electronics

MDPI

Article

Research on Plantar Signal Measurement and Foot Arch Classification

Jinyu Zhu [1], Baoqing Nie [1,2,*] and Chuanhao Yu [1]

[1] School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China; 2315401037@stu.suda.edu.cn (J.Z.); 2228410116@stu.suda.edu.cn (C.Y.)
[2] Technology Research Institute, Suzhou LZY Technology Company Ltd., Suzhou 215153, China
* Correspondence: qixinbq@suda.edu.cn

Abstract

The foot arch functions as a dynamic biomechanical system, maintained by the integrated actions of bones, ligaments, and muscles. A large body of clinical evidence indicates that, in addition to congenital foot deformities, acquired variations in the foot arch caused by factors such as poor gait, aging, weight, or injury can significantly affect quality of life. Early intervention upon detection of foot arch changes can help mitigate progression and prevent further deterioration. Despite the availability of multimodal sensor-integrated running platforms for gait analysis, such systems are inherently bulky and not conducive to routine walking measurement. To overcome the above limitations, this study employed a flexible plantar pressure insole with an integrated accelerometer and a dedicated acquisition circuit to capture plantar pressure and acceleration data. This smart insole system acquires plantar data, performs feature extraction via time–domain and wavelet analysis, and then employs machine learning to classify the foot arch type as a normal foot, flatfoot, or high-arched. A Random Forest classifier was then established to categorize foot arch types based on the collected data, which integrates numerous decision trees through bootstrap aggregation and random feature selection, with final classification determined by majority voting. A total of 30 volunteers participated, including 11 with normal arches, 11 with flat feet, and 8 with high arches. Compared with support vector machine, K nearest neighbors, and decision tree, the Random Forest achieved the highest recognition accuracy of 92%. This system reveals the patterns of plantar pressure distribution and acceleration fluctuations during walking across three foot arches and demonstrates that wavelet entropy can effectively quantify the changes in signal complexity included in foot arch differences. Compared with laboratory force plates, this system features lower cost and a smaller form factor, making it suitable for real-time monitoring. This system can lay the technical foundation for personalized foot orthopedics and health monitoring.

Keywords: foot arch; plantar pressure; flexible hybrid electronics; inertial measurement unit

Check for updates

Academic Editor: Enzo Pasquale Scilingo

Received: 6 April 2026
Revised: 5 May 2026
Accepted: 8 May 2026
Published: 11 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

As living standards improve, foot health is receiving increasing attention. The high incidence of foot diseases has a substantial impact on individuals' daily lives and occupational activities. Apart from congenital malformations, adult acquired flatfoot deformity (AAFD) and acquired pes cavus are frequently encountered in clinical practice. The etiology of AAFD includes degenerative changes in joints and soft tissues, diabetes mellitus, neuropathies, and tumors [1]. An acquired high arch is often associated with neuromuscular

disorders [2]. However, current clinically common foot arch assessment methods, such as footprinting, X-ray imaging, and instrumental measurements [3,4], are mostly static or semi-static, failing to capture the dynamic characteristics of the foot arch during daily walking. While laboratory-grade force plates and gait analysis can provide high-precision dynamic data, they are expensive and require a fixed installation space, rendering them unsuitable for daily real-time monitoring. Thus, portable real-time monitoring of foot arch dynamics during natural walking remains a key biomechanical and clinical challenge. Consequently, continuous monitoring of foot conditions is indispensable for the early recognition of pathological signals and the initiation of appropriate therapeutic measures.

Nikolaidou, M.E. and Boudolos, K.D. (2006) [5] proposed a classification method based on static footprints to distinguish different foot types. Using a K-means cluster analysis, they constructed a combined classification model and demonstrated that the arch index was the most reliable parameter among those evaluated [5]. The method is simple and low-cost, but it relies on static features and cannot detect dynamic foot arch changes. Yu-Chi Lee and Wang (2015) used 3D scanning data to classify the foot shapes of Taiwanese adults [6]. In their study, they used foot breadth in % foot length, ball of foot length in % FL, and arch height in % FL as feature parameters. Then, three foot shape types for males and females can be classified. Throughout the experimental procedure, ethnic disparities in foot shape were also noted. While offering high 3D measurement accuracy, this static method is also limited in its ability to capture arch changes. De Cock et al. acquired plantar pressure data from 215 healthy adults during jogging and applied K-means cluster analysis to identify four distinct pressure loading patterns based on forefoot relative impulse [7]. This work provides a new framework for classifying foot function during dynamic activity. The study by De Cock et al. relied on a plantar pressure measuring plate. However, this equipment is costly and not feasible for continuous monitoring in daily life. Xu et al. [8] employed a typical experimental paradigm; participants were first screened using the arch index. Data on plantar pressure and kinematics during running were acquired using a pressure distribution treadmill and inertial sensors. They included fatigue incrementally, comparing parameters (region pressure, contact time) across initial, steady, and fatigued phases, and used the maximum Lyapunov exponent. This study provides important insights into the biomechanical responses of the flatfoot under dynamic fatigue conditions. In a multimodal sensor-based gait classification study, Jlassi et al. [9] systematically compared different IMU configurations, pressure insoles, and their combination for outdoor walking condition recognition. This work compared multiple sensing schemes, showing that pressure insoles with IMUs boosted performance, with insoles being cheaper and lighter than running platforms. However, this focuses on terrain recognition rather than foot arch type classification. The integration of multimodal sensors into a smart insole for foot arch classification remains underexplored.

Departing from the static analysis characteristic of earlier work, this study leverages insights from multimodal sensor fusion—integrating plantar pressure with additional sensor modalities—to promote research. This study centers on plantar pressure under dynamic conditions and incorporates an arch-mounted accelerometer for synchronized kinematic measurement. Through this work, the proposed system is expected to become a next-generation smart portable biomedical electronic device [10]. The model established in this paper can analyze which plantar features are most indicative of foot arch type. A flexible hybrid electronic system with a dedicated measurement circuit was developed to acquire plantar pressure and acceleration data. After preprocessing, feature points were extracted for the construction of machine learning models. The optimal model was selected from multiple candidate algorithms to perform classification. The overall research process is illustrated in Figure 1. This system identifies distinct patterns of plantar pressure

and acceleration during gait across three foot arches and confirms that wavelet entropy reliably quantifies complexity changes included by arch variations. Moreover, its lower cost and smaller form factor relative to laboratory force plates make it an attractive option for real-time monitoring. A total of 30 volunteers participated, including 11 with normal arches, 11 with flat feet, and 8 with high arches. The Random Forest achieved the highest recognition accuracy of 92%. This system performs well in monitoring plantar changes.

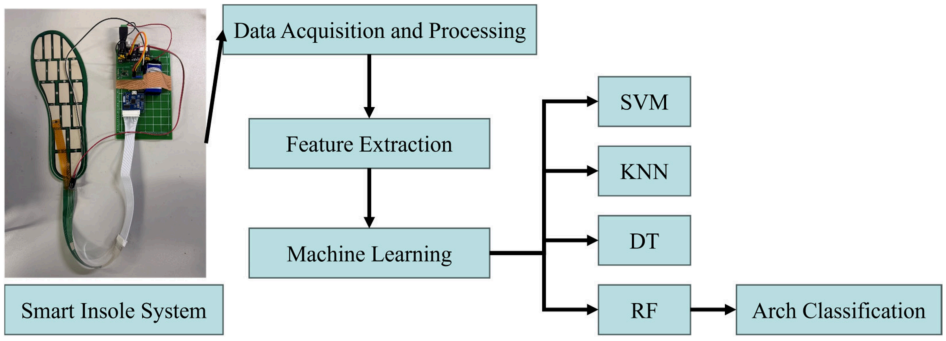


Figure 1. Overall pipeline from data acquisition to plantar arch classification.

2. Materials and Methods

2.1. Flexible Hybrid Electronics Fabrication and Signal Acquisition Circuits

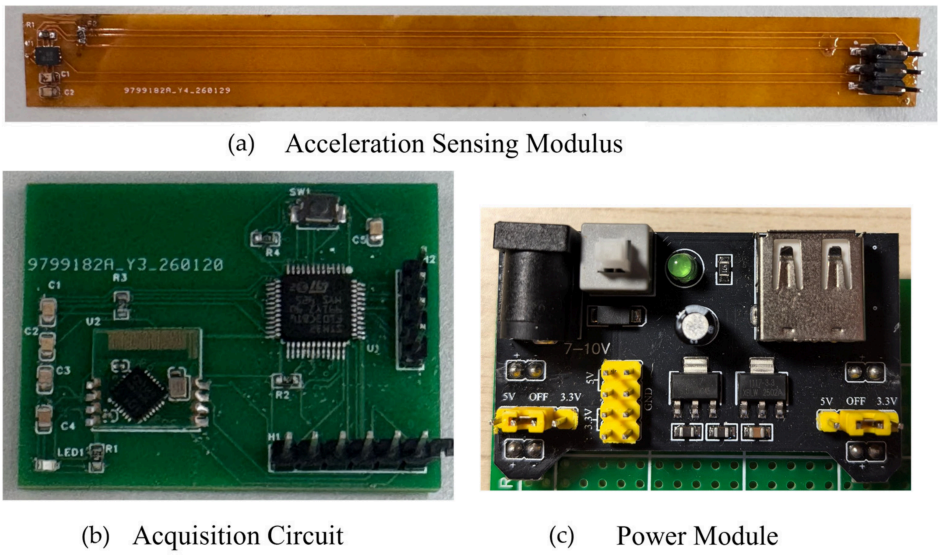
Differences in movement patterns among individuals with different foot types are clearly reflected in acceleration signals [11], which provide information not entirely reflected by pressure distribution alone. Unlike pressure data, which merely reflects the outcome of contact, acceleration sensors can capture the dynamic response of the arch—a structure that collapses and rebounds during running and jumping. Guided by the datasheet specifications, this paper develops a flexible FPC (Flexible Printed Circuit) for measuring acceleration data using the QMI8658A 6-axis inertial measurement unit (IMU) as the core component. The circuit was designed according to the measurement circuit in the data sheet. After fabricating the flexible printed circuit, the circuit was hand-soldered. It consists of a QMI8658A, two capacitors, and two resistors, with signals routed out via a ribbon cable. To capture the arch dynamics, the FPC was mounted on the arch area of the plantar pressure sensor. Finished flexible FPC and the relevant circuit are shown in Figure 2a–c. To ensure the robustness and reliability of the system under continuous operation, functional verification of the embedded acquisition architecture was performed. Although long-term (>8 h) continuous operation tests or thermal drift analysis were not performed in this study, we refer to recent methodological frameworks for the embedded acquisition architecture design and simulation-based validation, which demonstrated that circuit simulation, PCB integrity monitoring, and modular validation can greatly improve the reliability of biomedical monitoring devices [12].

The system hardware mainly includes the QMI8658A peripheral circuit, an STM32F103C8T6 MCU, a power module, and a CH9140 (BLE-SER-A-ANT) Bluetooth-to-serial module. The signal acquisition circuit reads the triaxial acceleration data and transmits it to a computer via Bluetooth serial communication.

2.2. Plantar Pressure Insole

This study employed a GT Electronic Technology European size 42 plantar pressure sensor and its standard acquisition module, which was then modified by adding an HC-06 Bluetooth module for wireless communication with a host PC. This array-distributed flexible thin-film pressure sensor is fabricated using precision printing technology. As shown in Figure 3a, the structure consists of top and bottom electrodes made of silver

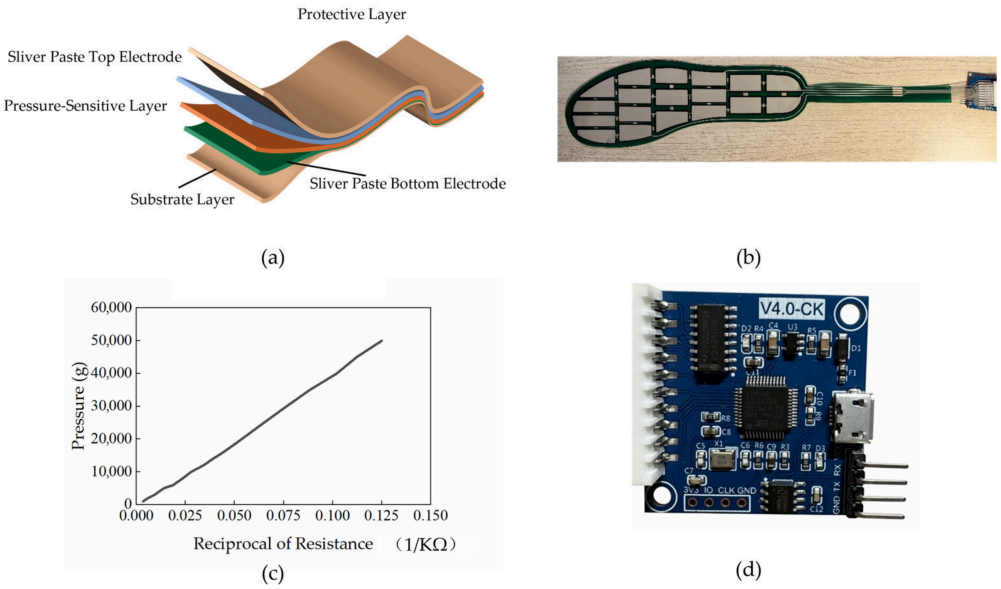
paste and a nano-piezoresistive pressure-sensitive layer. All layers are transferred onto a flexible film substrate and then dried and cured to form the insole presented in Figure 3b. When pressure is applied to the sensor, its resistance decreases as the pressure increases. We obtained the data relationship between resistance and pressure from the data sheet provided by the merchant. The reciprocal of resistance exhibits an approximately linear relationship with pressure, as shown in Figure 3c. The sensor features 18 independent sensing units in a distributed array. Each unit is individually designed to minimize mutual interference. Its single-point range is 0–70 k. Hysteresis is less than 6%. Drift is less than 8%. Repeatability is ±10%, and response time is less than 20 ms. This sensor is suitable for trend identification of plantar pressure distribution and foot arch type classification tasks. The circuit consists of a serial transceiver, sensor interface terminals, a Bluetooth transceiver module, an STM 32 chip, and multiple capacitors and resistors. Figure 3d is the pressure acquisition circuit.



(a) Acceleration Sensing Modulus

(b) Acquisition Circuit (c) Power Module

Figure 2. (**a**) The acceleration sensing modulus; (**b**) acquisition circuit of acceleration sensing modulus; (**c**) the power module of the circuit.



(a) (b)

(c) (d)

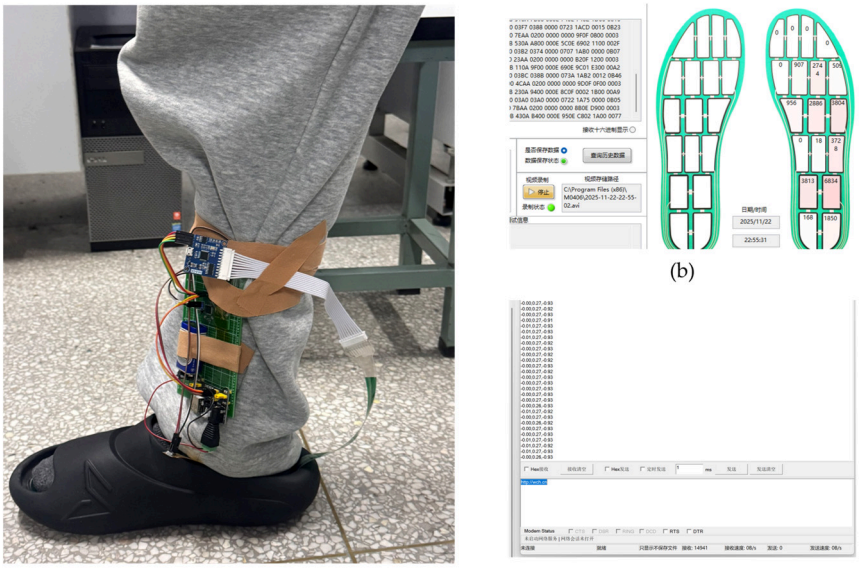
Figure 3. (**a**) The internal structure of insole; (**b**) physical insole; (**c**) sensor pressure characteristic curve; (**d**) acquisition circuit.

2.3. Experimental Procedure

The foot plantar data collection involved 30 male volunteers from the college, all with a shoe size 42 and aged 21–22. In this study, we only analyze the right foot. All participants understood and agreed to participate in this study and signed the informed consent form. Table 1 shows the basic physiological parameters of the participants. Before data collection, this study collected the volunteers' plantar information using footprinting and classified their arch types based on the footprints. During the experiment, each volunteer wore uniform dance shoes (to control the influence of sole material and thickness on plantar signals) and performed six normal walking steps on a horizontal surface in a quiet indoor environment to ensure no external interference. The actual measurement scene and host computer system are shown in Figure 4a–c.

Table 1. Basic physiological parameters of the volunteers.

Characteristic	Mean ± Standard Deviation
Height (cm)	173.8 ± 2.7
Weight (kg)	70.4 ± 8.54



(a) (b) (c)

Figure 4. (**a**) Actual measurement scenario; (**b**) pressure host system; (**c**) acceleration host system.

3. Results

3.1. Preprocessing

Under static conditions, plantar pressure serves as an indicator of the body's gravitational line and structural alignment of the skeleton. During dynamic activity, it provides a functional signal of the recruitment order of muscles and the quality of movement coordination. Numerous cases have demonstrated the significance of plantar pressure research in clinical medicine. For instance, Li Ke employed plantar pressure distribution analysis to forecast recurrence following high tibial osteotomy [13]. Yang et al. conducted a study to investigate the changes in plantar pressure after internal fixation in patients with supination–internal rotation type IV ankle fractures, providing scientific evidence for the development of a personal clinical rehab plan [14]. As the body's condition changes, the data on the foot arch change accordingly, which indicates its importance in research.

Accurate division of the gait cycle is a premise for subsequent calculations of gait parameters and biomechanical investigations [15]. Identifying gait subphases is vital for

fully assessing gait quality, with critical implications for diagnosing movement disorders. A complete gait cycle is generally defined as the period from the initial contact of one foot to the next heel strike of the same foot. Within the plantar pressure signal, the obvious pressure peak is generated at the moment of heel strike. This paper detects the peak in the total pressure of the right foot to identify the start of each gait cycle, then considers the signal between two start points as one cycle. This research used a single instrumented insole with 18 sensors to collect pressure data from the right foot at 20 Hz. The data were saved as a CSV file with units in grams–force. Since abnormal gait cycles may occur in the data, such anomalies must be removed for data cleaning. Then, based on pre-divided sensor points on the insole, segment the gait cycles and extract the relevant feature from the data using Python-3.12.6-amd64.

3.2. Time–Domain Features

The pressure data obtained in this study were visualized as contour maps. The map shows the pressure on the right foot. The values in Figure 5 are the pressure applied to the sensors, with the unit in grams. Then, we found that the hotspot locations during static standing differed among the participants with different arch types. To quantify this discrepancy, two categories of features were derived in the study. Figure 5a–c shows partial pressure maps and their footprint for the three arch types. It can be clearly seen that the footprint is consistent with the results reflected by the actual pressure map.

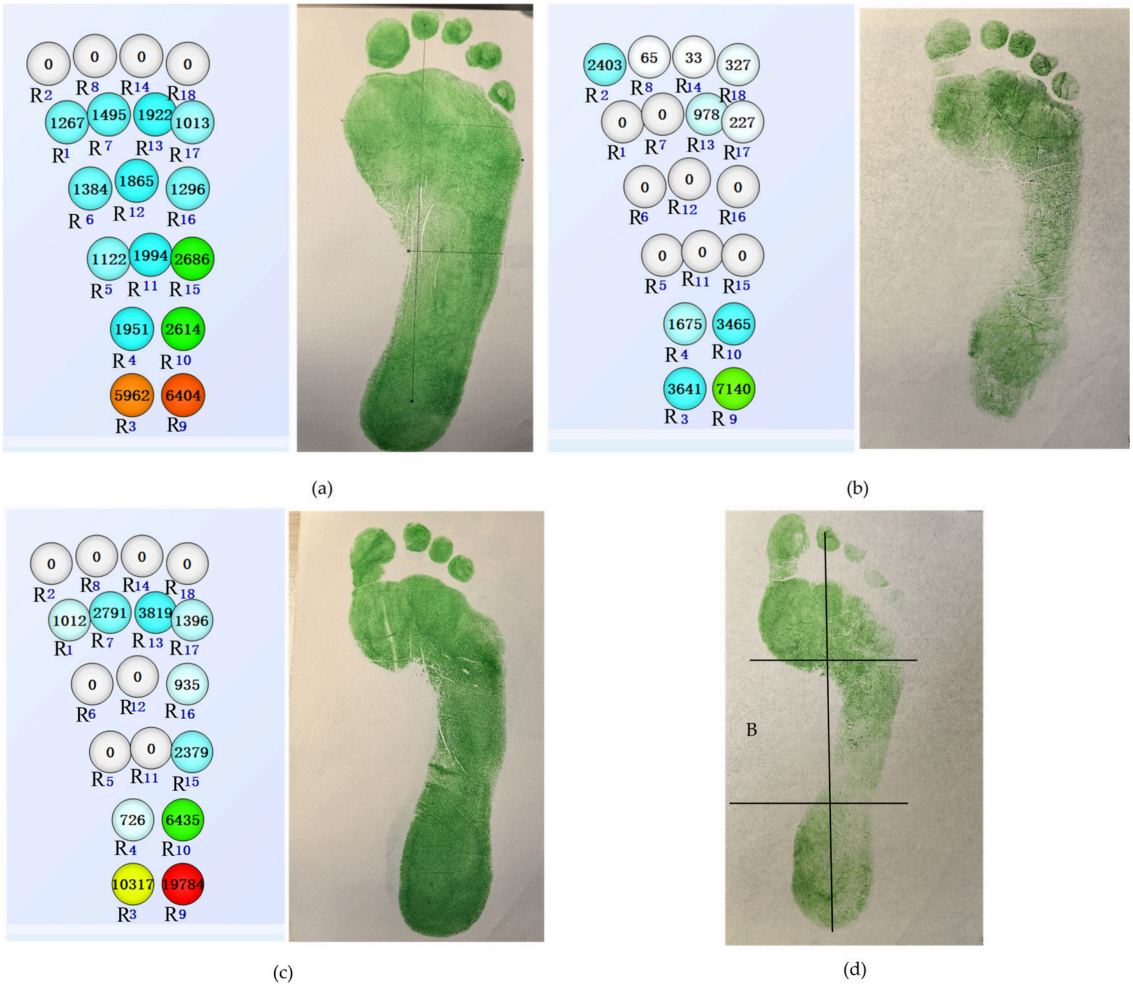


Figure 5. (**a**) Flatfoot right foot pressure distribution and its footprint; (**b**) High-arch right foot pressure distribution and its footprint; (**c**) Normal right foot pressure distribution during walking and its footprint; (**d**) Gold standard.

In Figure 5a,c, there is no data displayed at sensors R2, R8, R14, and R18 in the image, but there is periodic data during actual walking. The possible reason is that the volunteer lands on the heel first, not a sensor problem.

The distribution of peak pressure across sole regions is a key indicator for classifying arch types. The region of peak pressure varies by arch type. In normal feet, it is typically located at the heel. As Figure 5c shows, the pressure at the heel reaches as high as 30,000 g, whereas there is almost no data in the arch region, which is consistent with the expected behavior. In flat feet, it is often found in the midfoot. In Figure 5a, stable pressure values indeed appear in the arch region, and almost all sensors recorded data during the entire testing process. For high-arched feet, it is concentrated in the heel and later forefoot. As shown in Figure 5b, extremely high values are observed at the heel, while the arch region shows no data, which is in good agreement with actual observations. Furthermore, comparing the three types, some differences are also observed in the lateral foot region, with the high-arched foot showing a distinct pattern. Consequently, the peak region during static standing is used as a feature in this study. Then, this study uses Python code to find this feature. In this study, the classic footprint method combined with the Arch Index proposed by Cavanagh and Rodgers (1987) [3] was used for initial classification of participants' foot arch types. The procedure was as follows: each participant stood on one foot after applying inking material to the sole, and the footprint was captured on white paper. After scanning the footprint, the toe area was removed using image-processing software. The total length of the footprint was divided into three equal parts in Figure 5d, and the ratio of the area of the middle third (B) to the total area of the footprint was calculated as the Arch Index. The foot arch was classified into three types based on this ratio: normal arch (0.21~0.26), flat foot ($\geq$0.26), and high-arched foot (<0.21).

People frequently engage in activities such as walking and running. Therefore, analyzing only static pressure has drawbacks. Now, many people feel that the arch of the foot may evolve into a deformity during running. For this reason, this study focuses on analyzing the data collected during walking.

Jin et al. [16] noted that, during walking, the pressure percentages in different foot arches are different among boys with abnormal arches. This paper utilizes the pressure percentages as input features.

After observing all the pressure cloud maps, we chose special sensors R5, R6, and R15 to analyze based on gait data collected from six walking steps by drawing the pressure–time curves. Figure 6a,d,g shows the trend of sensors in the case of flatfoot. Figure 6b,e,h shows the trend of sensors in the case of a normal foot, and the others show the trend of sensors in a high-arched foot. It is evident that the stance phase duration is longer with a flatfoot [17], which is a manifestation of flatfoot collapse, comparing these figures. In contrast, R5 and R6 in the high-arched foot present almost no data; R15 in the flatfoot has less data than other types. In a normal foot, R5 and R6 present a few impulses. Owing to the elevated arch, people with high arches have less ground contact time and less pressure at the R15 sensor during a gait. As Figure 6d shows, a flat foot increases the area of plantar contact for its collapsed arches. However, normal feet show a cyclic change with load bearing during stance and unloading during swing, as in Figure 6h.

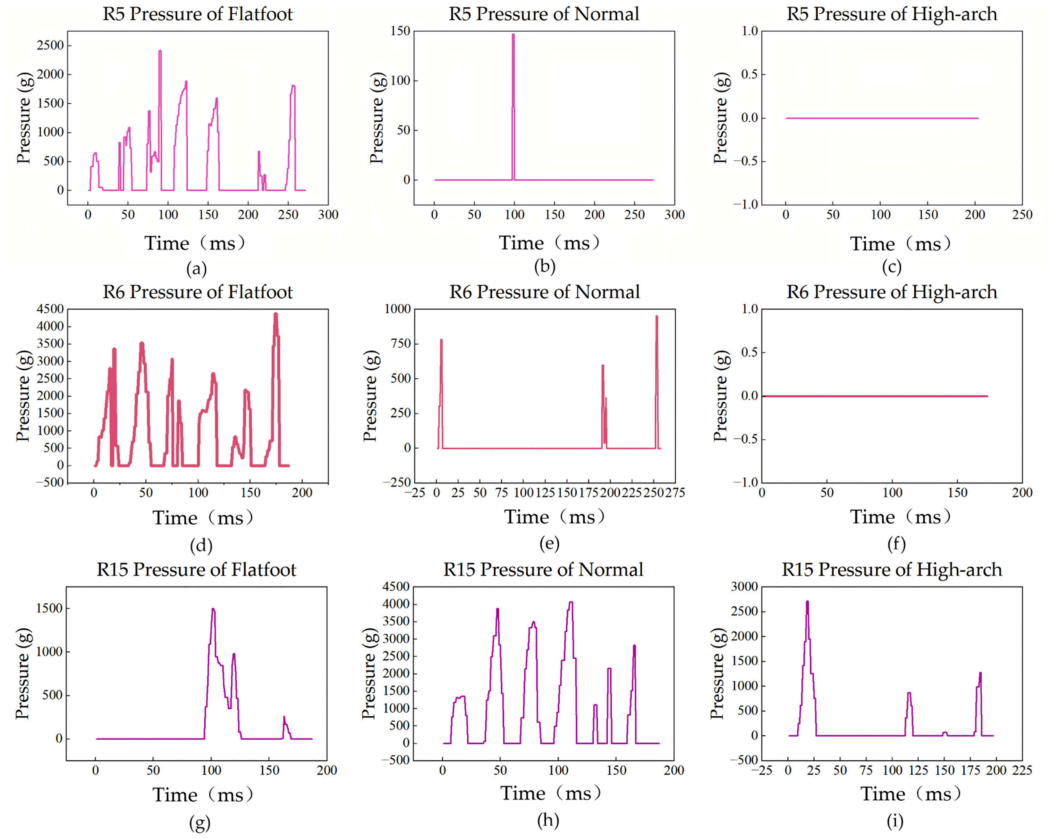


Figure 6. (**a**) The R5 data of person with flatfoot; (**b**) the R5 data of normal foot; (**c**) the R5 data of high-arched foot; (**d**) the R6 data of person with flatfoot; (**e**) the R6 data of normal foot; (**f**) the R6 data of high-arched foot; (**g**) the R15 data of person with flatfoot; (**h**) the R15 data of normal foot; (**i**) the R15 data of high-arched foot.

The pressure only reflects the transient impact. In contrast, impulse reflects the feedback effect produced by the forces in each plantar region interacting with the ground. The magnitude of the impulse is influenced by the plantar contact time and the pressure value, and it is also a direct factor in fatigue accumulation. The plantar pressure–time integral represents the cumulative effect of pressure on the foot during ground contact. Its magnitude is influenced by two factors, namely the peak pressure and contact time of every region, and it can reflect the magnitude of the load on the foot [18]. Li et al. calculated the maximum pressure time integral (PTI); different arches lead to different PTIs. Xiang et al. in their study showed that people with high arches have greater PTIs in the first, second, and third metatarsals and the medial heel [19]. Its calculation formula is as follows:

$$\mathrm{PTI} = \int_{t_1}^{t_2} F(t)\, dt \times \frac{9.8}{1000} \tag{1}$$

where $F(t)$ refers to a pressure signal with the unit of grams, and t_1 and t_2 refer to the start and end times of integration. The PTIs of different volunteers in one gait cycle are shown in Figure 7a–c. As shown in Figure 7a, the PTI values at R9, R3, R4, and R10 are relatively large. These four locations correspond to the heel region, which aligns with the findings of Xiang et al. Figure 7b reveals that, for the flatfoot, PTI is present in almost all regions. In addition, the arch region yields data that differ from those of the other two types. Furthermore, in Figure 7c, a high PTI is observed at the heel, and considerable PTI is also noted at R7, R13, and R1, which correspond to the second, third, and first metatarsal regions. These indicate that the normal foot receives a high proportion of PTI in these regions.

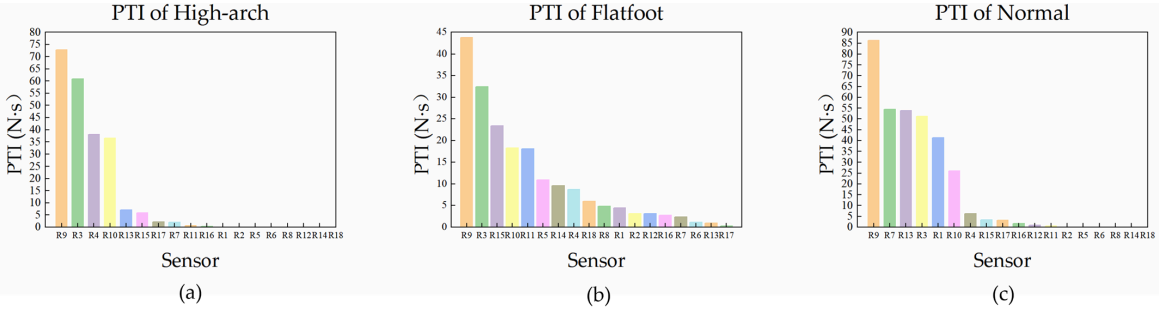


Figure 7. (**a**) The PTI of person with high arch in one gait cycle; (**b**) the PTI of flat feet; (**c**) the PTI of normal feet.

The raw acceleration data from the sensor's X, Y, and Z axes during normal walking are shown in Figure 8a–i. It can be clearly observed that the axis acceleration data follow the physical principle of dominant vertical gravitational acceleration during walking [20]. The curves show periodic patterns, corresponding to heel strike, foot support, and toe-off. A complete gait cycle can be divided into the stance and swing phases. This paper uses acceleration data to assist in dividing the gait cycle. The mean of the *x*-axis is about 0.5 in Figure 8a,d,g. The mean of the *y*-axis is about 0.2 in Figure 8b,e. Only Figure 8h shows a slightly different mean. And the *z*-axis is 1.05 in Figure 8c,f,i. Based on the mean, the stance phase is about 71% of the gait cycle from the *Y*-axis data in Figure 8b,e,h. We consider the period starting from the point where the data is not at the mean value and ending at the point where it reaches the value as the stance phase. The three arch types show no significant differences in mathematical characteristics. This finding is similar to those reported in the previously cited literature. The present study differs in that it simultaneously considers the *X*-axis and the *Y*-axis.

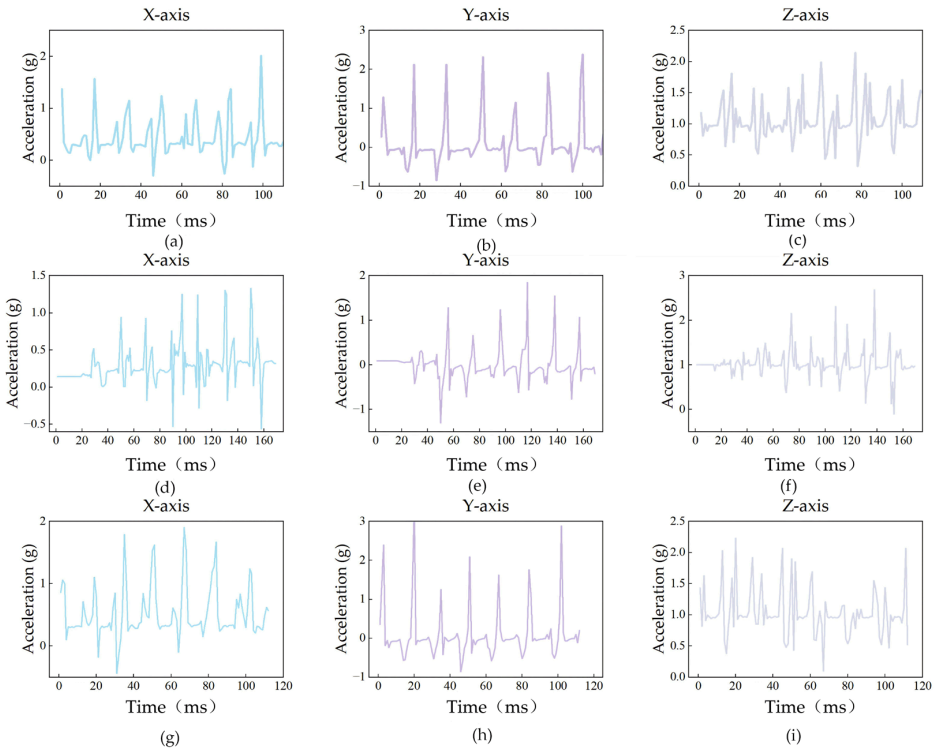


Figure 8. (**a**) The *x*-axis data of normal foot; (**b**) the *y*-axis data of normal foot; (**c**) the *z*-axis data of normal foot; (**d**) the *x*-axis data of flatfoot; (**e**) the *y*-axis data of flatfoot; (**f**) the *z*-axis data of flatfoot; (**g**) the *x*-axis data of high-arched foot; (**h**) the *y*-axis data of high-arched foot; (**i**) the *z*-axis data of high-arched foot.

3.3. Time–Frequency Domain Features

Gait-based time–domain analysis focuses on the motion attributes of gait patterns, which are typically reflected in the local landmarks of the overall waveform. Although parameters are frequently employed to describe gait time domain features with clear physical meanings, the number of features is unexpectedly small. In addition, these domain features only reflect pressure amplitude and cumulative effects but fail to capture the dynamic characteristics of the signal. The signal in the frequency domain changes significantly, which can be employed to classify different gait patterns. Yet, frequency–domain analysis alone has limitations in processing non-stationary signals. The pressure data sampling rate is 20 Hz. According to the Nyquist sampling theorem, the maximum frequency that can be analyzed is 10 Hz. This range covers the main energy frequency band of human gait activities. As plantar pressure is a non-stationary signal, it is suitable to use wavelet analysis, a time–frequency analysis. The procedure of wavelet analysis is illustrated in Figure 9a. The original pressure signal and the signal after gait cycle segmentation are shown in Figure 9b. Wavelet packet decomposition is employed for feature extraction. In each layer of decomposition, the pressure signal is convolved with a low-pass filter and a high-pass filter, respectively, followed by down-sampling, thereby decomposing the signal into different frequency subbands layer by layer. This process is implemented in Python using pre-packaged libraries. This paper employed the Daubechies 4 (db4) wavelet as the basis function, owing to its orthogonality, compact support, and regularity. In addition, this method is widely used in biomedical signal processing [21]. Compared with other wavelet families, db4 offered a slight advantage in convergence and stability during feature extraction. The decomposition level is set to 3, dividing the 0–10 Hz frequency band into 8 sub-bands of 1.25 Hz each, as the Figure 9c shown. This can not only cover the main frequency of the gait signal, but also avoid the problem caused by high dimensionality.

The raw signals were first smoothed using a low-pass filter. Gait cycles were segmented by identifying local maxima in the filtered signals. The three-layer wavelet packet decomposition is performed on the signal to obtain wavelet packet coefficients in eight frequency bands, as shown in the following formula:

$$d_i(j) = (F * \varnothing_i)\left(2^L j\right) \tag{2}$$

where $d_i(j)$ refers to the j-th wavelet packet coefficient in the i-th frequency sub-band, F refers to the original pressure signal, $\varnothing_i$ is the equivalent filter for the i-th frequency sub-band, $*$ is convolution operator, 2^L is down sampling factor, j refers to time point, and L is three in this study. The process is implemented internally by the code.

The time–domain reconstruction results of the wavelet packet decomposition are provided in Figure 10a–i, including the original pressure signal and the reconstructed waveforms corresponding to the eight frequency bands. This reconstruction plot is based on a single step of a volunteer's data for principle illustration. Band 1 (0–1.25 Hz) shows the slow pressure changes during stance in Figure 10b, while Band 8 (8.75–10 Hz) picks up the sharp impact when the heel hits the ground in Figure 10i. These demonstrate that wavelet packet decomposition effectively isolates the distinct frequency components inherent in the original signal.

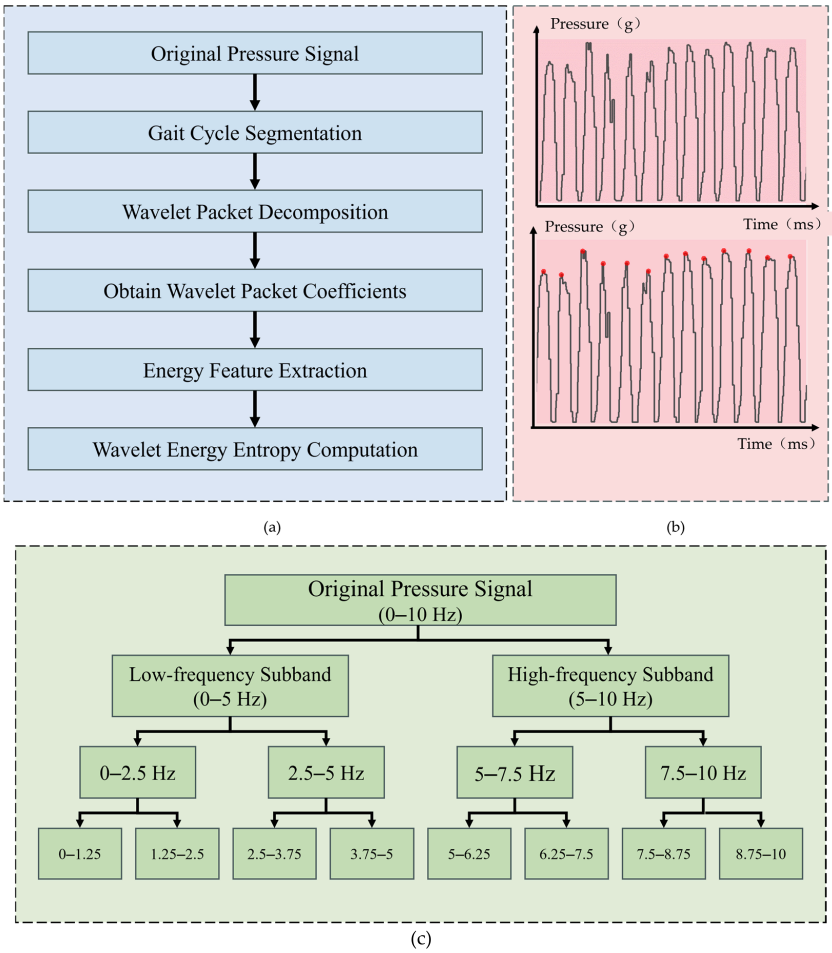


Figure 9. (**a**) The procedure of wavelet analysis; (**b**) the original signal and the signal after gait cycle segmentation; (**c**) the process of determining the frequency range of each subband in three-layer wavelet packet decomposition.

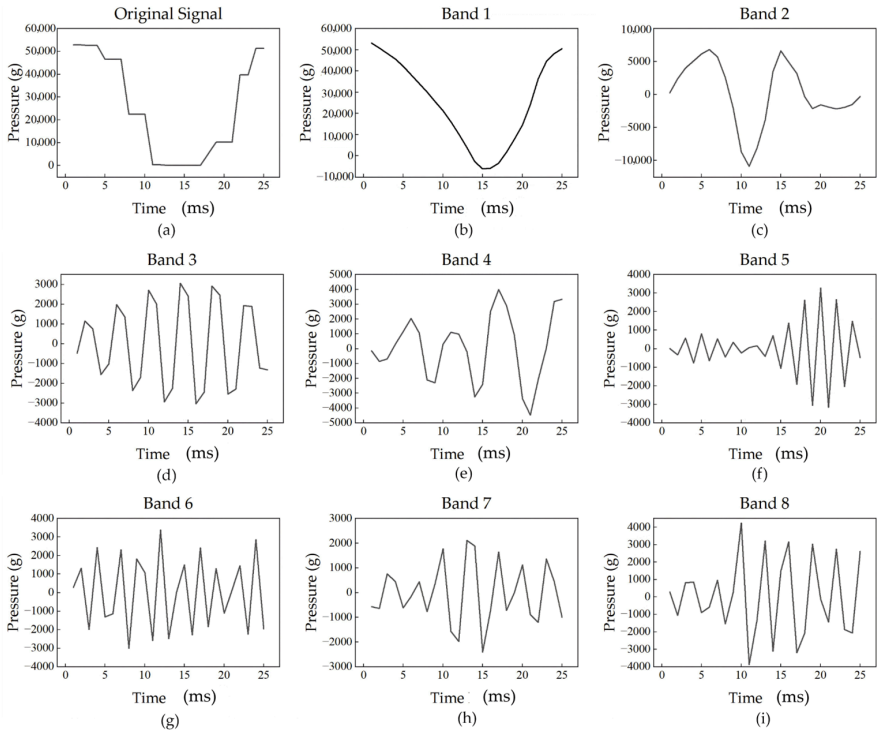


Figure 10. (**a**) The original pressure signal; (**b–i**) the signal in the eight frequency bands.

Wavelet packet decomposition (WPD) was applied to extract the energy features and the entropy features. The energy of each frequency sub-band is defined as the sum of the squares of its wavelet packet coefficients, which is expressed in the following equation.

$$E_i = \sum_{j=1}^{N_i} | d_i(j) |^2, i = 1, 2, 3 \ldots, 8 \tag{3}$$

where E_i is the sum of the squares of its wavelet packet coefficients and is also the total energy of the corresponding frequency band, N_i refers to the number of coefficients in the i-th band, $d_i(j)$ is obtained from the above calculation, and j is the number of the gait cycle.

To remove the influence of signal amplitude variations, each band is normalized as follows:

$$e_i = \frac{E_i}{\sum_{i=1}^{8} E_i}, i = 1, 2, 3 \ldots, 8 \tag{4}$$

Thus, the wavelet packet energy feature is obtained, which can reflect the condition of gait cycle energy in different bands. Then, based on the information entropy theory, the wavelet energy entropy is defined as [22]:

$$H = -\sum_{i=1}^{8} e_i \log_2 e_i \tag{5}$$

where the condition is $e_i \log_2 e_i = 0$ when $e_i = 0$. Energy entropy can reflect the degree of concentration of signal energy in the frequency domain. If the energy is concentrated in a few frequency sub-bands, the value is small, which means that the signal has a dominant frequency component. If the value is big, this means that energy is evenly distributed in many bands and the signal is very complex. After these operations, we obtained the energy entropy of many gait cycles. The mean and standard deviation (Std) of it are used as features:

$$Mean = \frac{1}{N} \sum_{k=1}^{N} H_k \tag{6}$$

where N is six because we chose six walking steps in the previous part of this paper.

$$Std = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} (H_k - Mean)^2} \tag{7}$$

where N is six.

After the theoretical derivation, a wavelet packet analysis was performed to obtain the energy distribution across different foot arch types. Then, we found that energy is concentrated in Band 1 for the normal feet in Figure 11c. Flat feet, although also having much energy in Band 1, display an extension across almost seven bands in Figure 11b. High arches exhibit energy across all bands in Figure 11a. These differences provide a basis for reliable arch type classification [23].

The extraction of the features employed in this study has been detailed in the preceding sections.

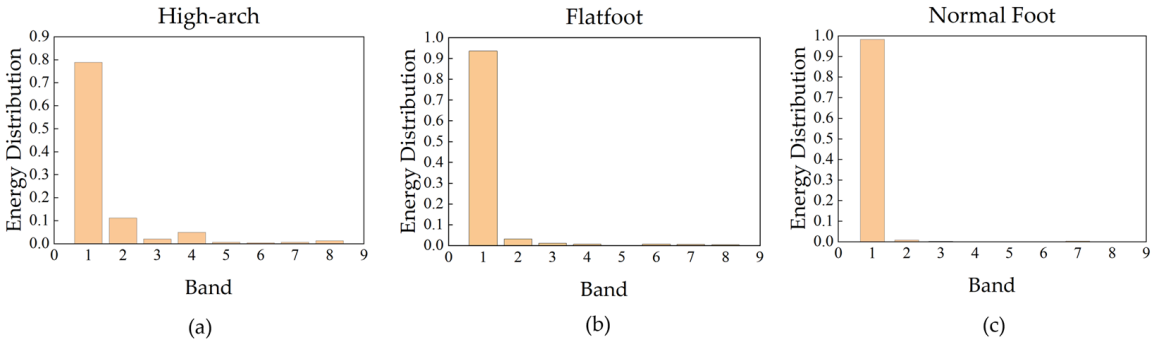


Figure 11. (**a**) The wavelet packet energy distribution of high arches; (**b**) the wavelet packet energy distribution of flat feet; (**c**) the wavelet packet energy distribution of normal feet.

4. Machine Learning

After the above work, we extracted static features, time–domain features, and time–frequency domain features. Since the extracted data included both numerical and non-numerical features, we applied Z-score normalization to all features to eliminate the impact of differing feature scales on distance-based algorithms [24], resulting in each feature having a mean of zero and a variance of one. The formula for Z-score standardization is as follows:

$$z = \frac{x - \mu}{\sigma} \tag{8}$$

Given that there are too many features, inputting all of them directly into the model would increase computational cost and processing time and could lead to the curse of dimensionality and overfitting. Therefore, we performed feature selection using a one-way analysis of variance (ANOVA) to retain the most contributing variables [25]. ANOVA is used to test whether there are significant differences in the means of a single feature across different arch types. The null hypothesis (H_0) assumes that the population means of the three arch types are equal, while the alternative hypothesis (H_1) assumes that the population means of the feature are not all equal across the three types. To assess this, we computed the F-statistic and its associated p-value, which indicate whether the observed differences are statistically significant. The relevant formulas are as follows:

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - u_i)^2 \tag{9}$$

Formula (9) is the within-group sum of squares (SSW). Formula (10) is the between-group sum of squares (SSB), where k is the number of three arch types. n_i is the sample size of group i. u_i is the group mean. x_{ij} denotes the j-th observed value of the feature within the i-th group:

$$SSB = \sum_{i=1}^{k} n_i (u_i - u)^2 \tag{10}$$

The mean squares are as follows:

$$MSB = \frac{SSB}{k-1}, \; MSW = \frac{SSW}{N-k} \tag{11}$$

F-statistic:

$$F = \frac{MSB}{MSW} \tag{12}$$

where k is the number of three arch types, n_i is the sample size of group i, N is the total sample size, u_i is the group mean, and u is the overall mean [26].

p-value is the area under the F-distribution curve to the right of the current F-value. If the *p*-value is less than the predetermined significance level (0.15), the null hypothesis is rejected, concluding that this feature exhibits a statistically significant difference among the three arch types. Then, we selected these features as an effective feature subset and used them as inputs for the model.

In this paper, we adopted a subject-wise split scheme to prevent data leakage and overestimation of classification accuracy. Specifically, the 30 participants were randomly divided into a training set and a test set, while ensuring through stratified sampling that the proportions of the three foot arch types were approximately the same in both sets. The training set was used for feature selection and model training, while the independent test set was used only for a single evaluation of the final model performance.

To further assess the classification capability of these features and contrast the effect of various machine learning systems, we chose four classic algorithms: support vector machine (SVM) [27], K-nearest neighbors (KNN), decision tree (DT), and Random Forest (RF). These three algorithms are widely used in the field of pattern recognition, and each has its own distinct advantages.

SVM is a supervised learning method in statistical learning. In the feature space, it draws different decision boundaries, seeking the boundary that maximizes the overall distance to all data points to find an optimal classification hyperplane that separates different classes. By widening the margin, the model achieves better generalization and is less likely to be thrown off by individual samples.

KNN is a non-parametric, instance-based classification method. For a given test sample, it finds the K known samples in the feature space that are closest to it, and then votes according to the classes of these K neighbors, selecting the class with the most votes as the prediction result.

DT is a tree-based supervised learning approach; it selects optimal features and optimal split thresholds to obtain subsets. To control model complexity and prevent overfitting, we use a grid search over max_depth, min_samples_split, min_samples_leaf, and criterion.

RF integrates the decisions of many trees, which can automatically handle high-dimensional data without needing complex selection [28]. The hyperparameter settings of the random forest model have a significant impact on classification performance [29]. The structure of RF is shown in Figure 12.

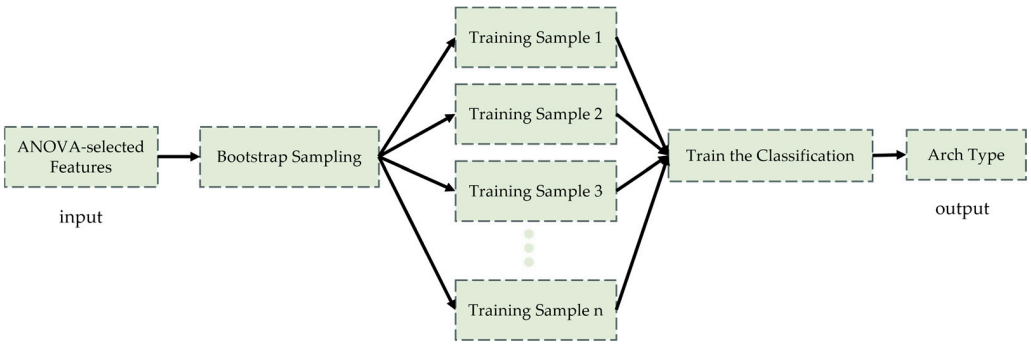


Figure 12. The structure of RF.

RF and DT were using grid search in combination. Based on the aforementioned procedures, the performance of different models is presented in Table 2.

From Table 2, we observed that KNN is inherently disadvantaged on small-sample, high-dimensional data; the accuracy was only 71%. Among them, RF achieves the highest accuracy. In addition, we compared the four algorithms in terms of recall, F1-score, and balanced accuracy. Under the Random Forest algorithm, one high-arched foot was misclas-

sified as flat, while all flat and normal feet were correctly identified. Since each foot arch type has its own F1-score, the value reported in the table is the macro average of the three classes. So, we chose the RF model to analyze and obtained class-wise scores, which are shown in Table 3.

Table 2. The performance of the models.

Model	Test Accuracy	Recall	F1-Score	Balanced Accuracy
SVM	0.86	0.89	0.87	0.89
KNN	0.71	0.67	0.68	0.67
DT	0.83	0.83	0.80	0.85
RF	0.92	0.91	0.92	0.91

Table 3. The class scores.

Class	Accuracy	Recall	F1-Score
Flat feet	0.95	0.95	0.95
High-arch	0.93	0.86	0.89
Normal feet	0.88	0.92	0.90

It is observed that the model has the lowest classification accuracy for normal feet. This may be because the characteristics of normal feet fall between the other two, leading to overlap in the feature space.

Additionally, during the RF modeling, the top-ten most important features were extracted. The corresponding proportions are shown in Figure 13. R6 (percentage 8.8% in the ten features) and R12 (percentage 4.7%) belong to the medial arch region, while R16 (percentage 10.3%) is on the lateral midfoot. These are crucial to classify different arch types, and the results are consistent with the theory. Band 3 captures the variability of low-frequency energy across gait cycles; Band 5 and Band 6 generally correspond to the impact peak at heel contact or rapid pressure changes. As mentioned in Section 3.3, different foot arch types exhibit different frequency band characteristics. According to the learning results, Bands 3, 5, and 6 contribute to the classification to a certain extent. From this figure, the pressure entropy mean (percentage 21.3%) and standard deviation (percentage 21%) are also important for this model.

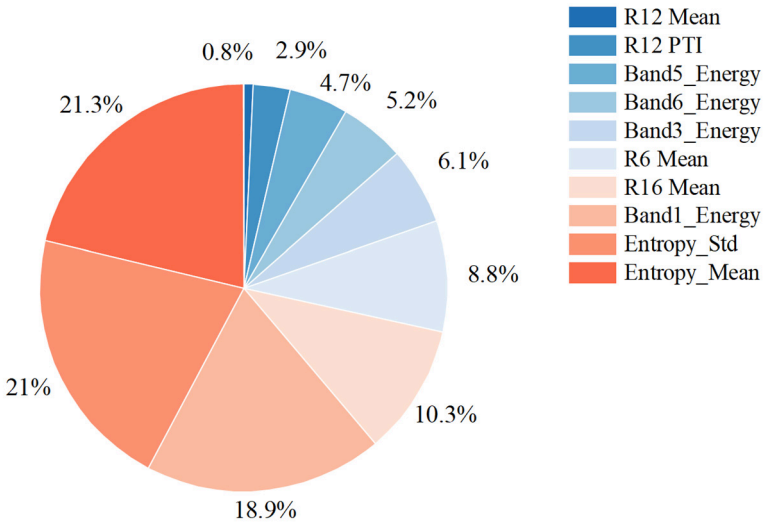


Figure 13. Important features of RF.

5. Conclusions

This paper presents a plantar-based arch classification system for daily foot monitoring. By designing flexible sensors and acquisition circuits, plantar acceleration and pressure data were obtained. In the study, over forty features were extracted from static, time–domain, and time–frequency domain. After feature selection, several machine learning models were trained. Based on ANOVA feature selection and standardization processing, the RF achieved the highest accuracy of 92%, outperformed the DT, and exhibited strong generalization ability. Furthermore, the results indicate that wavelet band energies (Bands 3, 5, and 6) and the mean pressure in some regions (R16, R6, and R12) are the discriminative features, which are consistent with the biomechanical differences among different arches during the stance phase of gait. Compared with existing static assessment methods or laboratory force plates, this system contributes by integrating multimodal sensors into a flexible insole for daily walking monitoring and, by using wavelet entropy, to quantify arch-induced signal complexity changes.

The current accuracy achieved in this work has room for improvement. The literature shows that, under few-shot conditions, the adaptive fused domain-cycling variational generative adversarial network can enable balanced training of the model [30]. Future research could integrate multiple modeling approaches with advanced medical knowledge to enhance the accuracy. Future research will also focus on expanding the sample size to cover different ages, body weights, and pathological conditions. It is expected to provide technical support for personalized foot orthopedics and health monitoring.

Author Contributions: Conceptualization, B.N. and J.Z.; methodology, B.N. and J.Z.; software, J.Z.; validation, J.Z. and C.Y.; formal analysis, J.Z.; investigation, C.Y.; resources, B.N.; data curation, J.Z. and C.Y.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z. and C.Y.; visualization, J.Z.; project administration, B.N.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: All subjects gave their informed consent for inclusion before they participated in the study. Ethics approval is not required for this type of study. The study was conducted following the local legislation: https://www.gov.cn/zhengce/zhengceku/2023-02/28/content_5743658.htm (accessed on 4 May 2026).

Informed Consent Statement: All subjects gave their informed consent for inclusion before they participated in the study.

Data Availability Statement: The raw/processed data required to reproduce the findings cannot be shared at this time because the data also form part of an ongoing study.

Conflicts of Interest: Author Baoqing Nie was employed by Suzhou LZY Technology Company, China. The remaining 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.

Abbreviations

The following abbreviations are used in this manuscript:

AAFD	Adult acquired flatfoot deformity
FPC	Flexible printed circuit
PTI	Pressure time integral
ANOVA	Analysis of variance
SSW	Within-group sum of squares
SSB	Between-group sum of squares
SVM	Support vector machine

KNN	K-nearest neighbors
DT	Decision tree
RF	Random Forest

References

1. Ye, J.P. Effect of Customized Insoles on Plantar Pressure in Adult-Acquired Flatfoot Stage II. Master's Thesis, Nanjing University of Chinese Medicine, Nanjing, China, 2019.
2. Qin, B.; Wu, S.; Zhang, H. Evaluation and Management of Cavus Foot in Adults: A Narrative Review. *J. Clin. Med.* **2022**, *11*, 3679. [CrossRef] [PubMed]
3. Cavanagh, P.R.; Rodgers, M.M. The arch index: A useful measure from footprints. *J. Biomech.* **1987**, *20*, 547–551. [CrossRef]
4. Williams, D.S.; McClay, I.S. Measurements used to characterize the foot and the medial longitudinal arch: Reliability and validity. *Phys. Ther.* **2000**, *80*, 864–871. [CrossRef]
5. Nikolaidou, M.E.; Boudolos, K.D. A footprint-based approach for the rational classification of foot types in young schoolchildren. *Foot* **2006**, *16*, 82–90. [CrossRef]
6. Lee, Y.C.; Wang, M.J. Taiwanese adult foot shape classification using 3D scanning data. *Ergonomics* **2015**, *58*, 513–523. [CrossRef]
7. 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 Posture* **2006**, *23*, 339–347. [CrossRef]
8. Xu, Z.; Gao, Z.; Zhou, Z.; Wang, Y.; Pan, J.; Xiang, L.; Song, Y.; Sun, D.; Radak, Z.; Cen, X. Impact of Fatigue on Spine Dynamic Stability and Gait Patterns in Runners with Moderate Flatfoot Versus Normal Arch. *Bioengineering* **2025**, *12*, 1256. [CrossRef] [PubMed]
9. Jlassi, O.; Emmerzaal, J.; Vinco, G.; Garcia, F.; Ley, C.; Grimm, B.; Dixon, P.C. Outdoor Walking Classification Based on Inertial Measurement Unit and Foot Pressure Sensor Data. *Sensors* **2026**, *26*, 232. [CrossRef]
10. Coluccio, M.L.; Pullano, S.A.; Vismara, M.F.M.; Coppedè, N.; Perozziello, G.; Candeloro, P.; Gentile, F.; Malara, N. Emerging Designs of Electronic Devices in Biomedicine. *Micromachines* **2020**, *11*, 123. [CrossRef] [PubMed]
11. Mei, Z.; Zhao, G.; Ivanov, K.; Guo, Y.; Zhu, Q.; Zhou, Y.; Wang, L. Sample entropy characteristics of movement for four foot types based on plantar centre of pressure during stance phase. *Biomed. Eng. Online* **2013**, *12*, 101. [CrossRef]
12. Laganà, F. Design and Simulation-Based Validation of an Embedded Acquisition Architecture for In Situ PCB Integrity Monitoring in Biomedical Devices. *Electronics* **2026**, *15*, 833. [CrossRef]
13. Li, K. Series Study on Changes of Limb Morphology and Plantar Pressure after High Tibial Osteotomy. Ph.D. Thesis, Southern Medical University, Guangzhou, China, 2024. [CrossRef]
14. Yang, C.; Zhang, T.; Du, J.L.; Liu, Z. Correlation analysis of plantar pressure change characteristics and joint function after internal fixation of supination-external rotation type IV ankle fracture. *Chin. J. Tradit. Med. Traumatol. Orthop.* **2025**, *33*, 22–28. [CrossRef]
15. Mohammed, S.; Samé, A.; Oukhellou, L.; Kong, K.; Huo, W.; Amirat, Y. Recognition of gait cycle phases using wearable sensors. *Robot. Auton. Syst.* **2016**, *75*, 50–59. [CrossRef]
16. Jin, Z.X.; He, H.; Yang, Y.; Zheng, Y.K.; Xiong, K.Y. Study on Static Balance Ability and Plantar Pressure Characteristics of Boys with Abnormal Foot Arch. *J. Chengdu Sport Univ.* **2018**, *44*, 116–121. [CrossRef]
17. Wang, X.T.; Xu, D.D.; Zhang, J.X.; Su, H.L.; Wang, Q. Stability of load-bearing cross barrier of different arch structures. *Chin. J. Tissue Eng. Res.* **2021**, *25*, 3838–3843. Available online: https://www.cjter.com/EN/10.12307/2021.089 (accessed on 7 May 2026).
18. Zhang, X.Y.; Li, Y.; Li, L.H.; Li, B. Effect of Height-increasing Insoles on Plantar Pressure Distribution and Heel Balance. *Leather Sci. Eng.* **2013**, *23*, 56–59. Available online: https://kns.cnki.net/kcms2/article/abstract?v=8kKd7LBMH3zTioWaYJ3XQAgtYFoquUe-4AukwRoRLoWIWNYYqfHh-fGwxm5ODDXGE02EZHVZLGg3YKAZ4iodfaoIzR7rLSdklKFE1qO3dUsQgNYERpR6LWQjSUX1s8QwCOJZwFpn9n_kXTLgx2crpIqUJKI3vWL8IwuUvQjXnoU=&uniplatform=NZKPT&language=CHS (accessed on 7 May 2026).
19. Li, Y.; Wu, J.; Zheng, C.; Huang, R.R.; Na, Y.H.; Xue, J.; Yang, F.; Su, B.; Wang, Z.S. Change of plantar dynamics in Chinese parachutists at landing. *Acad. J. Chin. PLA Med. Sch.* **2013**, *34*, 953–956. [CrossRef]
20. Yan, J.G. Human Activity Recognition Based on Foot-Mounted Accelerometer. Master's Thesis, Hefei University of Technology, Hefei, China, 2018.
21. Zhang, M.; Liu, D.; Wang, Q.; Zhao, B.; Bai, O.; Sun, J. Gait Pattern Recognition Based on Plantar Pressure Signals and Acceleration Signals. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 4008415. [CrossRef]
22. Li, J.X.; Ke, X.Z.; Guo, H. Application of Wavelet Variance and Wavelet Entropy in Signal Feature Extraction. *J. Xi'an Univ. Technol.* **2007**, *23*, 365–369. [CrossRef]
23. Li, N.; Hou, Y.Y.; Zhou, P.; Han, Y. Wavelet Packet Transform Analysis of Plantar Pressure Signals in Patients with Parkinson's Disease. *J. Clin. Rehabil. Tissue Eng. Res.* **2010**, *14*, 2377–2380. [CrossRef]
24. Wang, X.; Jiang, H.; Dong, Y.; Mu, M. Spatial-channel collaborative multi-scale graph interaction deep transfer learning for unsupervised rotating machinery fault diagnosis. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114691. [CrossRef]

25. Buchecker, M.; Müller, E.; Wegenkittl, S.; Sattlecker, G.; Stöggl, T. An entropy approach for evaluating adaptive motor learning processes while walking with unstable footwear. *Hum. Mov. Sci.* **2018**, *60*, 48–56. [CrossRef] [PubMed]
26. Li, Z.Y.; Wang, C.X. Feature selection method based on F-statistics and SVM and its application in diagnosis of erythematosquamous skin diseases. In Proceedings of the 2012 2nd International Conference on Future Computers in Education (ICFCE 2012), Shanghai, China, 1–2 June 2012; pp. 281–285.
27. Bao, L.; Intille, S.S. Activity Recognition from User-Annotated Acceleration Data. In *Pervasive Computing*; Ferscha, A., Mattern, F., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3001, pp. 1–17. [CrossRef]
28. Feng, T.Y. Research on Gait Recognition and Prediction Based on Deep Learning and Complex Network. Master's Thesis, Hebei University of Science and Technology, Shijiazhuang, China, 2025. [CrossRef]
29. Laganà, F.; Pratticò, D.; Quattrone, M.F.; Pullano, S.A.; Calcagno, S. Hybrid AI–Taguchi–ANOVA Approach for Thermographic Monitoring of Electronic Devices. *Eng* **2026**, *7*, 28. [CrossRef]
30. Wang, X.; Jiang, H.; Zeng, T.; Dong, Y. An adaptive fused domain-cycling variational generative adversarial network for machine fault diagnosis under data scarcity. *Inf. Fusion* **2025**, *126*, 103616. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.