RESEARCH

Hidden Markov Model based Stride Segmentation on Unsupervised Free-living Gait Data in Parkinson's Disease Patients

Nils Roth[1*], Arne Küderle[1], Martin Ullrich[1], Till Gladow[2], Franz Marxreiter[2], Jochen Klucken[2], Bjoern M. Eskofier[1] and Felix Kluge[1]

Correspondence: nils.roth@fau.de
Machine Learning and Data Analytics Lab, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
Full list of author information is available at the end of the article

Abstract

Background: To objectively assess a patient's gait, a robust identification of stride borders is one of the first steps in inertial sensor-based mobile gait analysis pipelines. While many different methods for stride segmentation have been presented in the literature, an out-of-lab evaluation of respective algorithms on free-living gait is still missing.

Method: To address this issue, we present a comprehensive free-living evaluation dataset, including 146.574 semi-automatic labeled strides of 28 Parkinson's Disease patients. This dataset was used to evaluate the segmentation performance of a new Hidden Markov Model (HMM) based stride segmentation approach compared to an available dynamic time warping (DTW) based method.

Results: The proposed HMM achieved a mean F1-score of 92.1 % and outperformed the DTW approach significantly. Further analysis revealed a dependency of segmentation performance to the number of strides within respective walking bouts. Shorter bouts (<30 strides) resulted in worse performance, which could be related to more heterogeneous gait and an increased diversity of different stride types in short free-living walking bouts. In contrast, the HMM reached F1-scores of more than 96.2 % for longer bouts (>50 strides). Furthermore, we showed that an HMM, which was trained on at-lab data only, could be transferred to a free-living context with a negligible decrease in performance.

Conclusion: The generalizability of the proposed HMM is a promising feature, as fully labeled free-living training data might not be available for many applications. To the best of our knowledge, this is the first evaluation of stride segmentation performance on a large scale free-living dataset. Our proposed HMM-based approach was able to address the increased complexity of free-living gait data, and thus will help to enable a robust assessment of stride parameters in future free-living gait analysis applications.

Keywords: HMM; IMU; machine learning; mobile gait analysis; stride borders; wearable sensors

Background

In typical clinical gait analysis settings, a physician or clinical expert observes a patient's gait in a hospital hallway, to assess for example postural instability or

motor function, which are major symptoms of the neurodegenerative Parkinson's Disease (PD) and strongly related to independence, quality of life or risk of falls [1]. Therefore, the analysis of a patient's gait is an important part of the clinical routine to assess a PD patients' disease state or monitor disease progression [2].

In recent years, wearable gait analysis systems found their way into those hospital settings to enable an objective assessment of a variety of gait characteristics and can therefore help to support clinical decision-making [3, 4]. Due to advancements in size, energy consumption, and usability of wearable sensors, mobile gait analysis systems are getting started to be used beyond the borders of laboratory assessments in home-monitoring as well as free-living scenarios [5]. Here, mobile gait analysis solutions enable a continuous and detailed insight into a patient's mobility and motor performance under more natural and realistic conditions compared to clinical snapshot assessments [5].

However, independent of the recording environment, a robust segmentation of individual strides from the continuous sensor data is one of the first steps in most wearable gait analysis systems and a crucial part of the underlying signal processing pipeline. Various different approaches have been proposed in the literature to solve the problem of stride segmentation for clinical gait analysis applications. Methods vary with sensor location, ranging from wrist, lower back and ankles to foot worn sensors as well as with sensor modalities like inertial measurement units (IMUs) or pressure sensors. Peak detection [6, 7], wavelet-based approaches [8, 9, 10, 11], template matching [12, 13] as well as probabilistic machine learning models like the Hidden Markov Model (HMM) [14, 15, 16, 17, 18, 19], and deep learning approaches like convolutional neural networks [20, 21] have been successfully applied in supervised laboratory conditions.

Although all of the aforementioned studies obtained good results, they were constrained in most cases by a controlled and supervised laboratory setting. Furthermore, most studies only provide a limited amount of heterogeneity within the dataset due to the inclusion of only healthy subjects, a rather small number of strides, or evaluation on standardized gait tests.

In contrast, a continuous full day free-living recording contains a variety of different unstructured activities as well as different stride types like initiation, termination, turnings, or transitions between activities. Influences of the external envi-

ronment like context, location, different undergrounds, or cognitive challenges are known to be confounding factors for certain gait parameters [22, 5]. For PD patients, day-to-day and intra-day fluctuations of motor symptoms (related to health and medication state) increase the heterogeneity and irregularity of strides even further [1].

Recent studies already highlighted the differences between laboratory and free-living gait analysis [23, 24, 25], but a technical evaluation of gait analysis pipelines on free-living data is still an open challenge. A major issue is the lack of free-living evaluation datasets, including ground truth information. For this reason, evaluation studies usually focus on at-lab recordings where for example pressurized mats [17] or motion capture systems [26] are used as gold-standard references. However, such system cannot be used for fully unconstrained free-living studies.

Martindale et al. [19] tried to address this issue by collecting data of several semi-realistic scenarios covering different activities with providing stride level annotations of a large cohort of 80 healthy subjects. They used video recordings combined with pressure insoles and a smart annotation approach to generate labels for all activities. However, the study still followed a supervised protocol with scripted scenarios and only healthy subjects.

Hickey et al. [27] also emphasized the lack of evaluation of gait analysis algorithms on free-living data. To address this issue, they equipped ten young and healthy subjects with a waist-mounted accelerometer together with a body-mounted camera filming the subject's feet on two consecutive days for one hour each. Subjects could then perform free-living activities in a completely unsupervised setting, while the video streams were used to generate reference labels. However, Hickey et al. only evaluated macro gait parameters like walking bout duration and step counts, but did not investigate actual stride border locations.

To address this lack of missing evaluation of stride segmentation algorithms on free-living data, we present a fully unconstrained and unsupervised free-living dataset including stride border annotations of 146.574 single strides of a cohort of 28 PD patients. Due to stigmatization and privacy concerns, we decided to not use body-worn cameras or similar reference systems but opted for a semi-automated manual labeling approach of the IMU data itself. For this semi-automated labeling approach, an existing template matching implementation based on subsequent Dy-

namic Time Warping (DTW), presented by Barth et al. [12], was combined with extensive manual post-processing.

Furthermore, a new HMM-based segmentation approach was implemented for evaluation, as a probabilistic model is a promising candidate to be robust and at the same time flexible enough to handle the expected variety and heterogeneity within free-living patient data. Haji Ghassemi et al. [28] compared different stride segmentation approaches for PD patients and found that an HMM-based segmentation could outperform template matching methods like DTW or simple peak detection methods, reaching F1-Scores above 95 % in laboratory settings. Additionally, HMMs showed the potential to handle also classification or sub-phase segmentation tasks. Martindale et al. [29] and Mannini et al. [17] presented HMMs, which can be applied beyond pure stride segmentation for gait phase segmentation and gait event detection by modelling swing and stance phase by individual models. Panahandeh et al. [30] used a combination of multiple individual HMMs for activity classification based on a cycle level identification. Furthermore, HMMs were successfully applied for at-lab gait analysis in hereditary spastic paraplegia (HSP) patients, where the heterogeneity of the disease required the need of personalized models and proofed their applicability in segmentation of severely impaired and strongly heterogeneous gait [29].

Although the first steps towards the development and evaluation of stride segmentation algorithms on free-living data have been taken, this topic remains an open field in research with the need for optimized and robust methods. In this paper, we make the following contributions to the field of mobile gait analysis: first, we present a comprehensive, free-living dataset of PD patients, including semi-automatically labeled stride borders. Second, a new custom HMM-based stride segmentation approach was developed. The model architecture was chosen to best fit the available training labels as well as to model the expected heterogeneity within free-living gait data. Additionally, to test the generalizability of the data-driven HMM-method, model training was performed first only on available laboratory data and second also on the free-living dataset. To estimate the relative performance of the proposed HMMs, their segmentation results were compared to a non-data-driven state-of-the-art DTW-based approach [12], which was initially used to support the manual annotation process of the dataset. Finally, we further investigated the influence of walking

bout length on segmentation performance to better understand the limitations and challenges associated with stride segmentation on such free-living datasets.

Materials and Methods

Data Acquisition

The dataset used for this work was acquired in the FallRiskPD study (DRKS-ID: DRKS00015085) and consists of a cohort of 28 PD patients (Table 1). The study was approved by the local ethics committee Re-No. 165_18B (Friedrich-Alexander-University Erlangen-Nuremberg, Germany). All patients gave written, informed consent, prior to the data collection, which can be divided into two subsets for this work:

At-lab Dataset

The patients visited the clinic for a routine checkup and performed standardized gait tests in a laboratory environment and under the supervision of clinical experts. Gait tests like the 4x10 m test (in self-selected, slow and fast speed), a 2 min walk test as well as the timed-up-and-go test were recorded using a mobile gait analysis system.

Free-living Dataset

After the clinical visit, the patients were equipped with the same mobile gait analysis system and sent home for two weeks for a continuous home-monitoring approach. During this period, the patients wore the sensor system in their wake-time as much as possible (indoors and outdoors), and followed their normal daily activities.

Table 1 Patient characteristics (N = 28). Parameters are either given by class or by mean ± standard deviation.

Characteristic	
Gender [f/m]	7 / 21
Age [years]	63.7 ± 7.2
Height [cm]	174.6 ± 8.8
Weight [kg]	77.1 ± 15.3
UPDRS-III	17.0 ± 8.1
Hoehn & Yahr	2.5 ± 0.6

All recordings for this study were acquired using the Mobile GaitLab (Portabiles HealthCare Technologies, Erlangen, Germany). The recording system consists of two IMUs (one per foot), attached to the instep position of orthopedic shoes

(Figure 1). Each sensor unit incorporates a 3D-accelerometer (range $\pm 16\,g$) and a 3D-gyroscope (range $\pm 1000\,deg/s$). The data were recorded at 102.4 Hz and left-right synchronized, such that both sensor units share a common time axis. For the first patients in the study ($N = 15$), the dataset had to be manually aligned on a bout level, while the rest of the dataset ($N = 13$) was recorded with an updated version of the Mobile GaitLab featuring hardware synchronization based on a proprietary 2.4 GHz protocol, which was presented in [31]. Both synchronization methods were confirmed to be sufficient for walking bout level definitions by visual inspection. Therefore, all recordings will be handled as a single dataset for this study without any further differentiation between the two utilized hardware versions.

All IMU data were calibrated using the Ferraris method to convert the raw measurement to meaningful physical units of m/s^2 for the accelerometer and deg/s for the gyroscope [32].

Coordinate System

To be able to perform stride segmentation on the left and right foot using a single segmentation pipeline, the individual sensor coordinate systems were transformed to a shared body-frame notation with the three main axes defined as medial to lateral (ml), posterior to anterior (pa), and superior to inferior (si) directions (Figure 1). This coordinate notation will be used throughout the whole manuscript. Additionally, the accelerometer si-axis was aligned to gravity based on static frames within each new walking bout. This ensures a consistent alignment between the sensor- and world frame over the full day recordings and accounts for intra-day coordinate system variations due to the patients putting their shoes on/off as well as attaching/detaching the sensor units.

Semi-automated Stride- and Walking Bout Annotation

To lower the overall annotation effort of the at-lab as well as the free-living dataset, a custom semi-automated labeling tool was used. Therefore, the tool featured an automatic pre-segmentation pipeline consisting of two major steps: First an extraction of activity windows, based on a gyro-norm threshold, and second a pre-segmentation of stride borders within those respective windows. For this purpose, an available DTW-based template matching approach introduced by Barth et al. [12] was used as it did not require any prior training. After applying this first "naïve", automatic

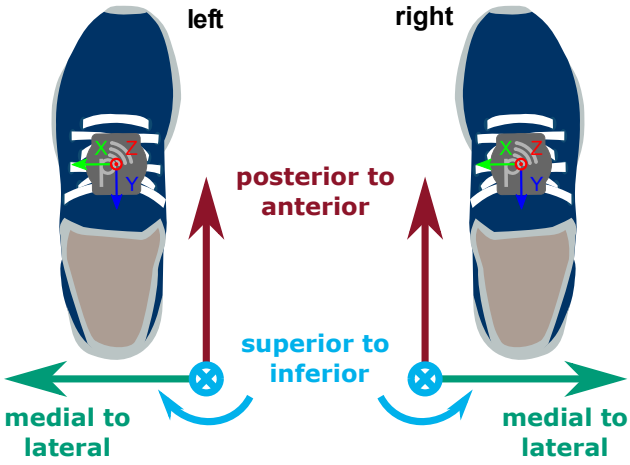

Figure 1 Shoe-/ body-frame coordinate system definition with IMU sensors attached to the instep position of the shoes.

pipeline, each sensor data stream together with the initially generated stride labels was entirely inspected by a human annotator and stride borders were manually added, corrected or deleted wherever necessary. Hence, the overall accuracy of the final reference labels was only dependent on the human annotator performance, this is why a basic yet easy to label stride border definition was chosen.

Stride Definition

The stride definition used for this work was chosen based on previous studies featuring foot-worn IMUs [12, 28], which enabled a consistent manual labeling of clearly visible sensor signal features. A stride was defined based on the angular velocity of the foot in the sagittal plane (gyr_{ml}). The negative peak just before the start of the swing phase was defined as the beginning of a stride, while the negative peak right after the stance phase was defined as its end (Figure 2).

To achieve consistent labels, the timings of the semi-automated DTW borders, as well as of the manually set borders were corrected by snapping the border to the respective signal minimum within a 200 ms centered window on the gyr_{ml}-axis. The strides were labeled for both the left and the right foot individually. However, the sensor streams were inspected on a shared time axis to be able to identify walking bouts and strides more easily. Manual stride border annotations used within this work were performed by the same single human annotator. A quantitative overview of the number of annotated strides, as well as walking bouts, is given in Table 2.

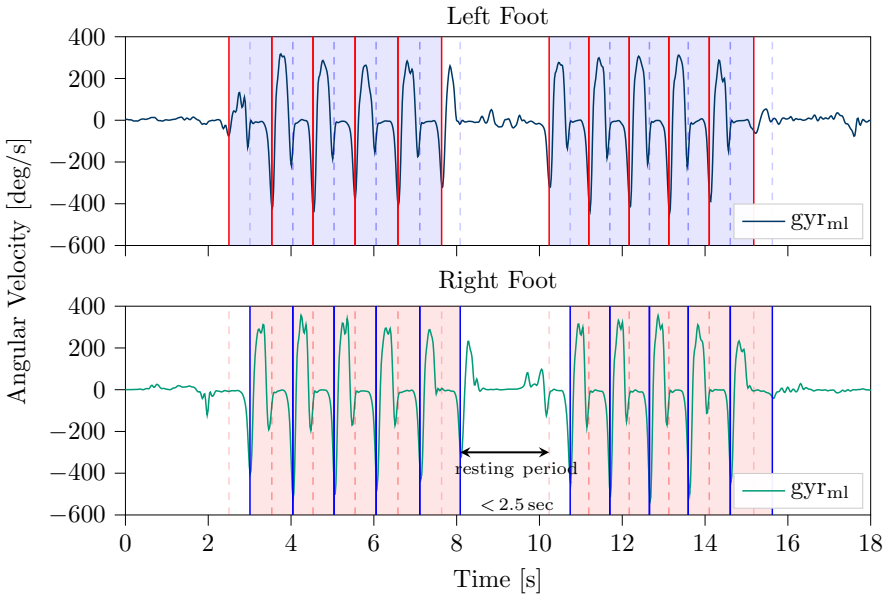

Figure 2 Example data of one walking bout with 20 manually annotated strides and a maximum resting period of 2.5 s. Solid lines represent semi-automatically labeled stride borders, while dashed lines correspond to the borders of the opposite foot.

Due to the extensive labeling amount required, the free-living dataset was restricted to one randomly selected day for each subject of the 14 day recording period.

Table 2 Number of strides and number of walking bouts per subject (N = 28) for the at-lab as well as the free-living dataset.

	N strides		N bouts	
	mean ± std	total	mean ± std	total
At-lab dataset	713 ± 176	19.964	13 ± 3	379
Free-living dataset	5234 ± 2915	146.574	189 ± 84	5318

Walking Bout Definition

In literature, walking bout definitions range from three steps to more than 60 s, based on the used sensor configuration and stride definition [5]. For this work, a walking bout was defined by a minimum number of four strides (based on the synchronized information of both feet) as at least two consecutive strides per foot are required for spatio-temporal parameter calculation using the aforementioned

stride border definition. Additionally, a maximum resting period of 2.5 s within a bout was allowed (Figure 2), as already suggested in other studies [33, 27, 34]. For this work only walking bouts following the given definition were included for the evaluation. Strides outside of walking bouts and non-gait activity windows were not considered.

Preprocessing and Feature Extraction

Preprocessing

For the segmentation task, the data were low-pass filtered using a fourth-order forward-backward Butterworth filter with a cut-off frequency of 10 Hz to minimize motion noise and at the same time, retain stride border features. Additionally, the data were downsampled by a factor of two to 51.2 Hz using a decimation filter to increase the sample to sample difference and lower the computational effort.

Feature Extraction

To include additional temporal information for the segmentation process, a sliding, centered window feature extraction was applied, and used as the input for the segmentation model. Windows were shifted by one sample to guarantee a maximum temporal resolution. Features considered for this work where: the raw data itself, the gradient of the linear regression fit, the signal variance, and the second-order polynomial fit, which were partially derived from literature [28]. Finally, z-score standardization was applied per feature axis per predefined walking bout. Different window sizes and feature combinations were tested during the parameter optimization step.

Hidden Markov Model

To model the sequential nature of human gait, a Hidden Markov Model (HMM) based approach was chosen for this work. An HMM is characterized by a doubly embedded stochastic process [35] of which one stochastic process is described by Markov chains and is referred to as hidden (in this work the sequential human gait model) because it can only be observed through a second stochastic process (the measured IMU data).

The proposed HMM architecture was chosen in a way to fit the available data and labels best. As only stride borders were available, the stride segmentation was

formulated as a two-class problem, where each class was modeled as an individual HMM: namely, strides, which correspond to the labeled data, as well as transitions, which represent the unlabeled part of the data (but might also include resting). Emission distributions for each HMM were represented using Gaussian mixture models (GMMs) to describe the hidden states. GMMs were chosen to model the expected heterogeneity of different transitions as well as stride characteristics, which are expected in free-living data. HMMs were trained in an unsupervised manner using the Baum-Welch (BW) algorithm, which iteratively refines model parameters using expectation–maximization. BW-training was limited to a maximum of ten iterations, which was found to be sufficient during preliminary testing. To be able to represent the multi-class structure of the given data, the trained models for strides and transitions were finally combined to the actual segmentation model (Figure 3). HMMs were implemented using the open source python package pomegranate v0.13.3 [36].

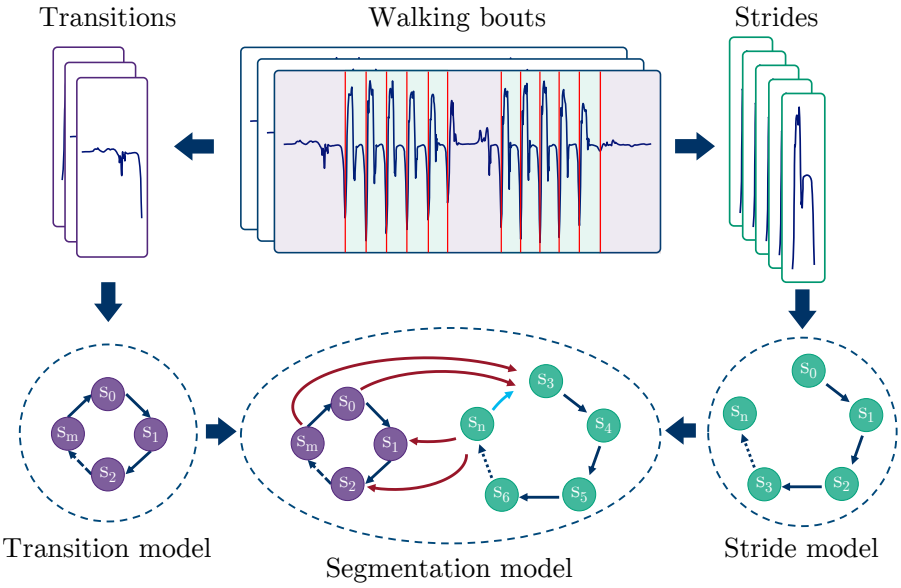

Figure 3 Schematic of a two-class HMM: The transition model and the stride model were trained separately on the respective parts of the training dataset. In a second step, the missing edges were calculated based on the complete training set to connect both models.

Stride Model Architecture

Based on the assumption, that a stride must follow a predefined and repetitive bio-mechanical order (swing and stance phase), the stride model was based on

a strict left-to-right Markov chain. Each hidden state was linked to a sub-phase of a stride. The time-granularity of those sub-phases was defined by the overall number of states within the model. This means that each stride was represented by a monotonic raising state sequence from s_0 to s_n with self transition probabilities $p_{n,n}$ and transition probabilities to the adjacent state $p_{n,n+1}$ as illustrated in Figure 4.

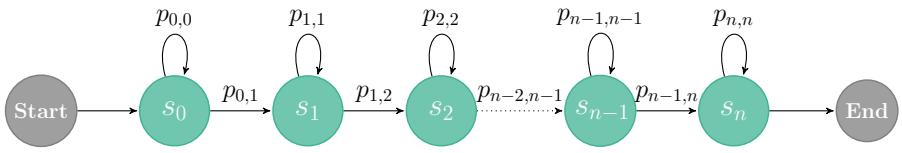

Figure 4 HMM structure of the stride sub-model: A left-right Markov chain from state s_0 to s_n, with transition probability $p_{n,n+1}$, and self transition probability $p_{n,n}$. The sequence is enforced to always start with state s_0 and end with s_n.

For initialization and training of the stride model, all individual strides (according to the manual stride border annotations) were extracted from each walking bout within the training dataset resulting in N training sequences for N labeled strides.

Hidden states were initialized by naïvely dividing each stride into n equally spaced sections to derive initial parameters for n GMMs, with n being the number of hidden states. Existing edges within the transition matrix were initialized uniformly. Transition probabilities, as well as, GMM parameters, were then optimized iteratively using the BW-algorithm in an unsupervised manner.

Transition Model Architecture

Although transitions can also be referred to as a simple "non-stride" or "null-class", a single hidden state would not be sufficient to model the rather complex nature of transitions to and from gait in free-living data. Therefore, the transition class was modeled as a separate HMM. To account for the fact, that a transition will not follow a strict repetitive bio-mechanical order for a single sequence (compared to a stride), a left-right Markov chain was extended by a transition between the last and the first state. Additionally, a transition sequence is allowed to start and end from any state, resulting in the model architecture as illustrated in Figure 5.

Parameter initialization and optimization of the transition model was performed in the same way as for the stride model, by naïve linear split and unsupervised BW-training.

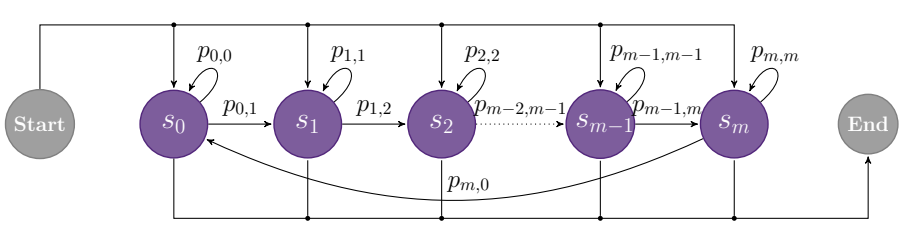

Figure 5 HMM structure of the transition sub-model: A left-right Markov chain from state s_0 to s_m, with transition probability $p_{m,m+1}$, and self transition probability $p_{m,m}$. Additionally, a transition from s_m to s_0 was added. Sequences may start and end from any state.

The training was performed on M transition sequences, where each walking bout of the training dataset included at least two (for start and end of the walking bout) or more transitions. Additional transitions within a valid walking bout were possible following the minimum resting period definition from Section .

Model Combination

To be able to classify strides and transitions simultaneously and find corresponding stride borders, both individually trained models had to be combined to a single flattened HMM. Therefore, their respective transition matrices were concatenated, so that the resulting matrix was of size $(n + m \times n + m)$ with n being the number of states of the stride model and m the number of states in the transition model. To find the edges, which connect both individual models, first, the hidden state sequence was predicted for each training sequence individually (per stride and per transition using the respective class model). Second, the predicted hidden states were merged to a continuous sequence for each walking bout. Based on this now fully hidden state labeled dataset, all possible state transitions could be derived to update the combined transition matrix. The already learned emission distributions were unchanged as these were already optimized during the previous BW-training.

Model parameters like number of states for the stride- and transition model as well as number of GMM components were optimized during grid search (Table 3).

Stride Border Prediction

To perform stride segmentation on unseen IMU data (see Figure 6), first, the most likely hidden state sequence for the respective sensor signal was predicted based on the combined segmentation model using the Viterbi algorithm for inference. Second,

any change from a transition state to a stride state or vice versa, plus transitions between the last and first state of the stride sub-model are considered as stride borders. To match the exact stride border definition of the manual labeling process, the borders were set to the minimum value of the gyr_{ml} axis in the region defined by the two hidden states, between which the transition occurred.

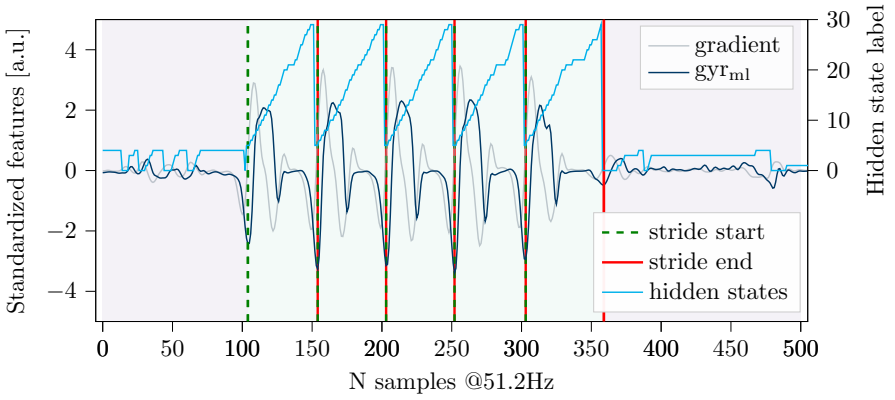

Figure 6 Input signal of a walking bout with predicted hidden state sequence and derived stride borders. States 0 to 4 represent transitions, while states 5 to 29 refer to the stride model.

Evaluation

Performance Metrics

The aim of the stride segmentation was to maximize the number of correctly segmented strides / true positives (TPs) while at the same time the number of missed strides / false negatives (FNs) and the number of falsely detected strides / false positives (FPs) should be minimized. These metrics are commonly represented in terms of precision and recall. As both precision and recall should be maximized for the evaluation of segmentation performance, the F1-score is used, which combines recall and precision by their harmonic mean:

$$precision = \frac{TP}{TP+FP} \quad , \quad recall = \frac{TP}{TP+FN}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Because the stride definition used for this study did not correspond to a specific stride event and small variations of stride borders will not have any impact on

a following event detection and stride parameter calculation, a stride border was considered valid if it was located within a centered ± 60 ms window (± 3 samples) around the ground truth border. Still, a stride was only counted as TP if both, the begin and the end, are within the described margins.

Evaluation on at-lab Data

For baseline evaluation, a 4-fold, nested leave-N-subjects-out cross-validation (CV) was performed on the at-lab dataset. The inner folds were used for grid-search and parameter optimization, while the outer folds were used for the performance evaluation. Inner and outer folds were split such that no subject data, which were used during parameter optimization, were used for evaluation to prevent the model from overfitting.

HMM Grid Search

The grid search parameters (Table 3) used for the model training were partly derived from previous literature [14, 37, 28] and partially from pre-study experiments. The sensor axis was fixed to $\mathrm{gyr_{ml}}$ for this study as it contained the main signal features, which were used to define the stride borders.

Table 3 Grid search values for HMM hyperparameter optimization. Grad = gradient of the linear regression fit, var = signal variance and poly fit = the first three coefficients of the second-order polynomial fit.

Parameters	Values
Window size [ms]	100, 220, 500
Feature combinations	[raw] / [raw, grad] / [raw, var, poly fit]
Number of GMM components	1, 3, 5, 8
Number of states for stride model	5, 10, 15, 20, 25
Number of states for transition model	3, 5, 8, 12

DTW Grid Search

The DTW implementation, which was used for the semi-automated labeling tool, was selected as a state-of-the-art reference method for the proposed HMM. For a fair comparison between the two methods, a parameter grid search was performed for the DTW approach on the at-lab dataset, as well. As for DTW, no actual training step is necessary, the parameter optimization and evaluation were performed only on the outer folds, where the same train-test split as for the HMM evaluation was

used. For DTW the maximum warping cost (*max. cost*) threshold was grid searched in a range from 2.0 to 5.0 in steps of 0.25 with the input feature fixed to the raw data of the gyr_{ml} axis.

For each fold, a new stride template was generated using all strides over all subjects of the train split. For the template generation, each stride was linearly interpolated to the mean number of samples over all strides. Afterwards, the mean per sample over all strides was calculated and used as the template.

Evaluation on Free-living Data

For the final evaluation on the free-living dataset, a repeated nested CV including a grid search was omitted due to the high computational demand required for training the model on free-living data. Instead, a leave-one-subject-out CV with a single optimized hyperparameter set was conducted.

Hyperparameter Selection

To select an optimal hyperparameter set for the HMM training, the mean F1-score per parameter set over all grid search folds (4x4 inner folds) of the previous nested CV on the at-lab dataset was considered. Thus, the parameter set, which reached the on average highest F1-score on the at-lab dataset, was chosen for the free-living evaluation. The optimal *max. cost* threshold for the evaluation of the DTW method on the free-living dataset was directly derived from the at-lab CV as well.

HMM Training Paradigm

To test the generalizability of the proposed HMM, two different training scenarios were applied:

At-lab Training

Here, the model was trained only on laboratory data recorded during the clinical visit (HMM_{lab}). This training scenario was chosen to test the possibility to transfer a lab-trained model to a free-living context. A laboratory setting is still one of the most commonly used gait analysis settings and often used to develop and evaluate models which shall later be used in unsupervised free-living conditions.

Free-living Training

In this scenario, the model was trained solely on the free-living dataset ($HMM_{free\text{-}living}$). Here, a much higher diversity of different stride types, like initiation-, termination-, or turning-strides, or shuffling-gait is expected within the dataset. As the number of walking bouts and, therefore, the number of strides can vary substantially between subjects for the free-living data, a randomly chosen subset of 50 walking bouts per patient was used for training. This was done to balance the influence of individual subjects on the free-living model.

Results

Evaluation on at-lab Data

Both DTW and HMM achieved promising segmentation results on the at-lab dataset (Table 4), while the HMM achieved a slightly better segmentation performance, reaching a F1-score of $96.2 \pm 1.1\,\%$ compared to $94.6 \pm 1.3\,\%$ for the DTW approach.

Table 4 Results of the stride segmentation performance, in controlled laboratory settings. All values are given as mean ± std over the 4 outer evaluation folds.

Method	Precision [%]	Recall [%]	F1-Score [%]
DTW	94.6 ± 1.3	92.1 ± 1.4	94.6 ± 1.3
HMM	96.1 ± 1.1	96.4 ± 1.4	96.2 ± 1.1

0.1 Evaluation on Free-living Data

Hyperparameter Selection

For the HMM method, the optimal hyperparameters set derived from the nested CV were: 8 components for the GMMs, 25 states for the stride model, and 5 states for the transition model, respectively. The best values for feature extraction were a window size of 220 ms as well as a combination of the raw data together with the linear gradient derived from the sliding window view. For DTW, the best performing *max. cost* threshold was 3.5.

Evaluation

Compared to the results of the at-lab dataset (Table 4), the segmentation performance on the free-living dataset dropped by almost 10 % for DTW and by 4 % for the HMM (Table 5). To identify significant differences, paired t-tests were performed for statistical analysis. While no relevant difference between the two training

paradigms (HMM_{lab} vs $HMM_{free\text{-}living}$) could be found, the HMM outperformed the DTW method significantly reaching a maximum F1-score of $92.4 \pm 4.1\,\%$ for the $HMM_{free\text{-}living}$ compared to only $85.1 \pm 9.0\,\%$ for the DTW approach ($p \leq 0.0001$). While the precision is similar for DTW as well as the HMMs, especially the recall value is considerably higher for the $HMM_{free\text{-}living}$ approach reaching $95.6 \pm 2.8\,\%$ compared to only $83.3 \pm 10.1\,\%$ for DTW.

Table 5 Results of the stride segmentation performance, for the free-living dataset. All values are given as mean ± std across the 28 subjects, over all strides per day, per patient.

Method	Precision [%]	Recall [%]	F1-Score [%]
DTW	87.2 ± 8.4	83.3 ± 10.1	85.1 ± 9.0
HMM_{lab}	89.4 ± 5.6	95.2 ± 3.0	92.2 ± 4.2
$HMM_{free\text{-}living}$	89.4 ± 5.7	95.6 ± 2.8	92.4 ± 4.1

Impact of Bout Length

Due to the complex nature of free-living data (like the influence of context, environment or disease fluctuations), the drop of segmentation performance between the at-lab recordings and free-living data was further investigated. Therefore, the different bout groups introduced by Del Din et al. [23] were adopted but converted into the respective number of strides. This resulted in the following walking bout subgroups: $4 \leq N \leq 15$, $15 < N \leq 30$, $30 < N \leq 50$, $50 < N \leq 100$, $100 < N \leq 200$, $N > 200$.

Performance metrics were re-calculated for each of the defined bout groups per patient and visualized by box plots in Figure 7 with detailed results being summarized in Table 6. Grouping the segmentation results by the number of strides per bout revealed a relation between segmentation performance and bout length for all methods. The segmentation F1-score for walking bouts ≤ 30 strides was below the overall free-living average, while bouts ≥ 50 strides were above average and reached comparable performance to the at-lab results for both the HMM as well as the DTW approach.

The HMM-based segmentation approach significantly outperformed the baseline DTW method for all bout groups ≤ 200 strides in terms of F1-score and recall. For the precision, the difference between HMM and DTW was only significant for the bout groups $4 \leq N \leq 15$ and $50 < N \leq 100$, while no significant difference was present between methods for bouts > 200 strides for all metrics.

Impact of Training Paradigm

Although significant differences between the two training paradigms for the HMM were found in some bout groups, the effect size in terms of Cohen´s d was always below 0.2. This indicates only small differences between HMM_{lab} and $HMM_{free\text{-}living}$, which is again reflected by the results in Table 6.

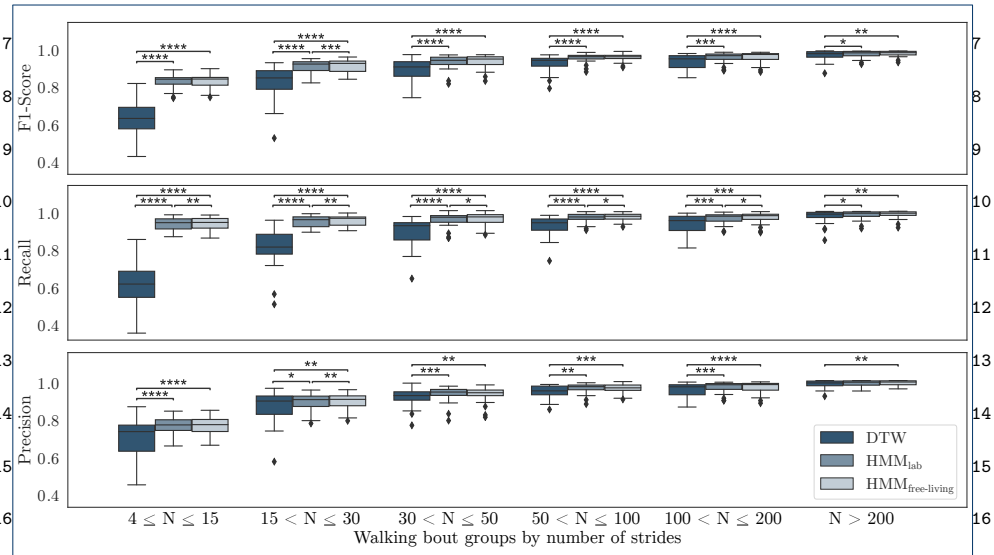

Figure 7 Segmentation performance on the free-living dataset, grouped by the number of strides per walking bout. Significant differences between methods were identified using paired t-tests with *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ and ****: $p \leq 0.0001$.

Table 6 Segmentation performance on the free-living dataset, grouped by the number of strides per walking bout.
All values are given as mean ± std.

N Strides per bout	Precision [%]			Recall [%]			F1-Score [%]		
	DTW	HMM$_{lab}$	HMM$_{free\text{-}living}$	DTW	HMM$_{lab}$	HMM$_{free\text{-}living}$	DTW	HMM$_{lab}$	HMM$_{free\text{-}living}$
$4 \leq N \leq 15$	69.2 ± 10.2	75.7 ± 4.5	75.5 ± 4.5	59.6 ± 12.4	92.7 ± 3.7	93.2 ± 3.7	63.5 ± 10.2	83.3 ± 3.7	83.4 ± 3.8
$15 < N \leq 30$	85.5 ± 8.9	88.2 ± 4.7	88.7 ± 4.7	80.5 ± 10.4	94.1 ± 3.3	94.7 ± 3.3	82.8 ± 9.1	91.1 ± 3.8	91.6 ± 3.5
$30 < N \leq 50$	90.8 ± 5.1	92.7 ± 4.2	92.6 ± 4.2	88.6 ± 7.9	95.0 ± 3.7	95.5 ± 3.7	89.6 ± 6.1	93.8 ± 3.7	94.0 ± 3.6
$50 < N \leq 100$	93.9 ± 3.8	95.7 ± 2.8	95.9 ± 2.8	91.8 ± 5.7	96.2 ± 2.5	96.5 ± 2.5	92.8 ± 4.6	95.9 ± 2.4	96.2 ± 2.0
$100 < N \leq 200$	95.1 ± 3.6	96.3 ± 3.0	96.4 ± 3.0	92.8 ± 5.1	95.8 ± 2.9	96.1 ± 2.9	93.9 ± 4.2	96.0 ± 2.9	96.2 ± 3.0
$N > 200$	97.9 ± 2.2	98.4 ± 1.8	98.7 ± 1.8	96.4 ± 4.0	97.4 ± 2.6	97.8 ± 2.6	97.2 ± 3.0	97.9 ± 2.1	98.2 ± 1.7

Discussion

To the best of our knowledge, this is the first study, which evaluated stride segmentation performance in foot worn IMU data not only on a large number of 19.964 strides in standardized laboratory settings but also on unsupervised, free-living PD patient data including a total of 146.574 annotated strides. To obtain the required

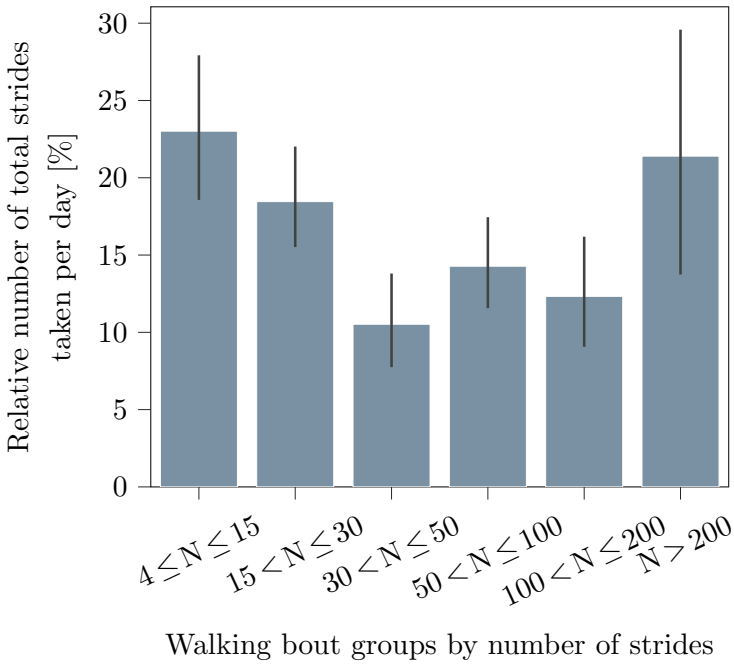

Figure 8 Distribution of the relative number of strides per day for the complete patient cohort. Bars represent the mean and the standard deviation per bout group.

stride borders on both datasets, a semi-automatic manual annotation approach was applied using an existing template matching method based on DTW [12].

Next, a simple yet robust HMM-based segmentation approach was developed. The proposed model was adapted to learn only from manual stride border annotations and best fit the requirements for free-living data analysis. Basic architecture parameters were derived from previous work, where HMM-based stride segmentation was successfully applied on laboratory data [14, 16, 38, 28]. Segmentation performance of the proposed HMM was compared to a state-of-the-art DTW based template matching method [12]. The evaluation of stride segmentation algorithms was continued beyond controlled laboratory settings by validating their performance on free-living PD patient data in a continuous home-monitoring setting. To better understand the impact of the context within free-living gait data, the final evaluation was extended by an in-depth analysis of the impact of bout length on the segmentation performance of the presented methods.

Performance on Laboratory Data

To evaluate the baseline performance and find optimal hyperparameters for the presented segmentation methods, a nested CV including an extensive grid search, was conducted on the laboratory dataset. At-lab performance in terms of F1-score was found to be promising for both methods, while the HMM achieved a higher value of $96.2 \pm 1.1\,\%$ compared to the DTW approach performing slightly worse with $94.6 \pm 1.3\,\%$. These results are almost in perfect agreement compared to previous studies in the literature [28]. However, since the at-lab dataset consists of mainly standardized gait test (performed under supervision in clinical hallways), these results cannot be directly transferred to free-living conditions.

Performance on Free-living Data

Hence, in the second part of the experiment, the segmentation performance was validated on continuous unsupervised free-living data. Here, two different training paradigms were considered for the HMM method. First, the model was only trained on laboratory data to test the generalizability of the approach as annotated free-living training data might not be available in many cases. Second, the model was trained on actual free-living data as a reference. Additionally, DTW was again applied as a baseline method.

As expected, the segmentation performance was worse for both HMM models as well as for the DTW approach, compared to the performance on the lab dataset. This is most likely related to the increased complexity and heterogeneity of free-living data compared to the controlled laboratory data.

Both HMMs performed almost identical in terms of mean F1-score reaching $92.2 \pm 4.2\,\%$ for the HMM_{lab} and $92.4 \pm 4.1\,\%$ for the $HMM_{free\text{-}living}$, respectively. Although performance dropped by approx. $4\,\%$ compared to the laboratory data for the HMM models, they significantly outperformed the reference DTW approach, which only could reach a F1-score of $85.1\,\%$ with a high standard deviation of $9.0\,\%$. These results are a first indication that the DTW approach (which relies on a single template) might not be flexible enough compared to the data-driven probabilistic HMM approach when analyzing free-living data.

Impact of Bout Length

To gain deeper insights into the stride segmentation performance on free-living patient data, results were grouped by previously proposed bout definitions [23]. This revealed a dependency of segmentation performance to bout length for the free-living dataset. Especially for shorter walking bouts with less than 30 strides, F1-scores dropped below average values (reported in Table 5) for all methods, while both HMMs still outperformed DTW significantly. For bouts with more than 50 strides, both, HMM and DTW, achieved a comparable performance as on the at-lab dataset reaching F1-scores of $96.2 \pm 2.0\,\%$ up to $98.2 \pm 1.7\,\%$ for $HMM_{free\text{-}living}$ and $92.8 \pm 4.5\,\%$ up to $97.2 \pm 3.0\,\%$ for DTW.

This relation of segmentation performance and bout length could be explained by an increased percentage of non-steady-strides for shorter bouts. Here, various different stride types like initiation-, termination- or turning-strides as well as shuffling are suspected of generating artifacts or blurring the signal features, which were required to define the stride borders. This could also be confirmed during the manual annotation approach, where stride borders were not always clearly defined particularly for short walking bouts. However, these shorter walking bouts with less than 30, mostly non-steady-strides, make up more than $40\,\%$ of total strides taken per day on average for the presented cohort (Figure 8). This also explains their considerable impact on the overall result. Although the importance of these non-steady strides is still poorly understood within free-living gait data, some studies already showed that this group of strides might yield additional clinical information [39, 40].

In contrast, for longer walking bouts, the proportion of such "non-steady" strides will be lower, while the regularity of strides is expected to rise with increased bout length. This, in turn, would result in a decreased complexity of the underlying IMU signals with clear stride border features and hence explain the better segmentation performance.

The overall F1-score performance of the HMM-based segmentation was dominated by its recall values, which yield significant differences between HMM and DTW up to bouts with ≤ 200 strides, while the precision was only noticeably different for the first bout group with < 15 strides.

This could be explained by the fact that a two-class model for the HMM method might not capture the entire complexity of free-living data and hence the HMM

switches into its stride sub-model too often. However, such false positives might be easily eliminated by a subsequent post-processing step like an event detection. Therefore, methods presented by Rampp et al. [41] could be used for a detection of heel-strike, toe-off, and mid-stance to validate a segmented stride candidate, which should boost the overall precision, while preserving the already high recall.

Impact of Training Paradigm

For the overall F1-score no significant difference between the two models $HMM_{free\text{-}living}$ and HMM_{lab} could be found. Although the $HMM_{free\text{-}living}$ achieved a significantly better recall value compared to the HMM_{lab}, the absolute difference between the two training paradigms was always below the standard deviation with Cohen's d values < 0.2 for all bout groups. The major disadvantage of training the HMM on free-living data is the extensive amount of manual labeling and verification, which was necessary to generate the required ground truth labels. In contrast, the HMM_{lab} has the advantage that reference stride borders required for training can be obtained easier and with less effort, for example, during routine clinical checkups. Therefore, the presented results might indicate a reasonable generalizability of the proposed HMM with the flexibility to handle completely unseen data from a different context due to its probabilistic approach. Thus, a lab trained model could be directly transferred to the free-living domain with some additional improvements regarding the precision using post-processing steps as mentioned before.

Limitations

Generating reference labels on completely unsupervised free-living data is still an open and unsolved challenge in IMU-based gait analysis. In this work, we tried to address this problem by a time costly manual annotation approach. As no alternative gold-standard reference system was available, annotation had to be performed directly on the IMU time-series signal. Although foot-mounted IMUs have the advantage of a high biomechanical resolution, stride borders were not always completely clear and, therefore, dependent on the subjective assessment of the annotator. Hence, the presence of isolated human annotation errors cannot be avoided entirely in the used reference labels. However, due to a large number of almost 150.000 annotated strides, the influence of human annotation errors should be on a negligible scale for the primary outcome of this work.

Furthermore, only predefined walking bouts were considered for the evaluation within this work. Walking bouts were derived from the manual stride border annotations, to avoid potential biases due to additional pre-processing steps (which will require individual evaluation in the future). This means that the presented segmentation model will be dependent on a robust gait sequence detection as a preceding step to be used for continuous free-living data. Methods based on the analysis of harmonic frequencies, as presented by Ullrich et al. [42], could solve this problem.

Conclusion

For this work, a comprehensive evaluation dataset for free-living stride segmentation was presented, including 146.574 semi-automated labeled strides of 28 PD patients. A custom HMM-based stride segmentation approach was introduced and evaluated together with an available state-of-the-art DTW-based reference method on laboratory as well as free-living data. The proposed HMM method was able to outperform the DTW-based approach on both datasets significantly. Further investigations revealed a relation between walking bout length and segmentation F1-score. Especially bouts with ≤ 30 strides showed a reduced performance below average, while bouts with > 50 strides reached a similar performance compared to results in the laboratory setting. Using free-living data compared to lab data for training did not improve the overall segmentation performance noticeably, which could indicate a good generalizability of the model. The HMM demonstrated its strength, especially for the recall metric, while the precision needs further improvement. This issue might be solved by additional post-processing steps to reliably identify false positive strides, for example, by a subsequent event detection or a simple stride-feature based classifier.

Although the presented HMM could improve segmentation performance significantly, compared to the DTW baseline approach, especially the shorter and more challenging walking bouts will require further technical attention. For this purpose, deep learning models based on convolutional neural networks (CNNs) [20, 21] or recently proposed recurrent neural networks (RNNs) [43] might be able to improve segmentation performance even further and help to model the expected heterogeneity within those specific bout groups. Hence, in future work, additional segmentation benchmarks should be carried out on free-living gait datasets to identify

other suitable stride segmentation methods for upcoming applications in the field of continuous free-living mobile gait analysis.

Nevertheless, our proposed probabilistic HMM-based stride segmentation approach proofed to be a promising candidate to handle the increased heterogeneity and complexity within free-living gait. The presented results highlight the challenges of stride segmentation on free-living datasets compared to controlled laboratory assessments as well as the influence of walking bout length on respective segmentation performance. Therefore, our presented work is an important step towards a robust and reliable assessment of stride parameters for future free-living gait analysis applications and respective clinical insights into neurological diseases.

Acknowledgements

The authors would like to thank T. Reichhardt and T. Greinwalder for their effort in recording the dataset as well as all participants of the study for their contributions.

Funding

This work was supported by the Bavarian Ministry for Economy, Regional Development & Energy via the Medical Valley Award 2017 (FallRiskPD Project). B. M. Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (grant number ES 434/8-1). F.K., M.U., A.K. and J.K. received funding from the IMI Mobilise-D project (grant agreement 820820).

Abbreviations

CV: Cross Validation; DTW: Dynamic Time Warping; FN: False Negative; FP: False Positive; HMM: Hidden Markov Model; HSP: Hereditary Spastic Paraplegia; IMU: Inertial Measurement Unit; ML: Medial to Lateral; PD: Parkinson's Disease; TP: True Positive;

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request and after approval by the ethical committee in case patient related data is requested.

Ethics approval and consent to participate

The study was approved by the local ethics committee Re-No. 165_18B (Friedrich-Alexander-University Erlangen-Nuremberg, Germany). All patients gave written, informed consent, prior to the data collection,

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Authors' contributions

N.R. performed the implementation and evaluation of the hidden Markov models, annotated the dataset, and wrote the paper. A.K. and M.U. supported the implementation and conception of the models. T.G., F.M. and J.K. designed the study protocol and supported the data acquisition. B.E. and F.K. supervised the overall conception and design of the work. All authors reviewed the paper and approved the final manuscript.

Author details

[1]Machine Learning and Data Analytics Lab, Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. [2]Department of Molecular Neurology, University Hospital of Erlangen, Erlangen, Germany.

References

1. Mirelman, A., Bonato, P., Camicioli, R., Ellis, T.D., Giladi, N., Hamilton, J.L., Hass, C.J., Hausdorff, J.M., Pelosin, E., Almeida, Q.J.: Gait impairments in parkinson's disease. The Lancet Neurology **18**(7), 697–708 (2019)
2. Klucken, J., Barth, J., Kugler, P., Schlachetzki, J., Henze, T., Marxreiter, F., Kohl, Z., Steidl, R., Hornegger, J., Eskofier, B., *et al.*: Unbiased and mobile gait analysis detects motor impairment in parkinson's disease. PloS one **8**(2), 56956 (2013)
3. Schlachetzki, J.C., Barth, J., Marxreiter, F., Gossler, J., Kohl, Z., Reinfelder, S., Gassner, H., Aminian, K., Eskofier, B.M., Winkler, J., *et al.*: Wearable sensors objectively measure gait parameters in parkinson's disease. PloS one **12**(10), 0183989 (2017)
4. Marxreiter, F., Gaßner, H., Borozdina, O., Barth, J., Kohl, Z., Schlachetzki, J.C., Thun-Hohenstein, C., Volc, D., Eskofier, B.M., Winkler, J., *et al.*: Sensor-based gait analysis of individualized improvement during apomorphine titration in parkinson's disease. Journal of neurology **265**(11), 2656–2665 (2018)
5. Del Din, S., Godfrey, A., Mazzà, C., Lord, S., Rochester, L.: Free-living monitoring of parkinson's disease: Lessons from the field. Movement Disorders **31**(9), 1293–1313 (2016)
6. Lueken, M., ten Kate, W., Batista, J.P., Ngo, C., Bollheimer, C., Leonhardt, S.: Peak detection algorithm for gait segmentation in long-term monitoring for stride time estimation using inertial measurement sensors. In: 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 1–4 (2019). IEEE
7. Lueken, M., Warner, R.T., Kate, T., Valenti, G., Batista, J., Bollheimer, C., Leonhardt, S., Ngo, C.: Estimation of stride time variability in unobtrusive long-term monitoring using inertial measurement sensors. IEEE Journal of Biomedical and Health Informatics (2020)
8. McCamley, J., Donati, M., Grimpampi, E., Mazza, C.: An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. Gait & posture **36**(2), 316–318 (2012)
9. Boutaayamou, M., Brüls, O., Denoël, V., Schwartz, C., Demonceau, M., Garraux, G., Verly, J.G.: Segmentation of gait cycles using foot-mounted 3d accelerometers. In: 2015 International Conference on 3D Imaging (IC3D), pp. 1–7 (2015). IEEE
10. Boutaayamou, M., Denoël, V., Brüls, O., Demonceau, M., Maquet, D., Forthomme, B., Croisier, J.-L., Schwartz, C., Verly, J.G., Garraux, G.: Algorithm for temporal gait analysis using wireless foot-mounted accelerometers. In: International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 236–254 (2016). Springer
11. Prateek, G., Mazzoni, P., Earhart, G.M., Nehorai, A.: Gait cycle validation and segmentation using inertial sensors. IEEE Transactions on Biomedical Engineering (2019)
12. Barth, J., Oberndorfer, C., Pasluosta, C., Schülein, S., Gassner, H., Reinfelder, S., Kugler, P., Schuldhaus, D., Winkler, J., Klucken, J., *et al.*: Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. Sensors **15**(3), 6419–6440 (2015)
13. Vienne-Jumeau, A., Oudre, L., Moreau, A., Quijoux, F., Edmond, S., Dandrieux, M., Legendre, E., Vidal, P.P., Ricard, D.: Personalized template-based step detection from inertial measurement units signals in multiple sclerosis. Frontiers in Neurology **11**, 261 (2020)
14. Pfau, T., Ferrari, M., Parsons, K., Wilson, A.: A hidden markov model-based stride segmentation technique applied to equine inertial sensor trunk movement data. Journal of biomechanics **41**(1), 216–220 (2008)
15. Mannini, A., Sabatini, A.M.: A hidden markov model-based technique for gait segmentation using a foot-mounted gyroscope. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4369–4373 (2011). IEEE
16. Mannini, A., Sabatini, A.M.: Gait phase detection and discrimination between walking-jogging activities using hidden markov models applied to foot motion data from a gyroscope. Gait & posture **36**(4), 657–661 (2012)
17. Mannini, A., Trojaniello, D., Della Croce, U., Sabatini, A.M.: Hidden markov model-based strategy for gait segmentation using inertial sensors: application to elderly, hemiparetic patients and huntington's disease patients. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5179–5182 (2015). IEEE
18. Martindale, C.F., Sprager, S., Eskofier, B.M.: Hidden markov model-based smart annotation for benchmark cyclic activity recognition database using wearables. Sensors **19**(8), 1820 (2019)

19. Martindale, C.F., Roth, N., Hannink, J., Sprager, S., Eskofier, B.M.: Smart annotation tool for multi-sensor gait-based daily activity data. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 549–554 (2018). IEEE
20. Gadaleta, M., Cisotto, G., Rossi, M., Rehman, R.Z.U., Rochester, L., Del Din, S.: Deep learning techniques for improving digital gait segmentation. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1834–1837 (2019). IEEE
21. Steinmetzer, T., Bönninger, I., Reckhardt, M., Reinhardt, F., Erk, D., Travieso, C.M.: Comparison of algorithms and classifiers for stride detection using wearables. Neural Computing and Applications, 1–12 (2019)
22. Ottosson, J., Lavesson, L., Pinzke, S., Grahn, P.: The significance of experiences of nature for people with parkinson's disease, with special focus on freezing of gait—the necessity for a biophilic environment. a multi-method single subject study. International journal of environmental research and public health **12**(7), 7274–7299 (2015)
23. Del Din, S., Godfrey, A., Galna, B., Lord, S., Rochester, L.: Free-living gait characteristics in ageing and parkinson's disease: impact of environment and ambulatory bout length. Journal of neuroengineering and rehabilitation **13**(1), 46 (2016)
24. Tamburini, P., Storm, F., Buckley, C., Bisi, M.C., Stagni, R., Mazzà, C.: Moving from laboratory to real life conditions: Influence on the assessment of variability and stability of gait. Gait & posture **59**, 248–252 (2018)
25. Storm, F.A., Nair, K., Clarke, A.J., Van der Meulen, J.M., Mazzà, C.: Free-living and laboratory gait characteristics in patients with multiple sclerosis. PloS one **13**(5), 0196463 (2018)
26. Zhao, H., Wang, Z., Qiu, S., Wang, J., Xu, F., Wang, Z., Shen, Y.: Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. Information Fusion **52**, 157–166 (2019)
27. Hickey, A., Del Din, S., Rochester, L., Godfrey, A.: Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. Physiological measurement **38**(1), 1 (2016)
28. Haji Ghassemi, N., Hannink, J., Martindale, C.F., Gaßner, H., Müller, M., Klucken, J., Eskofier, B.M.: Segmentation of gait sequences in sensor-based movement analysis: a comparison of methods in parkinson's disease. Sensors **18**(1), 145 (2018)
29. Martindale, C.F., Roth, N., Gaßner, H., List, J., Regensburger, M., Eskofier, B.M., Kohl, Z.: Technical validation of an automated mobile gait analysis system for hereditary spastic paraplegia patients. IEEE Journal of Biomedical and Health Informatics **24**(5), 1490–1499 (2019)
30. Panahandeh, G., Mohammadiha, N., Leijon, A., Händel, P.: Continuous hidden markov model for pedestrian activity classification and gait analysis. IEEE Transactions on Instrumentation and Measurement **62**(5), 1073–1083 (2013)
31. Roth, N., Martindale, C.F., Eskofier, B.M., Gaßner, H., Kohl, Z., Klucken, J.: Synchronized sensor insoles for clinical gait analysis in home-monitoring applications. Current Directions in Biomedical Engineering **4**(1), 433–437 (2018)
32. Ferraris, F., Grimaldi, U., Parvis, M.: Procedure for effortless in-field calibration of three-axis rate gyros and accelerometers. Sensors and Materials **7**, 311–311 (1995)
33. Aminian, K., Najafi, B., Büla, C., Leyvraz, P.-F., Robert, P.: Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. Journal of biomechanics **35**(5), 689–699 (2002)
34. Del Din, S., Galna, B., Godfrey, A., Bekkers, E.M., Pelosin, E., Nieuwhof, F., Mirelman, A., Hausdorff, J.M., Rochester, L.: Analysis of free-living gait in older adults with and without parkinson's disease and with and without a history of falls: identifying generic and disease-specific characteristics. The Journals of Gerontology: Series A **74**(4), 500–506 (2019)
35. Rabiner, L., Juang, B.: An introduction to hidden markov models. ieee assp magazine **3**(1), 4–16 (1986)
36. Schreiber, J.: Pomegranate: fast and flexible probabilistic modeling in python. The Journal of Machine Learning Research **18**(1), 5992–5997 (2017)
37. Martindale, C.F., Strauss, M., Gaßner, H., List, J., Müller, M., Klucken, J., Kohl, Z., Eskofier, B.M.: Segmentation of gait sequences using inertial sensor data in hereditary spastic paraplegia. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1266–1269 (2017). IEEE
38. Martindale, C.F., Hoenig, F., Strohrmann, C., Eskofier, B.M.: Smart annotation of cyclic data using hierarchical

hidden markov models. Sensors **17**(10), 2328 (2017)

39. Nguyen, A., Roth, N., Ghassemi, N.H., Hannink, J., Seel, T., Klucken, J., Gassner, H., Eskofier, B.M.: Development and clinical validation of inertial sensor-based gait-clustering methods in parkinson's disease. Journal of neuroengineering and rehabilitation **16**(1), 77 (2019)
40. Haji Ghassemi, N., Hannink, J., Roth, N., Gaßner, H., Marxreiter, F., Klucken, J., Eskofier, B.M.: Turning analysis during standardized test using on-shoe wearable sensors in parkinson's disease. Sensors **19**(14), 3103 (2019)
41. Rampp, A., Barth, J., Schülein, S., Gaßmann, K.-G., Klucken, J., Eskofier, B.M.: Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. IEEE transactions on biomedical engineering **62**(4), 1089–1097 (2014)
42. Ullrich, M., Küderle, A., Hannink, J., Del Din, S., Gassner, H., Marxreiter, F., Klucken, J., Eskofier, B.M., Kluge, F.: Detection of gait from continuous inertial sensor data using harmonic frequencies. IEEE Journal of Biomedical and Health Informatics (2020)
43. Martindale, C.F., Christlein, V., Klumpp, P., Eskofier, B.M.: Wearables-based multi-task gait and activity segmentation using recurrent neural networks. Neurocomputing (2020)