12 May 2022

Dear reviewer,

Thank you for taking the time and effort for reviewing our submitted manuscript. We appreciate the positive feedback on our works and feel that the issues pointed out helped us to further improve our manuscript.

Below we will go through the comments step-by-step, trying to address these as carefully as possible.

1. **In introduction, the authors should describe and refer what current deep learning methods are used for 'Gait deficits'. What are the advantages and disadvantages of these methods. The author should give more research progress to illustrate the innovation and necessity of this research in introduction.**
    - Thank you for pointing this out. We have tried to paint a more complete picture of deep learning approaches with regard to "gait deficits" by expanding the Introduction section as follows:
        1. *"[…] In the field of image analysis similar requirements have been successfully addressed by algorithms that share a common underlying methodology referred to as deep learning (DL) (LeCun, 2015; Hannink 2016; Ter Haar Romeny, 2019), for example in differentiating diseased from healthy cells (Iqbal, 2021). The main advantage of DL is that rather than relying on expert-defined, hand-crafted features, the algorithm learns relevant data representations automatically (LeCun, 2015; Eskofier, 2016). Furthermore, DL approaches allow for individualization of the algorithm to a specific patient (Stober, 2016; Yao, 2018; Horst, 2019). Previously, DL approaches for wearable IMU data have successfully been applied in classification of bradykinesia (Eskofier, 2016), detection of freezing of gait (Camps, 2018), and prediction of spatiotemporal gait parameters in people with osteoarthritis and total knee arthroplasty (Sharifi Renani, 2020). DL was used to detect gait events from marker-based motion capture, and showed improved performance when compared to conventional, often heuristics-based, algorithms (Kidzinski, 2019; Lempereur, 2020; Filtjens2020). Another study used a DL approach to detect gait events from either three IMUs (worn on the low back, and both ankles) or a single IMU (worn on the low back), and showed that the time error was considerably smaller for the deep learning algorithm than for a commonly applied wavelet-based approach (Gadaleta, 2019). […]"*
2. **The introduction describes the objectives but not the novelty. The authors should highlight the contributions in the introduction (not only in the conclusion).**
    - We have addressed this by adopting the description of the aim of our study: "*To our best knowledge, this is the first study that validates the performance of a DL approach for detecting gait events in a heterogeneous cohort of healthy and neurodegenerative diseases at multiple self-selected walking speeds from short*

walking distances using a single sensor setup that can be worn on either side, either laterally just above the ankle joint or proximally just below the knee joint."

3. **The structure of this paper should be briefly introduced in the 'Introduction'.**
    - We have added the following as an outline of the paper to the Introduction section: *"The structure of the paper is as follows: in the Material and Methods section the data collection, data pre-processing, and the model architecture are introduced. The Results section presents the results from gait event detection, and the subsequently extracted stride-specific gait parameters. In the Discussion, the results are set in relation to relevant literature, and finally in the Conclusions we outline the research content, results and innovations."*
4. **I would suggest that authors should use a summary table to describe for training set, validation and test set. It should be included the sample numbers, labels, sample size, etc.**
    - We thank the reviewer in particular for this comment, and have added an additional table that describes the number of participants, number of trials, and number of instances (examples) for the train, validation and test set. Furthermore, we have added an Excel spreadsheet to the GitHub repository (see; https://github.com/rmndrs89/my-gait-events-tcn, ds_overview.xlsx) that depicts exactly which data were used for each participant, for each trial (preferred, fast, slow speed), and for each sensor location (left ankle, right ankle, left shank, right shank).
5. **What exactly is the input to the network? Is it the raw signals from the accelerometer and gyroscope? Are they input separately or as composite features after fusion? What is the size of the input signal? The author should explain in detail. Is the input signal one-dimensional signal or image?**
    - We have written now explicitly that input to the network is $\boldsymbol{X}$, an array with raw sensor data from the 3-axis accelerometer and 3-axis gyroscope of a single sensor location. Furthermore, from Equation (1), it is given that:

$$\boldsymbol{X} = \begin{bmatrix} x_1[1] & x_2[1] & \cdots & x_D[1] \\ x_1[2] & x_2[2] & \cdots & x_D[2] \\ \vdots & \vdots & \ddots & \vdots \\ x_1[N] & x_2[N] & \cdots & x_D[N] \end{bmatrix}$$

    With $\boldsymbol{x}_d = [x_d[1] \quad x_d[2] \quad \cdots \quad x_d[N]]^{\mathrm{T}}$ with d referring to the dth sensor channel, and n referring to the nth time step.
6. **The structure and parameters (stride, kernel size, kernel number, etc.) of deep learning network should be described in detail in a table.**
    - We have added an additional table that describes the model hyperparameters for each of the relevant layers:

Table 4. Model layer hyperparameters.

Layer #	Layer type	Hyperparameters	Output shape
0	inputs		batch size × 400 × 6
1a	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 1	batch size × 400 × 16
1b	conv	no. of filters: 16 kernel size: 1 stride: 1 padding: same dilation: 1	batch size × 400 × 16
2	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 1	batch size × 400 × 16
3	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 2	batch size × 400 × 16
4	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 2	batch size × 400 × 16
5	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 4	batch size × 400 × 16
6	conv	no. of filters: 16 kernel size: 5 stride: 1 padding: same dilation: 4	batch size × 400 × 16
7a	dense	no. of units: 1	batch size × 400 × 1
7b	dense	no. of units: 1	batch size × 400 × 1

conv: convolutional layer

7. The results of this method should be compared with existing methods to illustrate the advantages of proposed method.

- We have added, in the Discussion section, some additional results from literature, for each of the following
    1. overall detection performance: "*[…] The values for recall, precision and F1 score are comparable to recall (≥ 85%), precision (≥ 95%), F1 score (≥ 91%) from studies that detected gait events in OA, people with PD, and people with MS [34] or adults and hemiplegic patients [80].*"
    2. time errors between the IMU-based and reference methods: *"[…] Previous studies that investigated the time error of IMU-based gait event detection reported a 95% confidence interval of [0.007, 0.013] s for IC, and [-0.005, 0.004] s for FC for young and older adults in treadmill and overground walking [81], or [-0.016, 0.001] s for IC or [0.037, 0.063] s for FC for typically developing children in overground walking [31]. Others reported the mean for the time error in healthy elderly subjects, subjects with PD, subjects with choreatic movement disorder, and hemiparetic subjects, and found a maximum mean error of 0.011 s at normal walking speed, and 0.022 s at faster speed [36]. Similarly, for healthy subjects a mean error of 0.017 s for ICs and -0.016 s for FCs were reported, while for a single transfemoral amputee the mean error was 0.012 s for ICs and -0.024 s for FCs for the intact limb [33]. The median and IQR of the time errors of the current study were in the same range as previously found by [53] and time errors were smaller than previously reported time errors from a continuous wavelet-based approach [80]. Hence, our proposed deep learning*

approach resulted in time errors that are in the same range or better than those from previous approaches, whilst not being restricted to an exact sensor location (left or right ankle, or shank) and sensor alignment."

3. time agreement for the stride-specific parameters: *"[…] The differences between the deep learning model-based stride parameters and the marker-based stride parameters were in a similar range as a recent study that compared IMU-derived stride parameters against stride parameters obtained with a pressure sensing walkway [29,53], and were also in the same range as results from a study that compared IMU-derived stride parameters with stride parameters obtained with an OMC systems [64]. The mean error was lower than the mean errors reported for stance and swing time (0.011 s and 0.011 s, respectively) across elderly subjects, subjects with PD, subjects with choreatic movement disorder, and hemiparetic subjects [36]"*.

8. The conclusion should be improved to outline the research content, results and innovations.

- We have adjusted the Conclusion section, such that it better highlights the innovations and potential applications of the current approach, compared to previous approaches: "*In this study we have validated a deep learning-based approach to detect gait events and subsequently extract clinically relevant stride-specific gait parameters from a single IMU worn either laterally above the ankle joint or proximal below the knee joint. Performance analysis showed an excellent detection rate, and low time errors in both event detection and stride parameter calculation for different walking speeds and across both healthy and neurological cohorts. Compared to relevant approaches that detected gait events from an ankle- or shank-worn IMU, the DL approach reached a performance that was on par or better, while it did not rely on expert-defined, hand-crafted features or empirically derived thresholds. The performance of the DL approach was not affected by the exact sensor placement and orientation, and hence is and will be less obtrusive for potential applications in long-term continuous monitoring. In contrast to previous approaches, it allows for individualization of the network to individual study participants and is easily transferable even to other sensor placement locations (e.g., a foot-worn, or low back-worn IMU) without the need for rethinking the set of decision rules and thresholds. Our next step is to further develop and validate these methods with real-life walking sequences in patients with neurodegenerative diseases."*

On behalf of the authors,

Best regards,

Robbin Romijnders