



OPEN Plantar pressure classification and feature extraction based on multiple fusion algorithms

Xiaotian Bai¹, Xiao Hou², Yiling Song¹, Zhengyan Tang¹, Hongfeng Huo^{3,4} & Jingmin Liu¹✉

Using multiple fusion algorithms to optimize the classification and feature extraction of plantar pressure during walking stance phase in healthy people, and explore the diversity of plantar pressure distribution. 243 healthy young male individuals was studied to collect data on plantar impulse and maximum pressure indices from ten distinct regions of the foot during walking. Principal component analysis was utilized to reduce the dimensionality of the data. Optimized clustering and feature extraction algorithms categorized the plantar pressure characteristics and extracted key indicators. Classification discriminant functions were developed using linear discriminant analysis. Analysis of variance compared the differences in features between various plantar pressure distribution patterns. Three types of plantar pressure distribution were identified by multiple fusion algorithms, and four indicators were extracted, including impulses of Toe1, Meta1, Meta5 and Midfoot. The average accuracy rates of original data and cross-validation were 89.70% and 88.50%. Based on one-way analysis of variance, the distribution types were ultimately determined as thumb extension type, midfoot-lateral forefoot push-off type, and normal type. Plantar pressure distribution during walking in healthy people can be categorized into thumb extension type, midfoot-lateral forefoot push-off type, and normal type. Among them, the impulses around the first metatarsophalangeal joint region, fifth metatarsal bone region and midfoot region showed better classification performance. It is recommended that future studies combine the current findings and use prospective studies to further analyze the relationship between gait characteristics and sports injuries.

Keywords Gait, Plantar pressure, Cluster analysis, Feature extraction

Walking is the most common form of human locomotion, and the foot, as the organ in direct contact with the ground, interacts with the surface to generate plantar pressure during ambulation^{1–3}. Research shows that during walking, different foot sole regions mainly perform cushioning and propulsion at different stance phases, with the foot regions working together to demonstrate distinct dynamic characteristics for each function¹. During the cushioning phase, different regions of the foot bear the load, leading to a gradual buildup of force over time. This can be reflected by the impulse and pressure, which indicate the key areas for foot cushioning and those at greater risk of injury^{2–4}. In the propelling phase, the impulse and maximum pressure of the foot's regions also evaluate the main supporting and key force points during foot push-off^{2,5}. The distribution of plantar pressure exhibits distinct characteristics across different populations. In patients with heel pain, plantar pressure in the rearfoot shifts more towards the forefoot during walking compared to the general population⁶. Individuals with flat feet show greater midfoot loading both in standing and walking^{2,7}, whereas those with high arches experience lower midfoot loading compared to normal feet^{2,8}. For individuals with hallux valgus, studies have found a lateral shift in forefoot plantar pressure during movement due to the abnormal structure of the big toe^{9,10}. Additionally, research has shown a negative correlation between peak plantar pressure and perceived foot comfort^{11,12}. Measuring peak plantar pressure during walking provides valuable insights for evaluating insole design. Thus, the dynamic characteristics of the plantar are key indicators for assessing foot function and informing shoe design.

In the analysis of abnormal foot types and foot diseases, existing research has developed relatively comprehensive testing methods and evaluation systems for plantar pressure^{7–9}. However, the above studies are based on the assumption that all normal foot types exhibit the similar plantar pressure distribution. As a result,

¹Department of Physical Education, Tsinghua University, Beijing, China. ²School of Sport Science, Beijing Sport University, Beijing, China. ³College of Physical Education, Hebei Normal University, Shijiazhuang, China. ⁴Key Laboratory of Bioinformatics Evaluation of Human Movement, Hebei Normal University, Shijiazhuang, China. ✉email: ljm_th@mail.tsinghua.edu.cn

they are grouped into a single control group. In reality, for the general population, the complexity of foot types and postures, along with the variability in foot structure and morphology among individuals, leads to a diverse distribution of plantar pressure^{2,13,14}. De Cock et al. found that during jogging, the support patterns of the forefoot in healthy young individuals can be categorized into four types¹⁵. This variability may be related to their running habits but also suggests the presence of diverse plantar pressure distribution patterns within healthy populations. The complexity of foot structure also means that there may be redundancy in the pressure data, necessitating the extraction of data that have the most significant classification effects to ensure the accuracy and practicality of gait analysis. Although some studies have classified foot types based on morphological characteristics^{16,17}, such studies focus more on the morphological classification of the foot, with limited exploration of foot function during common physical activities. Therefore, systematically extracting and classifying plantar pressure characteristics can not only improve the accuracy of foot health assessments, but also simplify the framework of gait analysis. This is particularly important when dealing with complex and diverse plantar pressure data, laying the foundation for more detailed studies of foot function difference in healthy populations.

To investigate the diversity of plantar pressure distribution during walking in the general population, this study first extracts the main features of plantar pressure during gait using principal component analysis. The K-means optimization algorithm is employed to classify gait plantar pressure patterns. On this basis, by minimizing intra-class variance and maximizing inter-class variance, a more intuitive linear discriminant function is constructed using the correlation coefficient method and Fisher linear discriminant analysis, with redundant indicators eliminated. The optimal classification features are extracted to provide a research foundation and theoretical basis for refining foot function studies and exploring the diversity of plantar pressure distribution.

Methods

Participants

Young male participants were recruited with the following inclusion criteria: participants had to be in optimal physical condition, with no history of lower limb surgery within the past three years, and exhibit a normal foot type (arch index between 0.21 and 0.26¹⁸). Individuals experiencing foot pain or who had engaged in vigorous physical activity within the preceding 48 h were excluded. A final sample of 292 male participants was obtained (age: 33.76 ± 4.59 years, height: 176.21 ± 5.33 cm, weight: 68.44 ± 5.23 kg). Prior to testing, participants were informed of the study details and signed an informed consent form. This study was approved by the Ethics Committee of Hebei Normal University (No. 2022LLSC026), and all procedures of this experiment were in accordance with the ethical standards laid out in the 1964 Declaration of Helsinki.

Data collection

After familiarizing with the protocol and warming up, participants had their plantar pressure data collected during walking using a high-frequency plantar pressure plate (RSscan International, Belgium, sampling frequency 126 Hz, minimum resolution 0.25 N, measuring range 1–60 N/cm², plate length 2 m with 1.5 m extended runways at start and end). Three trials of plantar pressure data were recorded for each participant at their self-selected walking speed. Participants were required to walk with their gaze directed forward and maintain a natural gait (starting with the heel strike and ending with the propulsion of the forefoot) during data collection. Refer to previous studies on walking speed^{19,20}, Data from trials with walking speeds between 1.1 and 1.3 m/s and containing complete bilateral footprints were included for processing.

Data processing

Data preprocessing

Based on the collected plantar pressure data, the foot regions were adjusted in the instrument's built-in plantar pressure analysis software (see Fig. 1). The peak pressures and impulses of different regions were recorded for both feet of each participant. The foot with higher total impulse across 10 regions during the stance phase was identified as the primary force-producing foot. A total of 20 parameters including peak pressures and impulses from the 10 regions of the primary force-producing foot during stance were compiled for all participants. Z-scores were calculated to exclude outliers (As shown in Eq. 1), with any plantar pressure parameter having a Z-score greater than 3 considered an outlier^{21,22}. Data without outliers were retained for further analysis, resulting in a final sample size of 243 participants.

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

where X is the value of the original data; μ is the data mean, σ is the data standard deviation.

Principal component extraction

The preprocessed 20 parameters underwent the Kaiser–Meyer–Olkin (KMO) and Bartlett's tests of sphericity to assess correlations and multicollinearity among the parameters. Principal components with eigenvalues greater than 1 were extracted for dimensionality reduction. To enhance interpretability, the principal components were rotated using the varimax method, forcing variable loadings toward extreme values (near 0 or ± 1) to reduce cross-loadings and clarify component-variable associations. The scores of the extracted principal components were retained.

Plantar pressure cluster number determination

The K-means++ clustering algorithm was applied to the scores of the extracted principal components. The number of clusters ranged from 2 to n (where n is the number of extracted principal components). Compared to

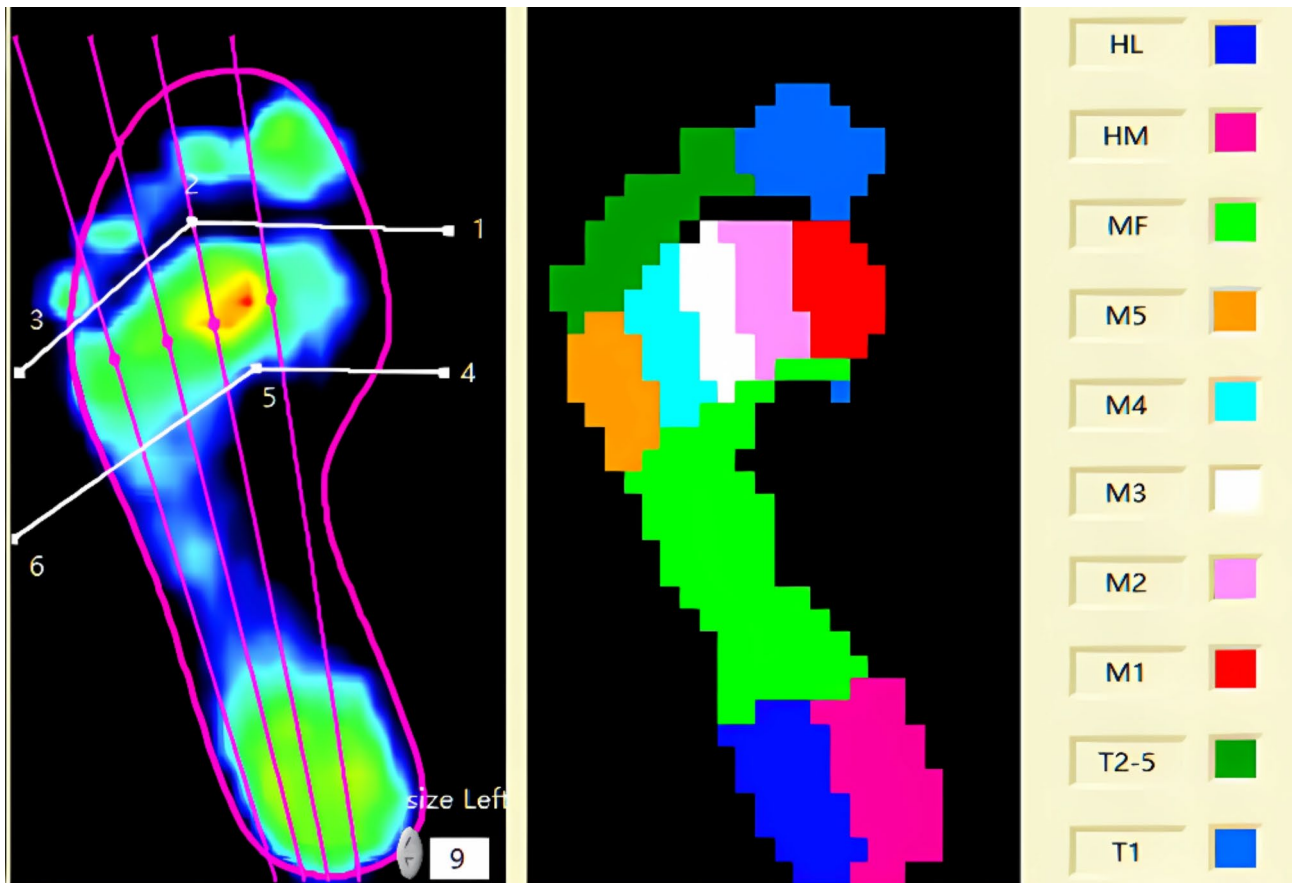


Fig. 1. Plantar pressure regions. (Note: HL = Lateral Heel, HM = Medial Heel, MF = Midfoot, M 1- M 5 correspond to the 1st to 5th Metatarsal regions, T 2-5 = 2nd to 5th Toes, T 1 = Hallux region).

traditional K-means, K-means++ optimizes the selection of initial cluster centers, thereby improving clustering quality²³. In this study, the maximum number of iterations was set to 10,000. The silhouette coefficient was introduced to optimize the iteration quality (Eq. 2). The silhouette coefficient ranges from -1 to 1, where values closer to 1 indicate better clustering and negative values suggest misclassification^{24,25}. The number of misclassified samples (silhouette coefficient < 0) and the average silhouette coefficient (Eq. 3) in each cluster were calculated to optimize each iteration.

The Elbow Method is a technique to determine the optimal number of clusters by finding the “elbow point” of the sum of squared errors (SSE) curve. SSE represents the sum of squared distances between each sample and its nearest cluster centroid (Eq. 4). As the number of clusters increases, SSE gradually decreases, but after the “elbow point,” the reduction in SSE becomes insignificant with additional clusters. This elbow point indicates the optimal number of clusters²⁶. In this study, the optimal number of clusters was determined by combining the Elbow Method and silhouette coefficients. The process of the optimized clustering algorithm is shown in Fig. 2.

$$s_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{2}$$

s_i is the silhouette coefficient of sample i , b_i is the distance from i to its cluster centroid, a_i is the distance from i to the nearest neighboring cluster centroid

$$s_k = \sum_{i=1}^n s_i \tag{3}$$

Where n is the number of samples, k is the number of clusters. s_k is the average silhouette coefficient for that number of clusters.

$$SSE = \sum_{k=1}^n \sum_{i \in C_k} |i - m_k|^2 \tag{4}$$

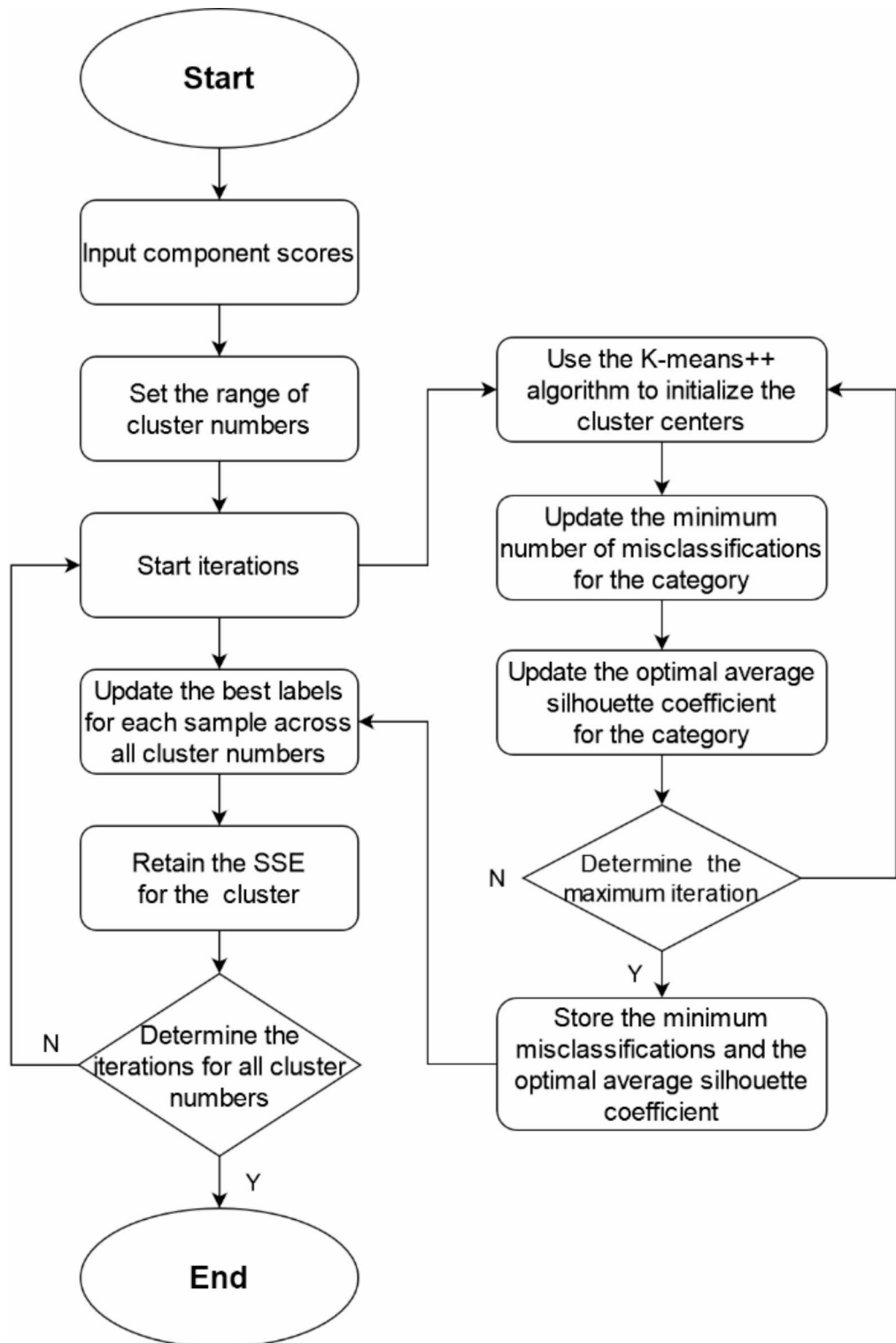


Fig. 2. Diagram of the optimized clustering algorithm.

where n is the number of clusters, C_k is the k -th cluster, i is a sample point in cluster C_k , m_k is the centroid of the k th cluster.

Plantar pressure feature selection

In this study, feature indexes are selected by combining correlation coefficients and the Fisher criterion^{27,28}. For plantar pressure indices, Pearson's correlation analysis is used to categorize the similarity of indicators, as shown in Eq. (5). If $r \geq 0.60$, it was considered that there was redundancy between the indicators, and the highly correlated indices are grouped accordingly.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

where \bar{x} and \bar{y} represent the mean values of the samples within the two feature sets, and the value of r ranges from $[-1, 1]$.

For indicators with high correlation, the Fisher criterion is used for feature selection in conjunction with the classification results mentioned above. The Fisher criterion extracts the optimal feature indicators for classification by calculating the ratio of between-class variance to within-class variance, achieving the best class discrimination^{27,29}. The calculation method for the Fisher criterion is as follows:

If n samples belong to c categories, the calculation of the between-class distance f_a is as shown in Eq. (6). The calculation of the between-class distance f_b for different indicators is as shown in Eq. (7). Finally, the Fisher value F for each indicator is obtained, as shown in Eq. (8).

$$f_a = \frac{1}{n} \sum_{i=1}^c \sum_{x \in n_i} (x_i^{(k)} - m_i^{(k)})^2 \quad (6)$$

where $x_i^{(k)}$ denotes the value of the k -th feature for the i -th sample, and $m_i^{(k)}$ is the mean value of the k -th feature for the i -th sample.

$$f_b = \sum_{i=1}^c \frac{n_i}{n} (m_i^{(k)} - m^{(k)})^2 \quad (7)$$

where n_i is the number of samples in the i -th class, n is the total number of samples, and $m^{(k)}$ is the mean value of the k -th feature across all samples.

$$F = \frac{f_b}{f_a} \quad (8)$$

Higher F values indicate better separability of features and higher classification performance for the system. Among highly similar indicators, the one with the highest F value, representing the optimal classification performance, is extracted while the remaining indicators are excluded. Under the premise of relative independence among indicators, this completes the extraction of optimal plantar pressure feature indicators.

Construction of plantar pressure classification discriminant

The principle of linear discriminant analysis is based on multivariate linear regression and discriminant analysis. In this study, stepwise selection was used to exclude redundant independent variables and construct an optimal predictive model^{29,30}. At each step, the contribution of an indicator to classification performance was evaluated based on the change in its F value. If $p < 0.05$, indicating a significant F statistic, the indicator was included in the discriminant function.

Based on the plantar pressure classification results, the extracted plantar pressure feature indicators were used as independent variables for linear discriminant analysis to obtain the classification function coefficients and constants for each included indicator. Discriminant functions were constructed for each class. The class corresponding to the maximum value calculated from the discriminant functions determined the class membership of a given sample. This process established the discriminant functions for classifying walking plantar pressure and identified the final set of feature indicators.

Statistical methods

The optimized K-means++ clustering algorithm was implemented using a custom script in MATLAB R2022b. Principal component analysis and linear discriminant analysis were performed using SPSS 25.0. For different plantar pressure classification results, independent samples t -tests or Wilcoxon rank-sum tests were used for between-group comparisons if there were two classes, according to the data distribution characteristics. If there were more than two classes, one-way analysis of variance or Kruskal–Wallis H tests were used for between-group comparisons. The significance level was set at 0.05.

Principal components	Initial eigenvalues			Sums of squared loadings			Rotation sums of squared loadings		
	Eigenvalues	Variance%	Cumulative %	Eigenvalues	Variance%	Cumulative %	Eigenvalues	Variance%	Cumulative %
1	5.34	26.71	26.71	5.34	26.71	26.71	3.43	17.15	17.15
2	4.22	21.10	47.80	4.22	21.10	47.80	3.24	16.19	33.33
3	1.96	9.79	57.59	1.96	9.79	57.59	2.94	14.71	48.04
4	1.88	9.39	66.99	1.88	9.39	66.99	2.58	12.89	60.94
5	1.53	7.64	74.63	1.53	7.64	74.63	2.22	11.12	72.06
6	1.27	6.37	81.00	1.27	6.37	81.00	1.79	8.94	81.00
7	0.97	4.86	85.86						
8	0.67	3.34	89.19						
9	0.61	3.07	92.26						
10	0.35	1.77	94.04						
11	0.27	1.37	95.40						
12	0.23	1.16	96.56						
13	0.17	0.83	97.39						
14	0.13	0.66	98.05						
15	0.11	0.54	98.59						
16	0.09	0.46	99.05						
17	0.08	0.39	99.44						
18	0.07	0.37	99.80						
19	0.02	0.11	99.92						
20	0.02	0.09	100.00						

Table 1. Principal component extraction results (n = 292).

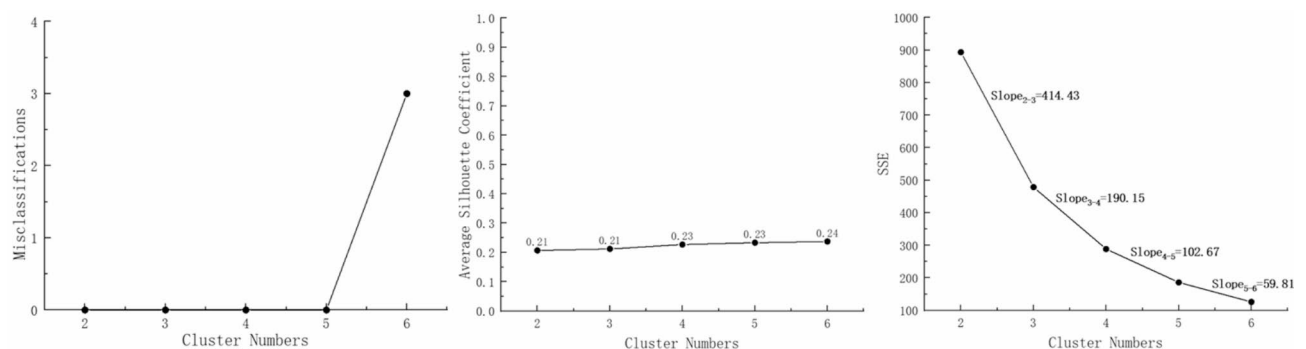


Fig. 3. Results of optimized clustering.

Results

Principal component extraction results

The KMO value for the 20 plantar pressure parameters was 0.621, and Bartlett's test of sphericity yielded an approximate chi-square of 4950.62 ($P < 0.001$), indicating correlations among the plantar pressure parameters during the stance phase and allowing for principal component extraction.

As shown in Table 1, among the 20 plantar pressure parameters, 6 principal components had eigenvalues greater than 1, accounting for 81.00% of the cumulative variance. Conduct further clustering analysis using the principal component scores derived from the 6 principal components above.

Plantar pressure clustering results

The optimized clustering algorithm was executed in MATLAB using the 243×6 matrix of scores from the 6 principal components for each sample. The clustering results are shown in Fig. 3.

Figure 3 shows that there were no misclassifications (samples with a silhouette coefficient < 0) for 2 to 5 clusters, while 6 clusters resulted in 3 misclassifications. The average silhouette coefficients ranged from 0.20 to 0.25 for 2 to 6 clusters, showing no significant difference. However, the SSE results reveal that the steepest drop (414.43) occurred between 2 and 3 clusters, indicating a slower rate of SSE reduction beyond 3 clusters as the number of clusters increased. Therefore, the optimal number of clusters was determined to be 3.

	Toe 1 impulse	Meta 1 impulse	Meta 2 impulse	Meta 5 impulse	Midfoot impulse	Heel lateral impulse	Toe 2–5 pressure
Toe 1 impulse	1.00	0.42	0.19	−0.23	−0.16	0.08	0.38
Meta 1 impulse		1.00	0.17	−0.26	−0.09	0.08	0.11
Meta 2 impulse			1.00	0.07	−0.03	0.15	0.14
Meta 5 impulse				1.00	0.37	0.16	0.00
Midfoot impulse					1.00	0.13	−0.07
Heel lateral impulse						1.00	0.23
Toe 2–5 pressure							1.00

Table 2. Correlation coefficient matrix of screened indicators (n = 292). Significant values are in bold.

Indicators	Class1	Class2	Class3
Toe 1 impulse	0.145	0.082	0.09
Meta 1 impulse	0.27	0.12	0.092
Meta 5 impulse	0.115	0.047	0.088
Midfoot impulse	0.091	0.348	0.066
Constant	− 13.789	− 12.973	− 4.422

Table 3. Linear discriminant function classification coefficients (n = 292).

		Label	Predicted group membership information			Accuracy (%)
			1	2	3	
Original data	Count	1	55	2	7	85.9
		2	3	42	6	82.4
		3	3	4	121	94.5
Cross-verification	Count	1	53	3	8	82.8
		2	3	42	6	82.4
		3	3	5	120	93.8

Table 4. Results of linear discriminant classification (n = 292).

Feature selection results

By integrating correlation coefficients and the Fisher criterion based on the clustering results of plantar pressure, 7 mutually independent indicators with strong classification performance were selected: Toe 1 Impulse, Meta 1 Impulse, Meta 2 Impulse, Meta 5 Impulse, Midfoot Impulse, Heel Lateral Pressure, and Toe 2–5 Pressure. The correlation coefficient matrix among these seven indicators is presented in Table 2, indicating that the indicators are relatively independent of each other.

Linear discriminant function results

Stepwise linear discriminant analysis was conducted on the 7 indicators from “Feature selection results” section of this paper. Finally, 4 indicators were retained: Toe 1 impulse, Meta 1 impulse, Meta 5 impulse, and Midfoot impulse. The classification function coefficients are shown in Table 3.

As showed in Table 3, the linear discriminant functions for the three classifications of plantar pressure are as follows:

$$\text{Class1} = 0.145 \times \text{Toe 1 Impulse} + 0.27 \times \text{Meta 1 Impulse} + 0.115 \times \text{Meta5 Impulse} + 0.091 \times \text{Midfoot Impulse} - 13.789 \quad (9)$$

$$\text{Class2} = 0.0082 \times \text{Toe 1 Impulse} + 0.12 \times \text{Meta 1 Impulse} + 0.047 \times \text{Meta5 Impulse} + 0.348 \times \text{Midfoot Impulse} - 12.973 \quad (10)$$

$$\text{Class3} = 0.09 \times \text{Toe 1 Impulse} + 0.092 \times \text{Meta 1 Impulse} + 0.088 \times \text{Meta5 Impulse} + 0.066 \times \text{Midfoot Impulse} - 4.422 \quad (11)$$

The classification of plantar pressure is judged according to the discriminant scores of Eq. (9)–(11), and the accuracy of the discriminant is shown in Table 4. As shown in Table 4, the average evaluation accuracy of the original data of the three classifications is 89.70%, and the average accuracy of the three classifications after cross-verification is 88.5%, indicating a high linear discriminant accuracy.

Plantar pressure classification characteristics

From Fig. 4, it can be observed that Class 1 had higher Toe1 impulse and Meta1 impulse compared to Class 2 and 3 ($P < 0.05$). Class 2 exhibited higher Midfoot impulse than Class 1 and 3 ($P < 0.05$), and higher Meta5 impulse than Class 1 ($P < 0.05$).

Based on the characteristics of the indices for each classification, it can be determined that individuals in Class 1 primarily bear weight on the thumb and first metatarsal during walking, hence named the “Thumb Extension Type”; those in Class 2 mainly exert force on the midfoot and fifth metatarsal, and are termed the “Midfoot-Lateral Forefoot Push-off Type”; Class 3 individuals do not exhibit distinct features compared to the first two classes and are thus named the “Normal Type”. The plantar pressure characteristics for the three types of walking are depicted in Fig. 5.

Discussion

Through multiple fusion algorithms, this study classified and extracted features from the peak pressures and impulses of 10 foot regions during the stance phase of walking in 243 participants. By combining principal component analysis with a clustering algorithm optimized based on silhouette coefficients, walking plantar pressure characteristics were categorized into three classes. Subsequently, redundant indicators were excluded through correlation coefficient screening and Fisher’s criterion. Linear discriminant functions were constructed for the three classifications. The results revealed that Toe 1 impulse, Meta 1 impulse, Meta 5 impulse, and Midfoot impulse, effectively distinguished plantar pressure characteristics during walking. Based on these features, an analysis of plantar pressure differences during the stance phase of different gait patterns was conducted, providing methodological support and theoretical references for the study of plantar pressure distribution characteristics and dynamic foot function analysis during walking.

Analysis of optimized clustering and feature extraction algorithms

Plantar pressure is an important dynamic indicator for investigating the biomechanical characteristics of the lower limbs, and it holds significant reference value for the diagnosis of foot diseases and the analysis of lower limb force lines^{2,8,9}. During physical activity, the dynamic indicators of certain foot regions exhibit coordinated characteristics to perform foot functions¹, suggesting potential redundancy among the dynamic indicators of different foot regions in plantar pressure analysis. PCA identifies the principal directions of variation in the

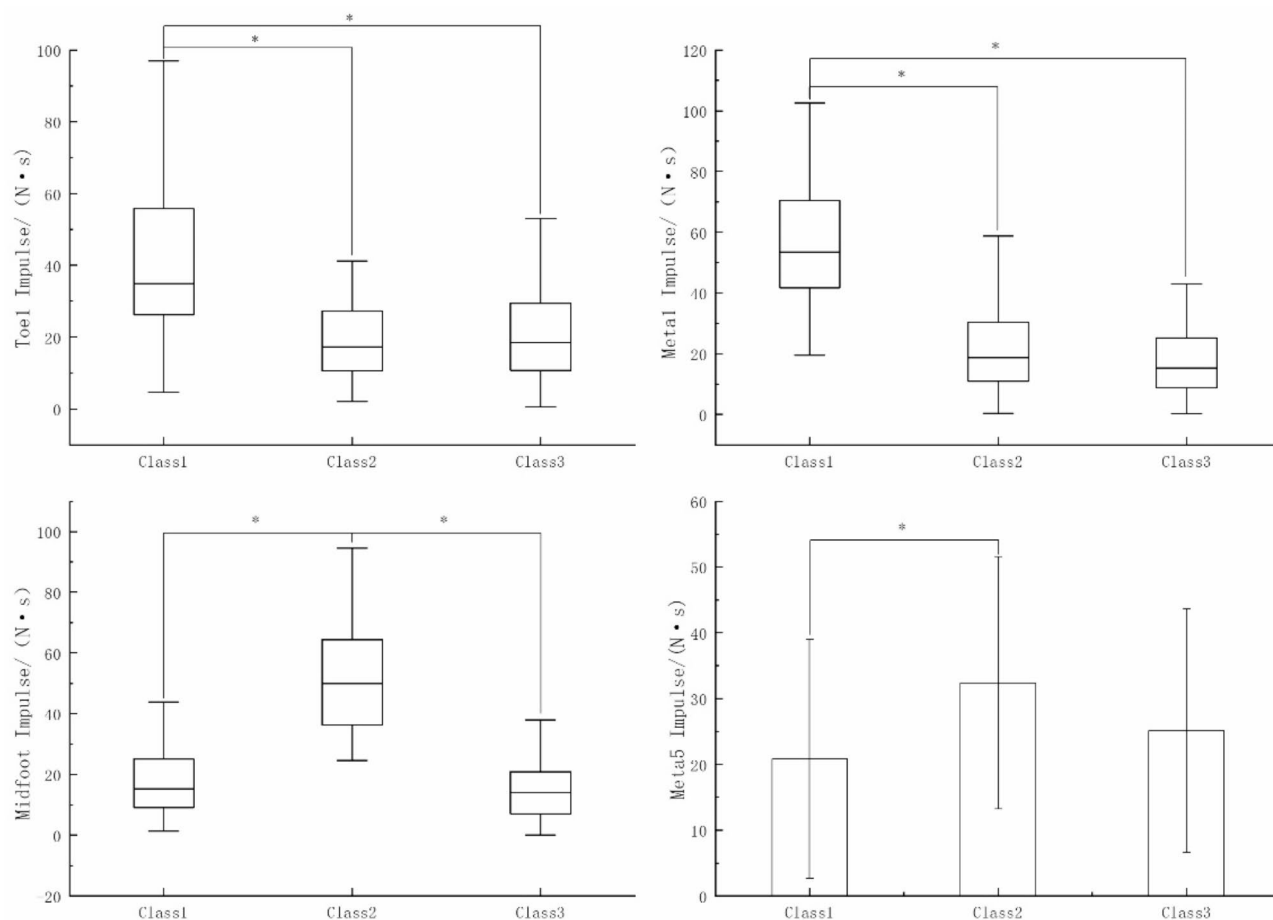


Fig. 4. Comparison results of each classification.

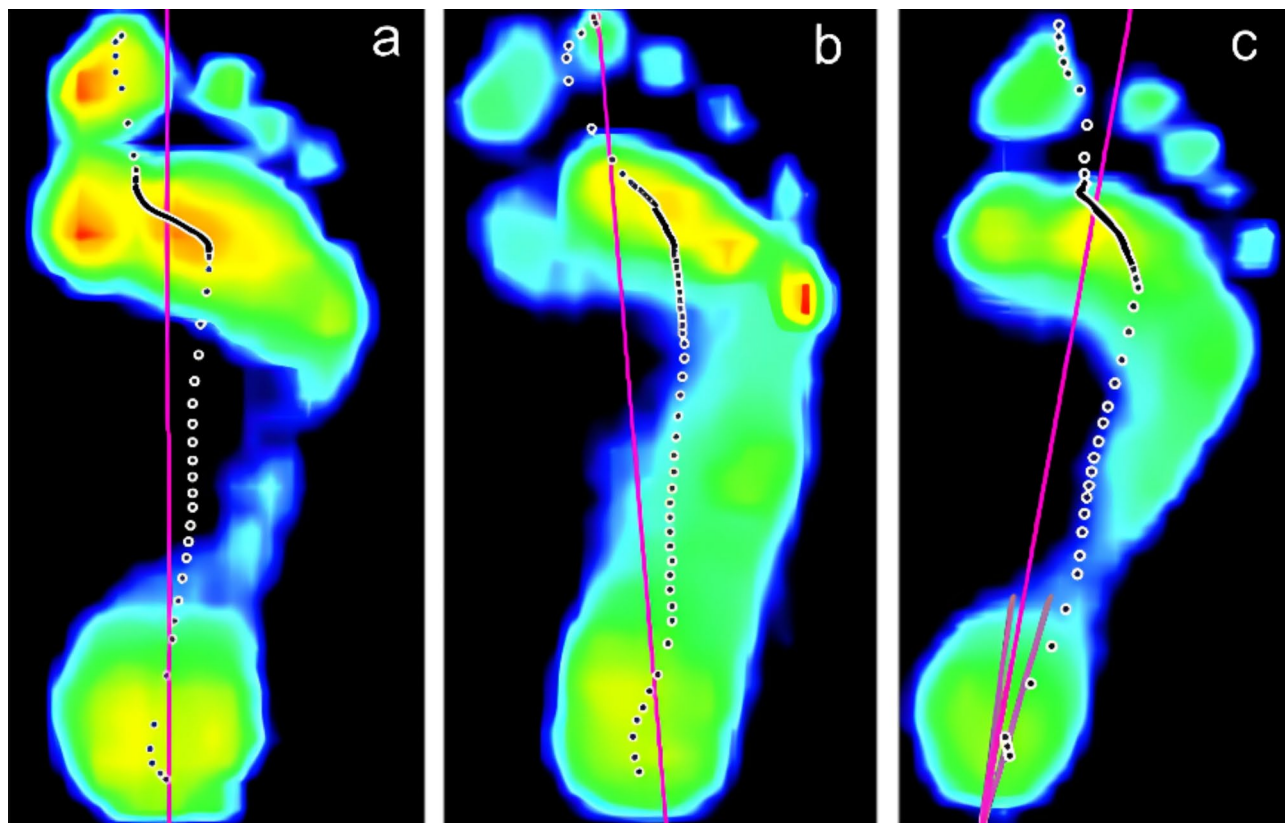


Fig. 5. Classifications of plantar pressure characteristics. (*Note: a* represents the thumb extension type, *b* represents the midfoot-lateral forefoot push-off type, and *c* represents the normal type).

data and projects high-dimensional data onto a lower-dimensional space, effectively reducing redundancy and multicollinearity among variables. This process simplifies the data structure while preserving most of the essential information. Therefore, when classifying plantar pressure, it is necessary to reduce the dimensionality of the complex dynamic information to extract effective principal components for analysis. In this study, principal component analysis was performed on 20 plantar pressure parameters, and the principal components were rotated using the maximum variance method to enhance their interpretability. Ultimately, 6 principal components accounting for 81% of the cumulative variance were extracted, preserving a significant amount of the original data information.

K-means is a common clustering analysis method, however, it is highly sensitive to the selection of initial cluster centers, which can lead the algorithm to converge to local optima³¹. To address this issue, K-means++ optimizes the selection of initial cluster centers in the K-means algorithm to improve clustering quality. Although K-means++ increases computational cost and complexity, it improves the distribution of initial cluster centers in the data space by selecting the point farthest from the current centers, helping to avoid convergence to local optima²³. In this study, based on the K-means++ algorithm, the silhouette coefficient was used to optimize the iterations^{24,25}. By minimizing the number of misclassified samples (those with negative silhouette coefficients) for different numbers of clusters and combining the average silhouette coefficient across samples for each number of clusters, clustering quality was improved. The elbow method was then applied to determine the optimal number of clusters²⁶, resulting in three classifications of plantar pressure characteristics during walking in the general population.

For the three classified feature sets, this study performed an initial feature screening by combining correlation coefficient selection and Fisher's criterion^{28,29}, to extract the indicators that yield the best classification performance from redundant features. For indicators with high Pearson correlation coefficients (threshold set at 0.60 in this study), the indicator with the maximum between-class difference and minimum within-class difference, as indicated by the Fisher value, was selected, while redundant indicators were excluded. This allowed the removal of redundant indicators while preserving complete information. Seven relatively independent pre-screened indicators were obtained: Toe 1 impulse, Meta 1 impulse, Meta 2 impulse, Meta 5 impulse, Midfoot impulse, Heel Lateral impulse, and Toe 2–5 pressure. However, although this method eliminated redundant indicators, it included indicators with minor classification impact to retain complete information. Therefore, this study employed stepwise linear discriminant analysis for the final screening, ultimately identifying four indicators: Toe 1 impulse, Meta 1 impulse, Meta 5 impulse, and Midfoot impulse. The three classifications based on linear discriminant analysis achieved an accuracy of 89.70% for the original data and 88.5% for cross-validation, which indicates that the multiple fusion algorithms used in this study maximized feature simplification while

maintaining optimal classification performance. To enhance classification performance, it is recommended to increase principal component extraction cumulative variance coverage, or alternatively, to implement nonlinear analytical methods for improved pattern recognition in complex datasets.

Analysis of different types of plantar pressure characteristics

By combining multiple fusion algorithms, this study screened out four indicators, Toe 1 impulse, Meta 1 impulse, Meta 5 impulse, and Midfoot impulse, from the initial 20 plantar pressure parameters. However, no peak pressure indicators demonstrated satisfactory classification performance among the selected indicators. Combined with previous research, peak pressures are typically associated with foot comfort or foot injuries, suggesting that peak pressures may exhibit strong individual differences and thus may not be suitable as classification references for normative data.

The 1st metatarsophalangeal joint and the hallux are crucial areas for energy absorption and propulsion during foot movement^{32–34}. The Toe 1 and Meta 1 impulse, as derived from this study, predominantly reflect the activity of the first metatarsophalangeal joint. During the push-off phase of walking, the load is transferred medially and the foot is propelled off the ground along the axes of the metatarsophalangeal joint and the hallux^{1,9}. It can be inferred that Class 1 exhibits a strong function of toe extension, hence the naming of the “Thumb Extension Type”. During the stance phase of walking, the foot arch compresses to bear the body’s weight, causing the soft tissues on the sole to lengthen and store elastic potential energy. In the push-off phase, the extension of the metatarsophalangeal joint further lengthens the plantar fascia. As the toes gradually lift off the ground, the elastic potential energy stored in the plantar fascia is released, which helps to improve the economy of movement during walking³⁵. Research has found that the stiffness of the metatarsophalangeal joint is also an important indicator for evaluating symptoms of plantar fasciitis³⁶. Fan et al.’s research also revealed that there are significant differences in the plantar impulse between young and elderly individuals. Their study believes that plantar impulse is an important indicator for assessing the function of the musculoskeletal system³⁷. This suggests that the “Thumb Extension Type” gait can make full use of the function of the metatarsophalangeal joint, thereby may exhibit better gait economy.

During the gait cycle, the center of pressure on the foot begins at the heel, transitions to the medial side of the forefoot, and is ultimately propelled off the ground by the hallux³⁸. Research on foot function indicates that during the stance and push-off phase of walking, the forefoot primarily bears load between the second and third metatarsals¹. This study observed that for Class 2, the gait characteristics show a significantly larger impulse in the lateral forefoot area compared to the other two classes, indicating a lateral bias in force distribution on the forefoot. Based on the above characteristics, Class 2 is named “Midfoot-Lateral Forefoot Push-off Type”. Regarding the arch of the foot, it descends to cushion the impact during weight-bearing and works in conjunction with the transverse arch to act as a rigid lever during extension^{35,39}. The findings suggest that individuals with the “Midfoot-Lateral Forefoot Push-off Type” gait exhibit a larger impulse in the midfoot area, which may be due to lower arch stiffness in this population, leading to poorer adaptability to weight-bearing and thus a larger impulse in the midfoot region. Studies have also found that patients with knee osteoarthritis (KOA) and those with flat feet exhibit significantly higher plantar impulse and peak pressure than individuals with normal feet, which suggests that abnormal foot posture and gait patterns may significantly influence plantar pressure distribution^{40–42}. However, since all subjects in this study had normal foot types, whether this gait pattern, which shows weaker arch function, would lead to foot injuries remains to be further confirmed with prospective studies.

This study has achieved classification and feature extraction of plantar pressure based on the dominant foot of healthy individuals, it is worth to note that the walking speed selected for this research (1.1–1.3 m/s) reflects the usual walking speed of young males²⁰. However, research indicates that optimal walking speed tends to decrease with age¹⁹, the plantar pressure during walking in the elderly population tends to shift towards the lateral side of the foot^{43,44}. Therefore, further studies are required to investigate plantar pressure distribution characteristics and functional laterality differences across different genders, ages, or walking speeds by incorporating non-dominant plantar pressure analysis. Additionally, to reduce computational resources for plantar pressure classification and provide a more intuitive linear discriminant model, this study employs Fisher’s linear discriminant analysis. For large sample datasets, it is recommended to use Support Vector Machines (SVM) or deep neural networks (DNN), combined with interpretable models such as layer-wise relevance propagation (LRP)^{45,46} to extract features from highly correlated redundant indicators that contribute optimally to classification and capture more complex nonlinear relationships within the data. In conclusion, it is recommended that future studies could expand the sample size and combine different foot conditions and lower limb injuries for prospective research to explore the relationship between different gait characteristics and sports injuries.

Conclusion

The plantar pressure among healthy individuals during walking can be categorized into three types: Thumb Extension Type, Midfoot-Lateral Forefoot Push-off Type, and Normal Type. Among these, the impulse in the areas around the 1st metatarsophalangeal joint, the 5th metatarsal, and the midfoot region exhibit satisfactory classification performance. It is recommended that future research should combine the results of this study and, through prospective studies, further analyze the relationship between different gait characteristics and sports injuries.

Data availability

The data supporting the findings of this study are available from the author Xiaotian Bai. However, due to the personal gait information of the subjects involved, the data from this study is not made publicly available.

Received: 15 March 2024; Accepted: 28 March 2025

Published online: 17 April 2025

References

- Bai, X., Huo, H. & Liu, J. Analysis of mechanical characteristics of walking and running foot functional units based on non-negative matrix factorization. *Front. Bioeng. Biotechnol.* **11**, 1201421 (2023).
- Hillstrom, H. J. et al. Foot type biomechanics part 1: Structure and function of the asymptomatic foot. *Gait Post.* **37**, 445–451 (2013).
- Hinz, P. et al. Analysis of pressure distribution below the metatarsals with different insoles in combat boots of the German army for prevention of March fractures. *Gait Post.* **27**, 535–538 (2008).
- Schepers, T., Van der Stoep, A., Van der Avert, H., Van Lieshout, E. M. & Patka, P. Plantar pressure analysis after percutaneous repair of displaced intra-articular calcaneal fractures. *Foot. Ankle Int.* **29**, 128–135 (2008).
- De Cock, A., De Clercq, D., Willems, T. & Witvrouw, E. Temporal characteristics of foot roll-over during barefoot jogging: Reference data for young adults. *Gait Post.* **21**, 432–439 (2005).
- Sullivan, J., Burns, J., Adams, R., Pappas, E. & Crosbie, J. Plantar heel pain and foot loading during normal walking. *Gait Post.* **41**, 688–693 (2015).
- Ledoux, W. R. & Hillstrom, H. J. The distributed plantar vertical force of neutrally aligned and pes planus feet. *Gait Post.* **15**, 1–9 (2002).
- Fernandez-Seguín, L. M. et al. Comparison of plantar pressures and contact area between normal and cavus foot. *Gait Post.* **39**, 789–792 (2014).
- Hofmann, U. K. et al. Transfer of plantar pressure from the medial to the central forefoot in patients with hallux valgus. *Bmc Musculoskelet. Disord.* **20**, 1–8 (2019).
- Koller, U. et al. Plantar pressure characteristics in hallux valgus feet. *J. Orthop. Res.* **32**, 1688–1693 (2014).
- Chen, H., Nigg, B. M. & Dekoning, J. Relationship between plantar pressure distribution under the foot and insole comfort. *Clin. Biomech.* **9**, 335–341 (1994).
- Lee, Y. H. & Hong, W. H. Effects of shoe inserts and heel height on foot pressure, impact force, and perceived comfort during walking. *Appl. Ergon.* **36**, 355–362 (2005).
- Arin-Bal, G., Livanelioglu, A., Leardini, A. & Belvedere, C. Correlations between plantar pressure and postural balance in healthy subjects and their comparison according to gender and limb dominance: A cross-sectional descriptive study. *Gait Post.* **108**, 124–131 (2024).
- Buldt, A. K., Allan, J. J., Landorf, K. B. & Menz, H. B. The relationship between foot posture and plantar pressure during walking in adults: A systematic review. *Gait Post.* **62**, 56–67 (2018).
- De Cock, A., Willems, T., Witvrouw, E., Vanrenterghem, J. & De Clercq, D. A functional foot type classification with cluster analysis based on plantar pressure distribution during jogging. *Gait Post.* **23**, 339–347 (2006).
- Lee, Y. C. & Wang, M. J. Taiwanese adult foot shape classification using 3D scanning data. *Ergonomics* **58**, 513–523 (2015).
- Xu, M., Li, J. X., Hong, Y. & Wang, L. Foot type classification for Chinese children and adolescents. *Kinesiology* **5**, 127–132 (2019).
- Cavanagh, P. R. & Rodgers, M. M. The arch index: A useful measure from footprints. *J. Biomech.* **20**, 547–551 (1987).
- Bohannon, R. W. Comfortable and maximum walking speed of adults aged 20–79 years: Reference values and determinants. *Age Ageing* **26**, 15–19 (1997).
- Murtagh, E. M., Mair, J. L., Aguiar, E., Tudor-Locke, C. & Murphy, M. H. Outdoor walking speeds of apparently healthy adults: A systematic review and metaanalysis. *Sports Med.* **8**, 1–31 (2020).
- Berger, A. & Kiefer, M. Comparison of different response time outlier exclusion methods: A simulation study. *Front. Psychol.* **12**, 675558 (2021).
- Shiffler, R. Maximum Z-scores and outliers. *Am. Stat.* **42**, 79–80 (1988).
- Arthur, D. & Vassilvitskii, S. K-Means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7–9, 2007* (2007).
- Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M. & Perona, I. An extensive comparative study of cluster validity indices. *Pattern Recognit.* **46**, 243–256 (2013).
- Nainggolan, R., Perangin-Angin, R., Simarmata, E. & Tarigan, A. F. *Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (Sse) Optimized by Using the Elbow Method* 12015–12016 (IOP Publishing Ltd, 2019).
- Wang, Y., Ji, J., Liang, P. & Wong, K. Feature selection of fMRI data based on normalized mutual information and fisher discriminant ratio. *J. X-Ray Sci. Technol.* **24**, 467–475 (2016).
- Adler, J. & Parmryd, I. Quantifying colocalization by correlation: the pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytom. Part A* **77**, 733–742 (2010).
- Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).
- Carolyn, R., Virginia, F. & Peter, L. Bias in error rate estimates in discriminant analysis when stepwise variable selection is employed. *Commun. Stat. Simul. Comput.* **20**, 1–22 (1991).
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* **622**, 178–210 (2023).
- Oleson, M., Adler, D. & Goldsmith, P. A comparison of forefoot stiffness in running and running shoe bending stiffness. *J. Biomech.* **38**, 1886–1894 (2005).
- Stefanyshyn, D. J. & Nigg, B. M. Mechanical energy contribution of the metatarsophalangeal joint to running and sprinting. *J. Biomech.* **30**, 1081–1085 (1997).
- Stefanyshyn, D. J. & Nigg, B. M. Contribution of the lower extremity joints to mechanical energy in running vertical jumps and running long jumps. *J. Sports. Sci.* **16**, 177–186 (1998).
- Kelly, L. A., Cresswell, A. G., Racinais, S., Whiteley, R. & Lichtwark, G. Intrinsic foot muscles have the capacity to control deformation of the longitudinal arch. *J. R. Soc. Interface* **11**, 20131188 (2014).
- Kim, Y., Bhatia, D., Lee, Y., Ryu, Y. & Park, H. Development and clinical evaluation of a novel foot stretching robot that simultaneously stretches plantar fascia and achilles tendon for treatment of plantar fasciitis. *IEEE. Trans. Biomed. Eng.* **69**, 2628–2637 (2022).
- Fan, Y. et al. Screening method based on walking plantar impulse for detecting musculoskeletal senescence and injury. *PLoS ONE* **8**, e83839 (2013).
- Pataky, T. C. et al. Vector field statistics for objective center-of-pressure trajectory analysis during gait, with evidence of scalar sensitivity to small coordinate system rotations. *Gait Post.* **40**, 255–258 (2014).
- Lichtwark, G. A. & Kelly, L. A. Ahead of the curve in the evolution of human feet. *Nature* **579**, 31–32 (2020).
- Panyarachun, P., Anghthong, C., Jindasakchai, P., Rajbhandari, P. & Rungrattanawilai, N. Abnormal foot pressure in older adults with knee osteoarthritis: A systematic review. *Eur. Rev. Med. Pharmacol. Sci.* **26**, 6236–6241 (2022).
- Zhang, Z., Wang, L., Hu, K. & Liu, Y. Characteristics of plantar loads during walking in patients with knee osteoarthritis. *Med. Sci. Monitor.* **23**, 5714–5719 (2017).

42. Buldt, A. K. et al. Foot posture is associated with plantar pressure during gait : A comparison of normal, planus and cavus feet. *Gait Post.* **62**, 235–240 (2018).
43. Hessert, M. J. et al. Foot pressure distribution during walking in young and old adults. *Bmc Geriatr.* **5**, 8 (2005).
44. Gimunová, M., Zvona, M. & Mikeska, O. The effect of aging and gender on plantar pressure distribution during the gait in elderly. *Acta Bioeng. Biomech.* **20**, 139–144 (2018).
45. Bach, S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**, e0130140 (2015).
46. Lopuschkin, S., Binder, A., Müller, K. & Samek, W. Understanding and comparing deep neural networks for age and gender classification. *IEEE* (2017).

Author contributions

XB and HH were in charge of designing the experiment. YS and ZT completed the data processing for this research. XB wrote the main manuscript. XH and JL revised the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-96440-6>.

Correspondence and requests for materials should be addressed to J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025