*Sensors and Materials*, Vol. 38, No. 9 (2026) 5121–5136
MYU Tokyo

S & M 4613



# Analysis of Effects of Incorporating Long-term Data into the Training Set for Gait Authentication Using Insole-type Gait Sensor

Etsushi Kumamoto,[1] Kyosuke Fukamachi,[1] Tomohito Yamamoto,[1] Susumu Sato,[2] and Takashi Kawanami[1*]

[1]Graduate Program in Information and Computer Engineering, Graduate School of Engineering, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan
[2]Humanities and Social Sciences Program, Academic Foundations Programs, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan

(Received April 1, 2026; accepted July 10, 2026)

***Keywords:*** gait authentication, insole sensor, long-term data, deep learning, continuous authentication

Gait authentication using wearable sensors has attracted attention as a promising approach for continuous authentication. However, in many existing studies, authentication models were trained only on short-term gait data collected under controlled experimental conditions, which may not reflect the diversity of gait patterns in daily life. In this study, we investigated the effect of incorporating long-term gait data collected in daily-life environments into the training dataset of a deep-learning-based gait authentication model using an insole-type gait sensor. Gait data were collected from 48 healthy participants using two measurement modes: a real-time mode for short-term data and a daily-life mode for long-term data. Two training conditions were compared: using only short-term data and using both short-term and long-term data. Authentication performance was evaluated for unseen users using the equal error rate (*EER*). Incorporating long-term data tended to improve the performance, reducing the average *EER* from 0.47 to 0.17% and the common-threshold *EER* from 0.71 to 0.28%. In addition, feature-space analysis using the Silhouette Score and Delta Cosine Similarity (Δ*CosSim*) showed improved separability of test embeddings, and SHapley Additive exPlanations (SHAP) based analysis revealed reduced dependence on roll-angle-related parameters and increased use of toe-in/out angle and minimum toe clearance.

## 1. Introduction

Biometric authentication is a technology for identity verification using physical traits (e.g., fingerprints, iris patterns, face, and DNA) or behavioral traits (e.g., gait, handwriting, and speaking patterns). Although biometric authentication based on physical traits is widely adopted and achieves high accuracy, it cannot prevent unauthorized use by a third party who gains access after the legitimate user has been authenticated. Therefore, the need for continuous

*Corresponding author: e-mail: t-kawanami@neptune.kanazawa-it.ac.jp
https://doi.org/10.18494/SAM6360

ISSN 0914-4935 
https://myukk.org/

authentication, where identity verification is continued over extended periods, has been recognized.(1)

Behavioral biometrics has attracted attention as an effective approach for realizing continuous authentication.(2) Among various behavioral traits, gait is known to possess individually distinguishable characteristics,(3) and because walking is a natural and routine activity, authentication can be performed without requiring conscious user actions. Owing to these properties, gait authentication has been regarded as a method well suited to continuous authentication,(4) and numerous studies have been conducted on this topic.(5–8)

Gait authentication methods can be broadly categorized into camera-based approaches(5,6) and wearable-sensor-based approaches.(7,8) Camera-based methods enable contactless identification within the camera's field of view; however, they are limited to specific areas and thus face challenges in continuous authentication applications. In contrast, wearable-sensor-based methods directly acquire gait data from sensors worn by individuals, imposing no spatial constraints and thus is more suitable for continuous authentication. Although existing studies using wearable sensors have reported high authentication accuracy, they share common challenges regarding data collection conditions. For example, Lee *et al.*(7) collected gait data from 20 participants over three days using nine-axis inertial measurement units (IMUs) attached to the wrist and thigh and achieved an authentication accuracy of 99.63% using a support vector machine. Moon *et al.*(8) collected approximately three minutes of gait data from 30 participants using a commercially available insole-type multimodal sensor (FootLogger) and achieved a personal identification accuracy exceeding 99% using a convolutional neural network–recurrent neural network (CNN-RNN) model. However, in most existing studies, gait authentication models were trained using only data collected over short periods under experimental conditions (hereafter, short-term data), and, to the best of our knowledge, few studies have effectively utilized data collected over extended periods in daily life (hereafter, long-term data). Furthermore, significant differences in gait parameters such as walking speed and stance phase time between experimental and daily living environments have been reported,(9,10) indicating that short-term data collected only under experimental conditions may not sufficiently capture the diverse walking patterns encountered in daily life. Therefore, incorporating long-term data collected during daily activities into training datasets may be effective in improving the generalization performance of gait authentication models.

In this study, we used a gait analysis sensor (hereafter, gait sensor) developed by NEC Corporation, which enables continuous use for over one year in daily life, to collect long-term data. We then investigated the effect of adding this data to the training dataset on authentication accuracy. Furthermore, we analyzed the contributing factors from the perspectives of feature vector separability and the importance of gait parameters in authentication.

## 2. Data, Materials, and Methods

### 2.1 Insole-type gait sensor

Figure 1 shows the gait sensor used in this study. This sensor is a commercially available product that facilitates the collection of gait data over extended periods. It is attached to a recess

Fig. 1. (Color online) Insole-type gait sensor.

on the bottom surface of an insole. When a user wears shoes with the insole inserted and begins walking, the gait parameters described below are measured. The measured gait parameters are transmitted to and stored in the cloud via a dedicated smartphone application.

The gait sensor measures 24 discrete gait parameters, as listed in Table 1. In this study, 23 gait parameters, excluding the frail level, which is not numerical data, were used as gait data. These gait parameters have been reported to correlate with gait parameters measured by motion capture systems,[11,12] and have also been used in medical research.[9,13]

Three types of gait sensor exist, differing in measurement frequency and timing. In this study, two types were used: the real-time measurement version and the daily measurement version.

The real-time measurement version is manually operated through the dedicated application and can acquire gait data for each step. Because it enables high-frequency data collection under controlled conditions, it is suitable for use in experimental settings.

The daily measurement version automatically detects walking activity and initiates measurement, acquiring gait data up to three times per day. Owing to its low power consumption, it does not require battery replacement and can be used continuously for more than one year, making it suitable for long-term measurement in daily living environments.

### 2.2 Collection of gait data

Gait data based on the gait parameters were collected from 48 healthy participants (both male and female). Of these, 40 participants used the real-time measurement version, and eight used the daily measurement version. The experiments involving human participants in this study were conducted with the approval of the Research Ethics Review Committee for Research Involving Human Subjects at Kanazawa Institute of Technology (Approval number: 2502011).

Among the participants who used the real-time measurement version, 31 were male and nine were female, with ages ranging from 14 to 75 years [mean: 33.5 years, (SD): 18.10 years]. Each participant attached the gait sensor to their own shoes and walked back and forth along a

Table 1
Gait parameters measurable by gait sensors.

| Description (Unit) | |
|---|---|
| Walking speed (km/h) | Maximum speed during the swing phase (m/s) |
| Stride length (cm) | Cadence (step/min) |
| Foot height (cm) | Peak swing angular velocity (deg/s) |
| Peak foot-sole angle in the dorsiflexion direction (deg) | Roll angle at heel contact (deg) |
| Peak foot-sole angle in the plantarflexion direction (deg) | Roll angle at toe-off (deg) |
| Stance phase time (s) | Toe-in/out angle (deg) |
| Loading response time (s) | Circumduction (cm) |
| Foot-flat time (s) | Foot clearance (cm) |
| Pushing time (s) | Minimum toe clearance (cm) |
| Swing phase time (s) | Center of pressure exclusion index (%) |
| Double support time 1 (s) | Hallux angle (deg) |
| Double support time 2 (s) | Frail level (-) |

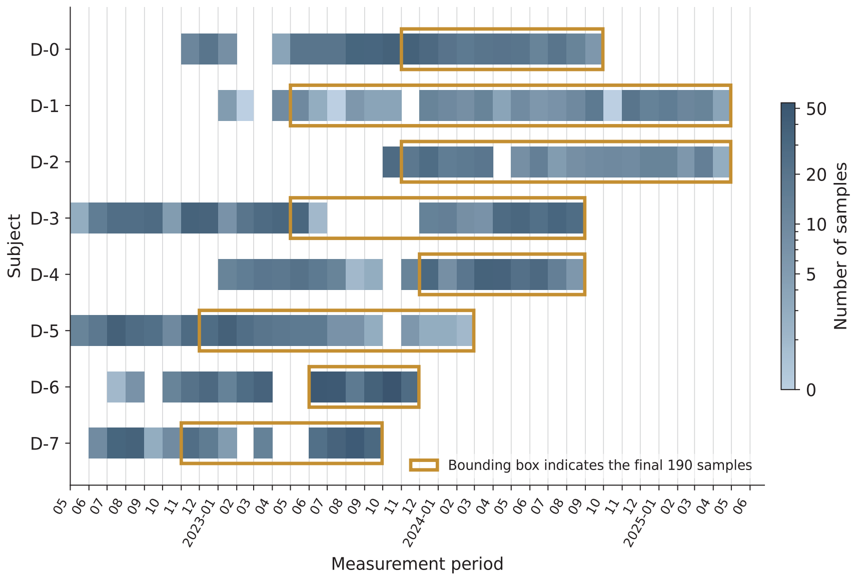
corridor inside a building at a self-selected walking speed. The walking distance varied among participants, and the measurement duration ranged from approximately 1 to 5 min depending on each participant's condition.

Among the participants who used the daily measurement version, seven were male and one was female, with ages ranging from 21 to 89 years (mean: 40.4 years, SD: 19.43 years). Each participant attached the gait sensor to their own shoes, and gait data were automatically measured during natural walking in daily life. Figure 2 presents the measurement periods and time-of-day distribution of the daily-life gait data. The time spans covered by the 190 most recent samples selected for model training ranged from 5.0 to 23.8 months across participants (mean: 12.2 months), with seven of the eight participants exceeding six months, as indicated by the bounding boxes in Fig. 2(a). The selected samples were recorded between 6:00 and 21:00 [Fig. 2(b)]. Thus, the selected data captured daytime gait patterns across multiple seasons.

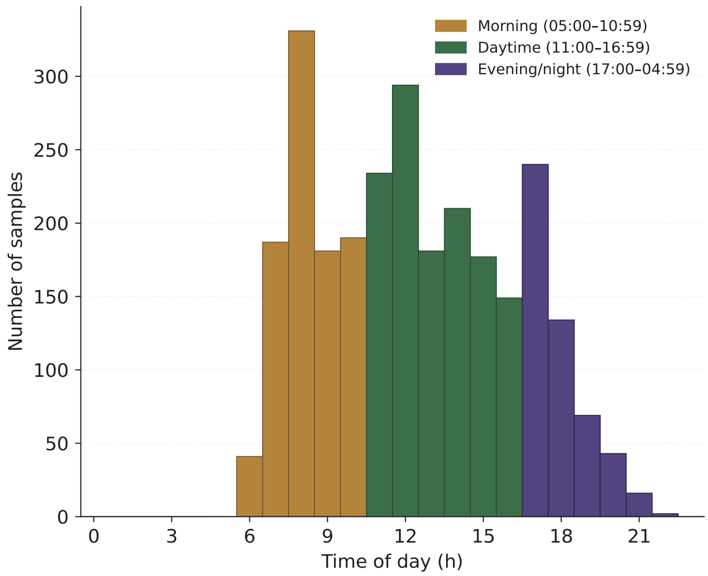
The collected gait data were classified into four groups (A–D) on the basis of the sensor type (real-time measurement version or daily measurement version) and the number of data samples per participant. Groups A–C correspond to the real-time measurement version, and Group D corresponds to the daily measurement version. The number of data samples was unified within each group: A: 20 participants × 190 samples, B: 10 participants × 20 samples, C: 10 participants × 70 samples, and D: eight participants × 190 samples. In addition, for Group D, the samples for each participant were arranged in chronological order, and the 190 most recent samples were selected [indicated by the bounding box in Fig. 2(a)].

### 2.3 Gait authentication model

Figure 3 shows the architecture of the gait authentication model. The model is a deep learning model that takes 46-dimensional bilateral combined data as input, which is formed by concatenating the gait parameters (23 dimensions) measured from consecutive left and right gait cycles, and learns highly discriminative feature representations through a classification task using ArcFace.[14] The intermediate layers consist of two fully connected layers with 256 and




(a)




(b)

Fig. 2. (Color online) Distribution of daily-life gait data measurements and the time spans containing the 190 most recent samples selected for model training. (a) Measurement period distribution. (b) Time-of-day distribution of the selected 190 samples.

512 nodes, respectively, followed by a 128-dimensional feature vector output, which is L2-normalized before being fed into the ArcFace layer. These hyperparameters were determined using Optuna[15] based on the authentication accuracy on the validation dataset. ArcFace is a metric learning method that uses cosine similarity (hereafter, *CosSim*) as a similarity measure while adding an angular margin between classes to learn feature representations where the

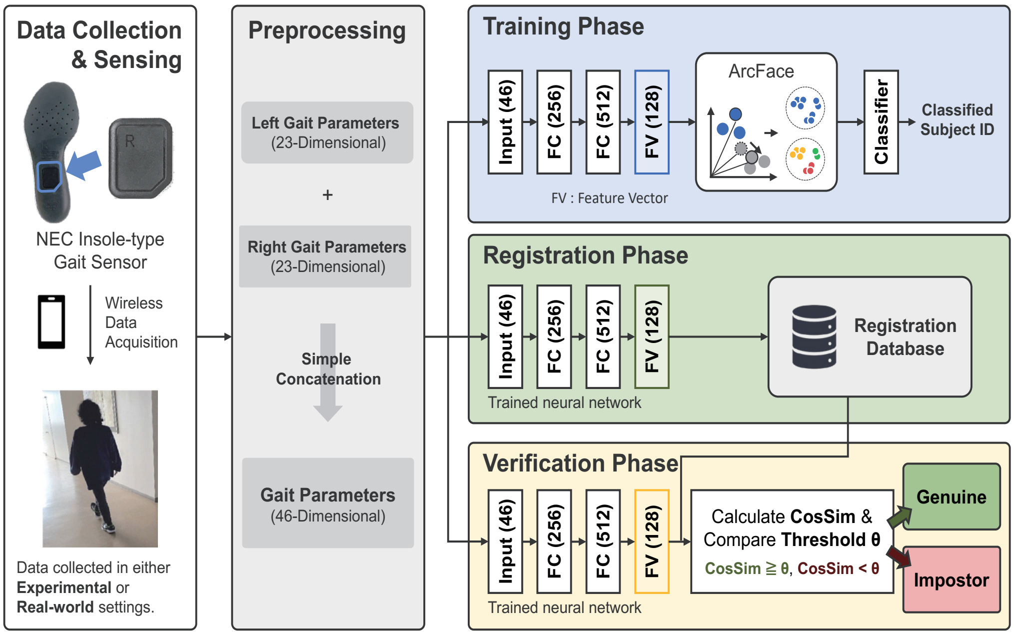



Fig. 3. (Color online) Personal authentication model structure and procedure.

distance between same-class feature vectors is small and the distance between different-class feature vectors is large.

The model operates in three phases—training, enrollment, and authentication—all using the same network but with different outputs. In the training phase, shown in the upper part of Fig. 3, the model is trained as an N-class classification task corresponding to N participants in the training dataset. This enables the model to learn feature representations that can distinguish the gait characteristics of each individual. In the enrollment phase, shown in the middle part of Fig. 3, multiple bilateral combined data samples of the target user (the genuine user) are fed into the model, and the mean of the resulting L2-normalized 128-dimensional feature vectors is computed and stored as the enrollment data. In the authentication phase, shown in the lower part of Fig. 3, a 128-dimensional feature vector is extracted from the input data in the same manner, and *CosSim* with the enrollment data is computed. The result is then compared with a predefined threshold to determine whether the input belongs to the genuine user. By using feature vectors in this manner, even unknown users who were not included in the training dataset can be immediately added as authentication targets simply by enrolling their data.

## 2.4 Experimental methodology

To investigate the effect of adding long-term data to the training dataset on authentication accuracy, two dataset configurations, shown in Fig. 4, were prepared for a comparative experiment. Under both conditions, the number of participants in the training data was unified to 20 to compare the effect of the presence or absence of long-term data.

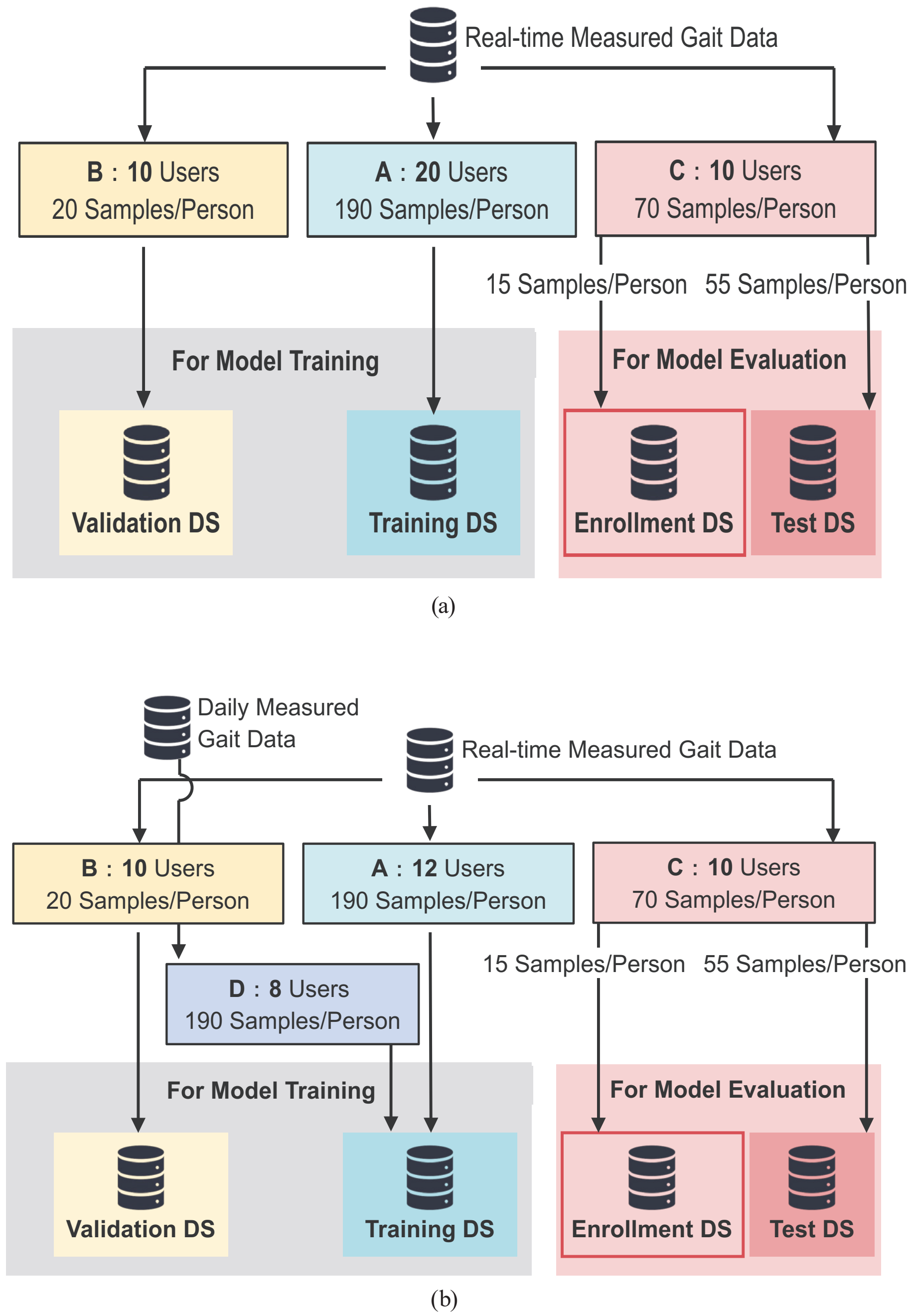



Fig. 4. (Color online) Dataset (DS) configurations for the comparative experiment. (a) Short-term data only and (b) short-term data combined with long-term data.

In Fig. 4(a), only short-term data collected with the real-time measurement sensor (20 participants from Group A) were used as the training data. In Fig. 4(b), to match the number of training participants with that in Fig. 4(a) while adding eight participants from Group D, the training data consisted of short-term data from 12 randomly selected participants from Group A combined with long-term data from the eight participants in Group D, totaling 20 participants.

Under both conditions, the 10 participants from Group B (20 samples each) were used as the validation dataset. For evaluation, unknown individuals not included in the model training (10 participants from Group C) were used as enrolled users (genuine users). For each participant in Group C, the 70 samples were split chronologically into 15 samples for enrollment and 55 samples for testing.

### 2.5 Evaluation methods

#### 2.5.1 Authentication accuracy evaluation

The equal error rate (*EER*) was used as the evaluation metric. The *EER* is the error rate at the point on the receiver operating characteristic (ROC) curve where the false acceptance rate (*FAR*) equals the false rejection rate (*FRR*). A lower value indicates higher personal authentication accuracy.

In the evaluation, for each participant in the evaluation dataset, the data of the enrolled genuine user were treated as legitimate data, and the data of other participants were treated as impostor data (data simulating unauthorized access attempts), from which the optimal threshold and *EER* were computed. Additionally, the *EER* obtained by applying the same threshold to all participants (*EER* with a common threshold) was also computed. Furthermore, participant-wise *EER*s under the two training conditions were compared using a two-sided Wilcoxon signed-rank test for paired observations.

#### 2.5.2 Feature vector separability evaluation

To quantitatively evaluate the changes in feature vectors resulting from the addition of long-term data to the training dataset, the Silhouette Score(16) and Δ*CosSim* were computed.

The Silhouette Score ranges from −1 to 1, with higher values indicating that each data point is more densely clustered within its own class and better separated from other classes.

Δ*CosSim* is a metric defined in this paper as the difference between the *CosSim* among feature vectors within the same class (intraclass *CosSim*) and the *CosSim* between feature vectors of different classes (interclass *CosSim*). A larger value indicates greater discriminability between classes while preserving within-class similarity.

To visually complement these quantitative metrics, t-distributed Stochastic Neighbor Embedding (t-SNE) was also applied to both the input gait parameter space and the 128-dimensional model output feature vectors under each training condition. t-SNE was implemented using scikit-learn (version 1.3.2) with perplexity = 30 and random_state = 42, and the same parameters were applied across all conditions.

#### 2.5.3 Gait parameter importance evaluation

To investigate how the model utilizes each gait parameter after the addition of long-term data to the training set, an analysis using SHapley Additive exPlanations (SHAP)(17) was conducted. Because the quantity to be explained is a similarity score rather than a class probability, the model-agnostic Kernel SHAP method (shap.KernelExplainer) was used. In this study, SHAP was applied to the authentication score, i.e., *CosSim* between the feature vector of the test data and the feature vector of the enrollment data. The enrollment feature vector of the genuine user was held fixed, and only the input of the test sample was perturbed. For the background (baseline) distribution, 100 samples randomly drawn from the corresponding training dataset for each condition were used. For left and right parameters with the same name, the mean of the SHAP values was computed. Furthermore, within each model, the absolute SHAP values of all parameters were normalized so that their sum equaled 1, enabling comparison as relative importance.

## 3. Results

### 3.1 Authentication accuracy

Table 2 shows the *EER* for each participant, the mean *EER*, and the *EER* with a common threshold under each condition. Under the condition with only short-term data, eight out of 10 participants achieved an *EER* below 1.00%, of which six achieved an *EER* of 0.00%. However, participants C-8 and C-9 showed high values of 1.11 and 2.00%, respectively. The mean *EER* was 0.47%, and the *EER* with a common threshold was 0.71%. In contrast, under the condition in which long-term data were added to the training dataset, all 10 participants achieved an *EER* below 1.00%, with seven recording an *EER* of 0.00%. In particular, participants C-8 and C-9, who had high *EER* values under the short-term-data-only condition, showed substantial improvements to 0.00 and 0.44%, respectively. The mean *EER* was 0.17%, and the *EER* with a common threshold was 0.28%, both of which were improved compared with those under the short-term-data-only condition. A two-sided Wilcoxon signed-rank test was conducted to compare the participant-wise *EER*s between the two training conditions. No statistically significant difference was observed between the short-term-data-only condition and the short-term-plus-long-term-data condition ($W = 2$, $p = 0.188$). Five of the ten participants had identical *EER*s under the two conditions, indicating that the statistical power of the comparison was limited.

Table 2
Authentication performance (*EER*) under different training conditions.

| Subject | *EER* (%) (short-term only) | *EER* (%) (short-term + long-term) |
|---|---|---|
| C-0 | 0.00 | 0.00 |
| C-1 | 0.00 | 0.46 |
| C-2 | 0.89 | 0.77 |
| C-3 | 0.00 | 0.00 |
| C-4 | 0.00 | 0.00 |
| C-5 | 0.67 | 0.00 |
| C-6 | 0.00 | 0.00 |
| C-7 | 0.00 | 0.00 |
| C-8 | 1.11 | 0.00 |
| C-9 | 2.00 | 0.44 |
| Mean *EER* (*SD*) | 0.47 (0.69) | 0.17 (0.28) |
| *EER* with a common threshold | 0.71 | 0.28 |

### 3.2 Feature vector separability

Table 3 shows the results of the above metrics under each condition. Regarding the training dataset, the Silhouette Score under the condition with long-term data was 0.2760, which was lower than the value of 0.3610 under the short-term-data-only condition. In contrast, regarding the feature vectors of the model output for the test data, the Silhouette Score under the condition with long-term data was 0.6420, which was higher than the value of 0.5950 under the short-term-data-only condition. Furthermore, $\Delta CosSim$ was 0.2198 under the short-term-data-only condition, compared with 0.3383 under the condition with long-term data.

### 3.3 Gait parameter importance based on SHAP

Figure 5 shows the relative importance of all 23 gait parameters in the model trained with only short-term data and the model with the addition of long-term data.

In the model trained with only short-term data, two parameters—roll angle at toe-off (25.6%) and roll angle at heel contact (16.8%)—accounted for 42.4% of the total importance. In contrast, in the model with the addition of long-term data, these values decreased to 21.6 and 12.9%, respectively, with a combined total of 34.5% (a decrease of 7.9 percentage points). Meanwhile, the importance of toe-in/out angle increased from 7.2 to 11.7%, and that of minimum toe clearance increased from 7.4 to 9.8%.

For participant C-8, whose accuracy improved substantially, the importance of roll angle at heel contact decreased from 20.8 to 12.5%, while that of minimum toe clearance increased from 7.2 to 15.5%. For participant C-9, the importance of roll angle at heel contact decreased from 19.9 to 9.1%, while that of toe-in/out angle increased from 16.5 to 29.4%. On the other hand, for participant C-1, the importance of circumduction increased from 9.0 to 25.1%, indicating a change in feature utilization; however, the *EER* slightly worsened from 0.00 to 0.46%.

## 4. Discussion

The objective of this study was to elucidate the effect of utilizing long-term data on gait authentication accuracy. The experimental results suggested that incorporating long-term data into the training dataset is promising for improving accuracy for participants who were difficult to distinguish using only short-term data. In this section, we discuss the factors associated with

Table 3
Feature vector separability metrics under different conditions.

| | Training Data | | Feature Vectors | |
|---|---|---|---|---|
| | Short-term only | Short-term + Long-term | Short-term only | Short-term + Long-term |
| Silhouette Score | 0.3610 | 0.2760 | 0.5950 | 0.6420 |
| Intraclass *CosSim* | 0.9997 | 0.9997 | 0.9299 | 0.8574 |
| Interclass *CosSim* | 0.9960 | 0.9970 | 0.7101 | 0.5191 |
| $\Delta CosSim$ | 0.0037 | 0.0027 | 0.2198 | 0.3383 |

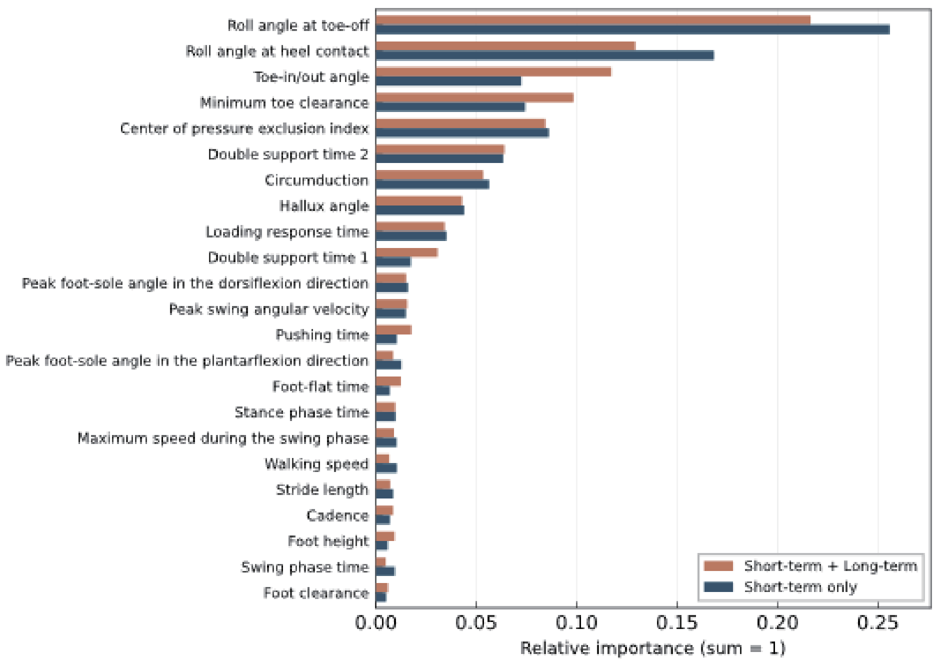



Fig. 5. (Color online) Relative importance of all 23 gait parameters based on SHAP.

the observed reduction in *EER* from two perspectives: structural changes in the feature space and changes in gait parameter importance.

### 4.1 Structural changes in feature space

Figure 6 shows the t-SNE visualizations of the gait parameter input space and the model output feature vectors under each training condition. Figures 5(a) and 5(b) show the input gait parameter space of the training data under the short-term-only and short-term+long-term conditions, respectively, Fig. 5(c) shows the input gait parameter space of the test data, and Figs. 5(d) and 5(e) show the 128-dimensional model output feature vector space of the test data under the corresponding conditions.

The results in Sect. 3.2 indicate that, under the condition incorporating long-term data, the class structure of the training dataset was more complex, whereas the Silhouette Score of the feature vectors output by the model for the test data improved and $\Delta CosSim$ increased. Comparison of Figs. 5(a) and 5(b) reveals that the class clusters in the training data are more dispersed and show greater inter-class overlap when long-term data are included. This broader distribution is consistent with the temporal variability contained in long-term daily-life data, although differences in participant composition may also have contributed to the observed class structure. Even for the same participant, gait patterns may vary over time owing to factors such as differences in walking environment, physical condition, and walking intent. Therefore, the broader class distributions observed in the training-data input space under the condition

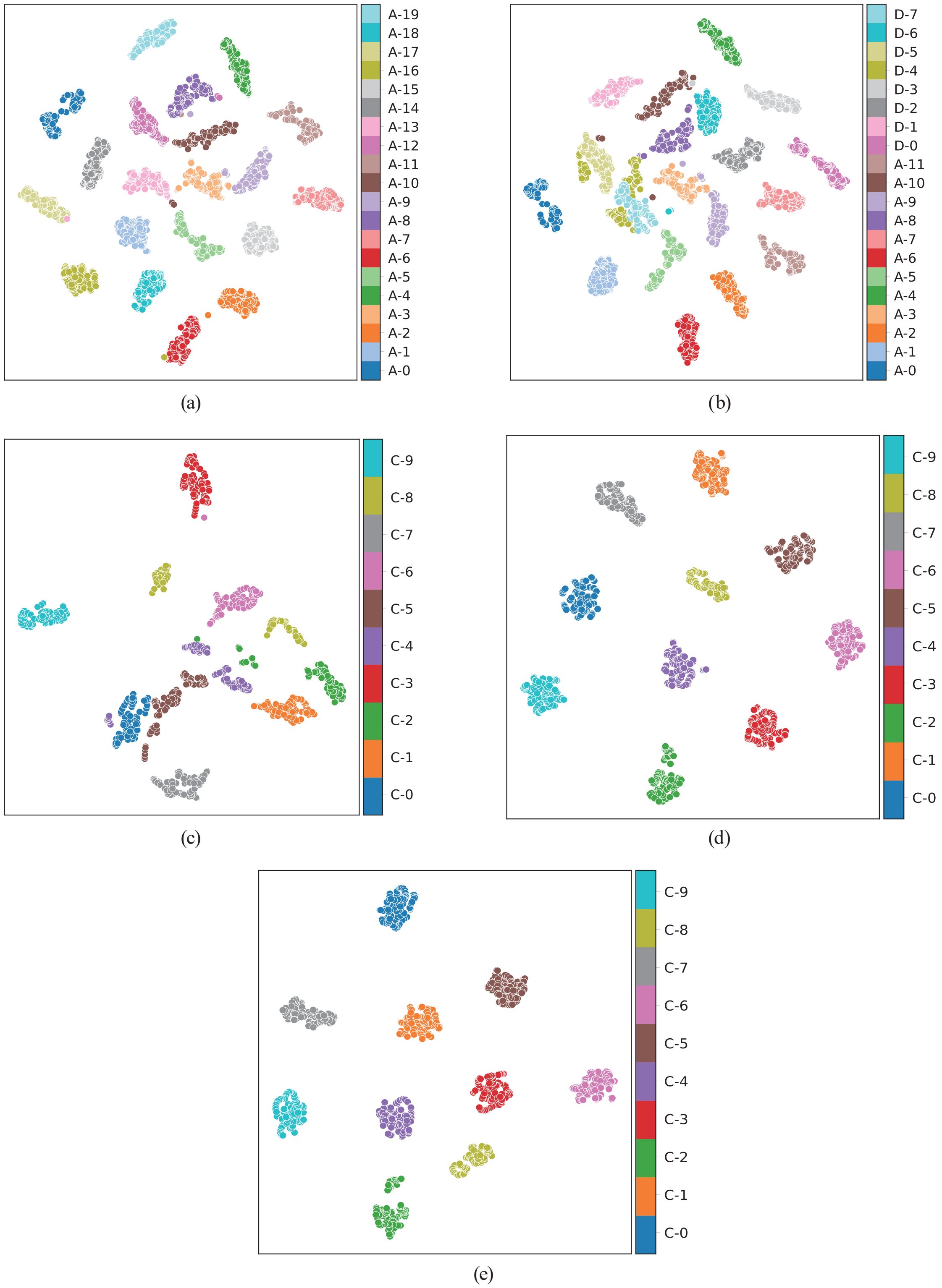



Fig. 6. (Color online) t-SNE visualizations of gait parameter input space and model output feature vector space. (a) Training data input space (short-term only). (b) Training data input space (short-term + long-term). (c) Test data input space. (d) Test data feature vector space (short-term only model). (e) Test data feature vector space (short-term + long-term model).

incorporating long-term data may reflect both within-individual temporal variability and differences in participant composition.

This suggests that the introduction of long-term data may have contributed to the model acquiring feature representations with clearer class separation, consistent with a possible enhancement of the model's generalization capability. Comparison of Figs. 5(d) and 5(e) reveals that the class clusters in the test data feature space show a slight improvement in class separation when the model is trained with long-term data, consistent with the quantitative improvements in Table 3. These findings suggest that the introduction of long-term data may have contributed to the model acquiring feature representations with greater generalization capability. These findings are also quantitatively consistent with the improvement in the Silhouette Score from 0.5950 to 0.6420 and the increase in $\Delta CosSim$ from 0.2198 to 0.3383 (Table 3).

### 4.2 Changes in gait parameter importance

The results in Sect. 3.3 suggest that in the model trained with the addition of long-term data, the dependence on a small number of specific parameters was reduced, and the model may have performed identification by combining a more diverse set of gait parameters.

This interpretation is consistent with findings from prior studies. Shimizu *et al.*[9] used the same insole-type sensor as in this study and reported that gait parameters related to foot angles, such as roll angle at heel contact and toe-in/out angle, exhibited significant differences between experimental and daily living environments. Renggli *et al.*[10] demonstrated that the toe-out angle increased significantly in real-world environments compared with experimental settings when using a wearable IMU. These findings indicate that features related to foot angles are susceptible to the effect of measurement environments, supporting the results of this study, in which the dependence on roll angle parameters decreased and the importance of toe-in/out angle increased after the introduction of long-term data.

On the other hand, cases were also observed, such as participant C-1, where changes in feature utilization were accompanied by a slight worsening of *EER*. This suggests that the effect of introducing long-term data varies among individuals, and the diversification of feature utilization does not always directly lead to performance improvement.

### 4.3 Limitations of this study

In this study, the number of test participants was limited to 10, and because the change in participant composition and the introduction of long-term data in the training dataset occurred simultaneously, the effect of long-term data could not be completely isolated and verified. Consequently, the reduction in *EER* observed in this study should be interpreted as reflecting the combined effects of long-term data and of differences in participant composition, and the magnitude of the effect attributable to long-term data alone cannot be quantitatively determined.

Furthermore, the SHAP analysis used in this study explains the degree to which each gait parameter contributed to the computation of the authentication score, but it does not demonstrate that the observed changes in feature utilization directly caused the improvement in authentication performance.

In addition, the statistical comparison of the participant-wise *EER*s was limited by the small number of test participants and by the fact that five participants had identical *EER*s under the two training conditions.

Therefore, additional experiments with controlled participant numbers and compositions in the training dataset are needed to rigorously elucidate the impact of long-term data introduction on changes in feature utilization and improvements in authentication accuracy. Specifically, both short-term data (obtained with the real-time measurement version) and long-term data (obtained with the daily measurement version) should be collected from the same participants. Comparing two conditions that differ only in the type of training data—short-term versus long-term—while keeping the participant identity, the number of participants, and the number of samples identical would make it possible to evaluate the effect of incorporating long-term data in isolation from the effect of participant composition. However, such verification requires collecting data over an extended period from the same participants, and because the collection of long-term data requires measurement periods ranging from several months to over a year per participant, it imposes temporal constraints on the construction of large-scale datasets.

In the future, verification involving a larger number of participants is necessary to confirm the generalizability of the findings obtained in this study.

The system proposed in this study is intended for application to continuous authentication by incorporating the diversity of gait patterns encountered in daily life into the training data. However, the operation of the system in cases where gait changes abruptly due to sudden events such as accidents or acute illness has not been explicitly addressed in this study. In the event of such a "gait-altering event," it is assumed that the user or administrator would explicitly perform re-enrollment to update the registered template based on the new gait pattern, with identity verification conducted through independent means such as password authentication or administrator confirmation to prevent impersonation and to serve as a fallback against sensor anomalies. The development of automated responses to such events remains a subject for future investigation.

## 5. Conclusions

In this study, we investigated the effect of utilizing long-term data in training on authentication accuracy in a deep-learning-based gait authentication model using an insole-type gait sensor developed by NEC Corporation, and discussed the contributing factors. The main findings are summarized as follows.

- The model trained with the addition of long-term data showed an improvement in the mean *EER* from 0.47 to 0.17% and in the *EER* with a common threshold from 0.71 to 0.28%. In particular, notable improvements were observed for participants who had high *EER* values under the short-term-data-only condition.

- As a change associated with the observed reduction in *EER*, the analysis of feature vectors confirmed that the Silhouette Score improved from 0.5950 to 0.6420 and $\Delta CosSim$ increased from 0.2198 to 0.3383 in the model trained with the addition of long-term data. Despite the increased complexity of the class structure in the training dataset itself, the separability of the feature representations output by the model improved, suggesting a possible enhancement of the model's generalization capability.
- SHAP analysis suggested that in the model trained with the addition of long-term data, the concentrated dependence on specific parameters was mitigated, and identification was performed using a more diverse set of parameters. This may reflect the diversity of walking patterns contained in the long-term data collected under daily living conditions, which influenced the model's feature utilization. It should be noted, however, that the SHAP analysis does not demonstrate that these changes in feature utilization directly caused the improvement in authentication performance.

These results indicate that incorporating long-term data collected during daily life into the training set is a promising approach for improving gait authentication accuracy, although a statistically significant difference could not be confirmed owing to the limited number of evaluation participants. In the future, we plan to investigate the factors underlying individual differences in the effects of incorporating long-term data, as well as data augmentation and training methods that can achieve authentication accuracy comparable to that obtained with long-term data while using only short-term data, and to extend these findings to gait authentication using raw sensor data obtained from the gait sensor.

## Acknowledgments

The authors would like to express their sincere gratitude to NEC Corporation, especially its Business Innovation Department, for providing the sensors, part of the dataset, and valuable advice throughout this study. The authors also thank all the participants involved in the data collection.

## References

1 G. Dahia, L. Jesus, and M. P. Segundo: WIREs Data Min. Knowl. Discov. **10** (2020) e1365. https://doi.org/10.1002/widm.1365
2 I. C. Stylios, S. Kokolakis, O. Thanou, and S. P. Chatzis: Inf. Fusion **66** (2021) 76. https://doi.org/10.1016/j.inffus.2020.08.021
3 J. E. Cutting and L. T. Kozlowski: Bull. Psychon. Soc. **9** (1977) 353. https://doi.org/10.3758/BF03337021
4 P. G. Şahan, S. Şahin, and F. K. Gülağız: J. Supercomput. **80** (2024) 18392. https://doi.org/10.1007/s11227-024-06172-z
5 H. Chao, Y. He, J. Zhang, and J. Feng: Proc. AAAI Conf. Artif. Intell. **33** (2019) 8126. https://doi.org/10.1609/aaai.v33i01.33018126
6 C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2020) 14225. https://doi.org/10.1109/CVPR42600.2020.01423
7 S. Lee, S. Lee, E. Park, J. Lee, and I. Y. Kim: IEEE Access **10** (2022) 120122. https://doi.org/10.1109/ACCESS.2022.3221813

8 J. Moon, N. H. Minaya, N. A. Le, H.-C. Park, and S.-I. Choi: Sensors **20** (2020) 4001. https://doi.org/10.3390/s20144001
9 H. Shimizu, T. Saito, S. Kashiyama, S. Kawamoto, S. Morino, M. Nagai-Tanima, and T. Aoyama: Smart Health **35** (2025) 100526. https://doi.org/10.1016/j.smhl.2024.100526
10 D. Renggli, C. Graf, N. Tachatos, N. Singh, M. Meboldt, W. R. Taylor, L. Stieglitz, and M. Schmid Daners: Front. Physiol. **11** (2020) 90. https://doi.org/10.3389/fphys.2020.00090
11 K. Fukushi, C. Huang, Z. Wang, H. Kajitani, F. Nihey, and K. Nakahara: IEEE Sens. J. **22** (2022) 9636. https://doi.org/10.1109/JSEN.2022.3164057
12 C. Huang, K. Fukushi, Z. Wang, F. Nihey, H. Kajitani, and K. Nakahara: Sensors **22** (2022) 351. https://doi.org/10.3390/s22010351
13 A. Bandara, H. Shimizu, D. Watanabe, T. Misa, S. Suzuki, K. Tanigawa, M. Nagai-Tanima, and T. Aoyama: Sci. Rep. **15** (2025) 8026. https://doi.org/10.1038/s41598-025-91617-5
14 J. Deng, J. Guo, N. Xue, and S. Zafeiriou: Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2019) 4690. https://doi.org/10.1109/CVPR.2019.00482
15 T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama: Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (ACM, 2019) 2623. https://doi.org/10.1145/3292500.3330701
16 P. J. Rousseeuw: J. Comput. Appl. Math. **20** (1987) 53. https://doi.org/10.1016/0377-0427(87)90125-7
17 S. M. Lundberg and S.-I. Lee: Proc. 31st Int. Conf. Neural Information Processing Systems (Curran Associates, 2017) 4768. https://doi.org/10.48550/arXiv.1705.07874