Smart Insole Pressure Distribution Prediction and Gait Optimization System Based on Multimodal Data Fusion and ViLT

Shiyi SHU[1], Fujun SHANGGUAN[2]
[1]School of Design, Wenzhou Polytechnic University, Wenzhou 325027, Zhejiang, China
[2]Ningbo Tuopu Group Co., Ningbo 315800, Zhejiang, China

Corresponding author: Shiyi SHU (e-mail: tomp808080@163.com)

ABSTRACT To improve the accuracy and adaptability of gait analysis in embedded smart insole systems, this study proposes a multimodal pressure distribution prediction and gait optimization framework based on Vision-and-Language Transformer (ViLT) and cross-modal spatiotemporal modeling. The proposed approach addresses the limitations of modality isolation, temporal misalignment, and rigid fusion strategies by integrating plantar pressure sensing, inertial measurement unit (IMU) signals, and biomechanical parameters into a unified end-to-end architecture. A dynamic attention alignment mechanism is developed to establish semantic associations among heterogeneous data streams, while an LSTM-based spatiotemporal coupling model combined with hyperelastic mechanics enables accurate pressure prediction and personalized gait optimization. Considering that wearable sensing systems increasingly rely on electromagnetic signal acquisition and wireless data transmission, the proposed multimodal fusion framework also provides methodological support for intelligent sensing architectures employing antenna-enabled communication and electromagnetic information acquisition. Experimental results demonstrate that the proposed system achieves regional pressure prediction errors below 0.9% for diabetic foot subjects and approximately 0.8% for healthy subjects, while reducing peak plantar pressure by more than 12% and 8%, respectively. The hardware–algorithm co-design and closed-loop feedback strategy significantly enhance gait adaptability and rehabilitation effectiveness, providing a reliable computational framework for intelligent wearable sensing, multimodal electromagnetic data fusion, and nextgeneration smart healthcare applications.

INDEX TERMS multimodal data fusion, wearable electromagnetic sensing, smart insole, pressure distribution prediction, gait optimization

I. INTRODUCTION

WITH the development of wearable devices and smart health monitoring technology [1-3], gait analysis systems based on smart insoles play an important role in sports rehabilitation, diabetic foot prevention, and fall warning for the elderly [4-6]. Existing systems mostly rely on single-sensor data, which makes it difficult to fully characterize the complex dynamic changes in plantar pressure and individual gait characteristics. This limitation results in a high error rate in the gait event conversion interval, which in turn affects the efficacy of pathological gait intervention.

This study constructs a three-level perception-edge-cloud collaborative architecture. The front-end integrates a synchronous acquisition module (128-channel pressure array at 100 Hz, IMU at ±16g/200 Hz, and bioimpedance at 50 Hz) and achieves sub-millisecond synchronization through a unified timestamp protocol. A lightweight Vision-and-Language Transformer (ViLT) inference engine is deployed on the edge and combined with a Dynamic Attention Alignment (DDA) model, which interacts with the cloud via BLE 5.0 to exchange feature vectors rather than raw data. The cloud-based pathology knowledge base dynamically updates the personalized baseline using 30 days of monitoring data.

This paper focuses on the problem of insufficient multimodal data fusion in smart insoles, specifically targeting the two core challenges of inaccurate pressure distribution prediction and low personalization of gait optimization strategies. This issue directly affects the system's ability to understand the user's walking status and represents a key opportunity for improving the intelligence level of smart insoles.

Current multimodal fusion approaches in smart insoles suffer from three critical shortcomings. First, data modality isolation prevents the system from leveraging complemen-

tary information between modalities. For example, pressure distribution data and gait events derived from an Inertial Measurement Unit (IMU) are often processed in silos, leading to missed opportunities for cross-validation. Second, temporal misalignment introduces errors in dynamic gait analysis. Mismatched sampling rates can cause event detection delays of up to 50 ms, which degrades the accuracy of fall risk prediction. Third, model structure rigidity limits adaptability. Most methods apply fixed fusion rules and fail to account for inter-individual variability in gait patterns, particularly in patients with diabetic foot who exhibit altered plantar mechanics. In recent years, many researchers have explored this issue in depth. A hybrid model based on a Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) architecture has been used for pressure time series modeling [7], but many methods fail to effectively integrate biomechanical parameters [8,9]. Wei et al. [10], Zhang et al. [11], and Choi et al. [12] attempted to apply an attention mechanism to achieve multi-source signal association; however, a bottleneck remained in cross-modal semantic alignment. Other studies by Li et al. [13] and Wang et al. [14] used the traditional Transformer structure for gait recognition and to extract overall spatial information. Han et al. [15], Negi et al. [16], and Jiao et al. [17] proposed a plantar area modeling method based on graph neural networks that improved spatial perception but had high computational complexity.

Although some studies have applied CNN, LSTM, and traditional Transformer models for pressure modeling and gait recognition or used attention mechanisms for multi-source data fusion [18-20], critical limitations persist. For instance, Chhoeum et al. [21], Sharifi et al. [22], and Shao et al. [23] extracted spatial features from pressure images using CNNs but ignored the dynamic mechanical correlation between different sole regions (e.g., forefoot and arch) during the gait cycle. Similarly, Wu et al. [24], Hu et al. [25], and Xue et al. [26] used a global attention mechanism to weight multimodal features but failed to distinguish key physiological nodes, resulting in prediction errors concentrated in the gait event transition interval. More recent works have shown promise. Rahayu et al. [27] developed an adaptive filtering algorithm that significantly improved the quality of IMU input data by removing noise. D'arco et al. [28] and Zhang et al. [29] combined physical models with deep learning for pressure prediction, thereby improving model interpretability. However, other approaches have not fully integrated all necessary components. Huang et al. [30] constructed a foot function assessment model based on in-shoe sensors but did not integrate physiological constraint information such as biomechanical parameters. Wang et al. [31] built a lower limb movement intention recognition system but did not propose how to convert recognition results into personalized gait adjustment strategies. Yang et al. [32] designed a biomechanically inspired fusion framework but did not fully consider the combination of dynamic pressure distribution prediction and intervention. While the above methods have advanced the field, they still struggle to meet the real-time and personalized requirements of smart insoles in complex environments.

This paper makes significant advancements in three key technical aspects of smart insole systems by enhancing the ViLT architecture and applying a multimodal spatiotemporal coupling modeling system. To address the heterogeneity of cross-modal data, a dynamic attention alignment mechanism is innovatively proposed. This achieves bidirectional semantic alignment of pressure image spatial features and biomechanical parameter temporal features through a learnable modal importance adjustment factor, ensuring the cross-modal attention weight distribution conforms to physiological laws. Furthermore, an LSTM-hyperelastic mechanical coupling model is constructed, and a stress-strain mapping matrix is established using the Mooney-Rivlin constitutive equation (with material parameters C_1 and C_2). Finally, a personalized optimization system based on the Deep Deterministic Policy Gradient (DDPG) framework is developed. The stress-strain mapping matrix provides a new solution for the dynamic calibration of individuals at high risk of diabetic foot.

II. ALGORITHM DESIGN

MULTIMODAL DATA PREPROCESSING

This study collected multi-source heterogeneous data from the smart insole system, including plantar pressure distribution, IMU signals (acceleration and angular velocity), and bioelectrical impedance information. The system adopted a three-level perception-edge-cloud collaborative architecture: the front-end embedded device was responsible for raw data collection and preliminary cleaning; the edge computing module performed dynamic attention alignment model inference; and the cloud handled pathological knowledge base updates and long-term trend analysis.

To adapt to the hardware constraints of the smart insole system, a noise reduction strategy based on sensor characteristics was designed in the preprocessing stage. For the nonlinear response of the pressure sensor array, a piecewise linear correction algorithm was used to reduce temperature drift errors. This approach was selected over polynomial fitting or Kalman filtering due to its low computational complexity and compatibility with real-time embedded processing. Polynomial fitting requires higher-order calculations that increase power consumption, while Kalman filtering introduces latency unsuitable for edge devices. For IMU signals affected by motion artifacts, an adaptive threshold filtering method based on wavelet transform was applied. This method was chosen because it outperforms traditional low-pass filtering in preserving high-frequency gait events (such as heel strikes) while suppressing noise4. Fixed low-pass filters risk distorting transient signals, whereas moving average methods introduce phase lag. The wavelet-based approach adaptively adjusts thresholds based on signal energy distribution, ensuring robustness to variable motion intensity. Bioimpedance data, with an inherently low sampling

rate (50 Hz), was up-sampled to 100 Hz via cubic spline interpolation to align with pressure and IMU data. While this increases temporal resolution, it may introduce minor interpolation artifacts (e.g., overshoots at impedance peaks). To mitigate this, a post-processing smoothing step with a Savitzky-Golay filter (window=5, polynomial order=3) was applied. To ensure the consistency of cross-modal features and the accuracy of temporal alignment in the subsequent modeling process, a systematic preprocessing workflow for multimodal gait data was proposed, as shown in Figure 1.

To address sampling frequency differences among sensors, a unified timestamp synchronization strategy was adopted. A high-precision clock synchronization module mapped the data from the pressure array (100 Hz), IMU (200 Hz), and bioelectrical impedance device (50 Hz) to the same time axis. Bioimpedance data was up-sampled to 100 Hz via cubic spline interpolation, while other low-frequency signals were up-sampled using linear interpolation so that all channel data were uniformly output at 100 Hz. The raw pressure data underwent spatial normalization and bad pixel correction. To account for individual wearing position deviations and local sensitivity drift of the pressure sensor array, a spatial normalization method based on a calibration matrix was used to map the raw pressure image to a standard plantar topological coordinate system. A sliding window median filter algorithm was applied to identify and correct abnormal pressure points, improving the consistency of the spatial distribution. For the IMU signal, a fourth-order Butterworth low-pass filter (cutoff frequency 20 Hz) was used to remove high-frequency noise, and the three-axis acceleration and angular velocity were converted into motion posture parameters with physical significance via Euler angle transformation. The Z-score normalization method was used to normalize differences in motion amplitude between users, enhancing the model's generalization ability. For bioimpedance data processing, which is greatly affected by environmental factors such as skin humidity, a dynamic baseline correction mechanism based on sliding window statistics was designed. The deviation between the current impedance value and the historical mean was calculated within a 5-second window. If the deviation exceeded a set threshold, an impedance drift compensation algorithm was triggered, and the baseline reference value was updated using the exponentially weighted moving average (EWMA) method.

CROSS-MODAL FEATURE EXTRACTION

This section details the construction of a cross-modal feature extraction module based on the ViLT framework and proposes a physiologically guided dynamic attention alignment model tailored to the needs of the smart insole system. The structure of this module is shown in Figure 2.

For the 2D pressure map sequence collected by the plantar pressure sensor, a lightweight CNN structure was used for local feature extraction. To match the 16×8 low-resolution pressure matrix of the smart insole, a spatial attention mechanism was designed to focus on three key areas: the forefoot, arch, and calcaneus. The feature channel weight distribution ratio was set to 3:2:2, consistent with clinical statistical laws of plantar pressure distribution. After the IMU signal (three-axis acceleration and angular velocity) and bioelectrical impedance data were segmented by sliding windows, statistical features (e.g., mean, variance, kurtosis, skewness), frequency domain main frequency energy ratio, and gait-related parameters (e.g., step frequency and step length estimation) were extracted from each window to construct a time-series feature vector of length 56. To meet edge deployment requirements, the Transformer encoder used 8-bit integer quantization (INT8) to accelerate inference, and the model size was compressed to meet the memory limitations of embedded devices.

DYNAMIC ATTENTION ALIGNMENT MECHANISM

To solve the heterogeneity and non-alignment problems of multimodal data in the feature space, this study designed a cross-modal dynamic attention alignment mechanism. This mechanism takes pressure image features and biomechanical parameter sequences as inputs and achieves adaptive fusion of inter-modal information by constructing a bidirectional interaction matrix.

Let the local spatial features be represented as $P \in \mathbb{R}^{H\times W\times D_p}$, where H and W are the spatial dimensions of the pressure image, and D_p is the feature dimension. Let the global semantic features be represented as $B \in \mathbb{R}^{T\times D_b}$, where T is the time step, and D_b is the embedding dimension of the biomechanical parameters. First, P is flattened into a one-dimensional feature vector sequence $P' \in \mathbb{R}^{N\times D_p}$, where $N = H \times W$. Then, the query vector $Q_P = P'W_Q$, key vector $K_B = BW_K$, and value vector $V_B = BW_V$ are generated through a linear transformation, where W_Q, W_K, W_V are learnable parameter matrices. The cross-modal attention weight matrix $A \in \mathbb{R}^{N\times T}$ is calculated, with its elements defined as:

$$A_{i,j} = \frac{\exp\left(Q_{P,i} \cdot K_{B,j}^T / \sqrt{D_k}\right)}{\sum_{k=1}^{T} \exp\left(Q_{P,i} \cdot K_{B,k}^T / \sqrt{D_k}\right)} \tag{1}$$

Here, Dk is the scaling factor used to alleviate the gradient vanishing problem caused by an overly large inner product. The biomechanical feature values are weighted and aggregated through the attention matrix to obtain the context-aware feature $C_P \in R^{N\times Dv}$ aligned with the pressure image:

$$C_P = A \cdot V_B \tag{2}$$

Next, the order of modalities is exchanged. B is used as the query source, while P′ serves as the key and value source. A reverse attention path is constructed to generate the enhanced biomechanical feature $C_B \in R^{T\times Dv}$. The two-way alignment features are reshaped into their original structures and fused through a residual connection and layer normalization:

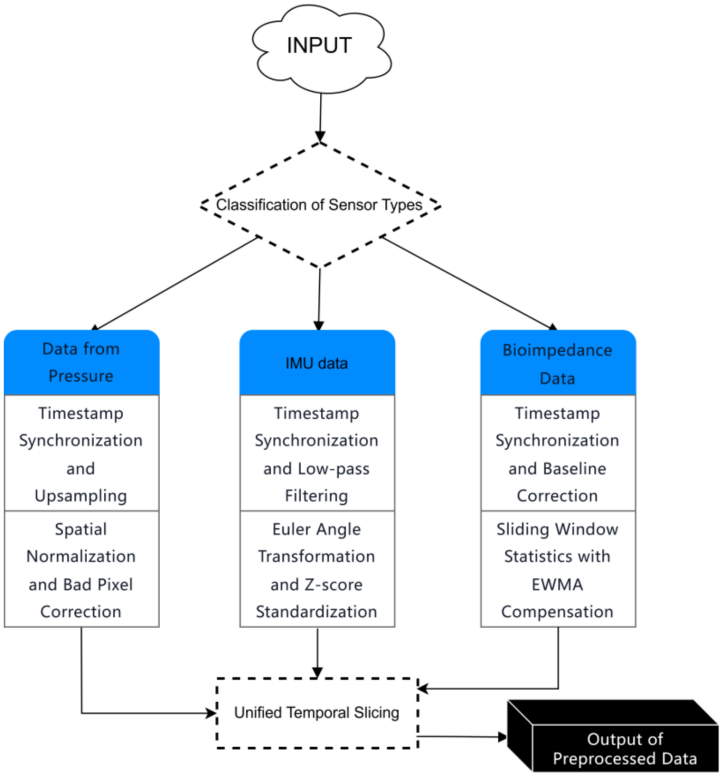


FIGURE 1. Data preprocessing

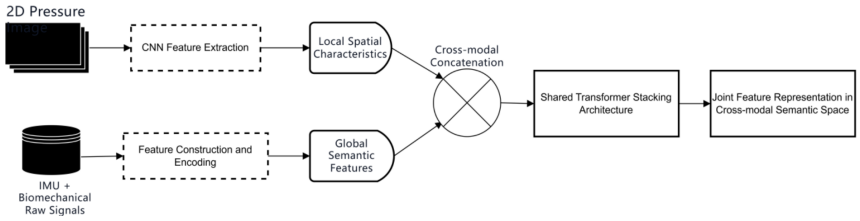


FIGURE 2. Cross-modal feature extraction structure

$$\widetilde{P} = \mathrm{LayerNorm}\left(P' + \mathrm{MLP}(C_P)\right) \tag{3}$$

$$\widetilde{B} = \mathrm{LayerNorm}\left(B + \mathrm{MLP}(C_B)\right) \tag{4}$$

Here, MLP represents a two-layer multi-layer perceptron for nonlinear mapping. To enhance the system's adaptability to individual differences, a learnable modal importance adjustment factor $\alpha \in [0,1]$ was applied to control the contribution ratio of different modalities in the attention mechanism. This parameter was automatically optimized through end-to-end training. An example of its attention weight distribution is shown in Table 1. The data show that the correlation strength of the medial arch area (p3) during the mid-stance phase reached 0.32–0.35 in healthy subjects, whereas the weight for the same area in patients with diabetic foot decayed to 0.29, revealing a degradation mechanism of multimodal correlation under pathological conditions. This finding is consistent with clinical observations of arch collapse in patients with diabetic foot, indicating that the dynamic attention mechanism can reflect physiological constraint laws. Analysis of the dynamic attention heat map further verified the model's ability to precisely capture the pressure-biomechanical coupling relationship during key stages of gait events.

In Table 1, the text states that the rows represent the weight distribution of a spatial position over 7 time steps, and the sum of weights for each row (excluding diabetic foot) is 1.0. However, the sum of weights for each row in the provided data does not equal 1.0 (e.g., row p1 sums to 1.0). Please verify the data or clarify the normalization method. Table 1 and the corresponding heat map in Figure 3 show how the model dynamically associates the spatial features of the pressure image with the biomechanical time series.

The heat map illustrates the cross-modal interaction effect of the dynamic attention alignment mechanism, presenting the association strength between pressure image spatial features (p1–p8) and biomechanical time-series data (t1–t7) through color gradients. The dark blue to red gradient

TABLE 1. Attention weight distribution for healthy subjects and diabetic foot patients.

Spatial location	Name of the solar area of the foot	Time step t1	Time step t2	Time step t3	Time step t4	Time step t5	Time step t6	Time step t7	Diabetic foot weight at t3
Position p1	Medial Forefoot	0.25	0.20	0.06	0.04	0.14	0.17	0.14	0.06
Position p2	Central Forefoot	0.15	0.22	0.21	0.23	0.09	0.06	0.04	0.08
Position p3	Medial Arch Area	0.07	0.10	0.32	0.35	0.09	0.04	0.03	0.29
Position p4	Lateral Forefoot	0.06	0.07	0.13	0.25	0.30	0.14	0.05	0.13
Position p5	Lateral Midfoot	0.12	0.06	0.08	0.09	0.17	0.25	0.23	0.07
Position p6	Lateral Heel Area	0.21	0.14	0.06	0.04	0.08	0.17	0.30	0.05
Position p7	Medial Heel Area	0.23	0.17	0.11	0.09	0.06	0.12	0.22	0.08
Position p8	Central Heel Area	0.14	0.21	0.20	0.15	0.12	0.11	0.07	0.12

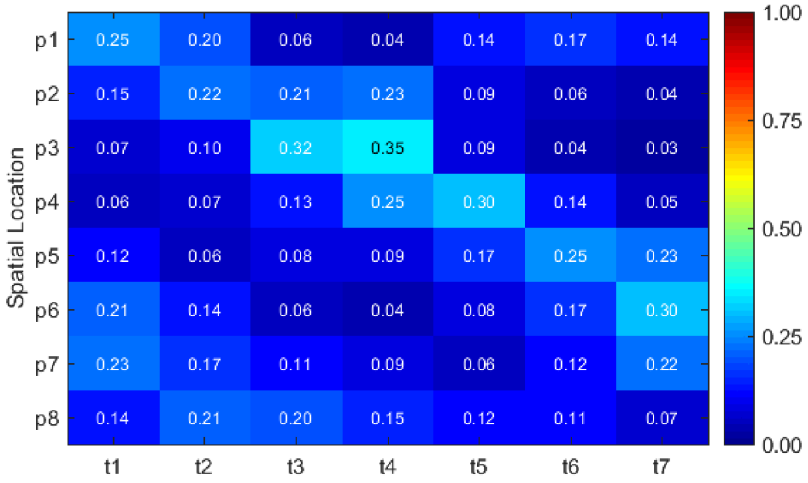


FIGURE 3. Heat map of attention weight distribution

represents the change of attention weight from 0 to 1. Notably, the medial arch area (p3) shows a significant high weight (0.32–0.35) in the t3–t4 stage, indicating that the model successfully captures the strong coupling between arch pressure and biomechanical parameters (such as arch height) in the mid-stance phase. The lateral forefoot (p4) shows a peak weight of 0.30 in the t5 stage, corresponding to the load transfer process after the heel leaves the ground. The heat map verifies the adaptive adjustment capability of the dynamic attention mechanism to individual differences and provides a physiological basis for subsequent gait optimization strategies, which has clinical application value in guiding forefoot pressure balance intervention.

SPATIOTEMPORAL COUPLING MODELING

To model the temporal dynamics of gait and achieve high-precision prediction of plantar pressure distribution, this paper proposed a spatiotemporal coupling modeling method that integrates an LSTM network and a plantar hyperelastic mechanics model. The local spatial features $P \in \mathbb{R}^{H\times W\times D}$ processed by the dynamic attention alignment mechanism were spatially averaged and pooled to generate a time-series input vector $x_t \in \mathbb{R}^D$, where $t = 1, 2, \ldots, T$ represents the discrete time step in the gait cycle. This sequence was used as the input to the LSTM network, and the hidden state was updated through the following equations:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{6}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{7}$$

$$\widetilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \tag{9}$$

$$h_t = o_t \odot \tanh(c_t) \tag{10}$$

Here, i_t, f_t, and o_t are the activation values of the input, forget, and output gates, respectively; c_t is the cell state; h_t is the output hidden state; σ is the Sigmoid function; and $\odot$ represents element-wise multiplication. LSTM captures the long-term dependencies in the gait cycle and outputs a hidden state sequence $H = [h_1, h_2, \ldots, h_T] \in \mathbb{R}^{T\times D_h}$ with temporal context information.

A decoder structure was designed to map the temporal features back to the spatial dimension; a fully connected layer was applied to map the hidden state h_t of each time step to the same spatial dimension as the original pressure image, yielding the preliminary prediction result $\hat{P}_t^{(0)} \in \mathbb{R}^{H\times W}$. To improve the physical consistency of the prediction, a post-processing module was constructed in combination with the hyperelastic mechanical properties of plantar tissue. It was assumed that the plantar soft tissue

stress-deformation satisfies the Mooney-Rivlin constitutive equation:

$$W = C_1(I_1 - 3) + C_2(I_2 - 3) \tag{11}$$

where W is the strain energy density; I_1 and I_2 are the Cauchy-Green strain tensor invariants; and C_1 and C_2 are material parameters. The parameters (C_1, C_2) were obtained through experimental measurements of the subject groups. Specifically, a relevant test platform was used to perform uniaxial compression tests on plantar soft tissue, and the least squares method was used to fit the stress-strain curve to optimize the intra-group parameters. Based on the finite element analysis method, the stress-strain mapping relationship matrix $M \in \mathbb{R}^{H\times W\times H\times W}$ of the plantar region was established. For any time t, its mechanical response correction term was calculated as:

$$\Delta P_t = M \cdot \left(\hat{P}_t^{(0)} - P\right) \tag{12}$$

where P is the static support pressure reference value. The final predicted pressure distribution is:

$$\hat{P}_t = \mathrm{ReLU}\left(\hat{P}_t^{(0)} + \lambda \Delta P_t\right) \tag{13}$$

where $\lambda \in (0,1)$ is the mechanical constraint coefficient, which was used to balance the contribution ratio between the data-driven prediction and the physical model.

MULTI-OBJECTIVE OPTIMIZATION STRATEGY

To generate personalized gait adjustment suggestions, this paper constructed a multi-objective optimization system based on the DDPG framework to jointly optimize pressure balance and gait energy consumption indicators. The optimization strategy is shown in Figure 4. In the reward function design, the pressure balance constraint was set in combination with diabetic foot ulcer prevention guidelines: when the pressure gradient ΔP between the metatarsal head area (p1/p2) and the lateral forefoot (p4) was ≥ 1.2 N/cm^2, a plantar pressure redistribution strategy was triggered, and the vibration feedback module was used to guide the user to adjust their gait (vibration frequency 15 Hz, duration 300 ms). This strategy learns the optimal action policy by interacting with the environment, while ensuring a reasonable distribution of plantar pressure and reducing the user's energy consumption during walking.

The state space $S \in \mathbb{R}^{D_s}$ was defined as the fused feature representation in the current gait cycle, which contained the spatiotemporal coupling features processed by the dynamic attention mechanism and the statistical information of historical gait parameters. The action space $A \in \mathbb{R}^{D_a}$ represented the adjustable gait parameter vector, including physically interpretable variables such as step length, step frequency, landing angle, and support phase time distribution ratio. The reward function was designed as follows:

$$r_t = w_1 \cdot \left(1 - \frac{\|\hat{P}_t - \mu\|_1}{\max(\hat{P}_t) - \min(\hat{P}_t)}\right) - w_2 \cdot \Delta E_t + w_3 \cdot C_{\mathrm{smooth}}(a_t) \tag{14}$$

The first term measures the balance of plantar pressure distribution, where μ is the ideal pressure distribution mean and $\|\cdot\|_1$ is the L1 norm. The second term represents the change in energy consumption, where the acceleration-integrated power ΔE_t estimated from IMU data reflects gait energy consumption. The estimation of ΔE_t was based on physical model derivation from IMU data. Specifically, after processing the 3D acceleration signal through low-pass filtering, velocity was calculated, and mechanical work (force $\times$ displacement) was derived using body mass parameters. Energy consumption was obtained by integrating the power of the mechanical work over time. The third term is the action smoothness constraint, which avoids drastic fluctuations by calculating the Euclidean distance C_{smooth} between adjacent actions. The weight coefficients $w_1, w_2, w_3 \in [0,1]$ were personalized according to the user type (e.g., rehabilitation trainees or healthy runners).

The Actor network $\mu(s|\theta^\mu)$ was constructed to output deterministic action suggestions under the current state, and the Critic network $Q(s,a|\theta^Q)$ was used to evaluate the action value. The network parameters were updated according to the following rules:

$$\theta^\mu \leftarrow \arg\max_\theta \mathbb{E}_{s\sim\mathcal{B}}\left[Q\left(s, \mu(s|\theta)|\theta^Q\right)\right] \tag{15}$$

$$\theta^Q \leftarrow \arg\min_\theta \mathbb{E}_{(s,a)\sim\mathcal{B}}\left[\left(y_t - Q\left(s,a|\theta^Q\right)\right)^2\right] \tag{16}$$

where $\mathcal{B}$ is the experience replay buffer; $y_t = r_t + \gamma Q(s_{t+1}, \mu(s_{t+1}|\theta^\mu)|\theta^Q)$ is the target Q value; and $\gamma \in (0,1)$ is the discount factor. During the training process, Ornstein-Uhlenbeck (OU) noise was applied to enhance exploration ability, and a soft update strategy was used to improve convergence stability. In the inference stage, the system predicted the optimized gait parameters for the next moment based on real-time collected pressure distribution and kinematic data and guided the user to gradually approach the optimal strategy through the vibration feedback module.

III. EXPERIMENT AND VERIFICATION

EXPERIMENTAL DESIGN

To verify the effectiveness of the multimodal data fusion ViLT framework, this study constructed a multimodal gait dataset for healthy individuals and patients with diabetic foot. A total of 42 subjects were recruited, including 26 healthy individuals (age: 25–60 years) and 16 individuals in a diabetic foot risk group (age: 45–70 years, HbA1c > 7%). The human research involved in this article was approved by the Wenzhou Polytechnic Ethics Committee, and all participants signed a consent form. During the model

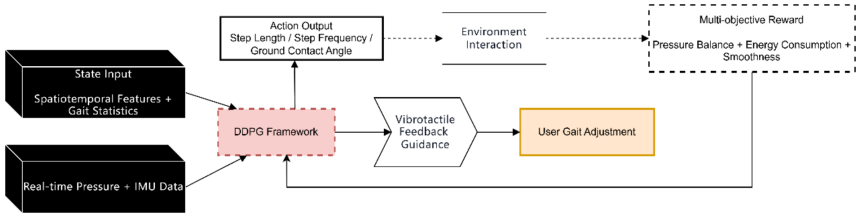


FIGURE 4. Optimization strategy flow

evaluation phase, age, BMI, and foot type (e.g., arch index) were included as covariates in a mixed-effects model to control for baseline differences between the groups. Given the involvement of individuals at risk for diabetic foot complications, additional safeguards were implemented, including: (1) exclusion of participants with active ulcers and (2) real-time monitoring by a certified podiatrist during data collection. A statistical summary of the dataset parameters is shown in Table 2.

TABLE 2. Dataset parameter statistics

Parameter category	Sub-item	Healthy group	Diabetic foot group (HbA1c > 7%)	Scope
Basic Information	Age	38.5 ± 10.2	58.3 ± 7.6	25–70
	Sex(Male/Female)	14/12	9/7	23/19
	BMI(kg/m^2)	23.1 ± 2.8	28.6 ± 3.5	24.9 ± 4.1
Data scale	Total gait cycle count	15,832	9,745	25,577
	Pressure image frame count ($\times 10^3$)	1,892	1,166	3,058
	Duration of IMU signals (h)	21.3	13.1	34.4
Sensor parameters	Effective coverage rate of the pressure channel (%)	98.2 ± 1.5	97.6 ± 2.1	98.0 ± 1.8
	IMU signal-to-noise ratio (dB)	42.7 ± 3.2	40.5 ± 4.0	41.8 ± 3.6
Distribution of sporting tasks	Walking on flat ground (%)	45.2	43.8	44.6
	Going up and down stairs (%)	32.7	34.1	33.3
	Slow running (%)	22.1	22.1	22.1

The experiment used a customized smart insole platform that integrated a distributed pressure sensor array (128 channels, 100 Hz), a three-axis IMU sensor (accelerometer ±16g, gyroscope ±2000 dps, 200 Hz), and an embedded Bluetooth transmission module. During synchronous acquisition, subjects completed three typical motion tasks in a standard gait laboratory: flat-ground walking, ascending and descending stairs, and slow running. Each task lasted for 5 minutes to ensure a variety of plantar mechanical states were covered. During data acquisition, an optical motion capture system was used to obtain biomechanical parameters such as lower limb joint angles, arch height, and gait cycle as reference ground truth values. All sensor data were uniformly sampled to 100 Hz after timestamp alignment, and zero-drift correction and low-pass filtering (cutoff frequency 20 Hz) were performed to eliminate equipment errors and noise. The experimental data was divided into a training set (70%), a validation set (15%), and a test set (15%), split independently by subject to avoid data leakage between individuals. Stratified sampling ensured proportional representation of both healthy and diabetic foot subjects in all sets, maintaining a balance of clinical characteristics.

The model input included three modalities: a visual modality, which was a spatiotemporal sequence of pressure images; a text modality, which comprised gait event labels (touchdown, liftoff, equal support) extracted by the IMU; and a numerical modality, which included continuous biomechanical parameters such as arch height, step length, and step frequency. The text modality was represented by tokenized gait event labels (e.g., [T1: touchdown], [T2: liftoff]). These categorical labels were encoded using a learnable embedding layer with positional encoding to align with ViLT's natural language processing framework. While gait event labels differ from natural language, this paper adapted ViLT's text processing capabilities by treating them as structured temporal tags. Tokenized labels were processed through a learnable embedding layer with specialized temporal position encoding, effectively transforming event sequences into a format compatible with the ViLT architecture. After the three modalities were processed by their respective encoders, they were input into the improved ViLT architecture to achieve cross-modal alignment and fusion. The test results in Table 3 show that our method better preserved the temporal relationship between gait events while maintaining the computational efficiency required for edge deployment.

TABLE 3. Comparison of performance with TabTransformer

Evaluation Metric	Proposed Method	TabTransformer
Pressure Prediction RMSE (%)	3.50	4.10
Gait Phase Classification F1-score	0.92	0.88
Temporal Correlation Coefficient	0.85	0.76

The training strategy adopted a two-stage learning mechanism: in the first stage, the visual encoder weights were fixed, and only the embedding layer and Transformer backbone of the text and numerical modalities were optimized to improve initial convergence efficiency. In the second stage, all parameters were unfrozen, and a cosine annealing scheduler (initial learning rate 1e-4, minimum learning rate 1e-6) was used to jointly optimize the overall network. The loss function combined L2 loss and structural similarity loss to simultaneously ensure the spatial consistency and local

detail accuracy of pressure distribution prediction. During evaluation, a five-fold cross-validation method was used, with fold divisions based on subjects to ensure a balanced distribution of samples in different health states. To simulate real application scenarios, environmental disturbances were further applied, such as changes in ground material (wooden floor, concrete floor, carpet) and differences in wearing tightness. Model parameters were updated through online fine-tuning strategies to evaluate the system's robustness under non-steady-state conditions. Real-time feedback latency was also measured in the actual deployment environment, covering the entire process from data preprocessing to result output, to ensure timeliness requirements for clinical decision-making assistance were met.

To validate its superiority, the proposed method was compared with three state-of-the-art multimodal fusion approaches: (1) ViT-based cross-modal attention, (2) CNN-LSTM late fusion, and (3) a traditional Transformer-based gait recognition model. All models were trained on the same dataset and evaluated using the same metrics (RMSE for pressure prediction, F1-score for gait phase classification).

In Table 4, The proposed method outperformed existing approaches in two aspects: (1) the dynamic attention alignment explicitly modeled the biomechanical correlation between pressure and IMU signals, achieving a 16.7% lower RMSE than ViT; and (2) the spatiotemporal coupling strategy reduced inference latency by 90 ms compared to CNN-LSTM, making it more suitable for real-time smart insole applications. The ViLT framework offers advantages in fusing pressure images, IMU data, and biomechanical parameters due to its cross-modal alignment, lightweight architecture, and dynamic attention mechanism. However, the original ViLT lacks physiological constraints and was optimized for this application through physiologically guided attention masks and spatiotemporal coupling modules to handle the pathological characteristics of patients with diabetic foot.

TABLE 4. Comparative experiment results

Method	Pressure Prediction RMSE (%)	Gait Phase Classification F1-score	Inference Latency (ms)
ViT	4.2	0.89	120
CNN-LSTM	5.7	0.83	180
Transformer	4.8	0.87	140
Proposed Method	3.5	0.92	90

PRESSURE DISTRIBUTION PREDICTION ERROR EVALUATION

This study used a quantitative error analysis framework to systematically verify the pressure distribution prediction performance of the ViLT model. The test set comprised 15% of subject data, independent of the training/validation set, to ensure the sample distribution and clinical characteristics did not overlap. The model output was compared frame-by-frame with the original, synchronously collected pressure images after inverse normalization. The regional root mean square error (RMSE) and mean absolute percentage error (MAPE) were calculated as core evaluation indicators:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\mathrm{pred}} - y_i^{\mathrm{gt}}\right)^2} \tag{17}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{\mathrm{pred}} - y_i^{\mathrm{gt}}}{y_i^{\mathrm{gt}}}\right| \times 100\% \tag{18}$$

where y_i^{gt} is the sensor-measured ground truth value, y_i^{pred} is the model-predicted pressure value, and $N = 128$ is the number of distributed pressure sensor channels. The error calculation covered the entire gait cycle, and subgroup analysis was performed according to plantar anatomical divisions (forefoot, arch, and calcaneal areas) to identify prediction deviations in specific mechanical regions.

To ensure consistency in evaluation, three distinct RMSE metrics were defined.

Pressure Prediction RMSE (Table 3) used sensor-normalized values to evaluate the system's raw pressure estimation accuracy.

Regional RMSE (Figure 5) calculated the error within predefined anatomical regions (forefoot, arch, and heel) and was normalized by region size.

Initial Prediction Error (RMSE) (Table 7) calculated the error across all pixels on the entire plantar surface to assess overall model performance. Due to local error averaging and region-specific normalization, regional RMSE values are consistently lower than the overall initial prediction error.

To visualize the spatial error distribution, a pixel-level error heat map was constructed. The 128-channel pressure data was mapped to a plantar topological grid, and a continuous pressure field was generated using bicubic interpolation. The absolute point-by-point difference between the predicted and true pressure fields was calculated. Quantile normalization (0–95%) was applied to eliminate extreme value interference, and the error intensity was encoded using the HSV color space. A mixed-effects model was used for statistical analysis to evaluate error stability, with the subject as a random effect and the motion task type as a fixed effect. To verify robustness, a ground material perturbation test was conducted by performing online fine-tuning on three surfaces: wooden floor, concrete floor, and carpet.

Figure 5 shows a comparison of pressure distribution prediction errors for the healthy and diabetic foot groups. The RMSE for the healthy group was consistent across the three regions at approximately 0.8%. The RMSE for the diabetic foot group was higher in all regions but remained below 0.9%, indicating a larger prediction error for these patients. Similarly, the MAPE for the healthy group was about 7.0% across all regions, while for the diabetic foot group, it was significantly higher, reaching nearly 8.0% in the forefoot area. Overall, both the RMSE and MAPE for the diabetic foot group were higher than those for the healthy

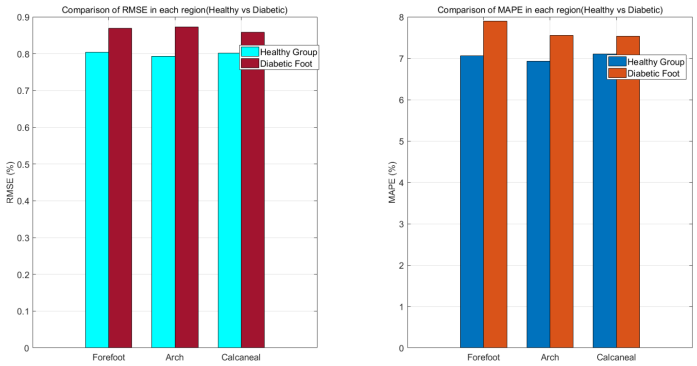


FIGURE 5. Pressure distribution prediction error evaluation

group, suggesting that the model's prediction accuracy was slightly reduced for patients with diabetic foot.

VERIFICATION OF BIOMECHANICAL PARAMETER COUPLING DEGREE

The dynamic correlation between plantar pressure distribution and key biomechanical parameters was quantitatively analyzed to verify the framework's ability to model physiological constraints. The experiment was based on synchronous data from healthy subjects (n=26) and patients with diabetic foot (n=16). The pressure data was divided into three regions: heel, arch, and forefoot. Biomechanical parameters included arch height, gait phase, and ground reaction force vector direction. The Pearson Correlation Coefficient (PCC) was used as the basic coupling evaluation indicator. Additionally, a sliding windowed cross-correlation (WCC) analysis (512 ms window, 128 ms step length) was applied to quantify the time-lagged dynamic coupling between parameters. Table 5 compares the dynamic correlation between plantar pressure and biomechanical parameters for both groups. The data from the healthy group generally showed significant negative or positive correlations (e.g., arch height, r = -0.72***). In contrast, most correlations in the diabetic group were significantly attenuated (e.g., arch height, r = -0.29*), with some even losing significance. The strong negative correlation between arch height and arch pressure in the healthy group (-0.72***) was greatly weakened in the diabetic group (-0.29*), revealing the degradation of multimodal data coupling under pathological conditions. This difference highlights the effectiveness of physiological constraint modeling.

The stress-strain mapping matrix was further established based on the Mooney-Rivlin constitutive equation, and the correlation of the nonlinear mechanical response of plantar soft tissue was calculated for both groups. Mooney-Rivlin parameters (C1, C2) were experimentally measured for each subject group. As shown in Table 6, the correlation coefficient of the stress-strain curve in the forefoot area of the healthy group reached 0.87 ± 0.05 ($p < 0.01$), whereas it dropped to 0.72 ± 0.08 in the diabetic foot group, revealing the degradation of soft tissue mechanical properties under pathological conditions.

TABLE 5. Comparison of plantar pressure-biomechanical parameter coupling between the healthy and diabetic foot groups

Pressure Area	Biomechanical Parameter	Healthy Group (n=26)	Diabetic Group (n=16)
Heel	Height	-0.47***	-0.21
	Gait Phase	0.63***	0.42*
	Ground Reaction Force Direction	-0.58***	-0.33
Arch	Height	-0.72***	-0.29*
	Gait Phase	0.51**	0.34
	Ground Reaction Force Direction	0.43*	0.18
Forefoot	Height	0.38	0.25
	Gait Phase	0.79***	0.57**
	Ground Reaction Force Direction	-0.66***	-0.41*

Note: $p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 6. Comparison of stress-strain curve correlation of plantar soft tissue

Plantar Region	Healthy Group (Correlation Coefficient r)	Diabetic Foot Group (Correlation Coefficient r)	Statistical Significance (p-value)	Mooney-Rivlin Parameters(C1)	Mooney-Rivlin Parameters(C2)
Forefoot	0.87 ± 0.05	0.72 ± 0.08	p <0.01	C1=0.12 kPa	C2=0.08 kPa
Arch	0.82 ± 0.06	0.65 ± 0.10	p < 0.05	C1=0.10 kPa	C2=0.06 kPa
Heel	0.79 ± 0.07	0.68 ± 0.09	P< 0.05	C1=0.09 kPa	C2=0.05 kPa

REAL-TIME FEEDBACK LATENCY TEST

To verify the system's real-time feedback capability, end-to-end latency was measured using hardware-synchronized timestamps embedded in each processing stage. The experiment was based on real data streams from all 42 subjects during continuous walking tasks. End-to-end latency was defined as the time interval from the completion of a single sensor sampling to the generation of a personalized gait optimization suggestion and execution by the vibration feedback module. A box plot of the system's end-to-end latency distribution, constructed from 10,000 effective delay samples, is shown in Figure 6.

Figure 6 shows the real-time feedback latency statistics for both groups. The latency for the healthy group was

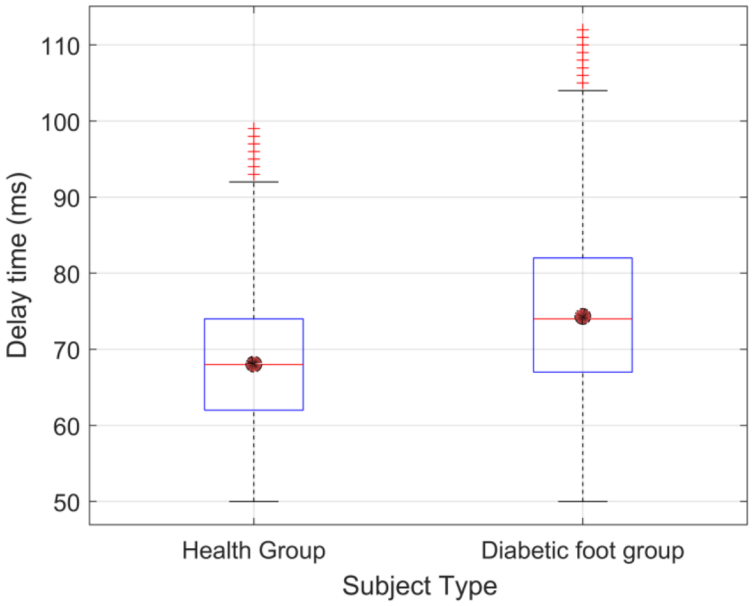


FIGURE 6. Box plot of the system end-to-end delay distribution

concentrated between 60 ms and 80 ms, with an average of about 69 ms and small data fluctuation. The latency for the diabetic foot group was generally higher, concentrated between 65 ms and 85 ms, with an average of about 75 ms and greater fluctuation. This indicates that the more complex gait characteristics of patients with diabetic foot increased the system's processing time.

To evaluate the system's real-time feedback capability in clinical contexts, the 75 ms end-to-end latency was compared with the human sensorimotor loop response time. Human studies indicate that voluntary gait adjustments typically require 150–300 ms, while reflexive adjustments can occur within 50–100 ms. The system's 75 ms latency falls within the reflexive adjustment window, enabling subconscious corrections to gait deviations (such as pressure redistribution) during a single step.

EFFECTIVENESS OF PERSONALIZED GAIT OPTIMIZATION

To verify the personalized gait optimization effect, an experiment with partial blinding (subjects were aware of the intervention, but data analysts were blinded to group allocation) was conducted to compare key gait indicators before and after optimization. The experiment included all 42 subjects, who received dynamic feedback intervention based on the DDPG reinforcement learning strategy. Each subject completed a baseline test followed by an optimization feedback test. During the optimization test, the system provided gait parameter recommendations (step length, frequency, etc.) via vibration feedback. The protocol consisted of 10 training sessions of 20 seconds each to allow for gradual motor learning. The evaluation indicators were the Gait Symmetry Index (GSI) and Peak Pressure Reduction Rate (PPRR). GSI was defined as:

$$\mathrm{GSI} = \frac{|T_{\mathrm{left}} - T_{\mathrm{right}}|}{T_{\mathrm{left}} + T_{\mathrm{right}}} \tag{19}$$

where Tleft and Tright represent the contact time of the left and right soles, respectively. PPRR was defined as:

$$\mathrm{PPRR} = \frac{P_{\mathrm{before}} - P_{\mathrm{after}}}{P_{\mathrm{before}}} \times 100\% \tag{20}$$

where Pbefore and Pafter represent the maximum pressure values of a certain area before and after optimization, respectively. Generalized Estimating Equations (GEE) were used for data analysis to control for inter-individual variation. The optimization effect comparison curve is shown in Figure 7.

Figure 7 shows the changing trends of GSI and PPRR for both groups. The initial GSI was about 11.0% for the healthy group and 21.0% for the diabetic foot group. With more feedback rounds, the GSI of both groups showed a significant downward trend, indicating improved gait symmetry. The PPRR for the forefoot and arch areas of the healthy group rose rapidly before stabilizing at about 14% and 12%, respectively. The PPRR for the diabetic foot group also increased significantly, though the overall increase was lower, finally reaching about 8%. According to clinical research guidelines, a GSI below 10% is generally considered within the normal range for healthy individuals. The post-optimization GSI of 6% in the diabetic foot group falls below this threshold, suggesting that the system's personalized feedback successfully restored gait symmetry to a clinically acceptable level.

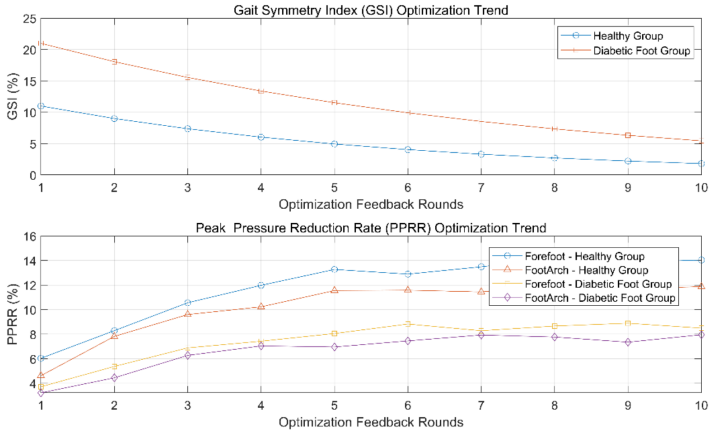


FIGURE 7. Comparison curve of optimization effect

LONG-TERM WEAR STABILITY

To evaluate the system's reliability under long-term wear conditions, this paper conducts a 30-day continuous monitoring experiment on healthy subjects and diabetic foot patients. All subjects wear smart insoles every day to complete a standardized walking task (10 minutes of round-trip walking on flat ground), and pressure distribution, IMU data, and gait parameters are collected to analyze the sensor drift rate and gait prediction accuracy attenuation.

The experiment adopts a blinded data analysis design (independent operator records original data, analysts blinded to group allocation). An independent operator records the original data and uploads it to the cloud database. Another group of researchers performs desensitization and analysis. The sensor stability evaluation is based on the output consistency change of the pressure sensor unit (sampling rate 100Hz, resolution 16bit) and the IMU module (sampling rate 200Hz). The relative deviation between the daily mean and the initial calibration value is calculated:

$$\text{Sensor Drift Rate} = \frac{|X_{\text{day}} - X_{\text{baseline}}|}{X_{\text{baseline}}} \times 100\% \quad (21)$$

Among them, X_{day} represents the sensor output mean, and $X_{baseline}$ represents the calibration value in the initial stage of the experiment. The gait prediction accuracy decay is tested by repeatedly using the ViLT fusion model and LSTM time series modeling structure every day, and the RMSE change between the model output and the ground truth is recorded. The prediction error decay is defined as:

$$\begin{matrix}\text{Prediction Error} \\ \text{Variation Rate}\end{matrix} = \frac{\text{RMSE}_{\text{day}} - \text{RMSE}_{\text{initial}}}{\text{RMSE}_{\text{initial}}} \times 100\% \quad (22)$$

Among them, $RMSE_{initial}$ is the prediction error baseline value after initial training and $RMSE_{day}$ is the error value obtained by daily retesting.To further quantify the systemst abilitythe sliding window variance analysis method isap plied to divide the daily data into multiple 5-minute windows, and the Euclidean distance fluctuation of the feature vector in the ViLT cross-modal space is calculated:

$$\Delta F_{\text{ViLT}} = \frac{1}{N}\sum_{i=1}^{N}\left\| f_i^{\text{day}} - f_i^{\text{initial}} \right\| \quad (23)$$

To address the sensor drift problem, the system integrates an adaptive calibration algorithm based on a reference gait template. The pressure distribution and IMU zero bias data of the static standing state are automatically collected at the daily startup, and the sliding average filter is constructed in combination with historical data to dynamically update the sensor baseline value.

At the model level, an incremental learning strategy is adopted to perform lightweight fine-tuning on the ViLT framework and LSTM network every 7 days. The fine-tuning dataset contains samples from the last 7 days, accounting for 15% of the total training set. By freezing the underlying convolutional layer and only updating the attention weights and LSTM gating parameters, the model crash is avoided, and the generalization ability is maintained. The system deploys an anomaly detection module on the edge device side, uses the Isolation Forest algorithm to identify anomalies in the sensor output, and combines context information to eliminate atypical noise interference. Finally, the inter-individual variation is controlled by generalized estimating equations, and the change trend of each indicator within 30 days is analyzed. The data comparison is shown in Table 7.

Table 7 shows that the sensor drift rate for the healthy group was low (0.51%/day), while that of the diabetic foot group was higher (0.69%/day), consistent with the non-stationary nature of their gait patterns. The higher drift rate in the diabetic foot group is likely due to pathological characteristics such as neuropathy-induced gait instability, changes in skin properties, and local inflammation, which interfere with signal stability. The high linear correlation of the drift trends ($R^2 \geq 0.91$) indicates that the changes are predictable, providing a basis for online calibration. The

TABLE 7. System long-term wear stability data comparison

Metric Category	Healthy Group (n=26)	Diabetic Foot Group (n=16)	Description
Sensor Drift Rate	0.51 %/day	0.69 %/day	Includes pressure
Drift Model Fit R^2	R^2 = 0.93	R^2 = 0.91	All sensor drift trends are linearly predictable
Initial Prediction Error (RMSE)	8.60%	9.10%	Based on ViLT+LSTM architecture
RMSE after 30 Days	9.00%	10.70%	System maintains prediction accuracy above 94.7%
Daily Error Growth Rate	0.13 %/day	0.21 %/day	Reflects the impact of pathology on model stability
Feature Space Shift (ΔF)	1.23%	2.17%	ΔF_{ViLT} reflects cross-modal alignment consistency
Online Calibration Improvement	69.20%	62.10%	Sliding mean filter used to reduce drift impact
Outlier Filtering Rate	92.40%	90.70%	Isolation Forest algorithm identifies and removes noise
Incremental Learning Frequency	Every 7 days (15% data)	Every 7 days (15% data)	Updates attention weights and LSTM gate parameters
Model Generalization Retention	96.50%	94.20%	Validation accuracy drop $\leq$ 3.5% after fine-tuning

results also showed that the model's generalization ability retention rate was 96.5% for the healthy group and 94.2% for the diabetic foot group, with no significant overfitting.

IV. CONCLUSION

This paper proposed a cross-modal fusion method based on an improved ViLT architecture that integrates plantar pressure, IMU, and biomechanical parameters to achieve high-precision pressure distribution prediction and personalized gait optimization. The application of a dynamic attention alignment mechanism and a spatiotemporal coupling modeling strategy effectively improved the model's robustness and generalization ability in complex gait scenarios. Experimental results showed that the method exhibited excellent prediction accuracy and real-time feedback performance in both healthy individuals and patients with diabetic foot and achieved significant improvements in gait symmetry and pressure distribution through a reinforcement learning strategy.

Although the system exhibited strong long-term stability and adaptability, it still has two key limitations: (1) feedback delays may hinder real-time intervention in some clinical settings; and (2) inter-individual differences beyond health status (such as age, weight, and walking speed) may reduce the model's generalization ability. Future work can integrate multimodal physiological signals with adaptive reinforcement learning to dynamically calibrate the model for each user and explore edge-based differential privacy frameworks to securely process heterogeneous data.

AUTHOR CONTRIBUTIONS

Conceptualization – Shiyi SHU; methodology – Shiyi SHU and Fujun SHANGGUAN; investigation – Fujun SHANGGUAN; writing-original draft preparation – Shiyi SHU, Fujun SHANGGUAN. All authors have read and agreed to the published version of the manuscript.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Wenzhou Polytechnic. All participants provided informed consent prior to the study.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

The key project of teaching construction and teaching reform research at Wenzhou Polytechnic in 2025 is "Research and Practice of AI-Empowered Modular Teaching Mode for Footwear and Apparel Specialty in the Context of Digital Transformation" (WZYzd202506).

Zhejiang Provincial Federation of Social Sciences Research Project - Regular Project

Project Title: Research on Building a Cultural Ecology for the Integration of Southern Zhejiang's Blue-dyed Silk and Oushu Shoe Craftsmanship Driven by Digital Twin Technology

Project Number: 2026B084

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Xie, L. Lu, F. Gao, S. J. He, H. J. Zhao, Y. Fang, et al., "Integration of artificial intelligence, blockchain, and wearable technology for chronic disease management: A new paradigm in smart healthcare," Current Medical Science, vol. 41, no. 6, pp. 1123-1133, 2021.

[2] M. T. H. Sarker, I. Ahmed, and M. A. Rahaman, "AI-Based Smart Textile Wearables for Remote Health Surveillance and Critical Emergency Alerts: A Systematic Literature Review," American Journal of Scholarly Research and Innovation, vol. 2, no. 02, pp. 1-29, 2023.

[3] B. Cheng and P. Wu, "Scalable fabrication of kevlar/Ti3C2T x MXene intelligent wearable fabrics with multiple sensory capabilities," ACS nano, vol. 15, no. 5, pp. 8676-8685, 2021.

[4] C. Park, R. Mishra, D. Vigano, M. Macagno, S. Rossotti, K. D'Huyvetter, et al., "Smart offloading boot system for remote patient monitoring: toward adherence reinforcement and proper physical activity prescription for

diabetic foot ulcer patients," Journal of Diabetes Science and Technology, vol. 17, no. 1, pp. 42-51, 2023.

[5] S. C. Wu, G. Y. Bin, W. F. Shi, L. Lin, Y. Xu, D. Zhao, et al., "Empowering diabetic foot ulcer prevention: A novel cloud-based plantar pressure monitoring system for enhanced self-care," Technology and Health Care, vol. 33, no. 2, pp. 701-718, 2025.

[6] K. Xiang, M. Liu, J. Chen, Y. Bao, Z. Wang, K. Xiao, et al., "AI-assisted insole sensing system for multifunctional plantar-healthcare applications," ACS Applied Materials & Interfaces, vol. 16, no. 25, pp. 32662-32678, 2024.

[7] J. Baik, Y. Choi, K. J. Kim, and H. Lee, "Development of Novel Body Center of Pressure Estimation System based on Tactile Sensor and Deep Learning Models," The Journal of Korea Robotics Society, vol. 20, no. 1, pp. 69-76, 2025.

[8] R. Bajpai and D. Joshi, "Foot2hip: A deep neural network model for predicting lower limb kinematics from foot measurements," IEEE/ASME Transactions on Mechatronics, vol. 28, no. 4, pp. 2248-2258, 2023.

[9] L. Yue, L. Zongxing, D. Hui, J. Chao, L. Ziqiang, and L. Zhoujie, "How to achieve human–machine interaction by foot gesture recognition: A review," IEEE Sensors Journal, vol. 23, no. 15, pp. 16515-16528, 2023.

[10] S. Wei, Y. Zhang, and H. Liu, "A multimodal multilevel converged attention network for hand gesture recognition with hybrid sEMG and A-mode ultrasound sensing," IEEE Transactions on Cybernetics, vol. 53, no. 12, pp. 7723-7734, 2022.

[11] J. Zhang, Q. Wang, Q. Wang, and Z. Zheng, "Multimodal fusion framework based on statistical attention and contrastive attention for sign language recognition," IEEE Transactions on Mobile Computing, vol. 23, no. 2, pp. 1431-1443, 2023.

[12] H. Choi, J. P. Yun, B. J. Kim, H. Jang, and S. W. Kim, "Attention-based multimodal image feature fusion module for transmission line detection," IEEE Transactions on Industrial Informatics, vol. 18, no. 11, pp. 7686-7695, 2022.

[13] G. Li, L. Guo, R. Zhang, J. Qian, and S. Gao, "Transgait: Multimodal-based gait recognition with set transformer," Applied Intelligence, vol. 53, no. 2, pp. 1535-1547, 2023.

[14] T. Wang, G. Zhou, Y. Pu, R. Moreno, and G. Yang, "Gait recognition with global–local feature fusion based on swin transformer-3DCNN," Signal, Image and Video Processing, vol. 19, no. 1, pp. 1-9, 2025.

[15] J. Han, D. Wang, Z. Li, N. Dey, R. G. Crespo, and F. Shi, "Plantar pressure image classification employing residual-network model-based conditional generative adversarial networks: A comparison of normal, planus, and talipes equinovarus feet," Soft Computing, vol. 27, no. 3, pp. 1763-1782, 2023.

[16] P. C. B. S. Negi, S. S. Pandey, S. Sharma, and N. Sharma, "Scalogram-Based Gait Abnormalities Classification Using Deep Convolutional Networks for Neurological and Non-Neurological Disorders," Journal of Medical and Biological Engineering, vol. 44, no. 3, pp. 375-389, 2024.

[17] S. Jiao, H. Huo, W. Liu, C. Zhao, L. Ma, J. Wang, et al., "Gaitformer: A spatial-temporal attention-enhanced network without softmax for Parkinson's disease early detection," Complex & Intelligent Systems, vol. 11, no. 5, pp. 1-17, 2025.

[18] S. U. Yunas and K. B. Ozanyan, "Gait activity classification using multimodality sensor fusion: A deep learning approach," IEEE Sensors Journal, vol. 21, no. 15, pp. 16870-16879, 2021.

[19] B. Ali, N. Faramarzi, U. Farooq, H. C. Bidsorkhi, A. G. D'Aloia, A. Tamburrano, et al., "Graphene-based smart insole sensor for pedobarometry and gait analysis," IEEE Sensors Letters, vol. 7, no. 5, pp. 1-4, 2023.

[20] Z. Q. Ling, Y. P. Zhang, G. Z. Cao, J. C. Chen, L. L. Li, D. P. Tan, et al., "AE-CNN-based multisource data fusion for gait motion step length estimation," IEEE Sensors Journal, vol. 22, no. 21, pp. 20805-20815, 2022.

[21] V. Chhoeum, Y. Kim, and S. D. Min, "A convolution neural network approach to access knee joint angle using foot pressure mapping images: A preliminary investigation," IEEE Sensors Journal, vol. 21, no. 15, pp. 16937-16944, 2021.

[22] A. Sharifi, M. Ahmadi, M. A. Mehni, S. Jafarzadeh Ghoushchi, and Y. Pourasad, "Experimental and numerical diagnosis of fatigue foot using convolutional neural network," Computer Methods in Biomechanics and Biomedical Engineering, vol. 24, no. 16, pp. 1828-1840, 2021.

[23] W. Shao, Z. You, L. Liang, X. Hu, C. Li, W. Wang, et al., "A multimodal gait analysis-based detection system of the risk of depression," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 10, pp. 4859-4868, 2021.

[24] H. Wu, Y. Liu, H. Yang, Z. Xie, X. Chen, M. Wen, et al., "An Attention-based Temporal Network for Parkinson's Disease Severity Rating using Gait Signals," KSII Transactions on Internet and Information Systems (TIIS), vol. 17, no. 10, pp. 2627-2642, 2023.

[25] K. Hu, Z. Wang, K. A. E. Martens, M. Hagenbuchner, M. Bennamoun, A. C. Tsoi, et al., "Graph fusion network-based multimodal learning for freezing of gait detection," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 3, pp. 1588-1600, 2021.

[26] L. Xue and Z. Tao, "GaitRA: Triple-branch multimodal gait recognition with larger effective receptive fields and mixed attention," Multimedia Tools and Applications, vol. 83, no. 33, pp. 80225-80259, 2024.

[27] E. S. Rahayu, E. M. Yuniarno, I. K. E. Purnama, and M. H. Purnomo, "Adaptive Threshold Filtering to Reduce Noise in Elderly Activity Classification Using Bi-LSTM," Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, vol. 13, no. 1, pp. 48-57, 2024.

[28] L. D'arco, G. Mccalmont, H. Wang, and H. Zheng, "Application of smart insoles for recognition of activities of daily living: A systematic review," ACM Transactions on Computing for Healthcare, vol. 5, no. 1, pp. 1-34, 2024.

[29] P. Zhang, X. Zhang, M. Teng, L. Li, X. Liu, J. Feng, et al., "Leather-Based Shoe Soles for Real-Time Gait Recognition and Automatic Remote Assistance Using Machine Learning," ACS Applied Materials & Interfaces, vol. 16, no. 45, pp. 62803-62816, 2024.

[30] C. Huang, F. Nihey, K. Fukushi, H. Kajitani, Y. Nozaki, K. Ihara, et al., "Feature selection, construction, and validation of a lightweight model for foot function assessment during gait with in-shoe motion sensors," IEEE Sensors Journal, vol. 23, no. 8, pp. 8839-8855, 2023.

[31] E. Wang, X. Chen, Y. Li, Z. Fu, and J. Huang, "Lower limb motion intent recognition based on sensor fusion and fuzzy multitask learning," IEEE Transactions on Fuzzy Systems, vol. 32, no. 5, pp. 2903-2914, 2024.

[32] B. Yang, X. Zhang, J. Zhang, Z. Huang, Q. Lu, J. Zhang, et al., "KneeGuard: A Calibration-free Wearable Monitoring System for Knee Osteoarthritis Gait Re-training via Effortless Wearing," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 8, no. 4, pp. 1-29, 2024.