results of these tests are shown in Table 5. The findings reveal significant correlations among the characteristics ($P < 0.001$), and the absolute values of the correlation coefficients between all characteristics are smaller than the square roots of their respective AVEs. This suggests that there is sufficient differentiation among the characteristics, demonstrating good discriminant validity of the scale.

4.7 Model construction

First, the 1866 data points are split into training and testing sets in a 3:1 ratio, with 23-dimensional features used as inputs. Three machine learning models are constructed: LR, SVM, and RF. Next, by selecting the optimal model, the aim is to identify whether there is a gap between college students' exercise willingness and behavior. To improve model performance, grid search and five-fold cross-validation are used to optimize the parameters of each model. Model evaluation metrics include accuracy, precision, recall, and F1-score. Finally, the features that contribute the most to the exercise intention-behavior gap in the models are analyzed based on the SHAP method, providing theoretical support and practical guidance for subsequent research.

4.8 Model configuration

The main hyperparameters for the machine learning models used in this study are determined via grid search combined

with five-fold cross-validation to ensure optimal performance. Specifically, the Support Vector Machine (SVM) employs an RBF kernel with a penalty parameter $C = 0.5$ and a kernel coefficient

$$\gamma = \frac{1}{\text{number of features}}.$$

The Random Forest model utilizes 100 trees with a maximum depth of 30. The Logistic Regression model is configured with the lbfgs solver, a maximum of 100 iterations, and an L2 regularization penalty. All experiments are conducted on a computing platform equipped with an Intel Xeon Gold 5218 CPU (16 cores), an NVIDIA Tesla V100 GPU (32 GB memory), and 128 GB of RAM, running Ubuntu 20.04 LTS. The Python environment used is version 3.8. This setup guarantees the reproducibility and reliability of the experimental results.

5 Result analysis

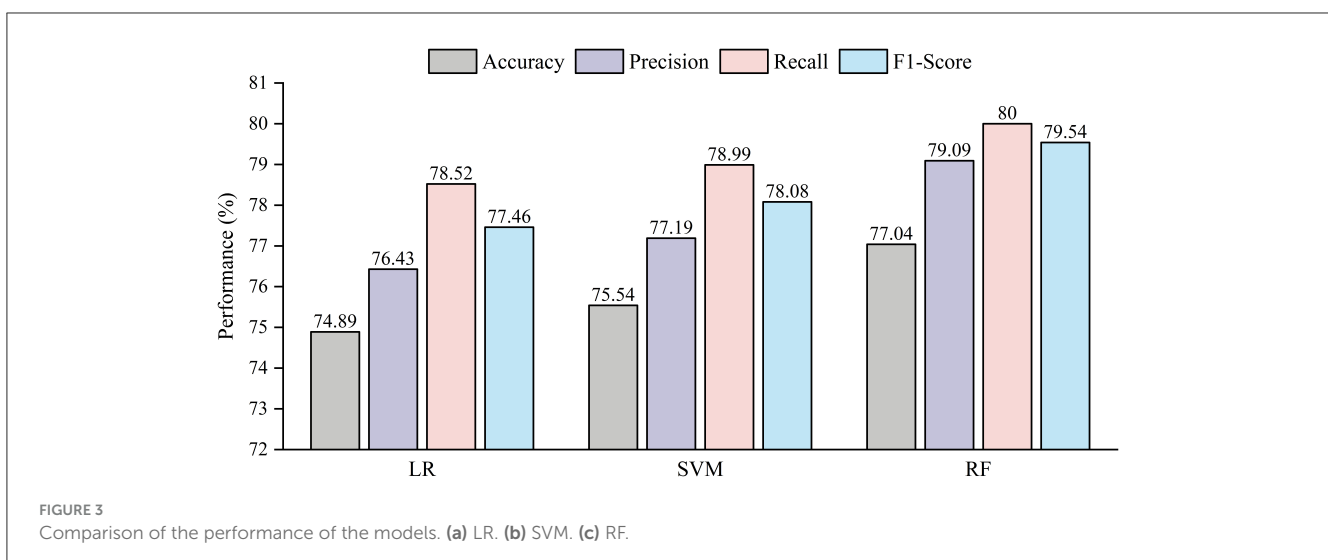
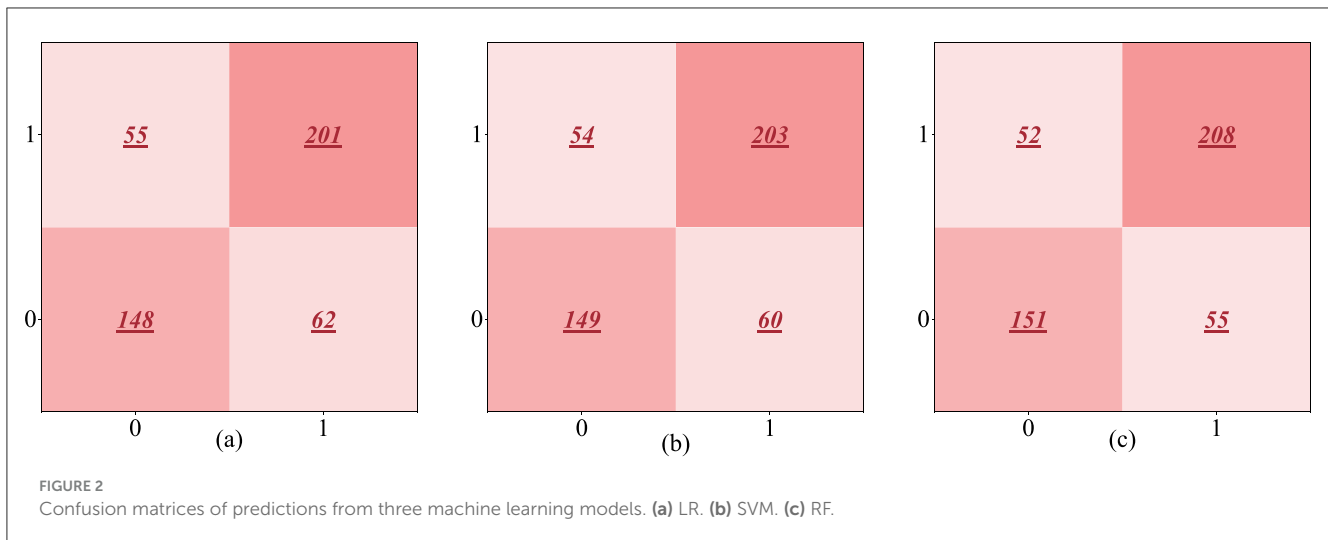
5.1 Prediction result analysis

To intuitively demonstrate the classification performance of the different models in predicting the intention-behavior gap in college students exercise behavior, we plot the confusion matrices corresponding to each model, as given in Figure 2. In these matrices, 0 represents the absence of the intention-behavior gap, while 1 indicates its presence. The color intensity reflects the number of predicted samples, with darker shades indicating a higher quantity. As shown in Figure 2, the three machine learning models effectively distinguish between the presence and absence of the exercise intention-behavior gap. This result indicates that AI-based modeling approaches successfully extract the underlying decision-making logic between exercise intention and behavior among college students by capturing multivariable interactions and the overall data structure.

In addition, Figure 3 further presents the performance of the three models in this prediction task, including accuracy, precision, recall, and F1-score. The results show that LR performs the worst, with an accuracy of 74.89%, precision of 76.43%, recall of 78.52%, and F1-score of 77.46%, indicating that the factors influencing the intention-behavior gap in college students exercise behavior exhibit strong nonlinearity. SVM performs better than LR, achieving an accuracy of 75.54%, precision of 77.19%, recall of 78.99%, and F1-score of 78.08%, suggesting its ability to capture certain nonlinear patterns in the data. Notably, RF achieves the best classification performance, with an accuracy of 77.04%, precision of 79.09%, recall of 80%, and F1-score of 79.54%, demonstrating its superior capacity to handle complex variable interactions. In summary, machine learning methods based on artificial intelligence, especially the RF model, demonstrate strong potential for predicting the intention-behavior gap in college students exercise behavior.

5.2 Feature importance analysis

To investigate the contribution of each feature variable to the intention-behavior gap between college students



exercise intentions and their actual behaviors, a SHAP-based bar chart of feature importance is constructed, as shown in Figure 4. Specifically, the average absolute SHAP values across all samples are computed to quantify the influence of each feature on the model's output. As illustrated in Figure 4, the features are ranked in descending order of importance as follows: PBa, AAt, grade, PBC, SNo, Pse, PBe, gender, SEf, and PSu. Among these, PBa exhibits the greatest influence on the intention-behavior gap, while PSu shows the least impact.

To further interpret the overall influence of each feature on the intention-behavior gap, a SHAP summary plot is generated, as shown in Figure 5. In this plot, the SHAP value greater than 0 indicate a narrowing of the gap, while value less than 0 suggest a widening. From Figure 5, increases in health belief perception and perceived behavioral control are generally associated with higher SHAP value, indicating that these factors contribute to reducing the

gap between intention and behavior. This observation aligns with findings from previous studies.

It is important to highlight that the impact of the grade in our study diverges from previous research. Our findings indicate that the intention-behavior gap diminishes as grade level increases. While earlier studies suggest that freshmen face challenges in adapting to their new environment, and seniors are constrained by pressures related to internships and job hunting, limiting their time and energy for physical activity, our results suggest a shift in this trend. With societal advancements and an increasing awareness of health benefits, more college students now recognize that regular physical activity not only enhances physical health but also alleviates stress. This recognition has fostered a more consistent engagement in exercise routines. Additionally, the SHAP value for the variables SEf and PSu are primarily clustered around zero, indicating that these variables exert minimal influence on the intention-behavior gap.

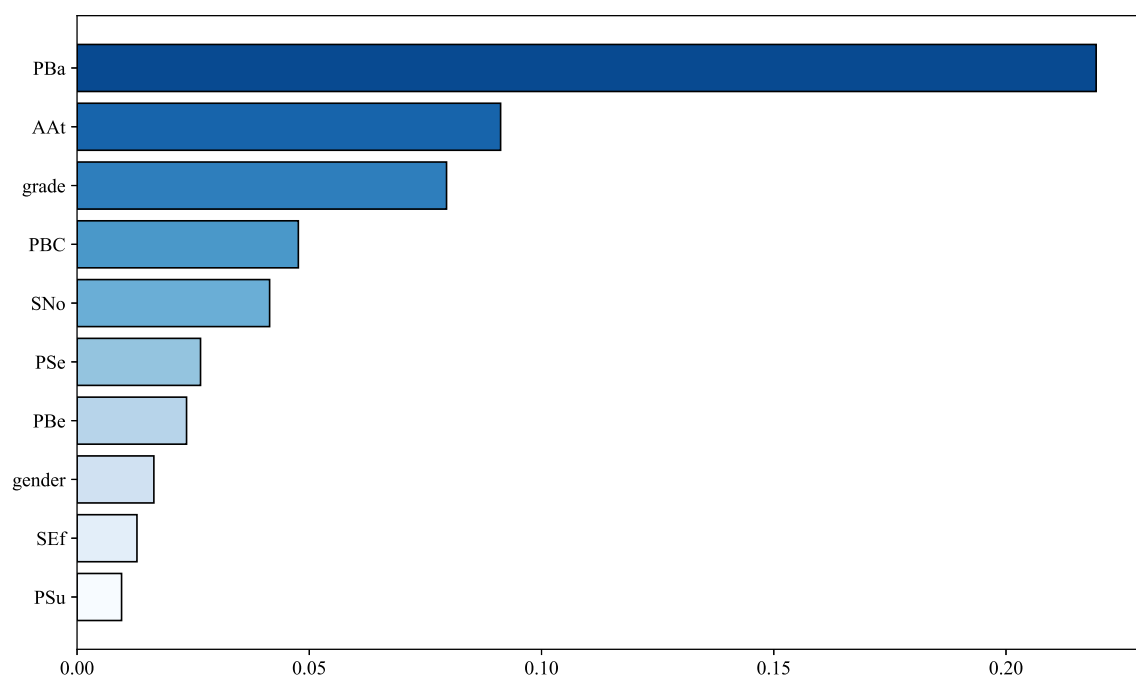


FIGURE 4
Feature bar chart based on the SHAP method.

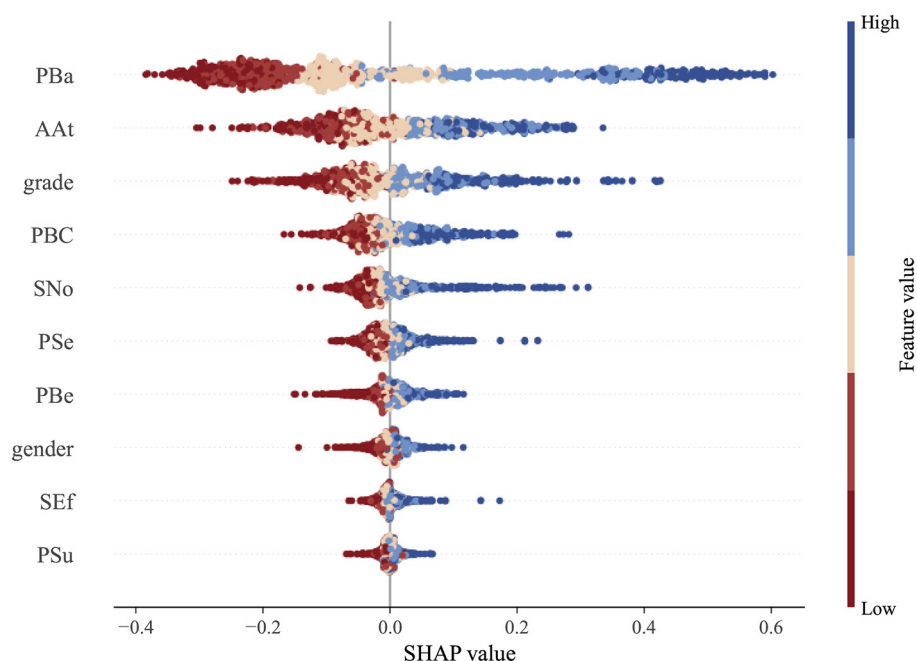
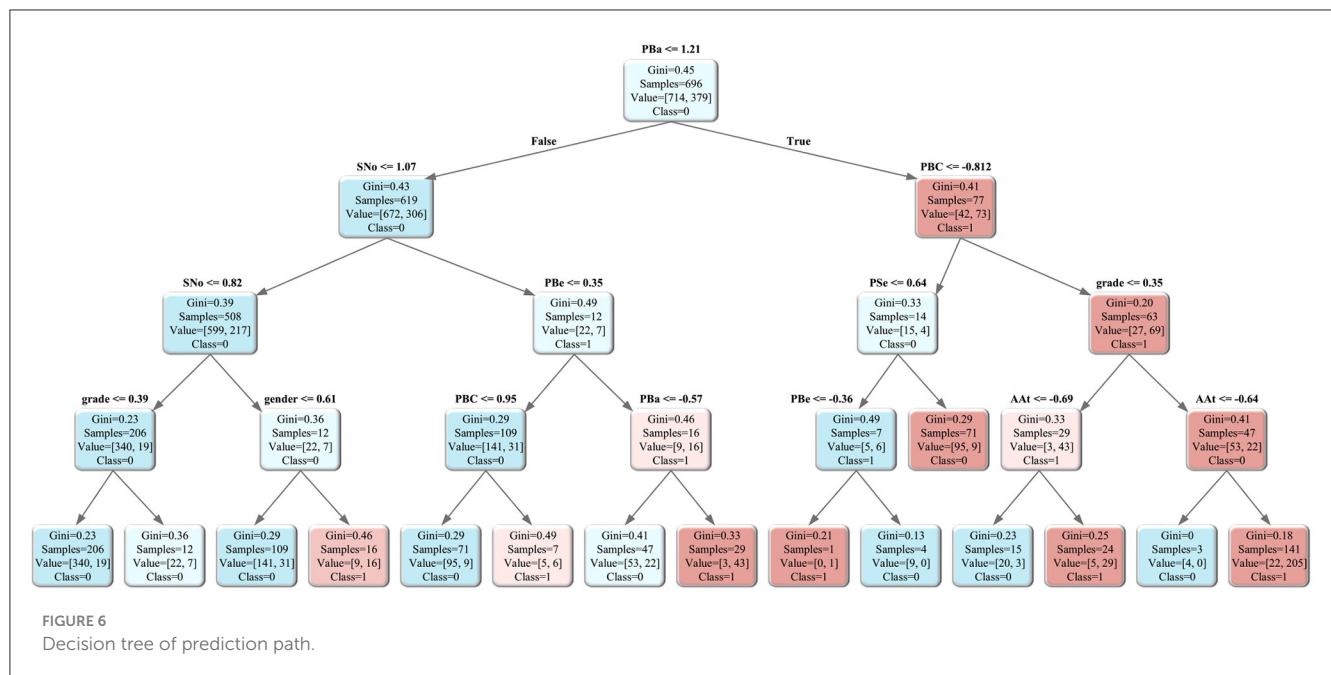


FIGURE 5
Feature summary plot based on the SHAP method.

5.3 Analysis of the impact of feature on the prediction path

Based on the analysis of key feature variables, a decision tree model is further plotted to illustrate the role of these

features in the exercise behavior transformation path, as shown in Figure 6. We tune the decision tree's parameters and find that a maximum depth of 4 best captures the key features identified in the feature importance analysis, which achieves the highest consistency between tree nodes and important variables. Therefore,



a maximum depth of 4 is selected as the final parameter for plotting. Under this setting, the Gini index-based partitioning results reveal that PBa is identified as the root node, with PBC and SNo serving as key child nodes. As shown in Figure 6, college students first consider perceived barriers to exercise when making decisions about exercise behavior. When perceived barriers are high, individuals tend to further refer to perceived subjective norms; however, when perceived barriers are low, the likelihood of an exercise intention-behavior gap decreases. Furthermore, if perceived behavioral control is weak, the probability of high-grade male students experiencing an intention-behavior gap in exercise is reduced. Overall, perceived barriers to exercise emerge as the primary predictor of the intention-behavior gap, while perceived behavioral control plays a crucial role in reducing this gap.

6 Conclusion

In this article, to accurately predict the intention-behavior gap in exercise among college students, we construct multiple machine learning models, including Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). By comparing the performance of different models, the best-performing RF model is selected, and the SHAP method is introduced for interpretability analysis. This analysis reveals the key feature variables that influence the gap between college students' exercise intentions and actual behaviors, providing intervention priorities and directions for narrowing the gap between exercise intentions and behaviors in real-world contexts. In addition, we introduce variables such as health belief perception and planned behavior perception into the research framework of the exercise intention-behavior gap, examining the multidimensional factors influencing this gap from a more holistic and systematic perspective, which has theoretical significance for the prediction and analysis of the influencing factors of the exercise intention-behavior gap.

Based on the conclusions, the following recommendations for addressing the exercise intention-behavior gap among college students are proposed:

1. Perceived barriers are a key factor influencing the existence of the intention-behavior gap in college students' exercise willingness. The ability of students with strong intentions to exercise to engage in physical activity and meet the recommended standards of physical activity largely depends on their perceptions of barriers to exercise behavior. Therefore, universities' efforts to promote and intervene in physical activity should begin by addressing and removing these perceived barriers. Specifically, universities should strategically allocate and optimize public sports resources, ensuring that students have access to sufficient exercise facilities. Moreover, universities should actively foster a supportive sports culture, which includes widespread dissemination of fundamental sports knowledge and the enrichment of physical education curricula. On this foundation, enhancing students' emotional attitudes toward the lack of physical activity will further facilitate the elimination of the intention-behavior gap.
2. Based on the explainable machine learning model, both grade level and gender exhibit a significant predictive effect on the intention-behavior gap among university students. The variations observed across different grade levels reflect the distinct challenges students face at various stages of their academic journey, including time pressures, shifting life rhythms, and evolving health awareness. These factors may influence students' motivation and capacity to engage in physical exercise. Therefore, it is essential for universities to extensively promote the benefits of physical activity for both physical and mental health. Thus, universities should widely promote the positive impact of physical exercise on physical and mental health, and enhance students' awareness of physical education by offering physical fitness development courses.

Furthermore, considering gender differences, universities should place a strong emphasis on ensuring the accessibility of sports participation for female students in the planning and management of sports facilities. This approach is crucial for enhancing their willingness to engage in physical exercise and increasing their overall participation in sports activities.

3. Based on the feature importance ranking of the model and the decision tree analysis, it is clear that perceived barriers play a primary role in predicting the exercise intention-behavior gap among university students. In contrast, the effects of emotional attitude, perceived benefits, and other feature variables appear to be less pronounced. Therefore, further research is needed to investigate the moderating effects of perceived barriers, emotional attitude, and perceived benefits on the relationship between exercise intention and behavior in university students.

7 Limitations

The data samples in this study are primarily collected from university students in China. Due to the relatively homogeneous regional and cultural background, the generalizability of the findings may be limited across the entire country or other cultural contexts. Variations in economic development, educational resources, and social culture across different areas in China may influence the intention-behavior gap in physical exercise among university students differently. In future work, we aim to expand the sample coverage to include a broader range of regions and types of universities globally to validate the model's generalizability and robustness. Moreover, factors such as urban-rural differences, economic conditions, and cultural backgrounds will be further explored to understand their impact mechanisms on students exercise intentions and behaviors. Additionally, longitudinal studies considering dynamic changes over time will be conducted to enhance the model's adaptability to real-world environments.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed

in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

CC: Conceptualization, Writing – original draft, Writing – review & editing. JY: Data curation, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Key Scientific Research Project Plan of Colleges and Universities in Henan Province “Research on the Obstacle Factors of Female College Students’ Physical Exercise in Private Colleges and Universities in Henan Province from the Perspective of Social Ecology” (No: 24A890002), part of the Young Backbone Teachers of Undergraduate Colleges and Universities in Henan Province “Research on the Obstacle Factors of Female College Students’ Physical Exercise in Private Colleges and Universities in Henan Province from the Perspective of Social Ecology” (No: 2023GGJS193), and Henan Province Soft Science Research “Project Name: Research and Practice on a New Model of Psychological Crisis Intervention for College Students from the Perspective of Deep Integration of “Physical Education, Sports, Psychology, Medicine, Health Care, and Music” – Health Promotion Integrated Cabin + Five-Tone Meridian Therapy” (No: 252400410637).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Zhou Z, Guo K, Guo S, Chen L. Relationship between physical exercise and college students' social adaptation: the chain mediating role of self-esteem and peer attachment. *Front Psychol.* (2024). 15:1453675. doi: 10.3389/fpsyg.2024.1453675
- Lv H, Zhang T, Li B, Wang R. The influence of control beliefs on the cardiovascular fitness of college students: the chain mediating effect of subjective exercise experience and exercise adherence. *BMC Public Health.* (2024) 24:1. doi: 10.1186/s12889-023-17509-3
- Li J, Wang L, Pan L, Hu Z, Yin R, Liu JF. Exercise motivation, physical exercise, and mental health among college students: examining the predictive power of five different types of exercise motivation. *Front Psychol.* (2024) 15:1356999. doi: 10.3389/fpsyg.2024.1356999
- Dong X, Huang F, Starratt G, Yang Z. Trend in health-related physical fitness for Chinese male first-year college students: 2013–2019. *Front Public Health.* (2023) 11:984511. doi: 10.3389/fpubh.2023.984511
- Sun J, Liang H, Wang L, Chen Z, Bi C, Zhang F. Secular trends of physical fitness for college students in Anhui Province over the past decade. *BMC Public Health.* (2025) 25:1. doi: 10.1186/s12889-025-21613-x
- Zhang Y, Jiang X. The effect of Baduanjin exercise on the physical and mental health of college students. *Medicine.* (2023) 102:897. doi: 10.1097/MD.00000000000034897
- Li K, Omar Dev RD, Li W. Physical activity and happiness of college students: chain mediating role of exercise attitude and sleep quality. *Front Public Health.* (2025) 13:1544194. doi: 10.3389/fpubh.2025.1544194
- Liu Y, Zhai X, Zhang Y, Jiang C, Zeng J, Yang M, et al. The promoting effect of exercise motivation on physical fitness in college students: a mediation effect model. *BMC Public Health.* (2023) 23:1. doi: 10.1186/s12889-023-17154-w
- Gao Y, Zhang J, He Z, Zhou Z. Feasibility and usability of an artificial intelligence-powered gamification intervention for enhancing physical activity among college students: quasi-experimental study. *JMIR Serious Games.* (2025) 13:e65498. doi: 10.2196/65498
- Bray SR, Harris S. Goal conflict and the physical activity intention-behavior relationship: Insights from a study of People's daily experiences. *Psychol Sport Exerc.* (2025) 78:102805. doi: 10.1016/j.psychsport.2025.102805
- Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. *Br J Sports Med.* (2023) 57:1265. doi: 10.1136/bjsports-2022-106640
- Wee ZQC, Dillon D. Increasing physical exercise through action and coping planning. *Int J Environ Res Public Health.* (2022) 19:3883. doi: 10.3390/ijerph19073883
- Lidong W, Xiuhong L, Kai Q, Donghai W. Examining the impact of perceived behavioral control and planning on closing the exercise intention-behavior gap: Insights from a meta-analytic structural equation modeling study. *Psychol Sport Exerc.* (2025) 78:102822. doi: 10.1016/j.psychsport.2025.102822
- More KR, Phillips LA. The utility of the integrated behavior change model as an extension of the theory of planned behavior. *Front Psychol.* (2022) 13:940777. doi: 10.3389/fpsyg.2022.940777
- Knapova L, Cho YW, Chow SM, Kuhnova J, Elavsky S. From intention to behavior: Within- and between-person moderators of the relationship between intention and physical activity. *Psychol Sport Exerc.* (2024) 71:102566. doi: 10.1016/j.psychsport.2023.102566
- Leung E, Urmitsky O. The narrow search effect and how broadening search promotes belief updating. *Proc Natl Acad Sci USA.* (2025) 122:e2408175122. doi: 10.1073/pnas.2408175122
- Kim S. Perceptions of discriminatory decisions of artificial intelligence: unpacking the role of individual characteristics. *Int J Hum Comput Stud.* (2025) 194:103387. doi: 10.1016/j.ijhcs.2024.103387
- Temenos A, Temenos N, Kaselimi M, Doulamis A, Doulamis N. Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. *IEEE Geosci Remote Sens Lett.* (2023) 20:1–5. doi: 10.1109/LGRS.2023.3251652
- Yang W, Li Y, Liu Y, Fan P, Yue W. Environmental factors for outdoor jogging in Beijing: Insights from using explainable spatial machine learning and massive trajectory data. *Landscape Urban Plann.* (2024) 243:104969. doi: 10.1016/j.landurbplan.2023.104969
- Khare SK, Acharya UR. An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals. *Comput Biol Med.* (2023) 155:106676. doi: 10.1016/j.compbio.2023.106676
- van der Heijden H. Decision support for selecting optimal logistic regression models. *Expert Syst. Appl.* (2012) 39:8573–83. doi: 10.1016/j.eswa.2012.01.168
- Ding S, Zhang N, Zhang X, Wu F. Twin support vector machine: theory, algorithm and applications. *Neural Comput Appl.* (2017) 28:3119–30. doi: 10.1007/s00521-016-2245-4
- Scornet E. Random forests and Kernel methods. *IEEE Trans Inf Theory.* (2016) 62:1485–500. doi: 10.1109/TIT.2016.2514489
- Conner M, Norman P. Understanding the intention-behavior gap: the role of intention strength. *Front Psychol.* (2022) 13:923464. doi: 10.3389/fpsyg.2022.923464
- Hagger MS, Cheung MWL, Ajzen I, Hamilton K. Perceived behavioral control moderating effects in the theory of planned behavior: a meta-analysis. *Health Psychol.* (2022) 41:155–67. doi: 10.1037/hea0001153
- Zhang CQ, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. *Health Psychol.* (2019) 38:623–7. doi: 10.1037/hea0000728
- Silva SB, Souza FO, Pinho PS, Santos DV. Health Belief Model in studies of influenza vaccination among health care workers. *Rev Bras Med Trab.* (2023) 21:e2022839. doi: 10.47626/1679-4435-2022-839
- Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. *Health Educ Q.* (1988) 15:175–83. doi: 10.1177/109019818801500203
- Ajzen I. The theory of planned behaviour: reactions and reflections. *Psychol Health.* (2011) 26:1113–27. doi: 10.1080/08870446.2011.613995
- Cheng O, Yam C, Cheung N, Lee P, Ngai M, Lin CY. Extended theory of planned behavior on eating and physical activity. *Am J Health Behav.* (2019) 43:569–81. doi: 10.5993/AJHB.43.3.11
- Yan L, Su H, Liu J, Wen X, Luo H, Yin Y, et al. Rapid detection of lung cancer based on serum Raman spectroscopy and a support vector machine: a case-control study. *BMC Cancer.* (2024) 24:1. doi: 10.1186/s12885-024-12578-y
- Roitman A, Edelstein Y, Katzir C, Ofir H, Peleg N, Doweck I, et al. Harnessing machine learning in diagnosing complex hoarseness cases. *Am J Otolaryng.* (2025) 46:104533. doi: 10.1016/j.amjoto.2024.104533
- Khalid MS, Quraishi IS, Nawaz MW, Sajjad H, Yaseen H, Mehmood A, et al. A low-cost PPG sensor-based empirical study on healthy aging based on changes in PPG morphology. *Physiol Meas.* (2025) 46:ada246. doi: 10.1088/1361-6579/ada246
- Wojtuch A, Jankowski R, Podlowska S. How can SHAP values help to shape metabolic stability of chemical compounds? *J Cheminf.* (2021) 13:74. doi: 10.1186/s13321-021-00542-y
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY: Curran Associates Inc. (2017). p. 4768–4777.
- Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res.* (1981) 18:39–50. doi: 10.1177/002224378101800104