It is made available under a [CC-BY-NC-ND 4.0 International license](http://creativecommons.org/licenses/by-nc-nd/4.0/) . **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2022.03.30.22272088;](https://doi.org/10.1101/2022.03.30.22272088) this version posted April 2, 2022. The copyright holder for this preprint

Using DeepLabCut for tracking body landmarks in videos of children with dyskinetic cerebral palsy: a working methodology

Helga Haberfehlner¹², Shankara S. van de Ven¹, Sven van der Burg³, Ignazio Aleo⁴, Laura A. Helga Haberfehlner⁴¹, Shankara S. van de Ven¹, Sven van der Burg³
Bonouvrié¹, Jaap Harlaar⁵, Annemieke I. Buizer^{1,6}, Marjolein M. van
¹ Amsterdam UMC, Department of Rehabilitation medicine, Amsterdam Movema
 , Ignazio Aleo⁴
der Krogt¹
ent Sciences , Laura A.

, Annemieke I. Buizer^{4,9}, Marjolein M. van der Krogt⁴
ht of Rehabilitation medicine, Amsterdam Movement Sciences
Sciences, KU Leuven, Campus Bruges
ngineering, TU Delft; Netherlands
msterdam UMC, Amsterdam, The Netherl ¹ Amsterdam UMC, Department of Rehabilitation medicine, Amsterdam Movement Sciences

2 Department of Rehabilitation Sciences, KU Leuven, Campus Bruges

3 Netherlands eScience Center

4 Moveshelf Labs B.V.

Bonouvrié

¹ Amsterdam

² Department

³ Netherlands

⁴ Moveshelf I

⁵ Department

⁶ Emma Chilc , Jaap Harlaar
1997 UMC, Departmen
105 Rehabilitation
105 Science Center
105 B.V.
105 Hospital, Appendence Science
106 5 Department Biomechanical Engineering, TU Delft; Netherlands

6 Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands

Abstract

Abstract
Markerle
postures.
within se
clinically
Method:
94 video postures. It could be a clinically feasible tool to objectively assess movement disorders
within severe dyskinetic cerebral palsy (CP). Here, we aim to evaluate tracking accuracy o
clinically recorded video data.
94 video

position severe dyskinetic cerebral palsy (CP). Here, we aim to evaluate tracking accuracy
clinically recorded video data.
Method:
94 video recordings of 33 participants (dyskinetic CP, 8-23 years; GMFCS IV-V, i.e. non-
 clinically recorded video data.
With severe dyskinetic CP, 8-23 years; GMFCS IV-V, i.e. non-
ambulatory) from a previous clinical trial were used. Twenty-second clips were cut during
lying down as this is a postion for thi Method:
94 video recordings of 33 parti
ambulatory) from a previous c
lying down as this is a postion
move. Video image resolution
We evaluated a model that wa
increasing number of manuall₎ Method:
94 video recordings of 33 participants (dyskinetic CP, 8-23 years; GMFCS IV-V, i.e. nongeneralizability of the model using the remaining 20%. For evaluation the mean absolute lying down as this is a postion for this group of children and young adults allows to freely
move. Video image resolution was 0.4 cm per pixel. Tracking was performed in DeepLabCu
We evaluated a model that was pre-trained nove. Video image resolution was 0.4 cm per pixel. Tracking was performed in DeepLabC
We evaluated a model that was pre-trained on a human healthy adult data set with an
increasing number of manually labeled frames (0, 1, We evaluated a model that was pre-trained on a human healthy adult data set with an
increasing number of manually labeled frames (0, 1, 2, 6, 10, 15 and 20 frames per video). To
assess generalizability, we used 80% of vide increasing number of manually labeled frames (0, 1, 2, 6, 10, 15 and 20 frames per vide
assess generalizability, we used 80% of videos for the model development and evaluate
generalizability of the model using the remainin increasing peralizability, we used 80% of videos for the model development and evaluated the
generalizability of the model using the remaining 20%. For evaluation the mean absolute
error (MAE) between DeepLabCut's predicti generalizability of the model using the remaining 20%. For evaluation the mean absolute
error (MAE) between DeepLabCut's prediction of the position of body points and manual
labels was calculated.
Results:
Using just the p

general (MAE) between DeepLabCut's prediction of the position of body points and manual
labels was calculated.
Results:
Using just the pre-trained adult human model yielded a MAE of 121 pixels. An MAE of 4.5
pixels (abou labels was calculated.
Results:
Using just the pre-trained adult human model yielded a MAE of 121 pixels. An MAE of 4.5
pixels (about 1.5 cm) could be achieved by adding 15-20 manual labels. When applied to
unseen video cl Results:

Using just the pre-train

pixels (about 1.5 cm) c

unseen video clips (i.e

trained on 20 frames p
 Conclusion:

Accuracy of tracking w Results:
Using just the pre-trained adult human model yielded a MAE of 121 pixels. An MAE of 4.5 pixels (about 1.5 cm) could be achieved by adding 15-20 manual labels. When applied to
unseen video clips (i.e. generalization set), the MAE was 33 pixels with a dedicated model
trained on 20 frames per videos.
Conclusion:

provided a clips (i.e. generalization set), the MAE was 33 pixels with a dedicated mode
trained on 20 frames per videos.
Conclusion:
Accuracy of tracking with a standard pre-trained model is insufficiently to automatical trained on 20 frames per videos.
 Conclusion:

Accuracy of tracking with a standard pre-trained model is insufficiently to automatically

assess movement disorders in dyskinetic CP. However, manually adding labels improv **Conclusion:**
Accuracy of tracking with a stand assess movement disorders in dy model performance substantially applicable to check the accuracy Conclusion:
Accuracy of
assess move
model perfo
applicable tc Accuracy movement disorders in dyskinetic CP. However, manually adding labels improves
model performance substantially. In addition, the methodology proposed within our stu
applicable to check the accuracy of DeepLabCut ap assess move in the control of the set of the methodology proposed within our study is
applicable to check the accuracy of DeepLabCut application within other clinical data set.
applicable to check the accuracy of DeepLabCu model performance substantially. In addition, the methodology proposed within other clinical data set.
applicable to check the accuracy of DeepLabCut application within other clinical data set. applicable to check the accuracy of DeepLabCut application with the accuracy of DeepLabCut application within
Contraction with our clinical data set. Accuracy of DeepLabCut application with the set of DeepLaCut applicatio

Movement di
Movement di
associated wi
To treat move
pumps and de
Accurate and Movement disorders in a yellication pale paly (CP), i.e. dyserved and interferential movements¹.
To treat movement disorders, neuromodulation treatments, including implanted medicine
pumps and deep deep brain stimulation associated with impaired muscle tone regulation and interfere with intentional movements*
To treat movement disorders, neuromodulation treatments, including implanted medicine
pumps and deep deep brain stimulation, aim to To treat move measurements of the pathological movements are essential for
Accurate and reliable measurements of the pathological movements are essential for
indication and evaluation of these treatments. However, it remai pumps and deep deep brain stimulation, aim to minimize the pathologic movements "
Accurate and reliable measurements of the pathological movements are essential for
indication and evaluation of these treatments. However, i .
o
c: Indication and evaluation of these treatments. However, it remains a huge challenge to capture the complexity of the dyskinetic movement disorder in an objective way.
Currently dystonia and choreoathetosis are assessed by capture the complexity of the dyskinetic movement disorder in an objective way.
Currently dystonia and choreoathetosis are assessed by a variety of clinical scales, most
which are video recorded and scored afterwards, suc Currently dystonia and choreoathetosis are assessed by a variety of clinical scales,
which are video recorded and scored afterwards, such as the Dyskinesia Impairme
or Barry-Albright Dystonia Scale³. Major drawbacks of t Entering a personal and scored afterwards, such as the Dyskinesia Impairment Scale²
or Barry-Albright Dystonia Scale³. Major drawbacks of these scales are that they measure a
one time point and mostly in a situation th which are video recorded and scored atterwards, such as the Dyskinesia Impairment Scale²
or Barry-Albright Dystonia Scale³. Major drawbacks of these scales are that they measure a
one time point and mostly in a situati $\frac{1}{2}$ or Barry-Albright Dystonia Scale". Major drawbacks of these scales are that they measure at
one time point and mostly in a situation that is unfamiliar to the patient. In addition,
outcomes of these scales are subjective, outcomes of these scales are subjective, i.e. dependent on the personal judgement a
experience of the rater. Therefore clinicians and researchers seek for other possibiliti
objectively assess motor function in dyskinetic experience of the rater. Therefore clinicians and researchers seek for other possibilities
objectively assess motor function in dyskinetic CP^4 . However, up to now there is no
established instrumented clinical assessment experience of the rater. Therefore and the researchers server clinic postablished instrumented clinical assessment method available for children and young adu
within non-ambulatory individuals with dyskinetic CP.
We sugges objectively assess motor function in dyskinetic CP". However, up to now there is no
established instrumented clinical assessment method available for children and you
within non-ambulatory individuals with dyskinetic CP.
W

established interaction and activitative constrained the station and young adults within non-ambulatory individuals with dyskinetic CP.
We suggest a machine learning approach to automatically classify dyskinetic
movement p We suggest a machine learning approach to au
movement patterns, using data extracted by markerle:
the long-term these videos can be recorded within the
As a first step, here we aim to assess accuracy of track
compared to m ent patterns, using data extracted by markerless motion tracking from 2D vide-
g-term these videos can be recorded within the home-environment by caregist
step, here we aim to assess accuracy of tracking body landmarks fro the long-term these videos can be recorded within the home-environment by caregivers.
As a first step, here we aim to assess accuracy of tracking body landmarks from 2D videos
compared to manual labeling in children and yo the long-term these videos compared to manual labeling in children and young adults with dyskinetic CP, using an op
source toolbox (DeepLabCut^{5,6}) with a model pre-trained on a human healthy adult data s
as backbone. Example 2014 and the step of tracking and young adults with dyskinetic CP, using an ope
source toolbox (DeepLabCut^{5,6}) with a model pre-trained on a human healthy adult data so
as backbone.
Methods source toolbox (DeepLabCut^{-y.}°) with a model pre-trained on a human healthy adult data set [,]
as backbone.
<mark>Methods</mark>
P*articipants*
Ninetyfour videos of 33 unique participants of a trial on the effects of intrathecal ba

Methods

compared to manual labeling in children and young adults with dyskinetic CP, using and young adults with dyskinetic CP, using an open-Participants

Methods
Participants
Ninetyfour vid
(IDYS trial Dut |
|
|
| (IDYS trial Dutch Trial Register NTR3642)⁸ from Amsterdam UMC, location VUmc, were use
for the current analysis. Videos were collected at baseline and at two follow-up
measurements within a year's time. Patients at base (IDYS trial Dutch Trial Register NTR3642)° from Amsterdam UMC, location VUmc, were used
for the current analysis. Videos were collected at baseline and at two follow-up
measurements within a year's time. Patients at baseli measurements within a year's time. Patients at baseline had following character
14.0 ± 3.9 (mean ± Standard deviation) years; weight: 32.7 ± 11.8 kg; height: 147
 $\frac{1}{2}$ 14.0 \pm 3.9 (mean \pm Standard deviation) years; weight: 32.7 \pm 11.8 kg; height: 147.1 \pm 20.4 $\frac{1}{2}$.0 $\frac{3}{2}$.9 $\frac{3}{2}$.9 $\frac{3}{2}$.7 $\frac{3}{2}$.7 $\frac{3}{2}$.7 $\frac{3}{2}$

2

cm; 8 females, 25 males; Females Function Classification Classification Class; III (N=3), IV (N=8), V (N=22). This
secondary analysis of the video data was approved by the local medical ethics committee.
Video data
Sequenc

Video data

V (N=20), Manual Ability Classification System (MACS), M(N=3), V (N=2), V (N=22).
Secondary analysis of the video data was approved by the local medical ethics committe
Video data
Sequences of twenty seconds (500 frames) w Sequences of twenty seconds (500 frames) were selected from videos in which children were
lying in rest on a mat. This is a position that enables this group of non-ambulatory individuals
to be assessed without external sup ーー こうしょう しょうしょう しょうしょう しょうしょう しょうしょう しょうしょう しょうしょう しょうしょう lying in rest on a mat. This is a position that enables this group of non-ambulatory individuals
to be assessed without external support. Within all videos the faces of the children and
caregivers were blurred using Mondri It is a mattem in rest on a mattem in the assessed without external support. Within all videos the faces of the children and
caregivers were blurred using Mondrian (version 1.1, Kinedata), a custom-made semi-
automatic pro to be a setternal that the matrice opperation in the face of the third matrice is the face semi-
automatic program. The face blurring was necessary to fulfill the privacy regulations of t
Amsterdam UMC for the succeeding a to the same size (720 width x 575 height pixels, which covered an area of about 3x2 meter. Amsterdam UMC for the succeeding analysis using DeepLabCut within a cloud environment
provided by SURFcumulus (SURF, Utrecht, the Netherlands) based on an Azure cloud
environment (Microsoft Azure, Europe West). Subsequentl Provided by SURFcumulus (SURF, Utrecht, the Netherlands) based on an Azure cloud
environment (Microsoft Azure, Europe West). Subsequently, the videos were all converted
to the same size (720 width x 575 height pixels, whic provironment (Microsoft Azure, Europe West). Subsequently, the videos were all conv
to the same size (720 width x 575 height pixels, which covered an area of about 3x2 n
i.e. an image resolution of 0.4 cm per pixel), and t to the same size (720 width x 575 height pixels, which covered an area of about 3x2 meter,
i.e. an image resolution of 0.4 cm per pixel), and the same video format (avi, svd2500 coded
using Any Video Converter (version 5.7

to the same size (120 minimized angle pixel), and the same video format (avi, svd2500 coded
using Any Video Converter (version 5.7.8, Anvsoft Inc).
We split the data in a development set and a generalization set. 80% of th i.e. an image resolution of 1.1 cm per pixel), and the same video format (avi) state of 1.1 cm pixel)
using Any Video Converter (version 5.7.8, Anvsoft Inc).
We split the data in a development set and a generalization set. We split the data in a development set and a ge
participants (i.e. 27 subjects with 76 related videos) we
development set and 20% of participants (i.e. six partic
generalization set. The development set was used to tra
of Me split the data in a development set and split the data in the spint set and 20% of participants (i.e. six participants with 18 related video ization set. The development set was used to train models with an increasing u participants (i.e. 27 subjects much a state master) in the cancel in the development set and 20% of participants (i.e. six participants with 18 related vide
generalization set. The development set was used to train models dependentlization set. The development set was used to train models with an increasing number of manually labeled frames (as explained in detail below). The videos of the generalization set were manually labeled as well, b generalization set. The development set was used to train model to train model and set were manually labeled as well, but kept apart from the model development process to show the potential of generalizability towards "uns of were manually labeled as well, but kept apart from the model development process to
show the potential of generalizability towards "unseen" videos. The process of splitting the
data and the following processing in DeepL show the potential of generalizability towards "unseen" videos. The process of splitting the
data and the following processing in DeepLabCut is visualized in Figure 1.
DeepLabCut
DeepLabCut is a recently developed open-sou

DeepLabCut

show the pollowing processing in DeepLabCut is visualized in Figure 1.
DeepLabCut
DeepLabCut is a recently developed open-source toolbox that allows training of a deep
neural network using pre-trained models with limited t data and the following processing in DeepLabCut is visualized in Figure 1.
DeepLabCut
DeepLabCut is a recently developed open-source toolbox that allows training of a deep
neural network using pre-trained models with limit |
|
|
| DeepLabCut is a recent of the training data to track user-defined operational methody points using transfer learning^{5,6}. We ran DeepLabCut (Version 2.1) using a single N
Tesla K80 GPU platform via Microsoft Azure's cloud neural network using trained models in the training and to train and training
body points using transfer learning^{5,6}. We ran DeepLabCut (Version 2.1) using a single NV
Tesla K80 GPU platform via Microsoft Azure's cloud w body points using transfer learning^{3,6}. We ran DeepLabCut (Version 2.1) using a single NVIDA
Tesla K80 GPU platform via Microsoft Azure's cloud with a 'Data science Virtual Machine -
Windows 2019)' blueprint. The conda e Testa KB1 GPU proteins and the set of the and the set of the state in the microsoft Wash Windows 2019)' blueprint. The conda environment (for GPU provided by DeepLabCut) was
used within a Jupyter notebook. Models were trai Windows 2019)' analysing the conditionment (for GPU provided by DeepLabCut) was
used within a Jupyter notebook. Models were trained using an available residual neural
Alternative of the condition of the condition of the co used with a Jupyter notebook. Models were trained using an available residual neural neural neural neural neur
In the contract residual neural neural
In

network with 101 layers (ResNet-101) pre-trained on the MPII Human pose dataset","," as
initial weights. This human model is available as built-in option in DeepLabCut. The same
body points included in MPII (i.e. wrists, e infinitive points included in MPII (i.e. wrists, elbows, shoulders, hips, knees and, ankles) were
manually labeled within the development set up to 20 frames per videos with exception c
chin and forehead (these body points manually labeled within the development set up to 20 frames per videos with exception of
chin and forehead (these body points could not be labeled due to blurring of the faces). On
frame per video was labeled within the ge chin and forehead (these body points could not be labeled due to blurring of the faces). One
frame per video was labeled within the generalization set. Frames for labeling were
automatically selected beforehand by DeepLabC frame per video was labeled within the generalization set. Frames for labeling were
automatically selected beforehand by DeepLabCut using k-means clustering to select frames
with a variety of postures within the datasets. frame per video was labeled with the generalization set video was clustering to select
automatically selected beforehand by DeepLabCut using k-means clustering to select
with a variety of postures within the datasets. The with a variety of postures within the datasets. The labeled frames of the development set
were randomly split in a training and test set (95% training dataset, 5% test dataset). The
amount of training and test data is pres were randomly split in a training and test set (95% training dataset, 5% test dataset). The
amount of training and test data is presented in Figure 1 for each dataset. Different model
were first trained on the development amount of training and test data is presented in Figure 1 for each dataset. Different mode
were first trained on the development set with an increasing number of manually labeled
frames (1, 2, 6, 10, 15 and 20) (Figure 1). were first trained on the development set with an increasing number of manually labeled
frames (1, 2, 6, 10, 15 and 20) (Figure 1). Training was performed using the default settings
of DeepLabcut e.g. shuffle is true. All Frames (1, 2, 6, 10, 15 and 20) (Figure 1). Training was performed using the default setting
of DeepLabcut e.g. shuffle is true. All models were trained up to 400.000 iterations with a
batch size of one. The human model wa of DeepLabcut e.g. shuffle is true. All models were trained up to 400.000 iterations with a
batch size of one. The human model was always taken as the intial weight within the
training. The graphs of cross-entropy loss wer batch size of one. The human model was always taken as the intial weight within the
training. The graphs of cross-entropy loss were inspected to determine convergence and
define minimal training iterations needed for our d batch size of one cannot and always taken as the main the gap maint are
training. The graphs of cross-entropy loss were inspected to determine convergence
define minimal training iterations needed for our dataset.
Evaluati

Evaluation

training. The graphs of cross-entropy loss were positional training insulated all models towards their own dataset, i.e. test and train error, and towards
We evaluated all models towards their own dataset, i.e. test and tr define minimal training for an increased for our dataset, i.e. te
Me evaluated all models towards their own dataset, i.e. te
generalization data set, i.e. generalization error (Figure 1).
nerformed within Deepl abCut by ca しいきょう generalization data set, i.e. generalization error (Figure 1). The model evaluation was
performed within DeepLabCut by calculation of the Euclidean distance for the pairs of x,y
coordinates (i.e. manually labeled versus pr performed within DeepLabCut by calculation of the Euclidean distance for the pairs of
coordinates (i.e. manually labeled versus predicted by the model). The mean of Euclid
distances (across all body points and frames) was performed within DeepDatesta, yearn with the Distributed with the pair Euclidean
coordinates (i.e. manually labeled versus predicted by the model). The mean of Euclidean
distances (across all body points and frames) was ta distances (across all body points and frames) was taken as the mean absolute error (MAE).
MAEs were calculated with and without a p-cutoff of 0.8 (i.e. leaving predicitions out with a
low likelihood to be correctly identif MAEs were calculated with and without a p-cutoff of 0.8 (i.e. leaving predicitions out with a

Results

Iow likelihood to be correctly identified by the model).
 **Results

Within the test set MAE decreased from 10.09 (1 labeled frame) to 4.49 pixels (20 labeled

frames) (Table 1). Applying a p-cutoff of 0.8 did not significa** low likelihood is a correct MAE decreased from 10.09 (1 labele
diametality in the test set MAE decreased from 10.09 (1 labele
).
MAEs were reached with the model with 15-20 addition
MAEs were reached with the model with 15 Frames) (Table 1). Applying a p-cutoff of 0.8 did not significantly affect the error. The lowes
MAEs were reached with the model with 15-20 additional labeled frames per video. Withir
the generalization set MAE decreased f MAEs were reached with the model with 15-20 additional labeled frames per video. Within
the generalization set MAE decreased from 107.04 (no labeled frame – i.e. pre-trained
model only) to 33.18 pixels (20 labeled frames) MAES WEED CERTIFY WITH THE MORE WITH THE LIFE DIMANDED WITH THE MORE PLATIFY WITH THE BEATHER THE MORE PORT OF
the generalization set MAE decreased from 107.04 (no labeled frame – i.e. pre-trained
model only) to 33.18 pixe the generalization set MAE decreased from 107.04 (no labeled frame – i.e. pre-trained model only) to 33.18 pixels (20 labeled frames) (Table 1). Applying a p-cutoff within the p-cutoff within the
Table 1). Applying a p-cutoff within the p-cutoff within the p-cutoff within the p-cutoff within the p-cutoff w

Discussion:

genberalization set improved the MAE towards 19.88 pixels (Table 1).
 Discussion:

DeepLabCut provides an accessible platform for tuning models towards an own video

dataset, which is necessary for pathologic movements n dataset, which is necessary for pathologic movements not included in standard human
datasets, especially in postions not commonly include in pretrained datasets such as lyi
down. Within this work we show how to evaluate th datasets, especially in postions not commonly include in pretrained datasets such as lyi
down. Within this work we show how to evaluate the accuracy of DeepLabCut applicati
a clinical data set. We assume that (1) spliting down. Within this work we show how to evaluate the accuracy of DeepLabCut application
a clinical data set. We assume that (1) spliting an available dataset within a development s
and generalization set, (2) applying an inc a clinical data set. We assume that (1) spliting an available dataset within a development set
and generalization set, (2) applying an increasing number of manual labels and (3)
independenly assess the tracking error (MAE) and generalization set, (2) applying an increasing number of manual labels and (3)
independenly assess the tracking error (MAE) is a suitable approach. The same approach
(Figure 1) can also be used towards new, different

independenly assess the tracking error (MAE) is a suitable approach. The same approach
(Figure 1) can also be used towards new, different datasets.
We achieved a tracking error of 1.5 cm, which did not further improve betw Figure 1) can also be used towards new, different datasets.
We achieved a tracking error of 1.5 cm, which did not further improve between 15
labeled frames per video. Within the assessment of dystonia and choreoathetosis w (Figure 1) can also be used towards new, different datasets.
We achieved a tracking error of 1.5 cm, which did not
labeled frames per video. Within the assessment of dystonia
suggest that general movement and posture featu The same of the seases in the seases we track in the sease of dystonia and choreoathetosis we
that general movement and posture features will be required, such as movement
ncy and distance of bodypoint towards the body cen suggest that general movement and posture features will be required, such as movem
frequency and distance of bodypoint towards the body center. Although the accuracy
effect this features, we assume that the accuracy is suf Frequency and distance of bodypoint towards the body center. Although the accuracy may
effect this features, we assume that the accuracy is sufficient for our planned approach to
use this body landmarks for the assessment

effect this features, we assume that the accuracy is sufficient for our planned approach to
use this body landmarks for the assessment of dystonia and choreoathetosis.
We did not systematically compare DeepLabCut with oth effect this body landmarks for the assessment of dystonia and choreoathetosis.
We did not systematically compare DeepLabCut with other options for human pose
estimation such as OpenPose¹¹ or the most recently developed E We did not systematically compare DeepLabCut with other options for
estimation such as OpenPose¹¹ or the most recently developed EfficientPose¹¹
pilotphase of our project, we also tried OpenPose but the results (assess We did not as OpenPose¹¹ or the most recently developed EfficientPose¹². Within a
ase of our project, we also tried OpenPose but the results (assessed by visual
ion of the overlayed videos) did not show sufficient trac estimation such as OpenPose⁴⁴ or the most recently developed EfficientPose⁴⁴. Within a
pilotphase of our project, we also tried OpenPose but the results (assessed by visual
inspection of the overlayed videos) did not s propertion of the overlayed videos) did not show sufficient tracking results for autom
tracking within our dataset. Anyhow, for all tracking options we deem it necessary to
evaluate accuracy of tracking when used in pathol inspection of the overlay cannot in the overlaying tracking options we deem it necessary to
evaluate accuracy of tracking when used in pathologies in childhood as models have mair
developed withing abled bodied adults with tracking within our dataset. Any forest and all tracking options to evaluate accuracy of tracking when used in pathologies in childhood as models have m
developed withing abled bodied adults without taking into account bon developed withing abled bodied adults without taking into account bony deformities or
environmental factors such as a wheelchair or walker in the picture. In addition,
"crowdsource labeling" as used for MPII dataset, might environmental factors such as a wheelchair or walker in the picture. In addition,
"crowdsource labeling" as used for MPII dataset, might not fulfill the requirements of
anatomical representation of the body part, as recent "Crowdsource labeling" as used for MPII dataset, might not fulfill the requiremer
anatomical representation of the body part, as recently highlighted in a review o
al (2021)¹³. Therefore, transfer learning using manual anatomical representation of the body part, as recently highlighted in a review of Cror
al (2021)¹³. Therefore, transfer learning using manual labeling can be also useful in sitt
where exact anatomical positions are nee

al (2021)¹³. Therefore, transfer learning using manual labeling can be also useful in situations
where exact anatomical positions are needed for further processing.
One major problem in the movement analysis in dyskineti al (2021)¹³. Therefore, transfer learning using manual labeling can be also useful in situations
where exact anatomical positions are needed for further processing.
One major problem in the movement analysis in dyskineti One major problem in the movement analysis in dyskinetic CP
using markerless motion tracking from 2D videos, historical data coul
development and extracted body position could be easily pooled with
this way, markerless mot ONET A THE MATHERS motion tracking from 2D videos, historical data could be used for model
One ment and extracted body position could be easily pooled without privacy issues. In
y, markerless motion tracking might open pos development and extracted body position could be easily pooled without privacy issues. In
this way, markerless motion tracking might open possibilities towards development of
this way, markerless motion tracking might open this way, markerless motion tracking might open possibilities towards development of this way, markerless motion tracking might open possibilities towards development of

5

conventional machine realing or deep realing approaches for an access mently developed
for generalized movement analysis¹⁴ and automatic gait assessment in CP¹⁵.
Conclusion:
Markerless motion tracking using Deepl abCut

movement and visitation as dysimilar discution methodologies are currently assisted
for generalized movement analysis¹⁴ and automatic gait assessment in CP¹⁵.
Conclusion:
Markerless motion tracking using DeepLabCut wit for generalized movement analysis⁴⁷ and automatic gait assessment in CP⁴⁵.
Conclusion:
Markerless motion tracking using DeepLabCut with a standard pre-trained m
insufficiently accurate to automatically track body point Conclusion:
Markerless r
insufficiently
dyskinetic CI
15-20 frame
proposed me Markerles in the tracking along 2 Leplan Laternin a standard pre-trained model
insufficiently accurate to automatically track body points in children and young adul
dyskinetic CP. However, accuracy substantially improves b dyskinetic CP. However, accuracy substantially improves by adding manual labels towards
15-20 frames per video. This opens up possibilities for near-automatic tracking of videos. Th
proposed methodology is applicable to as dy substantial contribution, accuracy substantially improves by adding manual labels to an interval 15-20 frames per video. This opens up possibilities for near-automatic tracking of videos. The proposed methodology is app 15-20 frames per video. The specific dip procedure of view and tracking accuracy for
proposed methodology is applicable to assess and improve this tracking accuracy for
different clinical populations. proposed methods of the control of the state of the s different clinical populations.

It is made available under a [CC-BY-NC-ND 4.0 International license](http://creativecommons.org/licenses/by-nc-nd/4.0/) . **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2022.03.30.22272088;](https://doi.org/10.1101/2022.03.30.22272088) this version posted April 2, 2022. The copyright holder for this preprint

- 1 Monbaliu, E. et al. Clinical presentation and management of dyskinetic cerebral palsy. *Lancet Neurol* **16**, 741-749 (2017).
- 2 Monbaliu, E. *et al.* The dyskinesia Impairment Scale: a new instrument to measure dystonia and choreoathetosis in dyskinetic cerebral palsy. *Dev Med Child Neurol* **54**, 278-283 (2012).
- 3 Barry, M. J., VanSwearingen, J. M. & Albright, A. L. Reliability and responsiveness of the Barry-Albright Dystonia Scale. *Dev Med Child Neurol* **41**, 404-411 (1999).
- 4 Haberfehlner, H. *et al.* Instrumented assessment of motor function in dyskinetic cerebral palsy: a systematic review. *J Neuroeng Rehabil* **17**, 39 (2020).
- 5 Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21**, 1281-1289, doi:10.1038/s41593-018-0209-y (2018).
- 6 Nath, T. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* **14**, 2152-2176, doi:10.1038/s41596-019-0176-0 (2019).
- 7 Andriluka, M., Pishchulin, L., Gehler, P. & Schiele, B. *http://human-pose.mpi-inf.mpg.de*.
- 8 Bonouvrie, L. A. *et al.* The Effect of Intrathecal Baclofen in Dyskinetic Cerebral Palsy: The IDYS Trial. *Ann Neurol* **86**, 79-90, doi:10.1002/ana.25498 (2019).
- 9 Andriluka, M., Pishchulin, L., Gehler, P. & Schiele, B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. *IEEE Conference on Computer Vision and Pattern Recognition* (2014).
- 10 Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M. & Schiele, B. DeeperCut: a deeper, stronger, and faster multi-person pose estimation model. *European Conference on Computer Vision 34–50* (2016).
- 11 Cao, Z., Simon, T., Wei, S.-E. & Sheikh, Y. Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291–7299 (2017).
- 12 Groos, D., Ramampiaro, H. & Ihlen, E. A. F. EfficientPose: Scalable single-person pose estimation. *Applied Intelligence* **51**, 2518-2533, doi:10.1007/s10489-020-01918-7 (2020).
- 13 Cronin, N. J. Using deep neural networks for kinematic analysis: Challenges and opportunities. *J Biomech* **123**, 110460, doi:10.1016/j.jbiomech.2021.110460 (2021).
- 14 Chambers, C. *et al.* Computer Vision to Automatically Assess Infant Neuromotor Risk. *IEEE Trans Neural Syst Rehabil Eng* **28**, 2431-2442, doi:10.1109/TNSRE.2020.3029121 (2020).
- 15 Kidzinski, L., Delp, S. & Schwartz, M. Automatic real-time gait event detection in children using deep neural networks. *PLoS One* **14**, e0211466, doi:10.1371/journal.pone.0211466 (2019).

The proje
(ZonMW,
Helga Hal
Seal of Ex
Seal of Ex The project is funded by the project intitive (IMDI) project number 104022005)
The project intitive (IMDI) project number 104022005)
Helga Haberfehlner is funded by the Postdoctoral Fellow Marie Skłodowska-Curie Actions
Se (Demain), and calculated above initiative (IMDI), project number of the Research Fostdoctoral Fellow Marie Skłodowska-Curie
Seal of Excellence of the Research Foundation – Flanders (SoE fellowship_12ZZW)
Seal of Excellence Seal of Excellence of the Research Foundation – Flanders (SoE fellowship_12ZZW22N).
Note of Excellence of the Research Foundation – Flanders (SoE fellowship_12ZZW22N). Seal of Excellence of the Research Foundation – Flanders (SoE fellowship_12ZZW22N).

Labeled	Train set	Test set	Test set	Generalization set	Generalization set
frames			with p-cutoff 0.8		with p-cutoff 0.8
Ω				107.04 pixels	121.18 pixels
1	1.17 pixels	10.09 pixels	10.23 pixels	37.24 pixels	28.72 pixels
2	1.11 pixels	5.83 pixels	5.83 pixels	39.00 pixels	28.89 pixels
6	1.65 pixels	5.14 pixels	5.06 pixels	32.82 pixels	23.51 pixels
10	1.8 pixels	5.86 pixels	5.18 pixels	34.68 pixels	25.91 pixels
15	2.36 pixels	4.49 pixels	4.48 pixels	36.26 pixels	28.33 pixels
20	2.71 pixels	4.49 pixels	4 48 pixels	33.18 pixels	19.88 pixels

Table 1.: Mean averaged error (MAE) in development set (train, test) and in the generalization set, evaluated with and without p-cutoff 0.8

