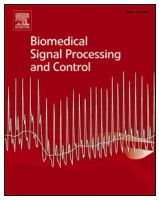
Biomedical Signal Processing and Control 104 (2025) 107406

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



# Age-related gait patterns classification using deep learning based on time-series data from one accelerometer

Xiaoping Zheng [a], Elisabeth Wilhelm [b], Egbert Otten [a], Michiel F. Reneman [c], Claudine J.C. Lamoth [a,*]

[a] University of Groningen, University Medical Center Groningen, Department of Human Movement Sciences, 9713 AV Groningen, the Netherlands
[b] University of Groningen, Faculty of Science and Engineering, 9747 AG Groningen, the Netherlands
[c] University of Groningen, University Medical Center Groningen, Department of Rehabilitation Medicine, 9751 ND Groningen, the Netherlands

ARTICLE INFO

*Keywords:*
Accelerometers
Ageing
Deep learning
Gait classification
Machine learning

ABSTRACT

Gait pattern classification is important for healthcare. Conventional machine learning (CML) approaches based on handcrafted gait features are widely used in gait classification. However, extracting features may lead to suboptimal performance by omitting useful features. End-to-end deep learning (DL) approaches eliminate the need for feature extraction. However, some state-of-the-art DL approaches have not been explored in gait analysis. Furthermore, no consensus exists regarding the window sizes of input acceleration, which affects classification accuracy. In this study, data were collected from one accelerometer during a 3-minute indoor walking task. A total of 267 participants were divided into adults (18–65 years) and older adults (>65) groups. To explore age-related gait patterns classification performance, 5 DL approaches based on raw data and 4 CML approaches based on handcrafted features were compared. The results show that DL outperformed CML, with all AUC (Area under the receiver operator curve) greater than 0.94 compared to the best CML approach of 0.83. This suggests that DL may have learned important gait features related to aging that have not yet been identified by previous research. Furthermore, windows of different sizes ranging from 128 to 5120 samples were tested. The best performance of DL was achieved at a window size of 1024 (including about 20 steps). These findings indicate that the differences and relationship between gait cycles are important factors for classifying age-related gait patterns. This study could contribute to the development of more accurate gait pattern classification and assist in detecting age-related gait patterns in clinical environments.

## 1. Introduction

Walking is one of the most common repetitive activities of humans. Gait patterns have been acknowledged as a potential biomarker for fall risk, Parkinson's disease and ageing [1]. Adaptive gait patterns of healthy individuals are the result of the delicate control and coordination of various systems, such as the central nervous and musculoskeletal systems. Therefore, gait analysis plays a crucial role in gait monitoring and abnormalities recognition, clinical interventions assessment, and rehabilitation programs [2,3].

Accelerometers have been widely used in gait assessment in clinical and daily-living environments [3]. From accelerometer signals, comprehensive sets of gait features are calculated, such as spatial, temporal, and dynamic outcomes [4–6] to characterize the alterations in gait patterns. For instance, results from gait analysis revealed that geriatric patients exhibit slower walking speeds accompanied by less regular, less predictable, and less local stable gait patterns compared with healthy older adults [7]. In older adults (>65 years), to avoid falls and increase stability, compensatory gait patterns have been observed, including lower walking speed, shorter step length, and larger step width compared to young controls [8]. The changes in gait patterns due to aging or pathology allow for classification of different populations based on gait features [9].

Between many gait features, temporal dependencies (e.g., between walking speed and step frequency) and non-linear interactions (e.g., between walking speed and local dynamic stability) exist [10,11]. Conventional machine learning (CML) approaches can capture the linear dependencies and non-linear interactions between gait features,

* Corresponding author.
*E-mail addresses:* x.zheng@umcg.nl (X. Zheng), e.wilhelm@rug.nl (E. Wilhelm), egbert.otten@umcg.nl (E. Otten), m.f.reneman@umcg.nl (M.F. Reneman), c.j.c.lamoth@umcg.nl (C.J.C. Lamoth).

https://doi.org/10.1016/j.bspc.2024.107406
Received 25 March 2024; Received in revised form 29 October 2024; Accepted 17 December 2024
Available online 23 January 2025

and have been comprehensively explored and successfully applied to classify pathological gait patterns and age-related gait patterns [12]. Support vector machine (SVM) and random forest (RF) are often-used supervised CML approaches in gait analysis [5,13]. A recent study explored the performance of 6 CML approaches (linear discriminant analysis, logistic regression (LR), naive bayes (NB), SVM, k-nearest neighbor (KNN), and RF) in classifying fallers from non-fallers based on gait features [14]. RF achieved an optimal classification accuracy of 98%. An artificial neural network (ANN) model classifies geriatric patients with an accuracy of 96% by incorporating the interaction between clinical and gait variables [15]. Similar, the accurate classification of freezing of gait, a common gait characteristic seen in patients with Parkinson's disease, has been studied by comparing 6 different CML learning approaches, namely KNN, RF, LR, NB, multilayer perceptron (MLP), and SVM [16]. The results show that SVM was the optimal classifier with 89.6% on the geometric mean of sensitivity (87.4%) and specificity (91.7%), compared to MLP, the second best classifier, which achieved 85.16%.

Although the above studies demonstrate that gait classification benefits from CML approaches, there are several drawbacks of these CML approaches for clinical application. First, they rely on handcrafted gait features for the input. The design and selection of handcrafted gait features requires expert knowledge and may omit some important features leading to suboptimal performance. Second, the calculation of gait features is performed most often offline and is laborious [4]. Third, to obtain stable and accurate gait features, such as maximum Lyapunov index as an index of stability, some strict signal requirements need to be met, e.g., regarding signal length and sample frequency, which may hamper clinical applications [17].

Unlike CML approaches, deep learning (DL) approaches include feature extraction as a part of the model and can learn suitable features from the raw data automatically. Apart from this, the multiple processing layers of DL approaches allow the progressive extraction of higher-level features from the accelerometer signal [18]. Moreover, DL approaches have been shown to outperform CML approaches in several fields (e.g., visual recognition and text analytics), especially in time series classification tasks [19]. In order to eliminate the need of handcrafted features and improve classification performance, DL approaches have been explored for clinical gait classification [20–24]. For instance, recurrent neural network (RNN) models (long short-term memory (LSTM) and bidirectional LSTM (BiLSTM)) have been applied on IMU (inertial measurement units; acceleration and gyro) data obtained during 1-minute walking to classify fallers and non-fallers in patients with multiple sclerosis [22]. With a window size of 1 min, the results of this study show that DL approach (BiLSTM) outperformed CML approaches, SVM and LR, with 0.88 area under the receiver operator curve (AUC), compared to the best performance of CML which was 0.79. For classification into fallers and non-fallers based on acceleration data during daily life, the performance of LSTM and gated recurrent unit (GRU) was compared using a 1.28-second window size. Both LSTM and GRU provided classification results around 0.96 accuracy [23].

The studies mentioned above show the potential of DL, especially RNN, for gait analysis. However, other state-of-the-art DL approaches (e.g., convolutional neural network (CNN) and hybrid neural network (HNN)) are not included [25]. RNN is capable of learning temporal relationships from accelerometer data, while CNN is widely known for its feature extraction capability and has been extensively used in time series classification tasks [25]. The combination of the deep structures of CNN and RNN called HNN, such as convolutional LSTM (ConvLSTM), may benefit from both advantages. These models have been successfully employed in some tasks similar to gait patterns classification, such as human activity recognition (HAR) based on inertial sensors [26]. Moreover, window size is discussed in other fields, such as in HAR [27] but not in gait analysis yet.

Considering the cyclic nature of walking, with repetitive gait cycles (spatio-temporal characteristics of swing and stand-phases), the choice of the window size and time span is crucial, since it will affect classification results. Additionally, in a given time window, different step frequencies from participants may lead to disparities in the number of steps, which may, in turn, impact other gait features, and result in an unfair comparison between participants. To mitigate these potential biases, acceleration data can be normalized based on participant's step frequency. This normalization process serves to standardize the number of steps within each time window, thereby facilitating more accurate and equitable comparisons between participants.

This study aims to compare different CML and DL approaches to classify age groups (adults vs. older adults) based on acceleration time-series obtained during 3-minute walking. More specifically, we will: 1) compare the classification performance of DL approaches (CNN, LSTM, GRU, and ConvLSTM) based on acceleration time-series data with CML approaches (RF, SVM, NB, and KNN) that use handcrafted gait features as input; 2) explore how window size affects the classification results in DL approaches; 3) study the effect of step frequency in DL classification by normalizing the data in the time window by step frequency.

## 2. Methods

### 2.1. Participants, equipment, and data collection

Accelerometer data obtained during walking from different studies [7,28–31] were pooled to create the present dataset. Herein 267 out of 394 participants were included in this study. Participants were excluded because of cognitive impairment (n = 104), no walking data (n = 19; power spectrum values smaller than 0.5HZ), and insufficient length of walking data (n = 4; less than 100s walking data within the 3-minute recording). Participants were divided into two sub-groups: the adult group (18–65) (74:56, Female: Male; mean age 38.3, SD: 15.5), the older adult group (>65) (60:77, Female: Male; mean age 77.1, SD: 6.1). Participants were asked to walk for 3 minutes at a comfortable walking speed. During walking, one 3-axes accelerometer (iPod, dynamic range ± 3g; Dynaport, dynamic range ± 8 g; or ActiGraph, dynamic range ± 6g; 100Hz sampling frequency) was fixed with a belt near the level of lumbar segment L3. In the standing upright position, the 3D accelerometer measures acceleration in three-dimensional space as follow:

- X-axis: Represents the vertical direction, pointing downward.
- Y-axis: Indicates the anteroposterior (back-to-front) direction, facing the walking direction.
- Z-axis: Represents the mediolateral (side-to-side) direction, extending from left to right perpendicular to the walking direction.

Data were collected at the University Medical Centre Groningen (2012–2022) and Slotervaart Hospital (2008–2018), with approval from their respective Medical Ethical Committees. Slotervaart Hospital closed in 2018, and all data collection there occurred prior to its closure. All participants provided written informed consent. This study was conducted in accordance with the principles of the Declaration of Helsinki.

### 2.2. Data preprocessing

A median filter (with window size of 5 samples) was used to remove spike noise and an additional low pass filter (4th order Butterworth cutoff frequency 20Hz) was applied to remove the high frequency noise. To remove the data collected during the sensor installation or uninstallation, the first and last 5s of data were discarded. To ensure that only walking data were included in the analysis, data recorded during turning were removed. During turns, gait speed decreases significantly, typically for at least several seconds, resulting in lower accelerometer readings across all three axes. By setting participant-specific experimental thresholds for sensor readings and time, turning data can be easily identified based on these lower values.

### 2.3. Segmentation and datasets splitting

In order to extract accurate gait features (e.g., maximum Lyapunov exponent and sample entropy), longer accelerometer signal series are needed. Therefore, 1024-sample window size was used. To enable a fair comparison between CML and DL approaches, the filtered walking accelerometer data of each participant were split into windows of 1024-samples.

Each participant was randomly sorted into training, testing, and validation sets (186: 54: 27). The splitting algorithm ensured that the same proportion of adults and older adults was assigned to each of the sets. The length of segments varied per participant. To ensure fair representation of each participant and to maximize data utilization, 10 segments were randomly sampled from each participant's data.

To study the impact of window size on the classification performance of DL approaches, various window sizes, including 128, 256, 1024, 2048, and 5120 samples, were compared.

To investigate the impact of step frequency on DL in gait classification, the approximate number of steps within each segment under the same window was kept consistent. Hence, the number of data points for one step of each participant should be calculated first. For participant $i$, the averaged step frequency $f_i$ was calculated using a fast Fourier transform based on the root mean square of the corresponding walking accelerometer data (across all segments). So, the length of one step in data points was defined as $s_i$,

$$s_i = \frac{1}{f_i} \times 100, \tag{1}$$

where 100 was the sampling frequency.

In 128-sample window size, the data points within the range of $[0, 1.2s_i]$ were selected. Because the step frequency tends to vary slightly during walking and to ensure that the data points of one complete step were included, the experimental coefficient 1.2 was used in this window size. Then these data points were interpolated by a 1-D smoothing spline (with a smoothing factor 0) into a size of 128 to replace the corresponding segments. For other window sizes, the coefficients were set as

$$coefficient = \frac{windowsize}{128} \times 1.2. \tag{2}$$

For example, with a window size of 256, a coefficient of 2.4 was applied, meaning that data points within the range of $[0, 2.4s_i]$ were selected and interpolated to a length of 256.

### 2.4. Deep learning

In this study, the entire dataset was normalized across each segment. Min-max normalization of each segment was performed separately for each axis.

1) Convolutional Neural Network (CNN) is one of the earliest successfully used DL approaches. Because of its excellent capacity for feature extraction, it has been widely used for human movement recognition [32]. CNN includes two parts: 1) convolution layer and pooling layer for feature extraction; 2) fully connected layer and detector layer for classification. These layers can be stacked to form a deep CNN.

2) Recurrent Neural Network (RNN) is a family of neural networks that have recurrent connections. The recurrent connections enable RNN to keep the "memory" from the input of the previous moment and use it to influence the output of the current input. Because it can learn sequential dependencies, RNN outperforms in dealing with sequential problems compared with general neural networks. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are the two mostly used variants of the RNN architecture in movement recognition [26]. LSTM has memory cells comprising inputs, forget, and outputs gates to control store or forget information. GRU has a similar gated structure to adaptively capture sequential dependencies, and it is more computationally efficient than LSTM. Bi-Directional LSTM (BiLSTM) is a variant of sequentially stacked LSTM layers. It has two hidden states that allow the model to use information from the past and the future, of each input, which has been shown to improve performance for some classification tasks [22].

3) Hybrid Models such as ConvLSTM combine the deep structures of CNN and LSTM. CNN are known for their feature extraction capability and LSTM is capable of learning temporal dynamics. In order to obtain the dual advantages of CNN and LSTM, hybrid models are proposed.

To further improve generalization performance, dropout and early stopping regularization techniques were used. Dropout removes non-output units randomly from the network. Thus, the dropout technique can reduce the scale and complexity of the neural network and eventually avoid overfitting. Early stopping is a simple regularization technique. It will stop the training when the validation error which is used as an estimate of the generalization error has no improvement in k epochs.

### 2.5. Conventional machine learning

In order to compare the performance of DL approaches based on acceleration signal with methods based on gait feature input, 4 different CML approaches [5] were employed, including nonlinear models (Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighborhood (KNN)) and tree ensemble models (Random Forest (RF)).

Herein 36 gait features were extracted from the accelerometer data and used to train the CML approaches. These gait features represent the pace (walking speed, stride length, stride time, stride frequency, and acceleration root mean square), regularity (stride regularity and gait symmetry index), smoothness (index of harmonicity and harmonic ratio), predictability (sample entropy), and stability (maximal Lyapunov exponent, maximal Lyapunov exponent normalized per stride by time) of gait and have been described previously [5,28]. The gait features were normalized using min–max normalization, followed by dimensionality reduction through kernel principal component analysis (KPCA). To increase generalization, the parameters from the training dataset were utilized to normalize the testing and validation datasets.

### 2.6. Fine-tuning

The classification performance of ML approaches, especially DL approaches, is highly sensitive to the hyperparameter setting. The conventionally used 5-fold cross validation and randomized search has been widely used to find the best hyperparameters for CML approaches. However, for DL approaches and big datasets, cross validation which repeats training iterations would lead to exploding computation costs. Hence, it is important to find the global optimum in a minimum number of steps. Bayesian Optimization (BO) has become a state-of-the-art solution [33]. It incorporates prior performance of the hyperparameters and updates the new hyperparameters to achieve better performance. Bayesian optimization uses an acquisition function that directs sampling to areas where an improvement over the current best observation is likely. Thus, in this study, BO was used to find the best hyperparameters based on validation data for the DL and CML approaches. For each approach, 18 parameter combination trails were searched.

### 2.7. Hyperparameter space

The details of the hyperparameter space of CML approaches for BO are as follows:

SVM) The "rbf" kernel was used. The box constraint parameter (C) was varied from 1 to 250. The degree was set from 1 to 50. The gamma was set from values 0.01 to 10 with increments of 0.05. The "True" and "False" shrinking were also considered.

NB) The portion of the largest variance of all features was set from $-11$ to $-7$ power of 10 with $-10$ power of 10 step.

KNN) The number of neighbors was varied from 1 to 15. Different algorithms which compute the nearest neighbors were explored, “ball_tree”, “kd_tree”, and “brute”. The weight functions “uniform” and “distance”, were explored.

RF) Different numbers of trees in the range of 10–1000 with increments of 10 were explored. For the trees, the hyperparameters were set as: the number of maximum depth varied from 3 to 25; the maximum number of leaf nodes was varied from 5 to 50. The minimum number of samples required to be at a leaf node was varied between 5 and 50.

The structure of DP approaches is shown in Fig. 1. After the input layer, the optimal number of the deep learning layer stack was tuned by BO. Each stack consists of a Deep layer (Conv1D, GRU, LSTM, Bidirectional-LSTM, or ConvLSTM2D layer) to extract features, a Batch Normalization layer to make the training faster and more stable, a Pooling layer to reduce the number of parameters, and a Dropout layer to avoid overfitting. Then, a Dense layer was used to fully connect the output of the previous layers and a flatten layer was used to flatten the dimensions of the output. Another Dropout layer was added after the flatten layer. The final layer was an output layer which was a combination of a Dense layer and a softmax activation. This Dense layer had 2 units to classify the output into 2 classes. The hyperparameter space of DL approaches for BO is shown in Table 1.

### 2.8. Evaluation

Classification performance was evaluated using the testing data, with commonly used evaluation metrics accuracy, recall (sensitivity), precision, and F1 score (harmonic mean of sensitivity and precision). The receiver operating characteristic curves were compiled, and the AUC was reported.

To allow for the verification and reproduction of results, the project repository has been made publicly available (https://github.com/xzheng93/Age-related_gait_classification). It includes the source code, the processed dataset, log files of all the experiments, and optimal models for each approach. The overall data processing pipeline is illustrated in Fig. 2.

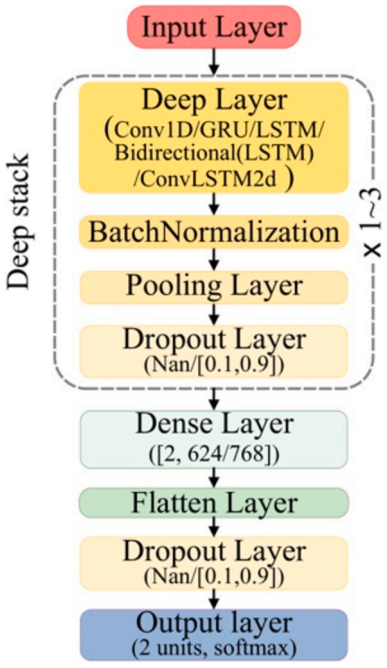



**Fig. 1.** General architecture of the deep learning approaches used in this study. A detailed description of the individual layers can be found in Table 1. Conv1D: 1-dimensional convolution neural network; GRU: gated recurrent unit; LSTM: long short-term memory; ConvLSTM2d: 2-dimensional convolution long short-term memory.

## 3. Results

After hyperparameter tuning based on the validation data, the best architectures and hyperparameters of each approach and the tuning logs were stored and can be found in https://github.com/xzheng93/Age-related_gait_classification/tree/main/result/logs.

### 3.1. Classification performance comparison for 1024 window size

The complete list of performance metrics calculated for each approaches based on a window size is depicted in Table 2. With F1-scores ranging from 0.86 to 0.9, the DL approaches outperformed the CML approaches (F1 scores ranging from 0.69 to 0.74).

To enable an in-depth look at the individual performance of the classifiers, their confusion matrices are presented in Fig. 3 and Fig. 4. GRU achieved the highest overall accuracy (89.3%) and was able to correctly classify 82.3% and 95.7% of the adult and older adult samples, respectively. With 73.9% and 73.5%, SVM and RF achieved the highest accuracies among the CML approaches. The difference in accuracy between the DL approach with the highest accuracy and the CML approach with the highest accuracy was 15.4%.

To further demonstrate the characteristics of the classifiers, the ROC curves are presented in Fig. 5. As depicted in Fig. 5, all DL approaches had an AUC higher than 0.94, while the CML classifiers only achieved AUC values of 0.75 to 0.83.

### 3.2. The effect of different window sizes and step frequency on DL

The mean step frequencies for the adult and older adult groups were 1.72 (SD: 0.08) and 1.97 (SD: 0.20) step/second, respectively. Given window sizes ranging from 128 to 5120 samples, each segment may include approximately 1, 4, 8, 20, 35, and 88 steps.

The different window sizes for the DL approaches were investigated and the AUC results are shown in Fig. 6 (a). Additional accuracy metrics can be found in Appendix A. To represent the performance of convolutional, recurrent, and hybrid models, CNN, GRU, and ConvLSTM were selected, respectively. When the window size increased from 128 to 1024, the performance of GRU and ConvLSTM improved, while the performance of CNN fluctuated with the window sizes. However, when the window sizes increased to 2048 and 5120, the AUC values decreased in all approaches, especially in CNN. This may be attributed to the decreased number of segments resulting from larger window sizes. Specifically, as the window size increased from 128 to 5120, the number of segments sharply decreased from 26,700 to 534.

The classification results based on step frequency normalization data are shown in Fig. 6 (b) (details in the Appendix B). All approaches exhibited a similar trend as with raw data. CNN and GRU achieved comparable results compared to raw data, while ConvLSTM showed slightly worse performance.

## 4. Discussion

This study compared the classification performance of 5 DL and 4 CML approaches in classifying age-related gait patterns in healthy adults based on acceleration time-series data collected by one 3D-accelerometer for a 3-minute walk. The results show that DL approaches achieved better classification performance by a remarkable margin (all AUC greater than 0.94 for DL vs. 0.83 for SVM (best in CML)). The study also explored the effective window size for DL and studied the effect of step frequency on DL classification. It was found that when the window size increased from 128 to 1024, the classification performance of GRU and ConvLSTM increased, while the performance of CNN fluctuated. However, when the window sizes were larger than 1024, the performance of all DL approaches decreased. Based on the normalized gait data, CNN

**Table 1**
Hyperparameters space for deep learning approaches.

| Layer | Parameters | | CNN | GRU | LSTM | BiLSTM | ConvLSTM |
|---|---|---|---|---|---|---|---|
| Input | | | | | | | |
| Stack | Stack number | | 1–3 | | | | |
| | Deep layer | Layer name | Conv1D | GRU | LSTM | Bidirectional (LSTM) | ConvLSTM2D |
| | | Unit/ Filter | [2, 768] | | | | [2, 624] |
| | | Kernel size | [1,15] | – | | | [1,8] |
| | | Activation | "ReLU" | | "tanh" | | |
| | Batch Normalization | | – | | | | |
| | Pooling | Kernel size | – | | | | [1,8] |
| | Dropout | Rate | Nan or [0.1, 0.9] | | | | |
| Dense | Unit | | [2, 768] | | | | [2, 624] |
| Flatten | | | – | | | | |
| Dropout | Rate | | Nan or [0.1, 0.9] | | | | |
| Output | Dense | | 2 units and "softmax" activation | | | | |
| | Learning rate | | [1e-5, 1e-2] | | | | |

CNN: convolutional neural network; GRU: gate recurrent unit; LSTM: long short-term memory; BiLSTM: bi-directional long short-term memory; ConvLSTM: convolutional long short-term memory; ReLU: rectified linear unit; tanh: hyperbolic tangent function.

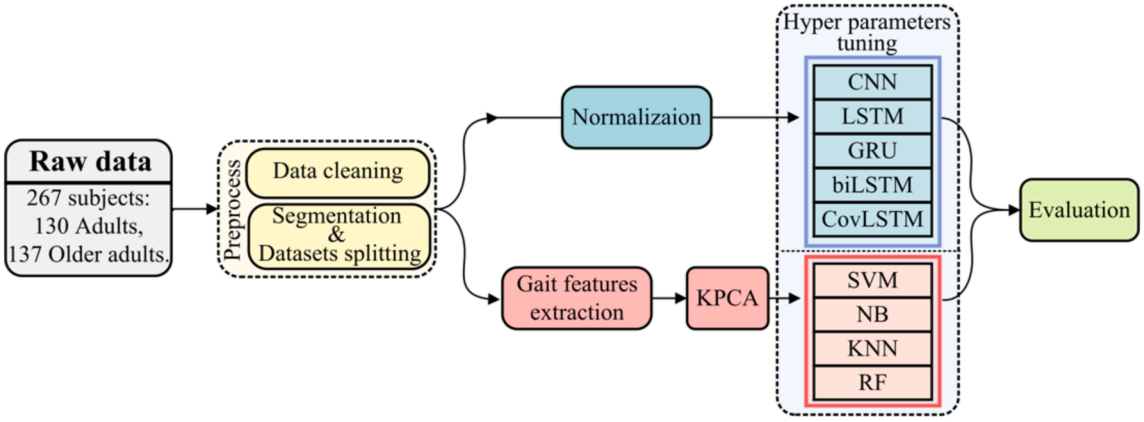



**Fig. 2.** Data processing pipeline for age-related gait patterns classification. KPCA: kernel principal component analysis; CNN: convolutional neural network; GRU: gate recurrent unit; LSTM: long short-term memory; BiLSTM: bi-directional long short-term memory; ConvLSTM: convolutional long short-term memory; SVM: support vector machine; NB: naive bayes; KNN: k-nearest neighborhood; RF: random forest.

**Table 2**
The performance metrics of DL and CML approaches.

| | Models | Acc | Pre | Recall | F1 | AUC |
|---|---|---|---|---|---|---|
| Deep learning | CNN | 0.88 | 0.83 | 0.98 | 0.90 | 0.96 |
| | LSTM | 0.85 | 0.81 | 0.94 | 0.87 | 0.94 |
| | GRU | 0.89 | 0.85 | 0.96 | 0.90 | 0.96 |
| | BiLSTM | 0.87 | 0.84 | 0.93 | 0.88 | 0.94 |
| | ConvLSTM | 0.85 | 0.85 | 0.86 | 0.86 | 0.95 |
| Conventional machine learning | SVM | 0.74 | 0.74 | 0.74 | 0.74 | 0.83 |
| | NB | 0.72 | 0.74 | 0.72 | 0.72 | 0.75 |
| | KNN | 0.68 | 0.79 | 0.68 | 0.69 | 0.76 |
| | RF | 0.74 | 0.74 | 0.74 | 0.74 | 0.82 |

CNN: convolutional neural network; GRU: gate recurrent unit; LSTM: long short-term memory; BiLSTM: bi-directional long short-term memory; ConvLSTM: convolutional long short-term memory; SVM: support vector machine; NB: naive bayes; KNN: k-nearest neighborhood; RF: random forest.

and GRU exhibited a similar trend and performance as with raw data, while ConvLSTM showed slightly worse performance.

In this study, DL approaches exhibited superior performance compared to CML approaches for the 1024-sample window size. This can be attributed to two main reasons. First, DL has superior learning capabilities. By using multi stacks of simple layers, DL is able to model a high degree of nonlinearity in the input data. Second, DL has better features extraction capacity and is able to select the most suitable features for classification because of its end-to-end characteristic which integrates feature extraction, selection, and classification within a neural network model [26]. DL can learn large amounts of features from raw data, including high-level features constructed from features learned in lower layers [18]. In contrast, CML approaches rely on handcrafted gait features that require domain expertise and may miss some useful features [34]. Although this study considered a comprehensive set of gait features, including pace, regularity, smoothness, and stability, the substantial performance gap between DL and CML suggests that important features reflecting changes in gait due to aging may not have been included in the handcrafted features. Furthermore, the better performance of DL suggests that handcrafted gait features are not necessary for age-related gait pattern classification.

Signal segmentation plays a crucial role for gait classification as it not only affects the classification performance [35], but also helps to understand how the data size and time span influence the classification results [27]. GRU is a kind of RNN approach and is designed to learn both short- and long-term dependent features [26]. ConvLSTM combines CNN and RNN structures and is expected to benefit from larger window sizes, since bigger window sizes provide richer features, especially time-dependent features, which are valuable for gait analysis. The AUC scores of GRU and ConvLSTM show remarkable increases when the window size increases from 128 (including about 1 step) to 1024 (including about 20 steps). Although walking is a repeating activity, the relationship and slight differences between gait cycles may disclose the

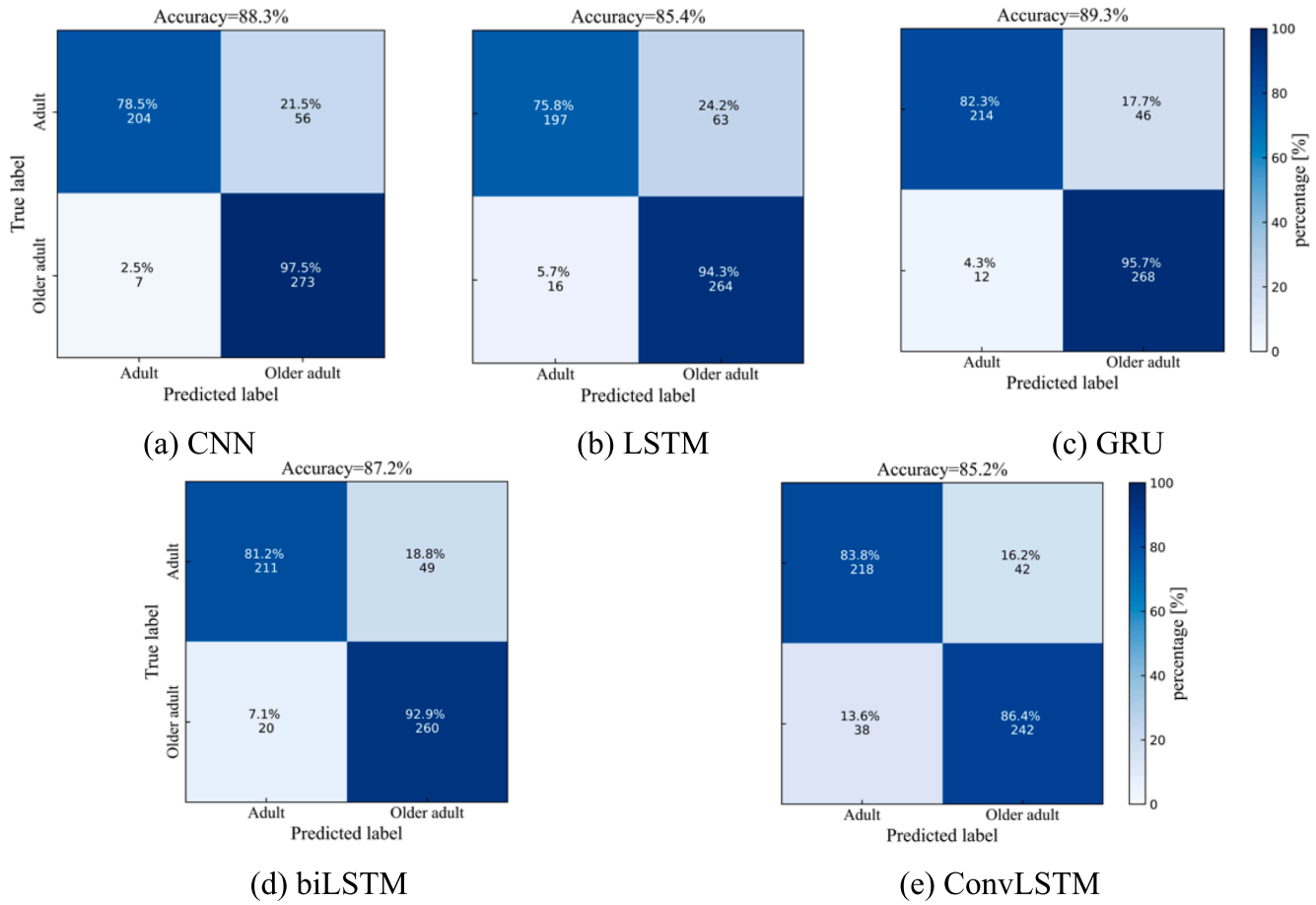



**Fig. 3.** Confusion matrices of DL approaches: (a) CNN, (b) LSTM, (c) GRU, (d) BiLSTM, (e) ConvLSTM. For testing, 54 participants and their corresponding 10 segments (n = 540) were used. A: adult group (n = 26); OA: older adult group (n = 28). CNN: convolutional neural network; GRU: gate recurrent unit; LSTM: long short-term memory; BiLSTM: bi-directional long short-term memory; ConvLSTM: convolutional long short-term memory.

postural control ability of participants. Changes in postural control due to aging are associated with alterations in sensory, musculoskeletal, and neuromuscular systems [36]. Gait analysis shows that because of the decrease of postural control capacity, older adults have less regularity and higher variability [37], lower local stability (higher largest Lyapunov exponent) [38], worse gait symmetry (lower symmetry index) [39], and less complexity (lower sample entropy) in gait [40]. These gait features are designed to measure the time dependent changes of gait cycles. Therefore, a larger window size contains this kind of information and may allow GRU and ConvLSTM to discriminate age groups. The sharp improvements of GRU and ConvLSTM from 128- to 1024-sample window size may indicate that the time dependent changes in gait cycles are important for age-related gait pattern discrimination.

CNN is well known for its local temporal and spatial features extraction capacity, allowing it to capture the repeating gait patterns in the data and achieve good performance. However, it cannot learn long-term dependencies from the data, which means that larger window sizes may not provide additional features for CNN. Additionally, the accelerometer data in this study were collected from participants walking in a straight and empty hallway without external perturbation. Thus, the gait cycle is highly repetitive and presumably predictable, as evident by the mean sample entropy values of 0.275. These may explain the small fluctuations in classification performance among window sizes ranging from 128 to 1024. For the 128-sample window size, CNN achieved accurate performance (AUC = 0.94). This result may indicate that the shape of one or two steps contains rich information that can be used to discriminate the age-related gait patterns. Indeed, gait features such as root mean square (gait intensity), rhythm (the proportion of stance and swing phase), and harmonic index (the smoothness of the acceleration curve) are designed to assess the shape of acceleration in gait, and have been utilized to characterize different age population [40]. Therefore, CNN specifically capture this type of information to discriminate the age-related gait patterns.

When the window size increased beyond 1024, the number of segments decreased, leading to a decline in classification performance for all approaches. Moreover, using a 5120-sample window size resulted in an insufficient number of segments (n = 523), which decreased the generalization of the model and led to a sharp decline in classification performance. Despite this, GRU and ConvLSTM demonstrated only slightly decreased performance, due to their ability to benefit from longer data segments compared to CNN.

CNN and GRU exhibited similar classification performance with both raw data and the normalization of segment data by step frequency. ConvLSTM exhibited slightly inferior performance when using normalized data compared to using raw data. This may be due to the normalization process which excluded step frequency information and only used partial data from the original segment. These results may indicate that data normalization based on step frequency may not be necessary for gait classification using DL.

In this study, all DL approaches achieved the best classification performance with a 1024-sample window size (approximately 10s, 20 steps). This window size may also be applicable for gait pattern classification in daily living environments, where walking bouts are typically short—60% of bouts last less than 30s, corresponding to approximately 24 to 60 steps, depending on a step frequency ranging from 0.8 to 2Hz [41].

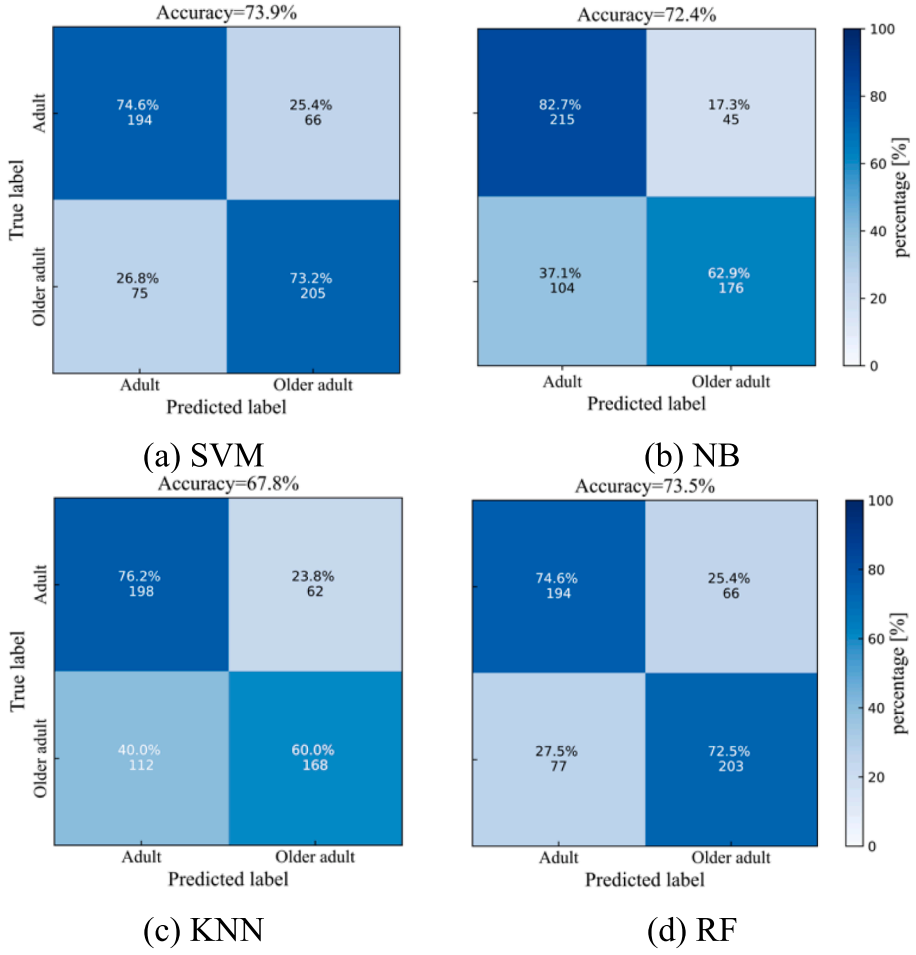



**Fig. 4.** Confusion matrices of CML approaches: (a) SVM, (b) NB, (c) KNN, (d) RF. For testing, 54 participants and their corresponding 10 segments (n = 540) were used. A: adult group (n = 26); OA: older adult group (n = 28). SVM: support vector machine; NB: naive bayes; KNN: k-nearest neighborhood; RF: random forest.

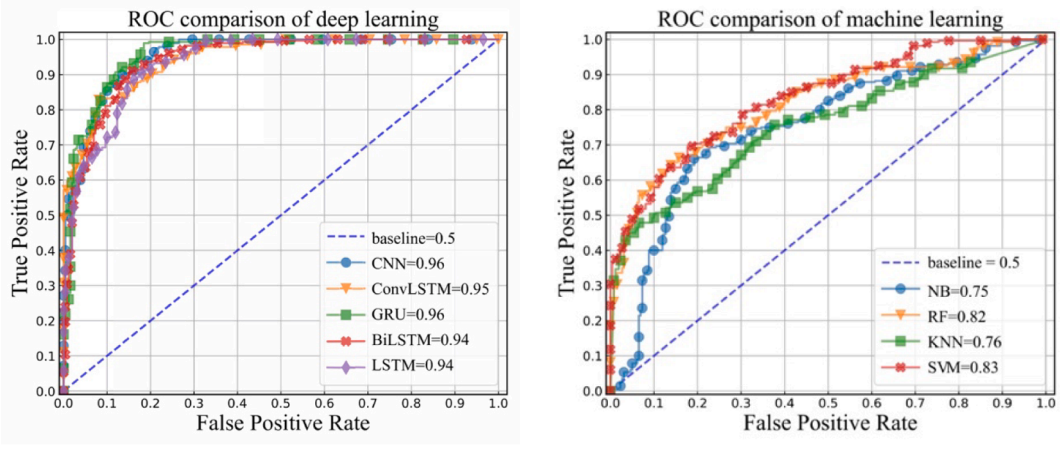



**Fig. 5.** Receiver operating characteristic curves (ROC) and area under curve (AUC) of (a) DL approaches and (b) CML approaches. CNN: convolutional neural network; GRU: gate recurrent unit; LSTM: long short-term memory; BiLSTM: bi-directional long short-term memory; ConvLSTM: convolutional long short-term memory; SVM: support vector machine; NB: naive bayes; KNN: k-nearest neighborhood; RF: random forest.

## 5. Limitation

In this study, to ensure a fair comparison, the same total amount of data was used for different window sizes, resulting in more segments for shorter window sizes. The varying number of input segments may potentially raise concerns about fair comparison. However, maintaining a consistent number of segments across window sizes would have led to unequal data input, as larger window sizes encompass more data. It

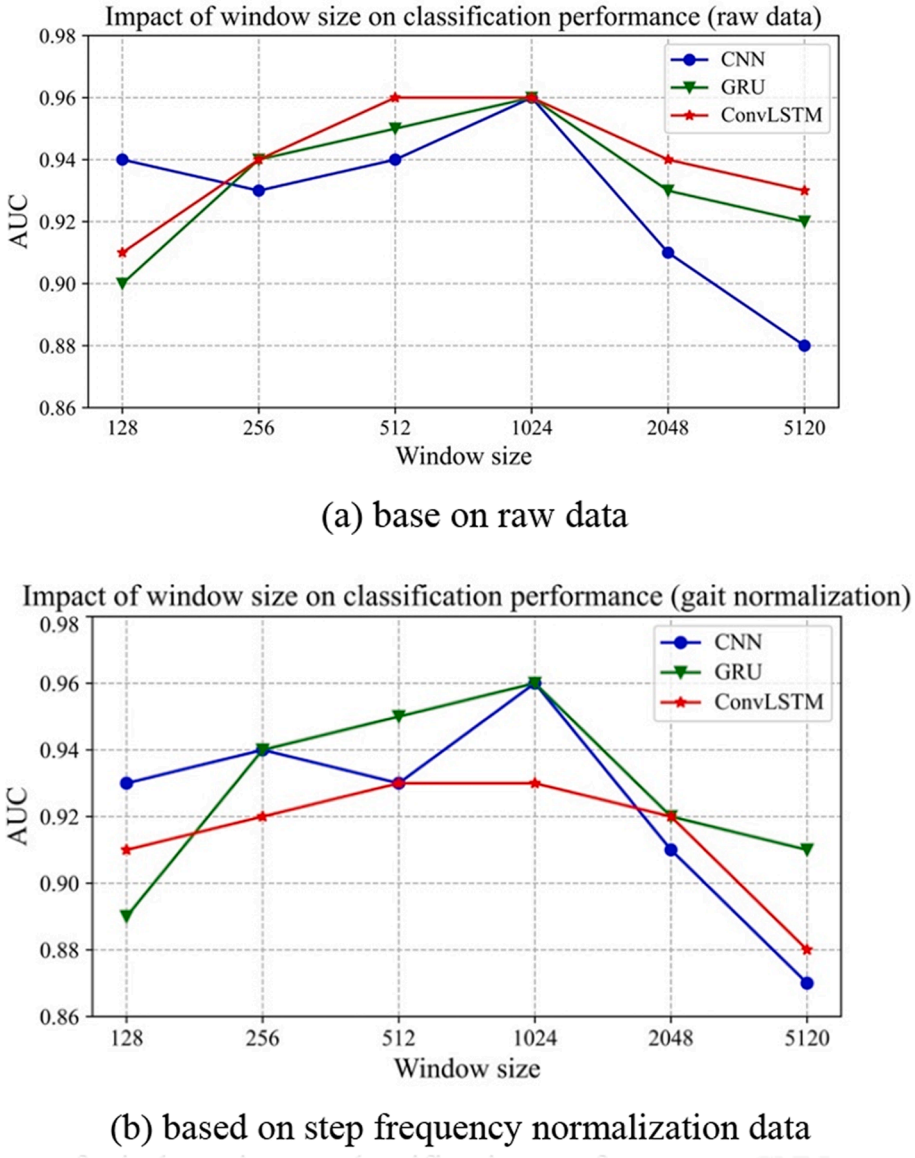



**Fig. 6.** Impact of window size on classification performance. CNN: convolutional neural network; GRU: gate recurrent unit; ConvLSTM: convolutional long short-term memory.

would create another issue of unfair comparison. Since the dataset size used in this study is limited, it is not possible to further deduce whether using a larger window size (e.g., 5120) would result in improved classification performance. Additionally, this research did not consider other sampling frequencies, such as 30Hz or lower. For the suggested window size which cover about 20 steps, a lower sampling frequency would result in shorter data segments, while a higher frequency would produce longer data segments. Future research is needed to better understand the impact of varying sampling frequencies on classification performance. Apart from this, the proposed approaches should be validated on data recorded outside a lab environment before they are deployed in a daily life environment. This step is necessary, because the environment will influence walking (e.g., stride length variability [42]).

Although DL approaches yield promising results in age-related gait pattern classification performance, the black-box nature of these approaches is often seen as a limitation for clinical application [37]. While the underlying mathematical principles of these approaches are understood, it is unclear why a particular prediction has been made. Additionally, features extracted by DL cannot be linked to physical explanations and models. Identifying age-related gait features that have not been covered by the current gait analysis may help to gain knowledge about age-related changes and assist clinicians in identifying people with altered gait patterns. Facing these challenges, the field of explainable artificial intelligence has gained increasing attention in recent years [43]. Further research should focus on determining the justification of classification predictions and making the prediction processes comprehensible to clinical experts. For the next step, popular methods, such as Interpretable Model-Agnostic Explanation [44] and SHapley Additive exPlanations [45], could be used to interpret and explain DL in performing gait pattern classification. Lastly, validation of these explainable models in real-world clinical settings will be crucial to ensure their reliability and acceptance in the medical community.

## 6. Conclusion

This study compared 4 CML and 5 DL approaches for classifying the age-related gait patterns of healthy adults, based on one accelerometer sensor. The results show that DL approaches based on raw accelerometer data outperformed the CML approaches based on handcrafted features. This suggests that DL may have captured certain gait features related to

aging that were not previously reported. The investigation into the impact of different window sizes on classification performance indicates that the shape of acceleration and the relationship and differences between gait cycles are important factors for age-related gait pattern classification. The results show that a 1024-sample window size was suitable for gait pattern classification as it yielded the best performance across all tested DL approaches. Notably, the study discovered that normalizing data by step frequency did not improve classification performance, suggesting that this step frequency normalization may not be necessary in a healthy population.

The present study highlights the potential of DL approaches for accurately classifying age-related gait patterns using accelerometer data. The findings suggest that subtle variations and interconnections between gait cycles, as well as the shape of one step acceleration are important factors for age-related pattern classification, contributing to a better understanding of the postural control changes associated with aging. This study could serve as one of the first stepping stones towards future studies towards the development of more accurate gait pattern classification, facilitating abnormal gait patterns diagnosis in the clinical environment and gait monitoring in the daily living environment.

## CRediT authorship contribution statement

**Xiaoping Zheng:** Writing – review & editing, Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Elisabeth Wilhelm:** Writing – review & editing, Visualization, Conceptualization. **Bert Otten:** Writing – review & editing, Supervision, Conceptualization. **Michiel F. Reneman:** Writing – review & editing, Supervision, Conceptualization. **Claudine J.C. Lamoth:** Writing – review & editing, Supervision, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

XZ was supported by China Scholarship Council-University of Groningen Scholarship [Grant No. 201906410084].

## Appendix A. . Classification performance of DL approaches for different window sizes based on raw data

.

**Table A1**
The performance metrics of DL for the 128 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.87 | 0.83 | 0.94 | 0.88 | 0.94 |
| GRU | 0.81 | 0.78 | 0.89 | 0.83 | 0.90 |
| ConvLSTM | 0.84 | 0.79 | 0.95 | 0.86 | 0.90 |

**Table A2**
The performance metrics of DL for the 256 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.88 | 0.83 | 0.96 | 0.89 | 0.93 |
| GRU | 0.87 | 0.82 | 0.95 | 0.88 | 0.94 |
| ConvLSTM | 0.88 | 0.84 | 0.95 | 0.89 | 0.93 |

**Table A3**
The performance metrics of DL for the 512 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.86 | 0.84 | 0.90 | 0.87 | 0.94 |
| GRU | 0.86 | 0.88 | 0.86 | 0.87 | 0.95 |
| ConvLSTM | 0.89 | 0.85 | 0.94 | 0.90 | 0.96 |

**Table A4**
The performance metrics of DL for the 2048 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.89 | 0.82 | 1 | 0.90 | 0.91 |
| GRU | 0.87] | 0.84 | 0.93 | 0.88 | 0.93 |
| ConvLSTM | 0.90 | 0.84 | 1 | 0.92 | 0.94 |

**Table A5**
The performance metrics of DL for the 5120 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.88 | 0.81 | 1 | 0.90 | 0.88 |
| GRU | 0.81 | 0.8 | 0.86 | 0.83 | 0.92 |
| ConvLSTM | 0.84 | 0.83 | 0.88 | 0.85 | 0.93 |

CNN: convolutional neural network; GRU: gate recurrent unit; ConvLSTM: convolutional long short-term memory.

## Appendix B

Classification performance of DL approaches for different window sizes based on gait normalization data.

**Table B1**
The performance metrics of DL for the 128 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.86 | 0.83 | 0.93 | 0.87 | 0.93 |
| GRU | 0.82 | 0.81 | 0.85 | 0.83 | 0.89 |
| ConvLSTM | 0.83 | 0.82 | 0.87 | 0.84 | 0.91 |

**Table B2**
The performance metrics of DL for the 256 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.88 | 0.83 | 0.95 | 0.89 | 0.94 |
| GRU | 0.87 | 0.82 | 0.96 | 0.89 | 0.94 |
| ConvLSTM | 0.84 | 0.82 | 0.89 | 0.86 | 0.92 |

**Table B3**
The performance metrics of DL for the 512 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.85 | 0.82 | 0.91 | 0.86 | 0.93 |
| GRU | 0.88 | 0.84 | 0.94 | 0.89 | 0.95 |
| ConvLSTM | 0.86 | 0.86 | 0.88 | 0.87 | 0.93 |

**Table B4**
The performance metrics of DL for the 1024 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.89 | 0.88 | 0.91 | 0.89 | 0.96 |
| GRU | 0.89 | 0.85 | 0.95 | 0.9 | 0.96 |
| ConvLSTM | 0.84 | 0.78 | 0.98 | 0.87 | 0.93 |

**Table B5**
The performance metrics of DL for the 2048 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.83 | 0.79 | 0.9 | 0.84 | 0.91 |
| GRU | 0.84 | 0.8 | 0.91 | 0.85 | 0.92 |
| ConvLSTM | 0.84 | 0.81 | 0.91 | 0.86 | 0.92 |

**Table B6**
The performance metrics of DL for the 5120 sample-window size.

| Models | Accuracy | Precision | Recall | F1-score | AUC |
|---|---|---|---|---|---|
| CNN | 0.81 | 0.77 | 0.89 | 0.83 | 0.87 |
| GRU | 0.84 | 0.79 | 0.95 | 0.86 | 0.91 |
| ConvLSTM | 0.82 | 0.84 | 0.82 | 0.83 | 0.88 |

CNN: convolutional neural network; GRU: gate recurrent unit; ConvLSTM: convolutional long short-term memory.

## Data availability

The authors do not have permission to share data.

## References


[1] M.A. Brodie, N.H. Lovell, C.G. Canning, H.B. Menz, K. Delbaere, S.J. Redmond, M. Latt, D.L. Sturnieks, J. Menant, S.T. Smith, Gait as a biomarker? Accelerometers reveal that reduced movement quality while walking is associated with Parkinson's disease, ageing and fall risk, 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 5968-5971.

[2] D. Jarchi, J. Pope, T.K. Lee, L. Tamjidi, A. Mirzaei, S. Sanei, A review on accelerometry-based gait analysis and emerging clinical applications, IEEE Rev. Biomed. Eng. 11 (2018) 177–194.

[3] D. Sethi, S. Bharti, C. Prakash, A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work, Artif. Intell. Med. 129 (2022), 102314.

[4] Y. Hutabarat, D. Owaki, M. Hayashibe, Recent advances in quantitative gait analysis using wearable sensors: a review, IEEE Sens. J. (2021).

[5] X. Zheng, M.F. Reneman, J.A. Echeita, R.H.S. Preuper, H. Kruitbosch, E. Otten, C. J. Lamoth, Association between central sensitization and gait in chronic low back pain: insights from a machine learning approach, Comput. Biol. Med. 144 (2022), 105329.

[6] Y.H. Zhou, R.Z.U. Rehman, C. Hansen, W. Maetzler, S. Del Din, L. Rochester, T. Hortobagyi, C.J.C. Lamoth, Classification of neurological patients to identify fallers based on spatial-temporal gait characteristics measured by a wearable device, Sensors 20 (2020).

[7] L. Kikkert, N. Vuillerme, J.P. van Campen, B.A. Appels, T. Hortobagyi, C.J. Lamoth, Gait characteristics and their discriminative power in geriatric patients with and without cognitive impairment, J. Neuroeng. Rehabil. 14 (2017).

[8] A. Aboutorabi, M. Arazpour, M. Bahramizadeh, S.W. Hutchins, R. Fadayevatan, The effect of aging on gait parameters in able-bodied older subjects: a literature review, Aging Clin. Exp. Res. 28 (2016) 393–405.

[9] A. Mannini, D. Trojaniello, A. Cereatti, A.M. Sabatini, A machine learning framework for gait classification using inertial sensors: application to elderly, post-stroke and huntington's disease patients, Sensors 16 (2016) 134.

[10] Y. Zhou, R. Romijnders, C. Hansen, J. van Campen, W. Maetzler, T. Hortobagyi, C. J.C. Lamoth, The detection of age groups by dynamic gait outcomes using machine learning approaches, Sci. Rep. 10 (2020).

[11] I. Hagoort, N. Vuillerme, T. Hortobágyi, C.J. Lamoth, Outcome-dependent effects of walking speed and age on quantitative and qualitative gait measures, Gait Posture 93 (2022) 39–46.

[12] C. Prakash, R. Kumar, N. Mittal, Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges, Artif. Intell. Rev. 49 (2018) 1–40.

[13] P. Khera, N. Kumar, Role of machine learning in gait analysis: a review, J. Med. Eng. Technol. 44 (2020) 441–467.

[14] R.Z.U. Rehman, Y. Zhou, S. Del Din, L. Alcock, C. Hansen, Y. Guan, T. Hortobágyi, W. Maetzler, L. Rochester, C.J. Lamoth, Gait analysis with wearables can accurately classify fallers from non-fallers: a step toward better management of neurological disorders, Sensors 20 (2020) 6992.

[15] Y. Zhou, J. van Campen, T. Hortobágyi, C.J. Lamoth, Artificial neural network to classify cognitive impairment using gait and clinical variables, Intelligence-Based Medicine 6 (2022), 100076.

[16] A. Samà, D. Rodríguez-Martín, C. Pérez-López, A. Català, S. Alcaine, B. Mestre, A. Prats, M.C. Crespo, À. Bayés, Determining the optimal features in freezing of gait detection through a single waist accelerometer in home environments, Pattern Recogn. Lett. 105 (2018) 135–143.

[17] S.M. Bruijn, J.H. van Dieën, O.G. Meijer, P.J. Beek, Statistical precision and sensitivity of measures of dynamic gait stability, J. Neurosci. Methods 178 (2009) 327–333.

[18] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[19] I.H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN Comput. Sci. 2 (2021) 1–20.

[20] Y. Matsushita, D.T. Tran, H. Yamazoe, J.-H. Lee, Recent use of deep learning techniques in clinical applications based on gait: a survey, J. Comput. Des. Eng. 8 (2021) 1499–1532.

[21] L. Borzì, L. Sigcha, D. Rodríguez-Martín, G. Olmo, Real-time detection of freezing of gait in Parkinson's disease using multi-head convolutional neural networks and a single inertial sensor, Artif. Intell. Med. 135 (2023), 102459.

[22] B.M. Meyer, L.J. Tulipani, R.D. Gurchiek, D.A. Allen, L. Adamowicz, D. Larie, A. J. Solomon, N. Cheney, R.S. McGinnis, Wearables and deep learning classify fall risk from gait in multiple sclerosis, IEEE J. Biomed. Health Inform. 25 (2020) 1824–1831.

[23] F. Luna-Perejón, M.J. Domínguez-Morales, A. Civit-Balcells, Wearable fall detector using recurrent neural networks, Sensors 19 (2019) 4885.

[24] M. Khokhlova, C. Migniot, A. Morozov, O. Sushkova, A. Dipanda, Normal and pathological gait classification LSTM model, Artif. Intell. Med. 94 (2019) 54–66.

[25] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: a review, Data Min. Knowl. Disc. 33 (2019) 917–963.

[26] F. Gu, M.-H. Chung, M. Chignell, S. Valaee, B. Zhou, X. Liu, A survey on deep learning for human activity recognition, ACM Computing Surveys (CSUR) 54 (2021) 1–34.

[27] M. Jaén-Vargas, K.M.R. Leiva, F. Fernandes, S.B. Gonçalves, M.T. Silva, D.S. Lopes, J.J.S. Olmedo, Effects of sliding window variation in the performance of acceleration-based human activity recognition using deep learning models, PeerJ Comput. Sci. 8 (2022) e1052.

[28] N.M. Kosse, S. Caljouw, D. Vervoort, N. Vuillerme, C.J. Lamoth, Validity and reliability of gait and postural control analysis using the tri-axial accelerometer of the iPod touch, Ann. Biomed. Eng. 43 (2015) 1935–1946.

[29] C.J. Lamoth, F.J. van Deudekom, J.P. van Campen, B.A. Appels, O.J. de Vries, M. Pijnappels, Gait stability and variability measures show effects of impaired cognition and dual tasking in frail people, J. Neuroeng. Rehabil. 8 (2011) 1–9.

[30] M.H. de Groot, H.C. van der Jagt-Willems, J.P. van Campen, W.F. Lems, J. H. Beijnen, C.J. Lamoth, A flexed posture in elderly patients is associated with impairments in postural control during walking, Gait Posture 39 (2014) 767–772.

[31] T. IJmker, C.J. Lamoth, Gait and cognition: the relationship between gait stability and variability with executive function in persons with and without dementia, Gait Posture 35 (2012) 126–130.

[32] A. Murad, J.Y. Pyun, Deep Recurrent Neural Networks for Human Activity Recognition, Sensors (basel) 17 (2017).

[33] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, S.-H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (2019) 26–40.

[34] F. Gu, K. Khoshelham, S. Valaee, J. Shang, R. Zhang, Locomotion activity recognition using stacked denoising autoencoders, IEEE Internet Things J. 5 (2018) 2085–2093.

[35] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, I. Rojas, Window size impact in human activity recognition, Sensors 14 (2014) 6474–6499.

[36] D. Borah, S. Wadhwa, U. Singh, S.L. Yadav, M. Bhattacharjee, V. Sindhu, Age related changes in postural stability, Indian J Physiol Pharmacol 51 (2007) 395–404.

[37] M.L. Callisaya, L. Blizzard, M.D. Schmidt, J.L. McGinley, V.K. Srikanth, Ageing and gait variability—a population-based study of older people, Age Ageing 39 (2010) 191–197.

[38] S. Mehdizadeh, The largest Lyapunov exponent of gait in young and elderly individuals: a systematic review, Gait Posture 60 (2018) 241–250.

[39] H. Kobayashi, W. Kakihana, T. Kimura, Combined effects of age and gender on gait symmetry and regularity assessed by autocorrelation of trunk acceleration, J. Neuroeng. Rehabil. 11 (2014) 1–6.

[40] N.M. Kosse, N. Vuillerme, T. Hortobagyi, C.J.C. Lamoth, Multiple gait parameters derived from iPod accelerometry predict age-related gait changes, Gait Posture 46 (2016) 112–117.

[41] M.S. Orendurff, J.A. Schoen, G.C. Bernatz, A.D. Segal, G.K. Klute, How humans walk: bout duration, steps per bout, and rest duration, J. Rehabil. Res. Dev. 45 (2008).

[42] E. Warmerdam, J.M. Hausdorff, A. Atrsaei, Y. Zhou, A. Mirelman, K. Aminian, A. J. Espay, C. Hansen, L.J. Evers, A. Keller, Long-term unsupervised mobility assessment in movement disorders, The Lancet Neurology 19 (2020) 462–470.

[43] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160.

[44] N.B. Kumarakulasinghe, T. Blomberg, J. Liu, A.S. Leao, P. Papapetrou, Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models, 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2020, pp. 7-12.

[45] X. Zheng, B. Otten, M.F. Reneman, C.J. Lamoth, Explaining Deep Learning Models for Age-related Gait Classification based on time series acceleration, Comput. Biol. Med. 184 (2015) 109338.