SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21, 207.
https://doi.org/10.62617/mcb.v21.207

Article

Behavioral prediction and health benefit evaluation of leisure sports activities based on multi-physiological characteristics analysis

Hui Wang, Bin Liu*

Shanghai Jian Qiao University, Shanghai 201306, China
*** Corresponding author:** Bin Liu, 01009@gench.edu.cn

CITATION

Wang H, Liu B. Behavioral prediction and health benefit evaluation of leisure sports activities based on multi-physiological characteristics analysis. Molecular & Cellular Biomechanics. 2024; 21: 207.
https://doi.org/10.62617/mcb.v21.207

ARTICLE INFO

Received: 24 June 2024
Accepted: 16 July 2024
Available online: 19 August 2024

Abstract: Leisure sports activities, as a healthy form of entertainment, have garnered increasing recognition. This paper introduces a data analysis model designed for behavior prediction and health benefit evaluation in leisure sports activities, utilizing multiple physiological features. The proposed model offers recommendations for leisure sports activities and provides health assessment results based on an array of physiological feature data. Constructed using a combination of Lasso (Least Absolute Shrinkage and Selection Operator) and GBDT (Gradient Boosting Decision Tree) regression models within the Stacking ensemble learning framework, the model leverages physiological feature data from the dataset for training. Experimental results reveal that the combined prediction model achieves a coefficient of determination of 0.9832, effectively mitigating the impact of pathological data on model fitting and demonstrating superior accuracy and stability compared to individual prediction models. Finally, this paper explores the future prospects of wearable devices for physiological feature data collection and the potential advancements in behavior prediction and health benefit evaluation methods based on such information.

Keywords: analysis of physiological characteristics; leisure sports; physical fitness and health; Lasso; GBDT

1. Introduction

In recent years, there has been a growing global preoccupation and appreciation for recreational sports activities. This trend is evident not only in developed countries but also in emerging nations [1]. The concept of leisure sports refers to various physical activities that people engage in during their leisure time. These activities not only provide physical health benefits but also contribute to social and psychological well-being. Extensive research has been conducted on the health benefits of leisure sports. Studies have shown that engaging in leisure sports can effectively reduce the risk of chronic diseases [2]. Despite the plethora of positive health effects associated with recreational physical activities, they may also have adverse impacts on individuals' well-being, such as the risk of injury, overexertion, and psychological stress related to performance pressure. To tackle these issues, monitoring devices can be employed to assess and predict individuals' exercise status, providing scientific recommendations that promote health benefits while mitigating potential harm.

Presently, wearable exercise monitoring devices are emerging in great numbers as hardware devices continue to shrink. Wearables offer numerous advantages over bulky machines, such as the provision of precise, real-time data feedback to aid users in better understanding their exercise status [3].

Upon obtaining the data, it is imperative to conduct a thorough and effective analysis of the physiological information in order to arrive at accurate outcomes. Thanks to the advancements in computing power, the academic community has witnessed an array of emerging data analysis methods. Research on the utilization of regression algorithms to analyze physiological feature information and predict bodily states has permeated multiple fields, such as medicine, bioinformatics, and health management [4]. Several of these achievements have already been extensively applied in clinical practice and health management. Some commonly used regression algorithms include linear regression, logistic regression, ridge regression, and Lasso regression [5]. Furthermore, the advancement of machine learning and deep learning technologies has enabled the application of regression algorithms to become more widespread and flexible.

The problem addressed in this paper pertains to regression analysis, which involves the use of models to predict specific numerical values. We leverage regression algorithms to predict the physiological state and health benefits of users during exercise. In this paper, we present a method for analyzing data collected by wearable fitness devices, providing recommendations for leisure sports activities and evaluating health benefits. This approach can effectively enhance the health benefits of leisure sports and prevent injuries resulting from excessive exercise.

We propose a method for predicting leisure sports behavior and evaluating health benefits based on the analysis of multiple physiological features, using a combination of Lasso and GBDT regression models based on Stacking ensemble learning. We validate the effectiveness of this method through practical application. By providing recommendations for exercise types and amounts before exercise, this method reduces the occurrence of injuries resulting from excessive or improper exercise. Experimental results demonstrate that the determination coefficient of the combined prediction model can reach 0.9832, and the number of injuries caused by improper exercise methods decreases by 41% when this method is applied. Based on multiple evaluation criteria, the health benefit index generated by the recommendations provided by this method shows a 21% improvement compared to exercise performed with the same duration under other circumstances.

The following is a summary of the main contributions of this article:

1) Based on the annotated data in the dataset, a regression model that utilizes Lasso in conjunction with GBDT was trained. By comparing the predicted results of the instances, the composite prediction model yields a mean absolute error (MAE) of 1.0168, a root mean square error (RMSE) of 1.1933, and a coefficient of determination (R2) of 0.9832.
2) Using this model, we have successfully developed a methodology for predicting behavior and evaluating health benefits. This methodology leverages physiological monitoring devices, such as those that track heart rate, blood oxygen saturation levels, and step count, to provide users with exercise recommendations.
3) Analyzed and discussed were the prospects for the development of physiological feature information gathering equipment, as well as prediction of leisure sports behavior and assessment of health benefits based on such information. It was concluded that with the miniaturization of electronic devices and the

advancement of machine learning techniques, individuals can readily achieve physiological feature monitoring and health benefit assessments.

2. Related works

The relationship between health issues in modern life and leisure sports activities is closely intertwined. With the development of society and the improvement of living standards, people's lifestyles have changed, and some health problems have emerged. As health issues become increasingly prominent, daily leisure sports activities are receiving more attention from people. Proper leisure sports activities can help people alleviate health problems and maintain physical and mental health.

The impact of leisure sports activities on the physical and mental health of contemporary people has received sufficient attention from academia, and related research is relatively mature. O'Donoghue et al. [6] used a regression-based approach to model and analyze the physical status of tennis player. The results showed that the model had certain predictive ability, but the accuracy needs to be improved. In addition, Wang [7] proposed a probability-based simple Bayesian distribution regression learning framework for studying the body recovery during sports training. Bayesian regression is capable of addressing the issue of limited sample sizes by leveraging prior knowledge to stabilize the model, thereby mitigating the risk of overfitting. However, it has been observed that the efficacy of this method declines when applied to high-dimensional data.

With the advancement of hardware technology and increased computing power, wearable devices have become possible for monitoring physical activity. In their study, Zhou et al. [8] examined the characteristics of spatio-temporal data sequences and utilized L1 regularization to sparsely weight and combine multiple STELM models for predicting athletic performance. The results indicate a high level of predictive accuracy, yet with a notable increase in computational complexity and a potential risk of excessive sparsity leading to information loss. Leveraging the voluminous data generated by wearable devices, Wang and Cai [9] have achieved promising results by applying big data techniques and enhancing the K-means ant clustering algorithm. This approach has effectively addressed the issue of missing data in remote health monitoring of semi-disabled elderly individuals. In the context of the widespread application of machine learning methods, the classic machine learning algorithm of BP neural networks has found extensive use in regression problems. Wang et al. [10] have applied the BP neural network to predict the brightness and color comfort of LCD screens as perceived by the human eye. Gao et al. [11] integrated the BP neural network with a genetic algorithm based on the results of ultrasound examination for fetal weight prediction. However, the training process of the backpropagation neural network is based on gradient descent, which is prone to getting trapped in local optima and unable to find the global optimum. Additionally, the training results of the backpropagation neural network are highly dependent on the initial selection of weights, as different initial weights can lead to different outcomes. Nevertheless, the approach used to address similar issues is worth considering in this study.

Regression algorithms are a type of machine learning algorithm used to predict continuous variables. During the monitoring of physiological characteristics, devices collect a vast amount of physiological data, which can be used to train regression algorithms to predict an individual's health status. Gregori et al. [12] evaluated the trend of cardiovascular disease and diabetes development and established a healthcare cost model for physiological feature analysis. The interpretability of the study is limited, and the findings are largely dependent on the choice of the model, which could lead to false results and conclusions. Kozlovskaia et al. [13] investigated the exercise habits of recreational runners in Australia, comparing the physical, lifestyle, and training characteristic data of male and female groups using the chi-square test. They evaluated the impact of running experience on weight and health status using the multivariate logistic regression method. Using the predictions from the model, researchers evaluated the trends in weight and health status among runners, which can aid in developing recreational sports programs that include leisurely running.

In addition to the aforementioned conventional methods, the Lasso regression model and the Gradient Boosting Decision Tree algorithm in machine learning can also be employed [14]. Based on the existing literature, it is evident that a single model can achieve a good fit for linear data. Nonetheless, when the variable data exhibits a large amplitude of fluctuations, the predictive error tends to deviate substantially.

These studies evince the wide-ranging potential of regression algorithms in analyzing physiological feature information and predicting health benefits. Moreover, they can furnish useful insights for clinical practice and health management. Nonetheless, further research and development are still needed to surmount the existing challenges and limitations, thereby enhancing the accuracy and reliability of the models.

To this end, we present a novel approach that leverages Stacking ensemble learning to combine Lasso and GBDT regression models, thereby facilitating the prediction of health benefits in leisure sports activities.

3. Methods

The prediction and health benefit assessment of leisure sports activities based on multi-physiological feature analysis is a health management approach that relies on data analysis and machine learning techniques. By monitoring and analyzing a variety of physiological features of individuals and employing machine learning algorithms, this approach is capable of predicting user behaviors in different leisure sports activities and evaluating their health benefits [15].

To be more refined, the fundamental procedure of this method comprises several sequential steps:

1) Data Acquisition: The study employs state-of-the-art equipment, including sensors, wristbands, and smartwatches, to monitor a multitude of physiological indicators in individuals, including heart rate, blood oxygen saturation, and step count.
2) Data Processing: The collected data undergoes rigorous cleaning, preprocessing, and feature extraction to obtain a comprehensive profile of the individual's physiological characteristics.
3) Behavior Prediction: The present study utilizes a Stacking ensemble learning approach, combining Lasso and GBDT models, to establish a sophisticated model

for predicting an individual's physiological status during leisure sports activities.

4) Health Benefit Evaluation: Based on the prediction results, the study assesses the health benefits of individuals participating in various leisure sports activities.

This methodology not only enables individuals to gain insights into their behavioral performance and health benefits during different leisure sports activities but also provides valuable reference for health management institutions to develop tailored health management plans [16].

The methodology proposed in this manuscript is presented in its entirety in **Figure 1**.

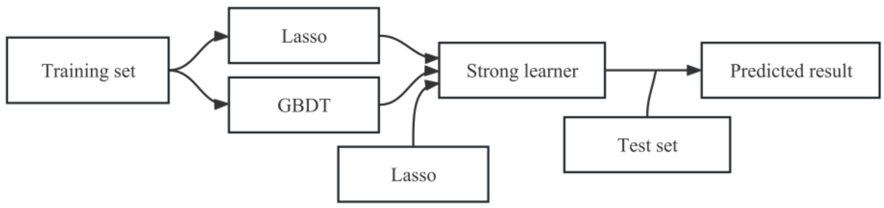

Figure 1. Processing flow.

Figure 1 illustrates the comprehensive workflow of the proposed methodology. Initially, the approach employs the Lasso regression and GBDT regression models as individual learners, utilizing a combination strategy known as Stacking. Lastly, the Lasso regression predictive model is integrated once again to forecast the predictions of individual learners, and its output serves as the ultimate prediction outcome.

3.1. Lasso

The Lasso algorithm, short for Least Absolute Shrinkage and Selection Operator, represents a linear regression method that achieves feature selection in high-dimensional datasets by introducing an L1 regularization term to the loss function, ultimately enforcing sparsity. The loss function in the Lasso algorithm comprises two parts: the squared error and the L1 regularization term. The squared error is used for data fitting, while the L1 regularization term constrains the model complexity, pushing the model coefficients towards zero and thus accomplishing feature selection [17]. The mathematical expression for the Lasso regression model is presented in Equation (1).

$$min\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_{ij}-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=0}^{p}\left|\beta_j\right|\right\} \tag{1}$$

In this context, n denotes the number of samples, p represents the number of features, x_{ij} denotes the $j-th$ feature value of the $i-th$ sample, y_{ij} denotes its corresponding label value, β_j represents the coefficient for the j-th feature, and λ is a hyperparameter used to control the strength of the regularization term.

The Lasso algorithm employs coordinate descent to optimize the model, updating only one coefficient at a time while keeping the others constant, until it converges. One of the benefits of the Lasso algorithm is that it can compress coefficients to reduce the model's complexity. Additionally, it can perform feature selection by removing features with small contributions to the prediction results [18].

3.2. GBDT

The GBDT regression model is an ensemble learning model based on the decision tree algorithm. It achieves high accuracy in predicting the target variable by iteratively training the model and performing weighted combinations [19].

As depicted in **Figure 2**, in the GBDT regression model, each decision tree serves as a weak learner, progressively enhancing the overall performance of the model through iterative processes. Specifically, the GBDT model first employs a simplistic model for prediction, then computes the error between the predicted and actual results, taking this error as the target for the next round of training, and fitting a new model to this error. This process continues until the error cannot be further reduced.

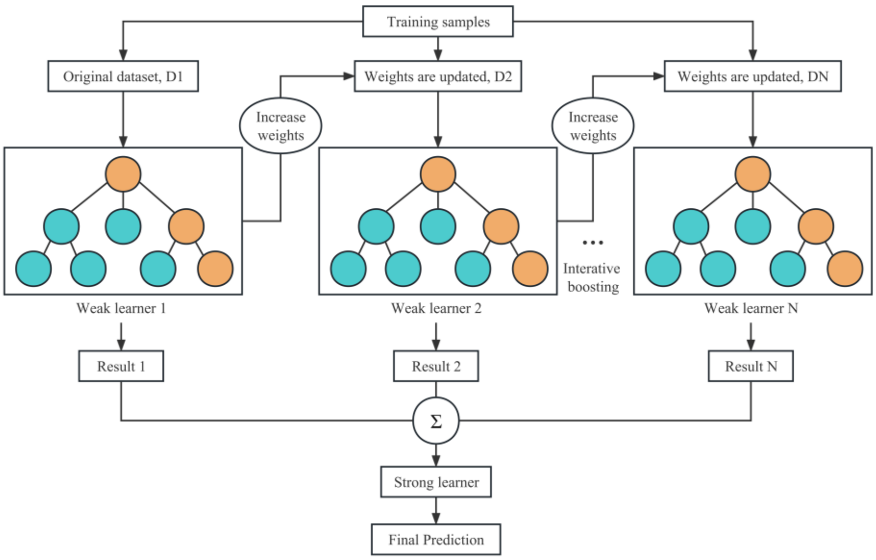

Figure 2. GBDT structure.

In each iteration, the GBDT model adjusts the sample weights based on the errors between the previous model's predictions and the actual results, to enable the subsequent training model to better fit the error. Furthermore, the GBDT model improves the generalization capability of the model and avoids overfitting by selecting and splitting features [20].

Ultimately, the gradient boosted decision tree model amalgamates all the decision trees obtained from the training process through a weighted combination, thereby yielding a final prediction. Given that each decision tree is a weak learner, and the combination process allows for mutual complementation, the GBDT model exhibits remarkable predictive and generalization capabilities.

3.3. Stacking

Stacking is a form of ensemble learning that integrates the predictions of multiple base models. It involves utilizing a meta model to blend these predictions and produce the ultimate prediction result [21].

The computational process of Stacking is illustrated in **Figure 3**. The Stacking method initially divides the dataset into several subsets, one of which is set aside as the testing set and the remaining subsets are used as the training set. In the first stage, each training set is employed to train a base model, which upon completion of training, becomes a primary learner. Each primary learner is then employed to make predictions on the testing set. In the second stage, all the predictions made in the first stage are combined to create a novel feature matrix, which serves as input to the meta-model. The feature matrix and the testing set are then utilized to train the meta-model. For testing data, the predictions made by each base model are merged to create a new feature matrix, which serves as input to the meta-model. The meta-model then generates the final prediction results.

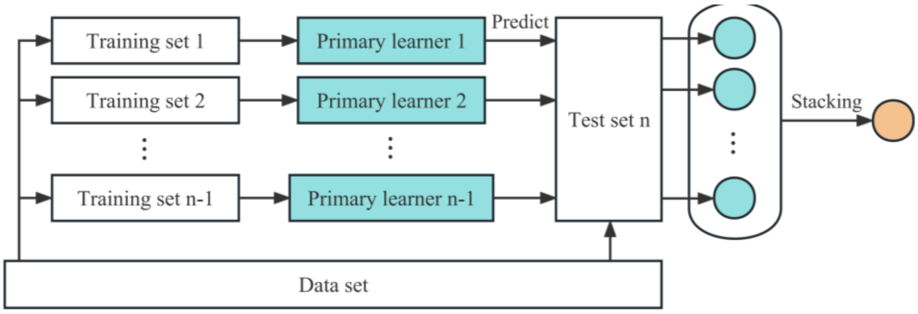

Figure 3. Stacking structure.

The advantage of Stacking lies in its ability to leverage the advantages of multiple base models, thereby enhancing the overall predictive performance [22]. Moreover, the meta-model can effectively utilize the strengths of each base model by providing weighted predictions for different base models. In this study, we have employed the Stacking technique to combine Lasso and GBDT models.

3.4. Evaluation method

Classification and prediction models' accuracy in predicting training sets cannot fully reflect the performance of the predictive models. Therefore, it is necessary to introduce evaluation metrics for correction. This article calculates the mean absolute error (*MAE*), root mean square error (*RMSE*), and coefficient of determination (R^2) for the prediction results of the three models, respectively. The Equations are presented below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y_i} - y_i| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2} \tag{3}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y_i} - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4}$$

In this Equation, the variable n denotes the total predicted time. The value of $\hat{y_i}$ represents the predicted result at time i, while y_i denotes the true value at that moment. Additionally, $\bar{y}_i$ corresponds to the mean value.

4. Experiment

4.1. Dataset preview

In our experiment, we utilized a comprehensive dataset comprising 2443 records of physiological measurements collected using state-of-the-art physiological monitoring devices. The dataset includes three key types of physiological data: heart rate, blood oxygen saturation, and step counts. Specifically, it consists of 801 heart rate measurements, 829 blood oxygen saturation records, and 813 step count observations. The physiological measurements were captured using the Fitbit Charge 4 for heart rate and step counts, and the Pulse Oximeter Model 50B for blood oxygen saturation.

The data collection process involved these advanced monitoring devices to ensure high precision and consistency in the measurements. To maintain the quality of the dataset, a rigorous preprocessing pipeline was employed. This process began with data cleaning to address any missing values, followed by normalization to standardize measurements across different physiological parameters. Outlier detection methods were also applied to identify and mitigate the impact of anomalous readings, thereby enhancing the overall reliability of the dataset.

Despite the thorough preprocessing, several potential biases may still affect the dataset. Variations in device calibration, particularly with the Fitbit Charge 4 and the Pulse Oximeter Model 50B, could introduce discrepancies in the recorded measurements. Additionally, individual health conditions of the participants could lead to variability in physiological readings, as different health states may influence heart rate, blood oxygen saturation, and step count. Environmental factors, such as temperature or humidity, could also affect the physiological measurements, further introducing potential sources of bias.

We partitioned the dataset into a training set, which accounted for 70% of the data, and a testing set, which comprised the remaining 30%. The training set was further divided into two subsets, each representing 35% of the overall dataset, and was utilized as training data for Lasso and GBDT, respectively. Our aim was to train regression models that could effectively predict users' exercise evaluation results. **Table 1** shows specific values selected for each parameter and aspect:

Table 1. Model parameters.

Aspect	Parameter Name	Reference Value
Model Parameter Selection	Regularization Strength (Lasso)	0.1
	Number of Boosting Stages (GBDT)	300
	Learning Rate (GBDT)	0.05
	Maximum Depth (GBDT)	7
	Number of Leaves (GBDT)	50
Training Process	Learning Rate (GBDT)	0.05
	Number of Boosting Rounds (GBDT)	300
	Early Stopping Rounds	15

Table 1. (*Continued*).

Aspect	Parameter Name	Reference Value
Validation Strategy	Number of Folds (Cross-Validation)	10-fold
	Stratified Sampling	Applied in classification tasks
	Repeated (Cross-Validation)	3 repetitions

Table 2 presents the distribution of the data among the different categories after the dataset was partitioned.

Table 2. The number of different categories of data.

Category	Training (Lasso)	Training (GBDT)	Testing	Total
Heart rate	280	281	240	801
Oxyhemoglobin saturation	290	290	249	829
Step number	284	285	244	813

In order to visually show the correlation of data, we used the t-SNE algorithm to reduce the dimensionality of the data vectorization results in the data set, and the visual results were shown in **Figure 4**.

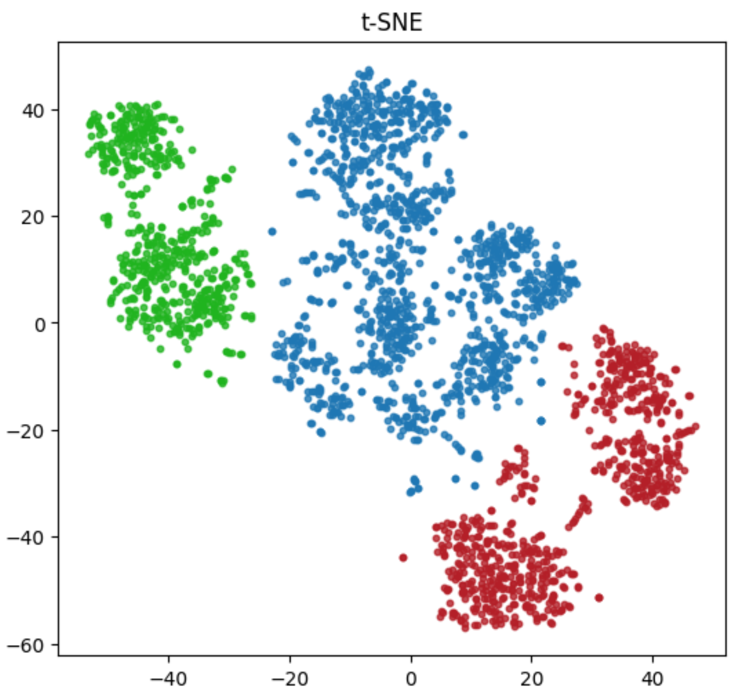

Figure 4. Visualization of t-SNE results.

The figure clearly illustrates the differential distribution of various data types. The green data points, which represent heart rate information, are located in the upper left corner of the chart. The blue data points located in the middle of the chart represent blood oxygen saturation information, while the red data points in the lower right corner represent step count information. Notably, there is a significant gap between the data points of heart rate and step count information, indicating a weak correlation between the two. However, it is possible that they are linked through blood oxygen saturation information.

4.2. Experimental method

Given the diverse units and large numerical disparities among the three physiological metrics, this study calculates a temporary evaluation value through weighted averaging of the data, as depicted in Equation (5).

$$H = \frac{r}{100} + d + \frac{s}{10000} \tag{5}$$

In this study, the variable r represents the numerical value of heart rate, measured in beats per minute (BPM), while *d* denotes the numerical value of blood oxygen saturation, measured as a percentage. The variable s represents the number of steps taken. The Lasso regression model employed in this study was assigned a weight of 20. Following the principles of the Lasso regression model, we conducted simulations on the predicted results, and the trends of these simulated results are presented in **Figure 5**.

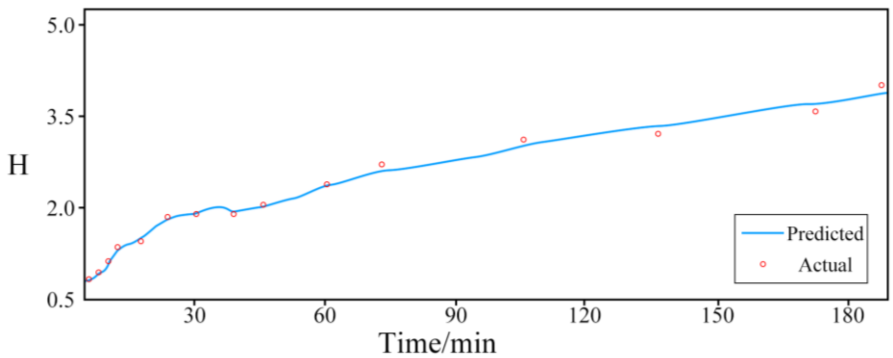

Figure 5. Comparison of Lasso predicted results with actual values (only partial actual values are shown).

From the figure, it can be observed that the Lasso regression model initially performs well in predicting the response variable, but as time progresses, the predicted values gradually deviate from the actual values. Initially, the predictions fit well with a small amount of data, but after about 60 min, a significant deviation between the predicted values and the actual values can be observed. One possible explanation for this phenomenon is that as time progresses, the increasing number of data points causes the earlier data to influence the later predictions. The Lasso model performs reasonably well, with a coefficient of determination above 0.8. However, the differences between the predictions at different times are significant, and the individual Lasso model fails to meet the accuracy requirements.

The GBDT regression model, which is based on a standard decision tree and incorporates gradient boosting, has been shown to improve the accuracy of prediction models by evolving from a single decision tree to multiple trees. In this study, the parameters used were *n* estimators = 3, max_depth = 2, and min_samples_split = 2. The experiment was repeated 500 times. The simulation predictions of the GBDT model are depicted in **Figure 6**.

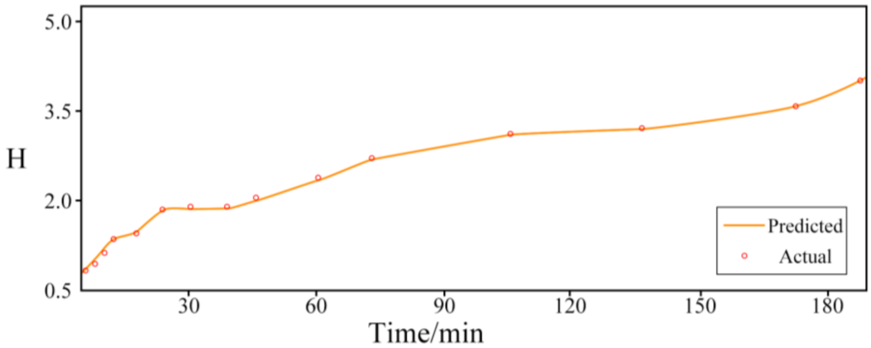

Figure 6. Comparison of GBDT predicted results with actual values (only partial actual values are shown).

From the figure, it is evident that the GBDT regression model fits well with the data beyond 70 minutes, but it fails to perform as well as the Lasso model before that time. One possible reason is that the GBDT model requires a substantial amount of data to fit the real data accurately. Although the coefficient of determination of the GBDT model is above 0.9, there exist several outliers in the early data, which require further improvement in the prediction stability of the GBDT regression model in predicting early data.

To achieve better predictions for all data, we combined Lasso and GBDT. However, the accuracy of the combined model needs further improvement, and the stability is stronger than the Lasso regression model. However, at times, the precision is lower than that of a single GBDT model. Therefore, we used the stacking method to combine Lasso and GBDT models and then used Lasso as the meta-model to integrate the predicted results. This model's stability and accuracy are substantially improved, as shown in **Figure 7**.

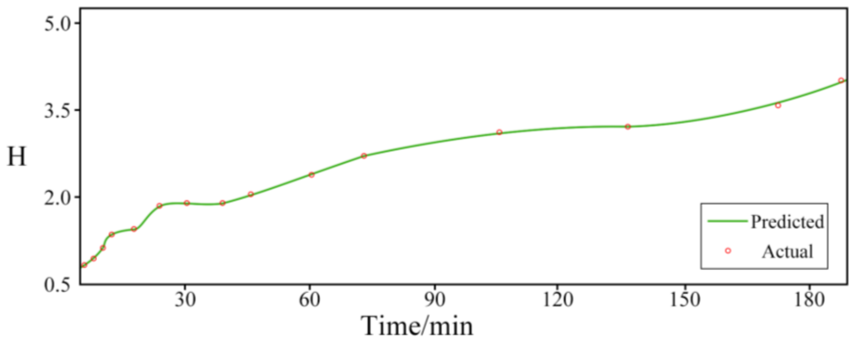

Figure 7. Comparison of Stacking predicted results with actual values (only partial actual values are shown).

From the figure, it is evident that the Stacking fused model exhibits superior overall data fitting performance. The Stacking model's coefficient of determination surpasses 0.98, effectively mitigating the adverse impact of a limited number of pathological data points on the fitting outcome.

4.3. Evaluation and results

To contrast the actual performance of various models, we conducted tests using multiple models. The evaluation results are presented in **Table 3**.

Table 3. Lasso and GBDT evaluation results.

Method	MAE	RMSE	R^2
Linear Regression	1.9141	1.9612	0.7569
Ridge Regression	1.7581	1.8021	0.8052
Lasso	1.6011	1.8427	0.8726
GBDT	1.3739	1.4853	0.9257
Stacking	1.0168	1.1933	0.9832

Table 3 presents the performance metrics of different models. The comparison reveals that the Stacking model remarkably enhances the coefficient of determination compared to individual models, while exhibiting higher stability. It is not prone to the impact of a small amount of pathological data that may affect the fitting performance. The composite model also reduces the mean absolute error and root mean squared error compared to the Linear Regression, Ridge Regression, Lasso Regression and GBDT Regression models. Therefore, the ensemble prediction model outperforms the individual models in terms of prediction accuracy and stability.

Figure 8 illustrates that the utilization of this approach resulted in a 41% reduction in injuries caused by improper exercising techniques among athletes. Moreover, adhering to the recommendations provided by this approach during the same amount of exercise time led to a 21% enhancement in the health benefit index based on multiple evaluation criteria.

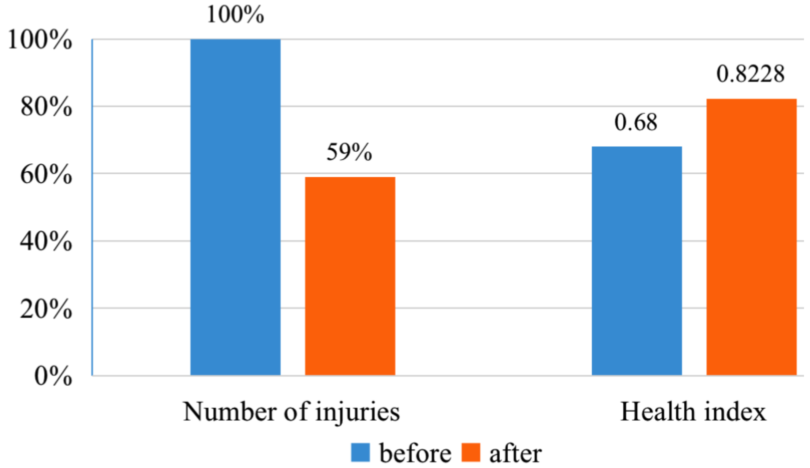

Figure 8. Comparison of application.

5. Discussion

The exploration of predicting and assessing the health benefits of leisure sports activities through multi-physiological feature analysis presents a promising avenue for enhancing personalized health recommendations. As biosensing technologies evolve and gain traction, an increasing number of studies focus on leveraging physiological features such as heart rate, respiration, and exercise posture to predict leisure sports activities and evaluate associated health benefits. Research has indicated that these physiological indicators can effectively differentiate between types of sports activities and exercise intensities, offering more tailored exercise recommendations [23].

The predictive model proposed in this study, which integrates Stacking ensemble learning with Lasso and GBDT combination regression, represents a novel approach in this domain. Our experimental results demonstrate that the Stacking method, which amalgamates multiple algorithms' strengths, addresses the limitations inherent in individual models. Specifically, while the Lasso algorithm provides a high degree of stability, it struggles with prediction accuracy. Conversely, the GBDT algorithm offers improved accuracy but lacks stability and fails to adequately fit early data. The synergy achieved through Stacking, complemented by a meta-model, enhances both accuracy and stability. The combined model achieved a mean absolute error (MAE) of 1.0168, a root mean squared error (RMSE) of 1.1933, and an R-squared value of 0.9832, indicating robust performance in predicting physiological characteristics and supporting precise exercise planning.

However, it is crucial to acknowledge the limitations of our study. First, the dataset used for training and validation may not be sufficiently diverse or extensive, potentially limiting the model's generalizability across different populations or exercise conditions. Second, the accuracy of the model is contingent upon the quality and granularity of physiological data collected. Variations in sensor accuracy or user compliance could impact the model's performance. Additionally, while our model demonstrates high accuracy and stability, real-world applications may encounter challenges related to the integration of physiological sensors into everyday sports activities and the variability of individual responses.

In-depth analysis of the results underscores the value of leveraging multi-physiological features for personalized exercise recommendations. Previous studies have highlighted the benefits of aerobic exercise in mitigating cardiovascular disease risks, improving metabolic rates, enhancing immune functions, and alleviating symptoms of mental health conditions such as anxiety and depression [24]. Our findings align with these observations, reinforcing the potential of our model to provide comprehensive health management recommendations.

Looking ahead, the advancement of technologies such as artificial intelligence and virtual reality holds significant potential for enhancing the prediction and assessment of leisure sports behaviors and health benefits. Machine learning and deep learning techniques can be further explored to extract more nuanced features from extensive physiological datasets, thereby refining prediction accuracy. Additionally, virtual reality can offer immersive sports experiences, while real-time monitoring through physiological sensors can deliver instant feedback and personalized recommendations.

Future research should focus on expanding datasets to include diverse demographics and exercise conditions, addressing sensor reliability, and exploring the integration of real-time physiological monitoring systems. By overcoming these challenges, we can enhance the applicability and effectiveness of predictive models in personalized health management and leisure sports optimization.

6. Conclusion

This article primarily investigates a model for predicting leisure sports behavior and evaluating health benefits based on multiple physiological features analysis. We discuss the potential of this method in practical applications. We have constructed a supervised learning-based algorithm for predicting physiological feature information and validated its effectiveness through comparative experiments. The experiment was conducted on a dataset collected by physiological monitoring devices. The results demonstrate that our proposed model outperforms other models in terms of accuracy and stability. The application of this method can reduce the number of injuries caused by improper exercise methods by 41% and increase the health benefits index based on multiple evaluation criteria by 21% for the same exercise time.

This study focuses on providing accurate exercise recommendations to users based on their physiological features. To monitor the user's physiological status and predict their physiological status during exercise, we utilized information obtained from wearable devices, effectively avoiding injuries caused by excessive exercise and improving health benefits. However, this study has some limitations. The amount of data we used was relatively small, and we only used historical data for simulation and prediction. Therefore, more extensive testing and research are needed to address the issue of physical differences among a large number of users. Additionally, we need to consider a range of technical and legal issues in the practical application of this method, such as privacy protection and data security. Future studies can further explore personalized exercise recommendations for different populations with varying physical conditions.

Author contributions: Conceptualization, HW and BL; methodology, HW and BL; validation, HW and BL; formal analysis, HW; data curation, BL; writing—original draft preparation, BL; writing—review and editing, HW and BL; visualization, BL; supervision, HW and BL; project administration, BL. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable comments on this paper.

Availability of data and materials: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Ling P. Interpretation of leisure sports in the pandemic situation of COVID 19. World Leisure Journal. 2020; 62(4): 319-321. doi: 10.1080/16078055.2020.1828786
2. Le Hénaff Y, Héas S. Engagement in leisure and physical activities: analysing the biographical disruptions of a rare chronic disease in France. Sociology of Health & Illness. 2019; 42(1): 65-79. doi: 10.1111/1467-9566.12987
3. Iqbal SMA, Mahgoub I, Du E, et al. Advances in healthcare wearable devices. npj Flexible Electronics. 2021; 5(1). doi: 10.1038/s41528-021-00107-x
4. Zhao Y, Shang Y, Song W, et al. Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. EClinicalMedicine. 2020; 25: 100463. doi: 10.1016/j.eclinm.2020.100463
5. Gambhir E, Jain R, Gupta A, et al. Regression Analysis of COVID-19 using Machine Learning Algorithms. In: Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC); 2020. doi: 10.1109/icosec49089.2020.9215356

6. O'Donoghue P, Cullinane A. A regression-based approach to interpreting sports performance. International Journal of Performance Analysis in Sport. 2011; 11(2): 295-307. doi: 10.1080/24748668.2011.11868549
7. Wang H. Retracted Article: Change of offshore surface water temperature characteristics based on Bayesian regression and physical recovery of sports training. Arabian Journal of Geosciences. 2021; 14(15). doi: 10.1007/s12517-021-07934-2
8. Zhou Y, Lu W, Zhang Y. Real-time monitoring of sports performance based on ensemble learning algorithm and neural network. Soft Computing; 2023. doi: 10.1007/s00500-023-08628-5
9. Wang ZR, Cai YG. Semi-disabled elderly remote health monitoring big data missing processing. Electronic Technology & Software Engineering; 2023.
10. Wang K, Ho CH, Tian C, et al. Optical health analysis of visual comfort for bright screen display based on back propagation neural network. Computer Methods and Programs in Biomedicine. 2020; 196: 105600. doi: 10.1016/j.cmpb.2020.105600
11. Gao H, Wu C, Huang D, et al. Prediction of fetal weight based on back propagation neural network optimized by genetic algorithm. Mathematical Biosciences and Engineering. 2021; 18(4): 4402-4410. doi: 10.3934/mbe.2021222
12. Gregori D, Petrinco M, Bo S, et al. Regression models for analyzing costs and their determinants in health care: an introductory review. International Journal for Quality in Health Care. 2011; 23(3): 331-341. doi: 10.1093/intqhc/mzr010
13. Kozlovskaia M, Vlahovich N, Rathbone E, et al. A profile of health, lifestyle and training habits of 4720 Australian recreational runners—The case for promoting running for health benefits. Health Promotion Journal of Australia. 2018; 30(2): 172-179. doi: 10.1002/hpja.30
14. Yang F, Li J, Liu S, et al. Lasso-GBDT-based Investment Forecasting for Distribution Transformer Replacement Projects. In: Proceedings of the 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS); 2023. doi: 10.1109/icicacs57338.2023.10100179
15. Bernal JL, Cummins S, Gasparrini A. Corrigendum to: Interrupted time series regression for the evaluation of public health interventions: a tutorial. International Journal of Epidemiology. 2020; 50(3): 1045-1045. doi: 10.1093/ije/dyaa118
16. Zhang YM, Wang H, Bai Y, et al. Bayesian dynamic regression for reconstructing missing data in structural health monitoring. Structural Health Monitoring. 2022; 21(5): 2097-2115. doi: 10.1177/14759217211053779
17. Shafiee S, Lied LM, Burud I, et al. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. Computers and Electronics in Agriculture. 2021; 183: 106036. doi: 10.1016/j.compag.2021.106036
18. Yazdi M, Golilarz NA, Nedjati A, et al. An improved lasso regression model for evaluating the efficiency of intervention actions in a system reliability analysis. Neural Computing and Applications. 2021; 33(13): 7913-7928. doi: 10.1007/s00521-020-05537-8
19. Seto H, Oyama A, Kitora S, et al. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. Scientific Reports. 2022; 12(1). doi: 10.1038/s41598-022-20149-z
20. Mansaray LR, Kanu AS, Yang L, et al. Dynamic modelling of rice leaf area index with quad-source optical imagery and machine learning regression models. Geocarto International. 2020; 37(3): 828-840. doi: 10.1080/10106049.2020.1745299
21. Koopialipoor M, Asteris PG, Salih Mohammed A, et al. Introducing stacking machine learning approaches for the prediction of rock deformation. Transportation Geotechnics. 2022; 34: 100756. doi: 10.1016/j.trgeo.2022.100756
22. Aboneh T, Rorissa A, Srinivasagan R. Stacking-Based Ensemble Learning Method for Multi-Spectral Image Classification. Technologies. 2022; 10(1): 17. doi: 10.3390/technologies10010017
23. Tennison I, Roschnik S, Ashby B, et al. Health care's response to climate change: a carbon footprint assessment of the NHS in England. The Lancet Planetary Health. 2021; 5(2): e84-e92. doi: 10.1016/S2542-5196(20)30271-0
24. Lenzen M, Malik A, Li M, et al. The environmental footprint of health care: a global assessment. The Lancet Planetary Health. 2020; 4(7): e271-e279. doi: 10.1016/S2542-5196(20)30121-2