



OPEN ACCESS

EDITED BY

Christian Maurer-Grubinger,
University Hospital Frankfurt, Germany

REVIEWED BY

Paolo Gaffurini,
University of Brescia, Italy
Datao Xu,
Ningbo University, China

*CORRESPONDENCE

Ming Zhang
✉ ming.zhang@polyu.edu.hk

RECEIVED 07 April 2025

ACCEPTED 21 August 2025

PUBLISHED 17 September 2025

CITATION

Bing F, Zhang G, Wei L and Zhang M (2025) A machine learning approach for saddle height classification in cycling.
Front. Sports Act. Living 7:1607212.
doi: 10.3389/fspor.2025.1607212

COPYRIGHT

© 2025 Bing, Zhang, Wei and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). 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.

A machine learning approach for saddle height classification in cycling

Fangbo Bing¹, Guoxin Zhang¹, Linjuan Wei¹ and Ming Zhang^{1,2*}

¹Department of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China, ²Research Institute for Sports and Technology, The Hong Kong Polytechnic University, Hong Kong SAR, China

Background: Saddle height is an important factor in bike fitting because it correlates with cycling efficiency and the risk of injuries. Conventional approaches use anthropometric parameters and joint angles as references to calculate the optimal saddle height, such as the greater trochanter height and knee flexion angle. However, these methods fail to consider individual dynamic differences in cycling.

Objective: This study proposed a machine learning (ML) model for calculating saddle height based on easily measured kinematic data.

Method: In total, 16 subjects participated in riding tests at three saddle heights. The motion capture system recorded the trajectories of markers attached to their lower limbs. Features were calculated using the hip, knee, and ankle joint angles. The optimal feature set was selected using forward sequential feature selection. The accuracies of four ML models were compared using leave-one-subject-out cross-validation.

Results: The optimal feature set contained 14 features related to the hip, knee, and ankle joint angles. The sagittal plane knee angle was the most sensitive to the saddle height, with a classification accuracy of 80%. The *k*-nearest neighbor model had the highest accuracy of 99.79% when using all the optimal features as inputs.

Conclusion: The proposed model compensates for the lack of consideration in traditional methods of individual dynamic variations in cycling, providing a more objective tool for data-driven personalization in bike fitting.

KEYWORDS

cycling, joint angle, lower limb, machine learning, saddle height

1 Introduction

Cycling is becoming more and more popular. However, the number of overuse injuries related to cycling has also increased. Proper bike fitting is important to reduce the risk of injuries and increase cycling efficiency (1). Traditional bike fitting methods rely on static measurements, empirical rules, and subjective feedback from cyclists, which may not fully account for individual biomechanical variations or dynamic riding conditions. Saddle height is one of the most studied variables in bike fitting because it has a greater impact on the range of motion (ROM) of the lower limb joints and muscles than other variables, such as handlebar height and crank length (2). A change of 2% in saddle height can significantly alter lower limb kinematics, affecting the extension and flexion angles of the hip and knee joints and their ROMs (3). Changes in saddle height of more than 4% can cause changes in oxygen uptake and

riding efficiency (2). Therefore, the lower limb joint angles are sensitive to alterations in saddle height. Previous studies have indicated that the saddle height should be set with a knee angle of 25°–35° when the crank is at the bottom dead center (BDC) (4, 5). However, static knee angles fail to match the dynamics of joint motion, especially at the 6 o'clock crank position, where the differences between the static and dynamic angles can reach $8.2 \pm 5^\circ$ (6). The peak joint loading during actual pedaling can reach two times the cyclist's body weight, which is much higher than that in the static situation (7). Moreover, cyclists tend to adjust their kinematics (e.g., pelvic rotation, ankle dorsiflexion) to compensate for suboptimal saddle heights, masking the true biomechanical relationship in static measurements (8). The force–length–velocity relationships of muscles vary among cyclists and are not detectable in static measurements. Some equations have been proposed to determine saddle height based on anthropometric measurements and joint angles (9, 10). However, these methods have not been verified in a diverse group of cyclists and may not be applicable to certain female cyclists.

In recent years, machine learning (ML) has been used in sports science, offering data-driven insights and personalized solutions (11). Previous studies have demonstrated the advantages of ML in solving practical problems in biomechanics (12, 13). Compared with motion capture systems and instrumented sensors, ML can reduce the cost and duration related to the evaluation of sports performance. Several studies have applied ML to cycling, including using long-short memory neural networks to predict heart rate (14), pulmonary oxygen uptake (15), and cadence (16). Moreover, power output in riding has been estimated without measurement by a gradient boosting algorithm (17) and a tree-based model with random forest (18). The cycling efficiency index, which reflects the cycling state, can be predicted by artificial neural networks with recursive feature elimination based on the lower limb joint kinematics, power, cadence, and individual mass (19). Pedal force is essential in assessing pedaling efficiency, but it requires equipment to measure. A neural network model was used to predict radial and mediolateral pedal forces based on power, cadence, and crank angle (20). However, vertical force could not be predicted, and the accuracy of the mediolateral force needs to be improved. In addition, a variety of competitions can benefit from cycling route optimization and race rank prediction using various ML models (21).

Notwithstanding the promising applications of ML in cycling, there are several restrictions. The dataset of most ML models is from professional cyclists and cycling races, limiting their applicability to amateur cyclists. Several studies involved fewer than 10 participants, restricting the applicability of these ML models (16, 17, 22). In addition, current studies focus on physiological metrics such as heart rate and oxygen uptake and overlook biomechanical factors such as joint kinematics and bicycle configurations.

Proper saddle height is crucial for injury prevention and pedaling efficiency, but ML applications in cycling have not adequately explored this aspect. The purpose of this study was

to develop an ML model that can accurately calculate saddle height based on lower limb joint angles during dynamic riding. Automatically distinguishing the appropriate saddle height will help cyclists optimize their riding posture for better performance and reduce joint stress to avoid overuse injuries.

2 Methods

This study consisted of two parts, as shown in Figure 1. In the cycling experiment part, subjects were recruited to perform riding tests at different saddle heights. The lower limb joint angles were calculated using Vicon Nexus 2.16. In the ML model development part, the model with the highest accuracy was constructed after comparing the performances of four ML models based on the selected optimal features.

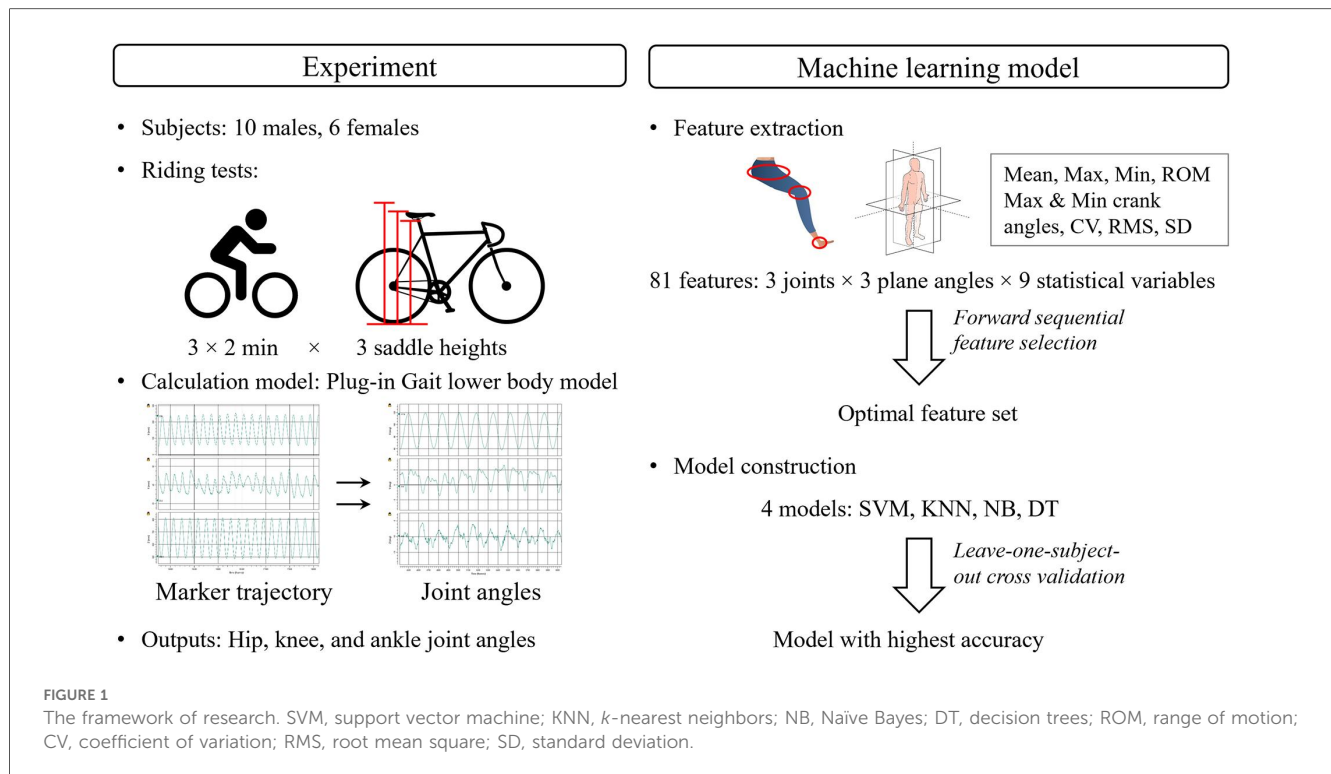
2.1 Participants

The inclusion criteria for the participants were healthy individuals between 20 and 30 years old, with a BMI between 19 and 24 kg/m² and a height of 165–180 cm for males or 155–175 cm for females. Furthermore, they were required to have reported riding more than once a week for longer than 30 min in daily life. In total, 16 amateur cyclists (10 males and 6 females, 24.64 ± 3.19 years, BMI of 21.34 ± 2.0 kg/m²) were recruited who reported that they had not been diagnosed with any musculoskeletal disease in the previous 6 months. All the participants signed informed consent forms after being informed about the experimental procedure and precautions. The experiment was approved by the university's Human Subjects Ethics Sub-Committee (Number: HSEARS20220615001).

2.2 Experiment protocol

The subjects wore uniform, tight-fitting sportswear and their own sneakers, with sole thicknesses not exceeding 3.5 cm. The riding tests were conducted on a mountain bike on a smart training platform (Tacx NEO 2T, Garmin, USA). The bike configuration was uniform except for the saddle height, which was set to low, moderate, and high levels. According to a previous study, saddle heights that were 95% of an individual's greater trochanter height (GTH) and 105% of their GTH were defined as low and high levels, respectively. A saddle height between 97% and 103% of one's GTH was defined as the moderate level.

After warm-up exercises and test riding, the participants performed three 2-min rides at low, moderate, and high saddle height levels, respectively. They were given plenty of rest time between each ride to avoid fatigue. In total, 16 reflective markers were attached to the participants as required by the lower limb model in the motion capture system (Vicon Motion Analysis Inc., Oxford, UK), and were placed at the anterior and posterior superior iliac spine, and the thigh, knee, tibia, ankle,



heel, and toes of the left and right lower limbs. The trajectories of these markers were recorded at 250 Hz by the system. The dynamic plug-in gait model was processed in Vicon Nexus 2.16 to calculate the kinematic results. The trajectory of the right ankle marker (RANK), which was placed on the lateral malleolus along an imaginary line that passed through the transmalleolar axis, and the joint angles of the hip, knee, and ankle of the right leg were output for further analysis.

2.3 Dataset

The trajectory of the RANK marker was first filtered by a zero-lag fourth-order low-pass filter with a cutoff frequency of 6 Hz. The interval between two adjacent *z*-coordinate maxima was defined as a pedaling cycle. Outlier data were excluded from the subsequent analysis. The lower limb joint angle data were divided according to the defined pedaling cycles. The synthetic minority oversampling technique was performed in Python 3.10 (Python Software Foundation, USA) to balance the number of datasets among the three groups. The number of resampled datasets was 72,354, with 24,118 in each saddle height category.

2.4 Feature extraction

A series of features were extracted from the joint angles, including the maximum and minimum angles, the corresponding crank angle for the maximum and minimum

angles (ranging from 0° to 360°), the root mean square (RMS) value, mean, standard deviation (SD), coefficient of variability (CV) (23), and the ROM (Figure 1). The hip, knee, and ankle joints each have three joint angles, namely the sagittal plane angle, coronal plane angle, and transverse plane angle (Figure 2). Therefore, 81 features (nine statistical types of features × three joints × three component angles) were extracted in every pedaling cycle, as summarized in Table 1. The features were normalized by their maximums. The final constructed feature array totaled 72,354 × 81. The label vector totaled 72,354 × 1, containing categories 1 (low saddle height), 2 (moderate saddle height), and 3 (high saddle height). The participant number vector totaled 72,354 × 1, containing numbers from 1 to 16 that were used to label the participant to which the features of each row belonged.

2.5 Machine learning model

Forward sequential feature selection in a wrapper fashion was used to select the optimal feature set from all the features. This approach begins with no features in the model and incrementally adds features based on their contribution to improving the accuracy of the classification model until the selection criteria are satisfied. The accuracy of the model was expressed as the misclassification rate, which was the number of misclassified samples as a percentage of the total number of samples (24). Five-fold cross-validation was adopted. The dataset was divided into five equally sized folds. When one fold was used as the test set, the remaining four folds formed the training set. The ML model was trained on the training set, and

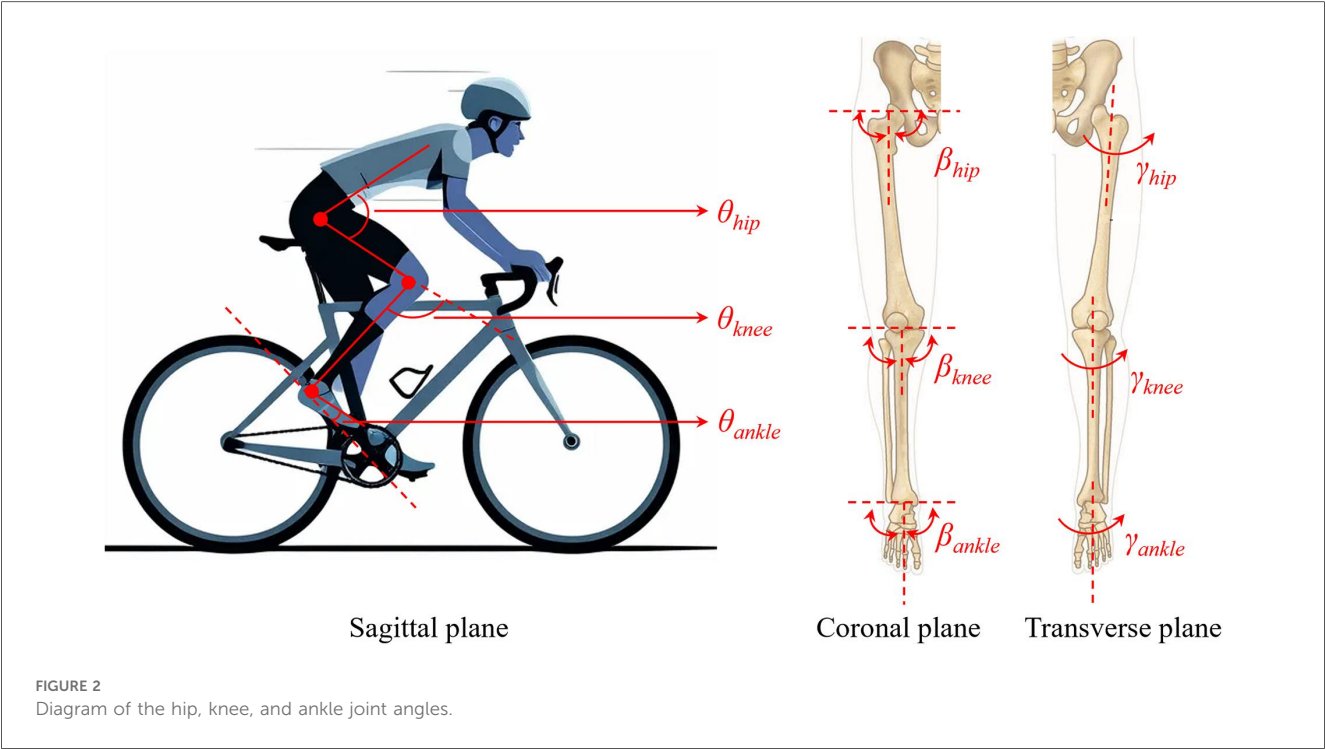


TABLE 1 Features extracted from joint angles.

Joint	Plane of angle	Features
Hip	Sagittal	$\theta_{hip, Max}$, $\theta_{hip, Min}$, $\theta_{hip, maxTiming}$, $\theta_{hip, minTiming}$, $\theta_{hip, RMS}$, $\theta_{hip, Mean}$, $\theta_{hip, SD}$, $\theta_{hip, CV}$, $\theta_{hip, ROM}$
	Coronal	$\beta_{hip, Max}$, $\beta_{hip, Min}$, $\beta_{hip, maxTiming}$, $\beta_{hip, minTiming}$, $\beta_{hip, RMS}$, $\beta_{hip, Mean}$, $\beta_{hip, SD}$, $\beta_{hip, CV}$, $\beta_{hip, ROM}$
	Transverse	$\gamma_{hip, Max}$, $\gamma_{hip, Min}$, $\gamma_{hip, maxTiming}$, $\gamma_{hip, minTiming}$, $\gamma_{hip, RMS}$, $\gamma_{hip, Mean}$, $\gamma_{hip, SD}$, $\gamma_{hip, CV}$, $\gamma_{hip, ROM}$
Knee	Sagittal	$\theta_{knee, Max}$, $\theta_{knee, Min}$, $\theta_{knee, maxTiming}$, $\theta_{knee, minTiming}$, $\theta_{knee, RMS}$, $\theta_{knee, Mean}$, $\theta_{knee, SD}$, $\theta_{knee, CV}$, $\theta_{knee, ROM}$
	Coronal	$\beta_{knee, Max}$, $\beta_{knee, Min}$, $\beta_{knee, maxTiming}$, $\beta_{knee, minTiming}$, $\beta_{knee, RMS}$, $\beta_{knee, Mean}$, $\beta_{knee, SD}$, $\beta_{knee, CV}$, $\beta_{knee, ROM}$
	Transverse	$\gamma_{knee, Max}$, $\gamma_{knee, Min}$, $\gamma_{knee, maxTiming}$, $\gamma_{knee, minTiming}$, $\gamma_{knee, RMS}$, $\gamma_{knee, Mean}$, $\gamma_{knee, SD}$, $\gamma_{knee, CV}$, $\gamma_{knee, ROM}$
Ankle	Sagittal	$\theta_{ankle, Max}$, $\theta_{ankle, Min}$, $\theta_{ankle, maxTiming}$, $\theta_{ankle, minTiming}$, $\theta_{ankle, RMS}$, $\theta_{ankle, Mean}$, $\theta_{ankle, SD}$, $\theta_{ankle, CV}$, $\theta_{ankle, ROM}$
	Coronal	$\beta_{ankle, Max}$, $\beta_{ankle, Min}$, $\beta_{ankle, maxTiming}$, $\beta_{ankle, minTiming}$, $\beta_{ankle, RMS}$, $\beta_{ankle, Mean}$, $\beta_{ankle, SD}$, $\beta_{ankle, CV}$, $\beta_{ankle, ROM}$
	Transverse	$\gamma_{ankle, Max}$, $\gamma_{ankle, Min}$, $\gamma_{ankle, maxTiming}$, $\gamma_{ankle, minTiming}$, $\gamma_{ankle, RMS}$, $\gamma_{ankle, Mean}$, $\gamma_{ankle, SD}$, $\gamma_{ankle, CV}$, $\gamma_{ankle, ROM}$

CV, coefficient of variation; Max, maximum; Min, minimum; RMS, root mean square; SD, standard deviation.

its accuracy was assessed by the test set. Five accuracy rates were obtained as each fold was used as the test set in turn. The average of the five accuracy rates was the final accuracy.

Support vector machine (SVM), *k*-nearest neighbors (KNN), naïve Bayes (NB), and decision trees (DT) models are commonly used in classification and prediction tasks (25). The accuracy of the four models was examined using leave-one-subject-out cross-validation (LOSOCV) based on the obtained

optimal feature set (24). Similarly, each subject's data were used as the test set in turn, and the remaining data from the 15 subjects were the training set. The performance of each ML model was assessed by the average accuracy from the 16 tests. The loss function was the lowest misclassification cost. Bayesian optimization was used.

2.6 Statistical analysis

The sample size of 16 was calculated using G*Power 3.1.9.7 (Universität Düsseldorf, Düsseldorf, Germany) based on a significance level of 0.05, statistical power of 0.8, and a medium effect size of 0.34 using the within-factor *F*-test with three repeated measures. Since the data did not meet the hypothesis of the normal distribution test and the assumption of the homogeneity of variance, the statistical differences in features between the three groups of saddle heights were assessed using the Friedman test with a significance level of $\alpha < 0.05$. A *post-hoc* pairwise comparison using the Wilcoxon signed-rank test with Bonferroni correction was conducted if significance was found in the Friedman test.

The classification accuracy rate based on a single feature in the optimal feature set was calculated by the ML model to characterize the contribution of each feature to the final performance of the model. The correlations between the individual features in the optimal set were measured using Pearson's correlation coefficient, *r*, which was defined as a strong ($|r| \geq 0.7$), moderate ($0.5 \leq |r| < 0.7$), low ($0.3 \leq |r| < 0.5$), or negligible correlation ($|r| < 0.3$) (26). The above analysis process, including processing the experimental data, dataset construction, feature extraction,

calculation of model accuracy, and statistical analysis, was conducted in MATLAB R2024a (MathWorks Inc., Natick, Massachusetts, USA).

3 Results

3.1 Optimal feature set

In total, 14 features were selected from 81 features to form the optimal set, including three ankle joint features ($\beta_{\text{ankle,SD}}$, $\gamma_{\text{ankle,RMS}}$, $\gamma_{\text{ankle,SD}}$), four hip joint features ($\theta_{\text{hip,SD}}$, $\beta_{\text{hip,Mean}}$, $\beta_{\text{hip,SD}}$, $\gamma_{\text{hip,RMS}}$), and seven knee joint features ($\theta_{\text{knee,Max}}$, $\theta_{\text{knee,Mean}}$, $\theta_{\text{knee,Range}}$, $\beta_{\text{knee,Mean}}$, $\beta_{\text{knee,SD}}$, $\gamma_{\text{knee,Mean}}$, $\gamma_{\text{knee,SD}}$). The selected features presented at least one set of statistical differences between the three saddle height levels, as shown in [Figure 3](#). The most notable changes were in $\theta_{\text{knee,Max}}$, $\theta_{\text{knee,Mean}}$, $\theta_{\text{knee,Range}}$, and $\beta_{\text{knee,SD}}$ ($p < 0.001$) as the saddle height increased. $\beta_{\text{knee,Mean}}$ was only statistically different between the low and high saddle height levels ($p < 0.001$). The classification accuracy of the saddle height levels based on one feature suggested a contribution of the selected feature to the final accuracy ([Figure 4](#)). $\theta_{\text{knee,Mean}}$ and $\theta_{\text{knee,Max}}$ achieved the highest accuracy rates of 80.19% and 79.58%, respectively. The top five features with the highest accuracy rankings were all related to knee joint angle. The bottom three features with the lowest accuracy rankings were $\beta_{\text{hip,SD}}$, $\beta_{\text{hip,Mean}}$, and $\theta_{\text{hip,SD}}$, with accuracy rates around 35%.

Most of the correlation coefficients between features in the optimal feature set were less than 0.5, meaning that most of the features had a low or negligible correlation with each other ([Figure 5](#)). Strong correlations existed between $\beta_{\text{ankle,SD}}$ and $\gamma_{\text{ankle,SD}}$ ($r = 0.86$), $\gamma_{\text{ankle,SD}}$ and $\gamma_{\text{hip,RMS}}$ ($r = 0.80$), $\beta_{\text{ankle,SD}}$ and $\gamma_{\text{hip,RMS}}$ ($r = 0.77$), and $\theta_{\text{knee,Max}}$ and $\theta_{\text{knee,Mean}}$ ($r = 0.77$).

3.2 Comparison of machine learning models

The classification accuracies at each saddle height level and the averaged accuracies were compared among the SVM, KNN, NB, and DT models ([Figure 6](#)). The KNN model achieved the highest average accuracy of 99.79% and outperformed the other three models. It also performed the best in the classification of each saddle height level with an accuracy of 99.96% for the low level, 99.52% for the moderate level, and 99.89% for the high level. The DT model was a bit inferior to the KNN model, with an average accuracy of 96.81%, which was higher than that of the SVM and NB models. The DT model had the lowest classification accuracy for the moderate level (93.16%) and the highest for the low level (99.47%). The average accuracy of the SVM model was 93.10%. The NB model had the lowest average accuracy (81.18%) among the four models, especially for the moderate level, with the accuracy of only 59.91%.

4 Discussion

An improper saddle height can lead to knee strain, lower back pain, and reduced power output (2, 27). Conventional methods of optimizing saddle height have predominantly relied on static anthropometric indices, such as leg length (28). However, these methods exhibit significant limitations because they neglect the dynamic interactions within the lower limb kinematic chain and interindividual heterogeneity in biomechanical responses, which lead to different definitions of the optimal saddle height. For instance, a study demonstrated that the knee joint angles of only 37% of subjects were within the recommended range of 25°–35° when using the 109% inseam method to set the saddle height (29). Static and dynamic knee angles were found to be significantly different by approximately 8° at the BDC position (4). Such contradictions underscore the methodological inadequacy of static models in capturing real-time cycling kinematics. Despite the growing acceptance of dynamic measurements, the required equipment and techniques are not always available, especially for daily training and outdoor cycling. An ML model was developed in this study to classify the saddle height level based on features extracted from the dynamically measured angles of lower limb joints. This compensates for the limitations of existing ML models in bike fitting, as most models focus on predictions of a cyclist's physiological parameters and competition performance.

The statistical analysis of features revealed statistically significant variations in lower limb joint angles in three dimensions across the three saddle height levels ([Figure 3](#)). These multiplanar kinematic changes corroborate previous findings that a reduction in saddle height increases ankle dorsiflexion, flexion and abduction of the knee, and flexion of the hip, while the ROMs of three lower limb joints also decrease (30–32). Furthermore, the most significant variation has been observed in the flexion-extension knee joint angle features because the knee joint angle has the largest ROM in the sagittal plane. This was confirmed in a previous study, as there were more significant changes in knee flexion angle and its ROM than in ankle and hip joints during cycling with varied saddle heights in both cyclists and triathletes (30). A 5% increase in saddle height caused a 25% increase in the ROM of the knee flexion angle (33). This explains our results, as the mean, maximum, and range values of the sagittal plane knee angle varied most significantly at different saddle heights. A previous study also showed a notable increase in the ROM of the ankle dorsiflexion angle from 27° to 41° as saddle height increased (33). However, our optimal feature set only included the RMS and SD values associated with the abduction-adduction and inversion-eversion ankle angles. This discrepancy may result from the fact that we used Pyro Platform shoes with front and rear heel constraints in the experiment, whereas they used common commercial cycling shoes. In addition, the mean of the abduction-adduction angle and the RMS of the external-internal rotation angle of the hip joint were included in the optimal feature set. The changes in joint kinematics in the transverse and coronal planes revealed by ML analysis may

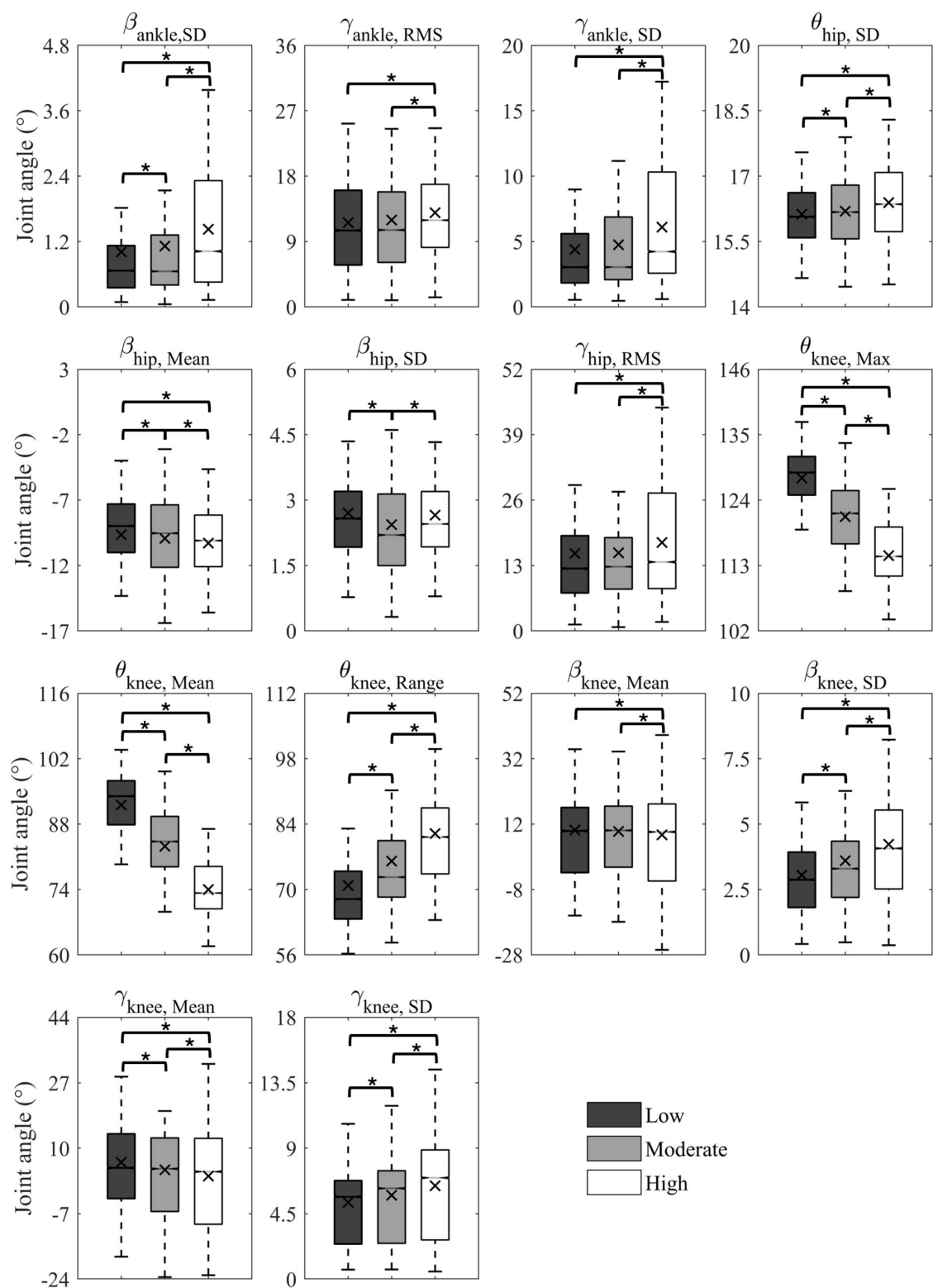


FIGURE 3 Statistical results of the selected features for the low, moderate, and high saddle heights.

provide new insights for future research, since most cycling studies have focused on the sagittal plane (34, 35). The fact that the highest classification accuracy was based on a single knee flexion angle feature emphasizes the influence of saddle height on knee kinematics (Figure 4). A lower saddle height resulted in a decreased sagittal plane angle and decreased ROM of the knee and further induced a greater knee extension moment rather than an abduction moment (31). Furthermore,

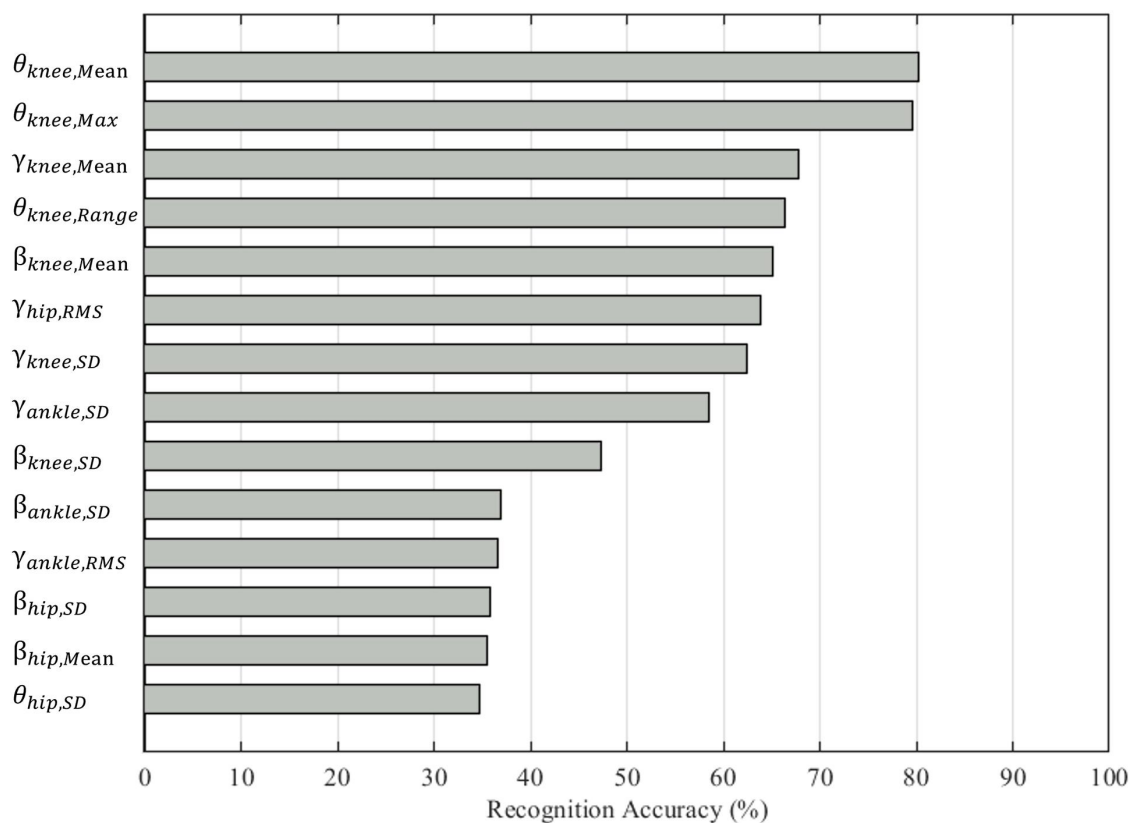
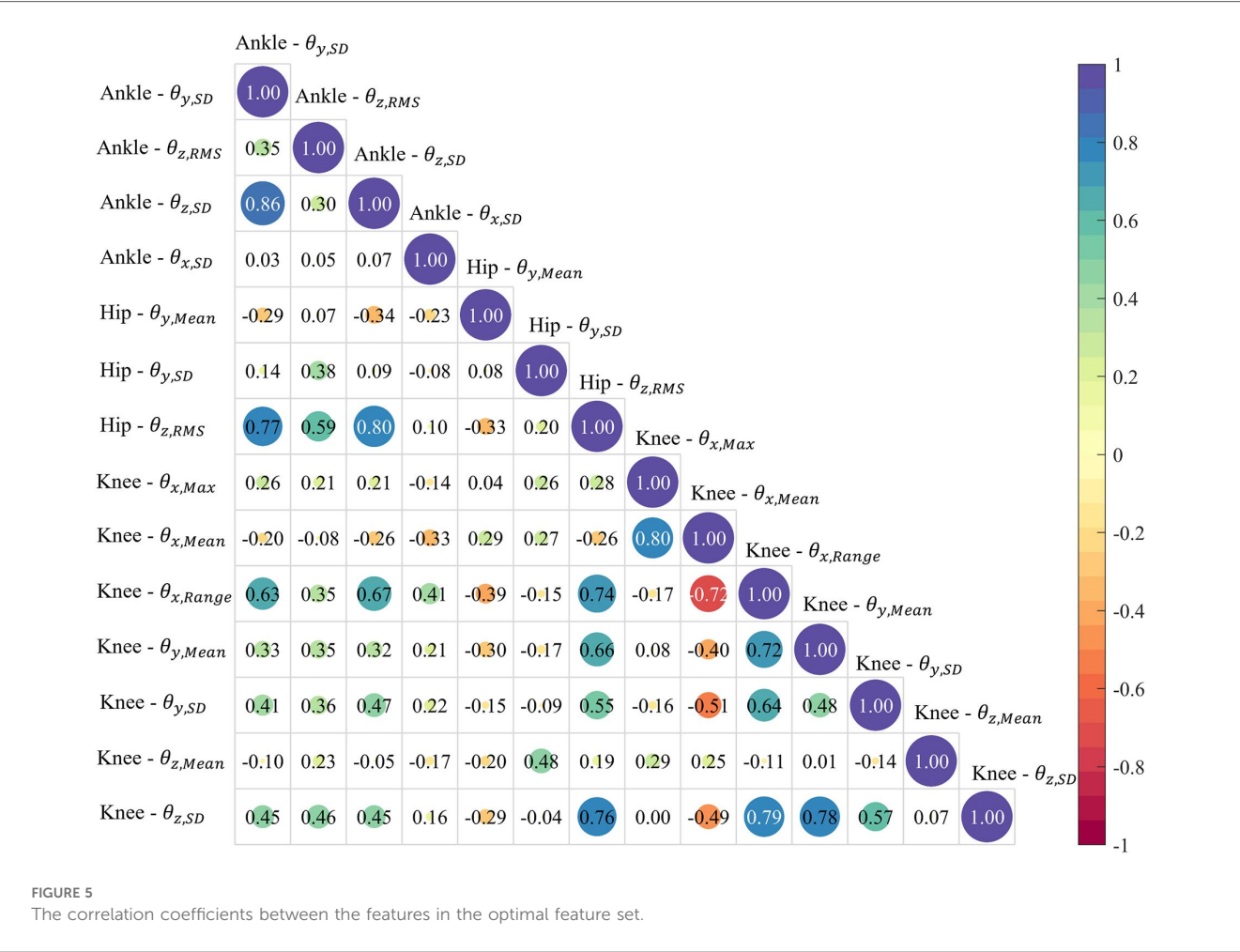


FIGURE 4
Classification accuracies based on individually selected features.

the abduction angle of the knee did not show significant changes in our study when the saddle height increased by less than 5%. The knee extension moment is an indicator of knee joint loading since it exhibits the same changing behavior at various saddle heights as the tibiofemoral compressive force (36). The SD values of the adduction/abduction and internal/external rotation angles of the knee in the optimal feature set increased at a higher saddle height, while their means declined. This may indicate that a high saddle height exacerbated oscillations and instability in the lower limbs. Therefore, saddle height adjustment is very important to prevent injuries to the knee joint.

The KNN model showed superior performance in saddle height classification compared to the other three ML models (Figure 6). However, all the models shared a common limitation: reduced accuracy in classifying the moderate saddle height level compared to the high and low levels. The NB model displayed particularly low accuracy of 59.91% for moderate heights. This may be related to the model's assumption of feature independence, but latent correlations always exist in human biomechanical datasets. Another critical reason is that the joint angles displayed more pronounced variations at extreme saddle heights. Cyclists may naturally exhibit greater movement variability when riding at moderate height deviations from their preferred position, as the biomechanical constraints

are less severe than at extreme heights. This adaptive behavior increases intraclass variation for moderate conditions. The moderate height condition (97%–103% of GTH) encompassed a wider range of saddle heights compared to the singular high (105% GTH) and low (95% GTH) conditions. This introduced greater variance in joint angles and could reduce the model's ability to identify consistent patterns for classification. Moreover, the use of the synthetic minority oversampling technique increased the data volume for the high and low saddle height groups while maintaining the original sample size for the moderate height group, which created an imbalance in data dispersion. The high/low saddle heights showed a lower numerical variance compared to the moderate group, making their classification comparatively easier. Despite the dataset only including lower limb joint angle features, the low correlation coefficients shown in Figure 5 indicated that feature selection successfully isolated complementary and non-redundant predictors. However, strong correlations were found between a subset of features. While multicollinearity may have distorted the interpretability of the model, its impact was mitigated by two factors. First, most features exhibited low correlations, preserving the diversity of input information. Second, the top-performing KNN model is non-parametric and relies on distance metrics rather than coefficient weights, reducing



sensitivity to inter-feature dependencies (37). In addition, the effects of saddle height on knee extension moments, oxygen uptake, and cycling efficiency have been identified (31, 38). Including more metrics, such as pedal force and power output, may enhance the model's sensitivity to subtle saddle height differences.

It is difficult to compare the accuracy rates of our model with other studies because we could not find similar ML models that identified saddle heights during dynamic cycling. However, the developed KNN model in this study achieved a high accuracy of 99.79%, which already demonstrated its superiority and the effectiveness of the approach. Several ML models have been developed to recognize cycling parameters such as cadence (16) and pedaling profiles (39) with high accuracies of more than 95%. Their data were acquired by inertial measurement units (IMUs). It has been reported that IMU measurement of joint angles has a lower error rate per pedaling cycle (40). Therefore, IMUs can be used instead of motion capture systems to measure joint angles in future studies, enabling real-world outdoor cycling experiments.

By utilizing lower limb joint angle features as inputs, the developed model achieved high saddle height classification accuracy, providing a more objective and personalized approach

by considering the dynamic effects in cycling. However, this study still has the following limitations. First, the recruited cyclists were not stratified by gender, age, or skill level, which could potentially affect the generalizability of the model. Gender disparity has been demonstrated in previous studies (41, 42), but it was neglected in this study. Due to differences in anthropometry, such as leg length and segment mass distribution, men and women may have different joint angles at the same saddle height. This may increase the dispersion of the original data used to train the model and subsequently affect the classification accuracy of the model. Second, the data were collected by a motion capture system and there is a discrepancy between cycling in the lab and actual outdoor cycling. Third, each feature in the optimal set was selected and verified, but some pairs of features still exhibited strong correlations. Future studies should classify participants based on their gender and cycling skills and use portable sensors, such as IMUs, to collect data during outdoor cycling (43). The accuracy and universality of the ML model can be improved by incorporating more kinematic and kinetic variables. Various ML models should be constructed for bike fitting for other configurations and cycling disciplines (e.g., road cycling vs. mountain biking).

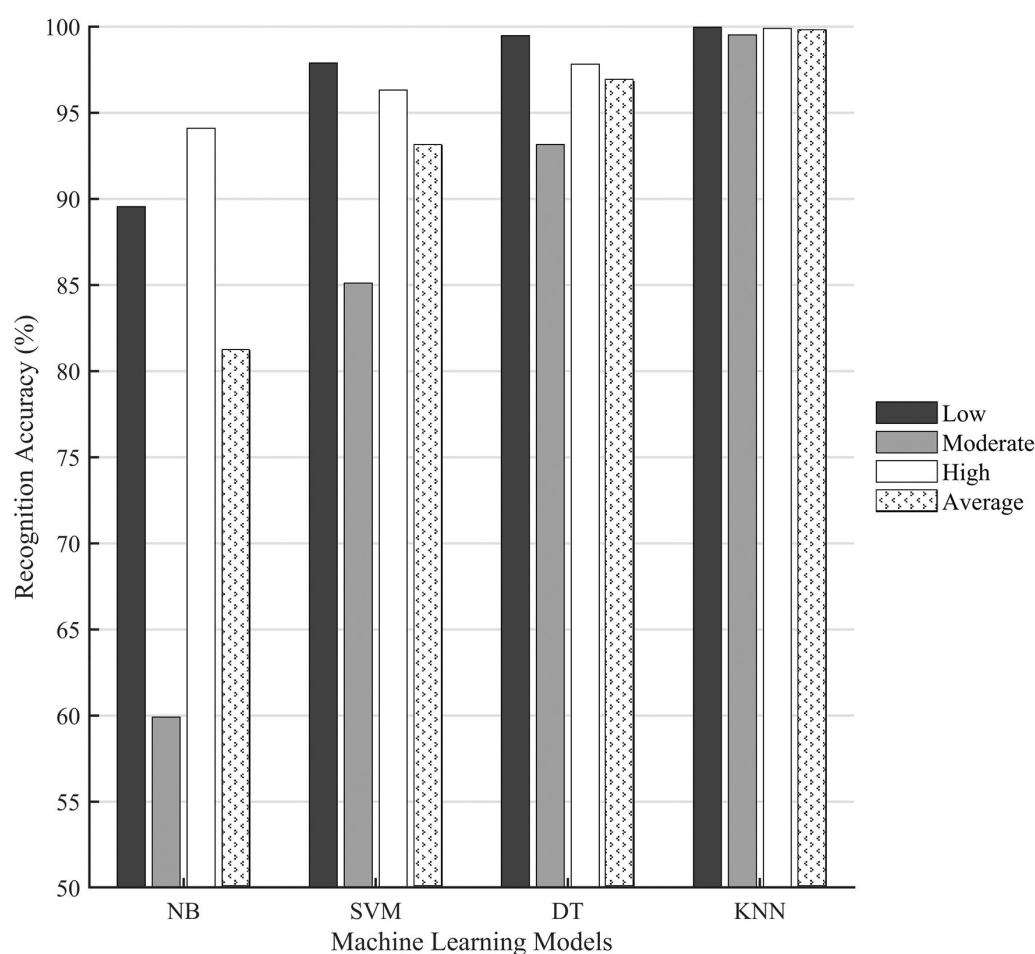


FIGURE 6

Comparison of the classification accuracies of four machine learning models. NB, Naïve Bayes; SVM, support vector machine; DT, decision tree; KNN, *k*-nearest neighbors.

5 Conclusion

This study developed a KNN machine learning model that had high accuracy when identifying saddle height levels using lower limb joint angle features. The four evaluated ML models showed lower accuracy for the moderate saddle height level compared to the low and high levels. The redundancy of the inputs and the correlations between the features were reduced by screening the optimal feature set. The sagittal plane knee joint angle was the variable most sensitive to saddle height, with a classification accuracy of 80.19% based on this feature. When the ankle and hip joint angles were included, the classification accuracy was improved to 99.79%. This approach highlights the potential for leveraging data-driven tools in cycling to provide personalized bike fitting and objective recommendations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study involving humans was approved by the Human Subjects Ethics Sub-Committee of Hong Kong Polytechnic University. The study was conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FB: Data curation, Methodology, Conceptualization, Investigation, Writing – original draft, Formal analysis. GZ: Software, Resources, Writing – review & editing. LW: Writing – review & editing, Visualization, Validation, Investigation. MZ: Project administration, Writing – review & editing, Supervision.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The authors acknowledge Hong Kong Polytechnic University for providing research funding support and are grateful to the participants.

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 Generative AI was used in the creation of this manuscript.

References

1. Quesada JIP, Kerr ZY, Bertucci WM, Carpes FP. The association of bike fitting with injury, comfort, and pain during cycling: an international retrospective survey. *Eur J Sport Sci.* (2019) 19(6):842–9. doi: 10.1080/17461391.2018.1556738
2. Bini R, Priego-Quesada J. Methods to determine saddle height in cycling and implications of changes in saddle height in performance and injury risk: a systematic review. *J Sports Sci.* (2022) 40(4):386–400. doi: 10.1080/02640414.2021.1994727
3. Ferrer-Roca V, Bescós R, Roig A, Galilea P, Valero O, García-López J. Acute effects of small changes in bicycle saddle height on gross efficiency and lower limb kinematics. *J Strength Cond Res.* (2014) 28(3):784–91. doi: 10.1519/JSC.0b013e3182a1f1a9
4. Millour G, Duc S, Puel F, Bertucci W. Comparison of static and dynamic methods based on knee kinematics to determine optimal saddle height in cycling. *Acta Bioeng Biomech.* (2019) 21(4):93–9. doi: 10.37190/ABB-01428-2019-02
5. Macedo RMB, Ulbricht L, Ripka W, Stadnik AMW, Ricieri DV. Ergonomics applied to the cycling with the utilization of two protocols for the saddle height adjust. *Occup Saf Hyg III.* (2015) 1:253–6. doi: 10.1201/b18042-50
6. Bini RR, Hume P. A comparison of static and dynamic measures of lower limb joint angles in cycling: application to bicycle fitting. *Hum Mov.* (2016) 17(1):36–42. doi: 10.1515/humo-2016-0005
7. Neptune RR, Kautz SA. Knee joint loading in forward versus backward pedaling implications for rehabilitation strategies. *Clin Biomech.* (2000) 15:528–35. doi: 10.1016/S0268-0033(00)00005-X
8. Rodrigo R, Bini ACT, Mota CB. Effects of saddle height, pedaling cadence, and workload on joint kinetics and kinematics during cycling. *J Sport Rehabil.* (2010) 19:1301–14. doi: 10.1123/jsr.19.3.301
9. Encarnación-Martínez A, Ventura F-R, García-López J. Influence of sex on current methods of adjusting saddle height in indoor cycling. *J Strength Cond Res.* (2021) 35(2):519–26. doi: 10.1519/JSC.0000000000002689
10. Gatti AA, Keir PJ, Noseworthy MD, Beauchamp MK, Maly MR. Equations to prescribe bicycle saddle height based on desired joint kinematics and bicycle geometry. *Eur J Sport Sci.* (2022) 22(3):344–53. doi: 10.1080/17461391.2021.1902570
11. Richter C, O'Reilly M, Delahunt E. Machine learning in sports science: challenges and opportunities. *Sports Biomech.* (2024) 23(8):961–7. doi: 10.1080/14763141.2021.1910334
12. Moissenet F, Chèze L, Dumas R. Individual muscle contributions to ground reaction and to joint contact, ligament and bone forces during normal gait. *Multibody Syst Dyn.* (2017) 40(2):193–211. doi: 10.1007/s11044-017-9564-9
13. Xu D, Zhou H, Quan W, Jiang X, Liang M, Li S, et al. A new method proposed for realizing human gait pattern recognition: inspirations for the application of sports and clinical gait analysis. *Gait Posture.* (2024) 107:293–305. doi: 10.1016/j.gaitpost.2023.10.019
14. Hilmkil A, Ivarsson O, Johansson M, Kuylensstierna D, van Erp T. Towards Machine Learning on Data from Professional Cyclists. *arXiv preprint arXiv:180800198* (2018).
15. Zignoli A, Fornasiero A, Ragni M, Pellegrini B, Schena F, Biral F, et al. Estimating an individual's oxygen uptake during cycling exercise with a recurrent neural network trained from easy-to-obtain inputs: a pilot study. *PLoS One.* (2020) 15(3):e0229466. doi: 10.1371/journal.pone.0229466
16. Hollaus B, Volmer JC, Fleischmann T. Cadence detection in road cycling using saddle tube motion and machine learning. *Sensors (Basel).* (2022) 22(16):6140. doi: 10.3390/s22166140
17. Lemaitre G, Lemaitre C. Estimate power without measuring it: a machine learning application. In: Zabala M, editor. *Science & Cycling Congress; 2018 Jul 4–5; Nantes, France.* Granada: Cycling Research Center (2018). p. 48–9.
18. Kataoka Y, Gray P. Real-Time power performance prediction in tour De France. *Machine Learning and Data Mining for Sports Analytics.* Lecture Notes in Computer Science (2019). p. 121–30
19. Torres A, Nougrou F, Domingue F. Machine learning models to predict kinetic variables in cycling. In: Zabala M, editor. *Science & Cycling Congress; 2024 Jun 26–27; Florence, Italy.* Granada: Cycling Research Center (2024). doi: 10.28985/1920.jsc.01
20. Ahmadi R, Rasoulani S, Veisari SF, Parsaei A, Heidary H, Herzog W, et al. A machine learning approach for predicting pedaling force profile in cycling. *Sensors (Basel).* (2024) 24(19):6440. doi: 10.3390/s24196440
21. Kholkin L, De Schepper T, Verdonck T, Latré S, editors. A machine learning approach for road cycling race performance prediction. *Machine Learning and Data Mining for Sports Analytics: 7th International Workshop; 2020 September 14–18; Ghent, Belgium:* Springer.
22. Cheshier SM, Martinotti C, Chapman DW, Rosalie SM, Charlton PC, Netto KJ. Automatic recognition of motor skills in triathlon: a novel tool for measuring movement cadence and cycling tasks. *J Funct Morphol Kinesiol.* (2024) 9(4):269. doi: 10.3390/jfmk9040269
23. Zhang G, Fu Y, Wei L, Bing F, Cai H, Chen TLW, et al. Continuously monitoring runners' adaptive strategies to prolonged running on an outdoor track with straight and curved paths: insights from the varying intersegment coordination variability and shock absorption during a full marathon. *J Sci Med Sport.* (2025):9. doi: 10.1016/j.jsams.2025.05.012
24. Zhang G, Hong TT, Li L, Zhang M. Automatic detection of fatigued gait patterns in older adults: an intelligent portable device integrating force and inertial measurements with machine learning. *Ann Biomed Eng.* (2024) 53:48–58. doi: 10.1007/s10439-024-03603-z
25. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. *Decis Anal J.* (2022) 3:10071. doi: 10.1016/j.dajour.2022.100071
26. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* (2012) 24(3):69–71.
27. Bini R, Hume PA, Croft JL. Effects of bicycle saddle height on knee injury risk and cycling performance. *Sports Med.* (2011) 41:463–76. doi: 10.2165/11588740-000000000-00000
28. Peveler W, Bishop P, Smith J, Richardson M, Whitehorn E. Comparing methods for setting saddle height in trained cyclists. *J Exerc Physiol Online.* (2005) 8(1):51–5.

29. Peveler WW, Green JM. Effects of saddle height on economy and anaerobic power in well-trained cyclists. *J Strength Cond Res.* (2011) 25(3):629–33. doi: 10.1519/JSC.0b013e3181d09e60
30. Bini RR, Hume PA, Kilding AE. Saddle height effects on pedal forces, joint mechanical work and kinematics of cyclists and triathletes. *Eur J Sport Sci.* (2014) 14(1):44–52. doi: 10.1080/17461391.2012.725105
31. Hummer E, Thorsen T, Zhang S. Does saddle height influence knee frontal-plane biomechanics during stationary cycling? *Knee.* (2021) 29:233–40. doi: 10.1016/j.knee.2021.01.026
32. Seo J-W, Choi J-S, Kang D-W, Bae J-H, Tack G-R. Relationship between lower-limb joint angle and muscle activity due to saddle height during cycle pedaling. *Korean J Sport Biomech.* (2012) 22(3):357–63. doi: 10.5103/kjsb.2012.22.3.357
33. Sanderson DJ, Amoroso AT. The influence of seat height on the mechanical function of the triceps surae muscles during steady-rate cycling. *J Electromyogr Kinesiol.* (2009) 19(6):e465–71. doi: 10.1016/j.jelekin.2008.09.011
34. Holliday W, Theo R, Fisher J, Swart J. Cycling: joint kinematics and muscle activity during differing intensities. *Sports Biomech.* (2023) 22(5):660–74. doi: 10.1080/14763141.2019.1640279
35. Gatti AA, Keir PJ, Noseworthy MD, Beauchamp MK, Maly MR. Hip and ankle kinematics are the most important predictors of knee joint loading during bicycling. *J Sci Med Sport.* (2021) 24(1):98–104. doi: 10.1016/j.jsams.2020.07.001
36. Holmes JC, Pruitt AL, Whalen NJ. Lower extremity overuse in bicycling. *Clin Sports Med.* (1994) 13(1):187–203. doi: 10.1016/S0278-5919(20)30363-X
37. Goyal R, Chandra P, Singh Y. Suitability of Knn regression in the development of interaction based software fault prediction models. *IERI Proc.* (2014) 6:15–21. doi: 10.1016/j.ieri.2014.03.004
38. Ferrer-Roca V, Andreu R, Galilea P, García-López J. Influence of saddle height on lower limb kinematics in well-trained cyclists. *J Strength Cond Res.* (2012) 26(11):3025–9. doi: 10.1519/JSC.0b013e318245c09d
39. Xu JY, Nan X, Ebken V, Wang Y, Pottier GJ, Kaiser WJ. Integrated inertial sensors and mobile computing for real-time cycling performance guidance via pedaling profile classification. *IEEE J Biomed Health Inform.* (2015) 19(2):440–5. doi: 10.1109/JBHI.2014.2322871
40. Obradović S, Stančin S. Inertial measurement units' reliability for measuring knee joint angle during road cycling. *Electronics (Basel).* (2023) 12(3):751. doi: 10.3390/electronics12030751
41. Beltrami FG, Schaer CE, Spengler CM. Sex differences in quadriceps and inspiratory muscle fatigability following high-intensity cycling. *J Sci Med Sport.* (2023) 26(3):208–13. doi: 10.1016/j.jsams.2023.02.006
42. Glace BW, Kremenec IJ, McHugh MP. Sex differences in central and peripheral mechanisms of fatigue in cyclists. *Eur J Appl Physiol.* (2012) 113(4):1091–8. doi: 10.1007/s00421-012-2516-4
43. Zhang G, Chen TL-W, Wei L, Bing F, Cai H, Liu Y, et al. Investigating rearfoot asymmetry in male marathon runners: dual imus reveals biomechanical trade-offs related to performance maintenance. *J Biomech.* (2025) 191:112906. doi: 10.1016/j.jbiomech.2025.112906