Machine learning-based distinction of left and right foot contacts in lower back inertial sensor gait data

Martin Ullrich¹, Arne Küderle¹, Luca Reggi², Andrea Cereatti³, Bjoern M. Eskoer¹, and Felix Kluge¹

Abstract—Digital gait measures derived from wearable inertial sensors have been shown to support the treatment of patients with motor impairments. From a technical perspective, the detection of left and right initial foot contacts (ICs) is essential for the computation of stride-by-stride outcome measures including gait asymmetry. However, in a majority of studies only one sensor close to the center of mass is used, complicating the assignment of detected ICs to the respective foot. Therefore, we developed an algorithm including supervised machine learning (ML) models for the robust classification of left and right ICs using multiple features from the gyroscope located at the lower back. The approach was tested on a data set including 40 participants (ten healthy controls, ten hemiparetic, ten Parkinson’s disease, and ten Huntington’s disease patients) and reached an accuracy of 96.3% for the overall data set and up to 100.0% for the Parkinson’s sub data set. These results were compared to a state-of-the-art algorithm. The ML approaches outperformed this traditional algorithm in all subgroups. Our study contributes to an improved classification of left and right ICs in inertial sensor signals recorded at the lower back and thus enables a reliable computation of clinically relevant mobility measures.

I. INTRODUCTION

The objective analysis of gait using wearable sensor systems is gaining importance in the treatment of patients with motor impairments [1]. Various studies have identified gait parameters captured by wearable inertial measurement units (IMUs) as digital measures for gait performance in different diseases like Parkinson’s disease (PD), Huntington’s disease (HD), and post-stroke Hemiparesis (HE) [2]–[4].

From a technical perspective, an important aspect in the signal processing pipeline for IMU-based gait analysis is the segmentation of the recorded data into meaningful portions. One basic unit of gait measurements is a stride, which describes the period from the initial foot contact (IC) of one foot until the next IC of the same foot in a gait cycle [5]. Stride-level information is for example essential for defining walking bouts [6], [7], the measurement of gait variability [8], or the computation of gait asymmetry [9], where differences between left and right strides are investigated.

In their review about the use of wearable motion sensors for gait assessment, Brognoara et al. reported that a majority of studies prefers the use of only one single sensor unit, which is most often attached close to the center of mass of the participant (e.g. on the lower back) [10]. For this sensor setup, the detection of ICs is very well investigated [11]–[14]. Still, to convert a sequence of ICs into strides according to the definition given above, ICs of the same foot need to be grouped and hence the laterality of each IC must be determined. However, this information is usually not directly available from IC detection algorithms and, therefore, stride segmentation is not trivial. Breaks, turns or missed ICs in the detection contradict the assumption of a steady alternating sequence of left and right ICs. Thus, only the explicit determination of the laterality of the ICs allows to define stride borders and additionally enable to differentiate between left and right strides which is, for example, crucial for the investigation of gait asymmetry.

To the best of our knowledge, there is only one publication describing an approach for the distinction of left and right ICs detected from a lower back-mounted sensor: McCamley et al. used the sign (positive or negative) of the lowpass filtered gyroscope signal of the vertical axis at the sample of the detected IC to determine the foot it was performed with [11]. However, the laterality assignment was not separately evaluated in their article. Furthermore, only young and healthy participants took part in their study and the transferability to patients with movement disorders and thus potentially higher gait variability needs to be investigated.

Considering the high relevance of stride level parameters, the left-right distinction is a crucial part of a lower back sensor-based gait analysis pipeline and needs further investigations, given the lack of suitable well evaluated algorithms. Therefore, the goal of this study is twofold: First, we propose potential improvements for the algorithm by McCamley et al. and evaluate the algorithm on data from patients with movement disorders. Second, we present a new approach for the left-right distinction based on supervised machine learning (ML) including a cross-validation. For the experiments, a data set including 40 participants with and without movement impairments, that was previously recorded and presented by Trojaniello et al. [15], was analyzed. The results of our study contribute to a better understanding and an increased reliability of stride-based gait analysis in single-sensor settings.

II. METHODS

A. Data set

The data set contained four groups (healthy controls (HC), patients with Parkinson’s disease (PD), patients with Huntington’s disease (HD), and hemiparetic patients (HE))
TABLE I
DEMOGRAPHIC DATA, GIVEN IN MEAN (SD). A DETAILED DESCRIPTION
OF THE DATA SET IS PRESENTED IN [15].

<table>
<thead>
<tr>
<th>Disease</th>
<th>All</th>
<th>HC</th>
<th>PD</th>
<th>HD</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>severity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age [yrs]</td>
<td>63.6</td>
<td>69.7</td>
<td>73.8</td>
<td>52.2</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>(12.7)</td>
<td>(5.8)</td>
<td>(5.7)</td>
<td>(13.6)</td>
<td>(12.1)</td>
</tr>
<tr>
<td>Height [cm]</td>
<td>166</td>
<td>162</td>
<td>166</td>
<td>166</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>(7)</td>
<td>(7)</td>
<td>(10)</td>
<td>(8)</td>
<td>(6)</td>
</tr>
<tr>
<td>Mass [kg]</td>
<td>68.7</td>
<td>63.6</td>
<td>67.7</td>
<td>61.5</td>
<td>82.1</td>
</tr>
<tr>
<td></td>
<td>(13.0)</td>
<td>(5.7)</td>
<td>(9.3)</td>
<td>(11.2)</td>
<td>(17.2)</td>
</tr>
<tr>
<td>Sex [m/f]</td>
<td>21 / 19</td>
<td>4 / 6</td>
<td>5.5 / 5</td>
<td>7 / 3</td>
<td></td>
</tr>
</tbody>
</table>

aUnified Parkinsons Disease Rating Scale (UPDRS)
bUnified Huntington's Disease Rating Scale (UHDRS)
cFunctional Ambulatory Category (FAC)

with ten participants each (Table I) [15]. The data set was
analyzed as a whole, but also subdivided based on the groups.
In total, this resulted in five data sets that were investigated
separately and will be called All, HC, PD, HD, and HE in
the following.

For the data acquisition, participants were equipped with
an IMU (Opal, APDM Inc., Portland, OR, USA) attached on
the lower back, including a 3-axis accelerometer and a 3-axis
gyroscope recording at 128 Hz. A 7 m long instrumented mat
(GAITRite Electronic Walkway, CIR Systems Inc., Franklin,
NJ, USA) was used to acquire reference data with a sampling
rate of 120 Hz. The sensor and the mat were synchronized
and only the data recorded during the passes on the mat were
considered. All participants performed straight walking back
and forth at comfortable and at higher speed for one minute
each along a 12-meter walkway with the instrumented mat
placed centrally in the area between the turning points [15].

The true time points and left-right labels for all ICs
were available from the instrumented mat. They served as
ground truth label for the training and the evaluation of the
algorithms as described in the following. The IC time points
for this study were used from the ground truth which allowed
the isolation of the task of laterality assignment.

B. McCamley Algorithm and Extensions

In the algorithm by McCamley et al., the gyroscope signal
of the vertical axis (gyr_v) is used to differentiate left and
right ICs [11]. In a pre-processing step, the signal mean is
subtracted and a lowpass filter (4th order Butterworth filter
with 2 Hz cutoff frequency) is applied. Afterwards, the sign
of the filtered gyr_v value at the IC time point n is considered:
In case of a positive value of gyr_v IC_n, the IC is assigned to
the right foot and in case of a negative value it is assigned to
the left foot. The filtered gyr_v signal is suitable, as it behaves
like a periodic wave alternating between left and right ICs
(Fig. 1).

Additionally, we observed the gyroscope signal of the
anterior-posterior axis (gyr_ap) to behave similarly periodic
with a constant phase shift w.r.t. gyr_v after application of
the lowpass filter described above. Therefore, it is also a
suitable input signal for the McCamley algorithm, when
inverting the sign (Fig. 1). As a third potential input signal,
a combination of the filtered signals from the vertical and
the anterior-posterior signals, gyr_comb, was computed by
subtracting the filtered gyr_ap signal from the filtered gyr_v
signal with the goal to enhance the sinusoidal shape of the
signal and improve the robustness of the laterality assignment
based on the sign (Fig. 1):

\[
gyr_{comb} = gyr_v - gyr_{ap}
\]

C. Machine Learning Approaches

1) Feature extraction: Based on the algorithm of McCam-
ley et al., a set of features was extracted from the filtered
gyr_v and gyr_ap signals. Additionally to the filtered raw signal
values (gyr_v IC_n), we calculated the first and second
derivative (gyr_v IC_n) and gyr_v IC_n respectively) at
the time points of the individual n ICs. For a total number
of N ICs in the data set, this resulted in a N \times 6 feature
matrix. Each feature was min-max-normalized to the range
of [0, 1]. The ground truth labels for each IC (left or right)
were provided by the gold standard data.

2) Classification: In the context of ML, the left-right
distinction is essentially a binary classification. We tested
four ML approaches for this task: Support Vector Machines
with linear and radial basis function kernel (SVM-lin, SVM-
rbf), k-Nearest-Neighbors (kNN), and Random Forest Classi-
fiers (RFC). All classifiers were implemented using Python
(version 3.7) and scikit-learn (version 0.23.1) [16].

The classifiers were evaluated in individual leave-one-
participant-out cross-validations for the entire data set as
well as for the respective sub sets. An inner 5-fold cross
validation with grid search (for SVM-lin, SVM-rbf, and
kNN) or randomized search (for RFC) was performed to find
optimized training parameters within a standard parameter
space [17]. For the outer and the inner cross-validation, the
accuracy was used as the optimization target parameter.

The accuracy was computed by counting the number of
agreements between predicted left and right labels and the
ground truth labels and dividing it over the total number of
ICs:

\[
\text{accuracy} = \frac{\sum_{n=1}^{N} \text{pred}_n = \text{true}_n}{N}
\]
D. Comparative analysis

In order to compare the performances of the different approaches, statistical tests were performed. The non-normally distributed accuracy values of the cross-validations of the different McCamley configurations and the ML approaches were compared using Wilcoxon signed-rank tests.

III. RESULTS

In total, 4377 ICs (HC: 1185, PD: 1187, HD: 1066, HE: 939) were available in the data set. An excerpt of the feature space is shown for one participant in Fig. 2. Overall, the best performance was achieved for the PD sub data set and the RFC with a mean accuracy of 100.0% in the leave-one-participant-out cross-validation (Table II). There was no clear tendency regarding the performance of different input signals between sub data sets in the McCamley results. Each $gyr_v$, $gyr_{ap}$ and $gyr_{comb}$ at least once performed best for one of the sub data sets. For all sub data sets, except for HE, the ML approaches reached a significantly higher mean accuracy than the McCamley approaches (exemplarily shown for RFC in Fig. 3). In general, the ML approaches tended to show a smaller interquartile range (IQR) than the McCamley approaches over the cross-validation.

IV. DISCUSSION

In this study, we investigated ML-based algorithms for the distinction of left and right ICs from inertial sensor data measured at the lower back. The pre-processing and the feature extraction were based on the work of McCamley et al. [11], who proposed an algorithm based on thresholding of one single gyroscope axis ($gyr_v$). We also expanded the approach by McCamley et al. by two other signal configurations ($gyr_{ap}$, $gyr_{comb}$) and performed comparisons between the approaches.

---

**TABLE II**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>All</th>
<th>HC</th>
<th>PD</th>
<th>HD</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCamley-$gyr_v$</td>
<td>72.0 (49.6)</td>
<td>84.9 (16.9)</td>
<td>70.7 (50.5)</td>
<td>54.2 (28.9)</td>
<td>78.0 (31.5)</td>
</tr>
<tr>
<td>McCamley-$gyr_{ap}$</td>
<td>87.1 (10.1)</td>
<td>96.1 (4.5)</td>
<td>91.6 (5.0)</td>
<td>93.2 (10.5)</td>
<td>67.3 (41.4)</td>
</tr>
<tr>
<td>McCamley-$gyr_{comb}$</td>
<td>87.9 (13.3)</td>
<td>97.3 (3.0)</td>
<td>92.1 (6.1)</td>
<td>87.0 (14.5)</td>
<td>75.2 (29.4)</td>
</tr>
<tr>
<td>SVM-lin</td>
<td>96.1 (2.2)</td>
<td>99.3 (0.0)</td>
<td>99.9 (0.0)</td>
<td>99.0 (0.8)</td>
<td>79.6 (40.1)</td>
</tr>
<tr>
<td>SVM-rbf</td>
<td>96.2 (1.3)</td>
<td>99.3 (0.0)</td>
<td>99.9 (0.0)</td>
<td>99.1 (0.8)</td>
<td>73.4 (27.1)</td>
</tr>
<tr>
<td>kNN</td>
<td>96.3 (1.8)</td>
<td>99.3 (0.0)</td>
<td>99.4 (0.5)</td>
<td>98.9 (0.8)</td>
<td>82.5 (30.7)</td>
</tr>
<tr>
<td>RFC</td>
<td>95.5 (2.2)</td>
<td>99.0 (0.0)</td>
<td>100.0 (0.0)</td>
<td>98.8 (1.9)</td>
<td>86.4 (30.6)</td>
</tr>
</tbody>
</table>

According to our observations, not only the $gyr_v$ signal, but also the $gyr_{ap}$ signal and a combination of the aforementioned ($gyr_{comb}$) are alternating in a gait sequence with respect to the left and right ICs (Fig. 1). Our results indicate that there are different favorable input signals for different sub data sets, reflecting that different gait impairments exist depending on the respective health condition [18]. This is a valuable extension of the work of McCamley et al., who used only data from young and healthy participants. As the accuracy depends on the clinical population, an a-priori decision per subject for the input signal axis is required (Table II).

The use of ML models allows the simultaneous application of multiple features as input for the binary classification. Besides the signal values, also the values of the first and second derivative at the IC time points were used in our experiments and two different sensor axes were exploited,
resulting in a six-dimensional feature space. The pair plots underline that even if gyroscope data was not separable by means of just the sign of the filtered raw signal, it was possible to separate the classes when using multiple features and sensor axes (Fig. 2). Further visual inspection of pair plots revealed distinct inter-participant differences regarding the separability of the classes by either the \textit{gyr}_x or the \textit{gyr}_x \textit{ap} data on their own. Using a feature vector with different features from the \textit{gyr}_x and the \textit{gyr}_x \textit{ap} axes automatically provides a signal combination, that is more valuable for the classification than selecting only one of the gyroscope axes a-priori (like in the McCamley algorithm) or computing a hand-crafted signal combination like \textit{gyr}_comb.

The performance of the ML approaches can be described as superior compared to the algorithm by McCamley et al. (Table II, Fig. 3). Within each subgroup, the different ML approaches achieved similar accuracy values, indicating that not a specific model, but the separability of the data in higher dimensional space was responsible for the improvements compared to the algorithm by McCamley et al. (Table II). Furthermore, there was no tendency that the group specific models performed better than the model trained on all data, which hints to a good generalizability over clinical populations.

Although the results indicate that the use of ML models for the distinction of left and right ICs is very promising, a limitation of this study is that the time points of the ICs were not derived from a dedicated IMU algorithm, but from an external reference system. Even though the instrumented mat is an accepted gold standard system for gait analysis, it could be possible that the IC time points have an offset compared to the results of a usual IC detection algorithm for IMU signals. However, for this study, the isolation of the task of left-right distinction was desired and the specific origin of the reference ICs can be considered irrelevant for the presented results.

Furthermore, we evaluated the algorithms on an in-lab data set that was heterogeneous regarding the study population but not with respect to the gait movements. It can be expected that the classification will be more challenging for gait movements from real-life scenarios including curvilinear and inclined walking, as well as spurious, unsteady gait bouts.

Hence, in future investigations, the experiments should be extended by assigning ICs detected by an actual IMU-based algorithm. Additionally, a validation on real-world data would be desirable in order to estimate the robustness of the algorithm on signals measured in challenging setups. In this context, an adaptation of the filters during preprocessing and an extension of the feature set should also be considered.

To conclude, our study showed that the optimal sensor axis for the distinction of left and right ICs with the algorithm of McCamley et al. was affected by the investigated study population. Even higher accuracy was achieved by applying ML algorithms with six-dimensional input data. Our results will contribute to a reliable stride-based gait analysis using inertial data measured at the lower back and thus support clinical decisions in the context of movement disorders.

Acknowledgment

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820820. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA. 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). The authors would like to thank Silvia Del Din for her input during the discussions about this study.

References