A novel markerless multi-camera tool to early predict neurodevelopmental diseases like cerebral palsy in infants-at-risk

Author: Roos Bos, master student of Neuroscience & Cognition at Utrecht University.
Supervisor: Laurens Witter, staff member of DDOD at Wilhelmina's Children's Hospital in Utrecht.

Abstract

General movement assessment (GMA) is the most predictive tool for physical impairments like Cerebral palsy (CP). However, GMA has its limitations which might be diminished with the utilisation of modern deep-learning tools. The software packages DeepLabCut (DLC) and Anipose offer a markerless multi-camera approach to build an extensive and dependable open-source database. Fifteen congenital heart patients and prematures, infants-at-risk for CP, were performing a markerless GMA surrounded by three cameras. With a MATLAB 2D analysis, we show the achievement of DLC on markerless labelling of limbs on par with human labelling. Retraining the neural networks offer refinement of more challenging markers like smaller joints. The 3D reconstruction, obtained with Anipose, showed good tracking as indicated by a constant length of rigid bodies. Furthermore, our preliminary MATLAB results show the possibility to analyse positional data in 3D and kinematics of limbs and joints. We compared those aspects of one infant, with the neurological and MOS-score outcomes. Those findings established that a 3D reconstruction can reveal different precise kinematic parameters, but the database and the parameters should both be expanded.

Key words: Cerebral Palsy, Infants-at-risk, General Movement Assessment, Deep-learning, Artificial intelligence, DeepLabCut, Anipose

Background

Cerebral palsy (CP) is defined as a group of disorders of movement and posture development causing activity limitation that are attributed to non-progressive disturbances that occurred in the developing foetal or infant brain[1]. The prevalence of CP is 2 out of 1.000 live births [2]. The most common increased risks to develop CP are being born at a low weight or before 28 weeks gestational age (GA), neonatal arterial ischemic stroke (NAIS), hypoxic-Ischemic encephalopathy (HIE) or perinatal asphyxia [3,4]. CP diagnosis is quite a challenge because most of those infants-at-risk are developing a normal motor outcome [3]. The probability to develop CP in infants born preterm ranges from 5 to 15% and up to 40% in infants born at 22 to 25 weeks GA [4]. Among infants born extremely premature, with major brain injury like asphyxia or NAIS, only a third will develop CP. In addition, 29% of infants born at term with CP were found to have a diagnosis of HIE at birth [3,4]. Besides that, infants have activity limitations, because they could not perform functional tasks like writing or climbing stairs. The presentation of motor type and topography classifications between infants and older children differ and there is a lack of definitive biomarkers at this stage. Therefore, an early diagnosis of CP is quite a challenge and there is no clear definition of CP in an infant who is a few months old. For this reason, severe symptoms of CP can only be diagnosed after 11 months but mild cases can also be diagnosed up to 24 months of age [5]. Even with those difficulties, it is crucial to find early intervention within the first year of the infant's life. It has been proven that it has an impact on the neurodevelopment and functional attainment, which improves neuroplasticity of the developing brain and motor outcome in childhood later-on[6].

There are standardised tools to detect the risk of motor developmental diseases like CP within the first year[6,7]. Those tools include, for instance, cranial ultrasound, magnetic resonance imaging (MRI), and neurological examination. On neonatal MRI, predictors of CP may be identified as an asymmetrical development or injury of the posterior limb of the internal capsule (PLIC), middle cerebral peduncle or white matter lesions in the corticospinal tracts (CST)[6,8]. However, several studies have proven that abnormalities in general movements (GMs) are the most predictive for the development of CP later-on, with a sensitivity of 98% and a specificity of 91% [7]. General movement assessment (GMA) looks at spontaneous movements in foetuses, preterm and term-born infants during their first few months of life, defined as gross movements of body parts like the neck, arm and trunk with no rhythmic pattern or regulatory sequence. GMs change in intensity, force, and speed and can be classified in different definitions depending on the age of the infant [9]. At term age and shortly after, the GMs called writhing movements (WMs), while the Fidgety movements (FMs) last between 6 till 8 weeks. FMs become pronounced at 12-16 weeks and vanish around 20 weeks of post term age (PTA)[9].

The Prechtl GMA, based on the gestalt perception, is the most commonly used GMA. Einspieler adapted the Prechtl to a more detailed GMA with different score sheets for separate body parts like the upper and lower extremities, the neck and the trunk[7]. Secondly, Einspieler came with another novelty, the motor optimality score **(**MOS) which also contributes to understanding later CP prognosis, but also includes early predictors which distinguish the diverse types and the CP severity [10]. In general, the unobtrusive way of GMA screening in infants is a huge advantage. It lasts 3 to 5 minutes and interrupting the infant is not necessary, because it consists of an observational video screening. The method is therefore very applicable and inexpensive. Although those novelties are important, GMA still lacks in various aspects. It is a subjective tool because of the enormous variability of human judgements. Furthermore, GMAs are committed by certified assessors. Expertise training of those assessors limits this tool's implementation as a daily clinical routine. There are still more techniques needed to scale up this valuable tool whereby the modern deep learning technology could be helpful [11–13]. Automatic motion analysis has already been carried out on multiple animals and humans[14]. An automatic model for GMA would make outcomes more efficient and objective because it is a repetitive system and independent from the assessor's tiredness or emotional state[9]. Additionally, generation of an automatic tool would help researchers to compare their studies and build an extensive and dependable open-source database. So far, studies already focus on an automatic GMA model [15–21]. However, none of

them did a markerless 3D analysis which supplies more precise kinematic parameters than a 2D analysis [22].

In this paper, we implemented markerless 3D analysis on approximately 3 months old infants-at-risk performing a GMA during FMs. We have included fifteen infants who performed a GMA. Besides this, we also performed another assessment called hand assessment (HAI) on thirty-one infants-at-risk. The HAI contains a playful session to assess hand movements in 6 to 9 months old infants-at-risk. We used the software package DeepLabCut (DLC) to label markers on the videos of the fifteen GMA performances. The software package Anipose is used to obtain a 3D reconstruction out of the DLC output. Before that, we critically analysed the 2D output of DLC in MATLAB, which shows that using a multi-camera approach combined with multiple deep neural networks, achievement of limbs labelling on par with human labelling is reached. Digits, smaller joints, are more challenging for the neural network, but become better whenever the neural networks are reputationally refined. The 3D reconstruction enables a 3D scatterplot, calculation of the length of rigid bodies and information about the limb and joint kinematics. Above this, we compared those aspects with the neurological outcome assessed by a fellow-neonatologist, and with the manual Einspieler MOS-score assessed by occupational therapists.

The aim and hypothesis

The aim of this study is to assess whether precise movement kinematics, during the FMs stage in approximately 3 months old infants-at-risk for cerebral palsy, could be detected from a markerless multi-camera approach combined with the software packages DLC and Anipose[11,12,23]. We hypothesised that these precise movement kinematics correlate with the manual MOS-score of the movements and with the integrity of the CST as assessed from the MRI scans. Such insights are crucial to set individual therapy goals and thus optimise the infant's functional potential.

List of abbreviations

AOV	-	Angle Of View
BW	-	Birth Weight
CA	-	Corrected Age
CP	-	Cerebral Palsy
DLC	-	DeepLabCut
FM	-	Fidgety Movements
GA	-	Gestational Age
GMA	-	General Movement Assessment
HAI	-	Hand Assessment for Infants
HIE	-	Hypoxic-Ischemic Encephalopathy
MOS	-	Motor Optimality Score
MRI	-	Magnetic Resonance Imaging
NAIS	-	Neonatal Arterial Ischemic Stroke
PLIC	-	Posterior Limb of Internal Capsule
PTA	-	Post Term Age
SMA	-	Spinal Muscular Atrophy
STD	-	Standard Deviation
WM	-	Writhing Movements

Contents

Material & Methods

Participants

This study holds an interdisciplinary collaboration of the DDOD laboratory, the Center for Child Development, Exercise and Physical Literacy and the neonatology department of the Wilhelmina's Children hospital in Utrecht. One hundred participants were carefully selected from the neonatology- and the congenital heart outpatient clinic to perform a GMA or a HAI. For the GMA, fifteen participants were included and did perform the task. Inclusions, exclusions, and specific numbers are given in appendix 1. Besides, appendix 2 contains a Layman's summary, describing the project specifically for parents. The group of fifteen participants consist of ten congenital heart patients (gestational age (GA) range 37 – 40 weeks), three prematures (GA range 25 – 31 weeks) and two near-prematures defined as near-term (GA range 33 – 34 weeks). A detailed overview is given in table 1.

Video measurements

Data collection is performed with usage of the software package Bonsai, which enables synchronous multi-camera recording. Three high-quality IDS UI-3060CP-C-HQ_Rev_2 cameras, with resolution of 1936 x 1216 pixels and acquisition speed of 100 Hz, have been utilised. The participants underwent a GMA with a duration of 4-10 minutes without interruptions and within a low-stimulus environment. Figure 1 illustrates the set-up of a markerless GMA performance. The Brazelton's states of Arousal [24] are used to quantify the behavioural state of the infant during the measurements. The author refers to appendix 3 for a detailed description of each state of arousal. The video duration, their corrected age at the time of measurement (CA) and their state of arousal is also noted in table 1. As already mentioned in the background, it is important to measure the infants between 12-20 weeks PTA to detect the FMs during GMA measurement. All the infants have a PTA within this range, so have performed a GMA during their FMs period. The infants born preterm (5,8,9,10 and 14) must be corrected for their age (CA) and all do fall within the range as a result.

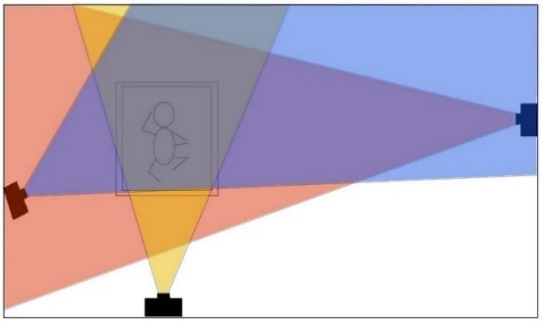


Figure 1. The setup of the markerless GMA performance. Three cameras are pointing to the box wherein an infant is performing a GMA. The red coloured area is the angle of view (AOV) of the right camera, the yellow AOV defines the front camera, while the blue one defines the left camera.

Neurological outcome

A fellow-neonatologist, from the neonatology department of the Wilhelmina's Children hospital in Utrecht, critically analysed the brain MRI-scans of the fifteen included infants. The MRI scans were all performed any days after the infant's GA. The fellow-neonatologist concluded for each infant if there were abnormalities, the location of this etiology and the associated risk for motor impairments like CP. A detailed overview of this is given in table 4.

Manual MOS-score of GMA

Two occupational therapists, from the Center for Child Development, Exercise and Physical Literacy at the Wilhelmina Children's Hospital, manually scored the fifteen children on their movement quality and spontaneous motility by using the MOS-score [25]. The scoring was based on the front camera video of each infant, which are the same used for the 3D-model. The assessors were blinded from the 3D video-analysis or neurological outcome when scoring.

Table 1. Characteristics and video information of the fifteen included participants

Infant	Category	Gender	GA (Week + Day)	BW (Gram)	Video (min)	PTA (weeks)	CA (weeks)	State of arousal
1	Cardio	Male	40 + 0	3350	10	14.0	14.0	Active awake
2	Cardio	Male	39 + 1	4000	6	16.1	15.3	Crying
3	Cardio	Female	39 + 3	2400	8	13.7	13.1	Quiet alert
4	Cardio	Female	40 + 0	3700	10	13.3	13.3	Active awake
5	Preterm	Male	31 + 0	1850	6	28.6	19.6	Crying
6	Cardio	Female	39 + 5	3020	10	14.6	14.3	Active awake
7	Cardio	Male	38 + 5	3000	8	13.6	12.3	Active awake
8	Near-term	Female	33 + 2	1965	10	23.0	16.3	Active awake
9	Near-term	Male	34 + 5	2844	10	24.1	18.9	Active awake
10	Preterm	Male	25 + 4	670	10	31.0	16.6	Active awake
11	Cardio	Male	38 + 2	3536	10	13.6	11.9	Active awake
12	Cardio	Male	38 + 6	3078	10	12.9	11.7	Active awake
13	Cardio	Male	37 + 1	2657	6	14.6	11.7	Active awake
14	Preterm	Male	31 + 6	1670	7	25.6	17.4	Active awake
15	Cardio	Male	38 + 6	3980	4	18.0	16.9	Quiet alert

2D-video analysis

For the quantitative GMAs, videos were annotated using the deep learning DLC tool [11,23]. Markers were carefully selected and manually applied upon the digits and limbs at twenty frames of each infant's video. Those manual placed markers provide DLC neural network training to track the positions of those markers at the remaining frames [26]. To have a clear overview of the working mechanism, the DLC protocol published a working flow, illustrated in figure 2 of the paper[26]. To provide a specific protocol for this research project, a manual has been created which explains each of those steps, with a given time-indication, in more detail (appendix 4).

To enable comparisons with the manual MOS-score of the GMAs, we applied markers such as the nose, eyes, bigger joints (e.g., shoulder, hip, elbow, knee) and smaller joints (e.g., toe, MCP- and top digits). Those markers were trained by DLC to reconstruct movement of the arms, trunk, head, legs, hands, and feet. First, a literature search in choice of markers in earlier GMA studies was performed[15,16,18,27,28]. Due to the complexity of human motion, the infant's movement can only be obtained in different body segments. Those multiple rigid bodies could be defined as star elements which do not vibrate or deform. To obtain limb kinematics (e.g., thigh), at least three markers (e.g., hip, knee, and ankle) are necessary to track the rigid body over time. The joints could be seen as switching points between the rigid bodies of the infant's body, which provide calculation of both rigid bodies and it minimises friction or marker detection errors. To calculate movement of all rigid bodies, we placed markers on the joints, but also on the end of body parts to still have three markers for the lower legs and digits (e.g., toe and top digits). The direction of the infant's head can be determined by the position of the nose and/or eyes. In addition, upper-limb digit markers were also applied to get comparable parameters within the MOS-score [25] and to test whether those movements could also be a reliable predictive motor outcome in infants around 3 months of age [29]. An overview of the selected markers, and the rigid bodies in between, are given in figure 2. Another requirement of human movement detection is that the markers should not be lined up in a row, but they should be spread out across the surface to provide 3D distributions. All earlier GMA 2D videos do lack this essential requirement, however our research project provides calculations of movement

capabilities in 3D distribution. Finally, all markers need to have unique markers configuration to keep them separate, which can be easily achieved through the DLC tool by labelling them.

To inspect the labelling errors, MATLAB has been used to critically analyse the 2D output of DLC. Eight infants performing a HAI were analysed in MATLAB to test whether it would be better to train the videos of each different camera together or separately in DLC. After this, three separate networks have been trained for thirteen infants performing the GMA (infant 1-13 of table 1). The output of those three neural networks were screened in MATLAB and retrained afterwards. In the network refinement stage, infant 14 and 15 were added. This means that those infants are only trained once.

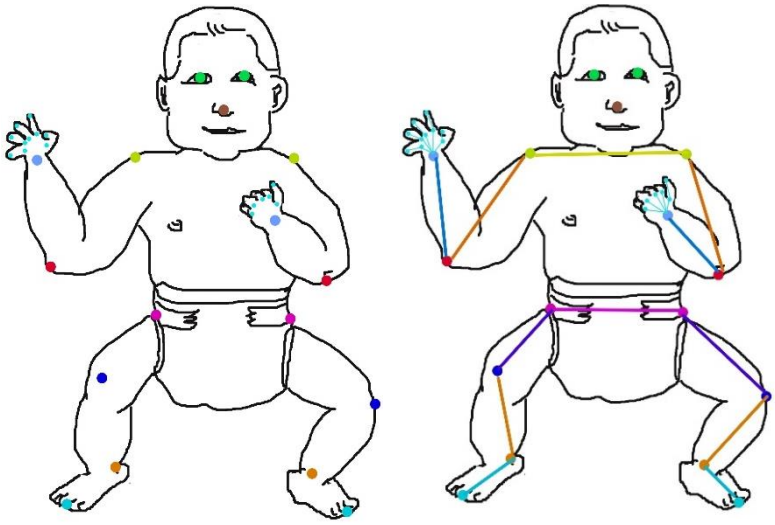


Figure 2. An overview of the selected markers used in DLC. On the left subfigure the markers are illustrated on the infant's body. On the right subfigure the rigid bodies are also presented in between the selected markers.

3D-video analysis

The software package Anipose is using the satisfactory output of DLC to create a 3D reconstruction of the quantitative GMAs, and will enable 3D kinematic analysis based on our videography[12]. MATLAB has been used to critically analyse the 3D distribution in space, the length of the rigid bodies, the kinematics, and the joint kinematics. A median filter has been applied to the data before analysis, to remove noise.

Results

A deep learning DLC approach has been utilised to place markers on video-frames. To evaluate the achievement of labelling by DLC (deep learning video analysis tool[11,23]), a MATLAB analysis is performed that allows visualisation of the 2D data and the associated certainty of the created neural network.

To first receive information about neural network building approaches, the HAI database is used to test whether building separate neural networks for each camera's angle of view (AOV) has a higher accuracy than one neural network for all three AOVs combined. The 3D combined neural network contains all twenty-four videos of eight infants performing the HAI (three videos per infant). The three separate networks for the left-, right- and front-camera hold each eight videos (one video per infant) and are additionally trained. DLC generates for each neural network a train- and test-error to quantify the certainty of marker detection. The train- and test-error are given by the average distances, in pixels, between the markers placed by the assessor and by DLC (detailed descriptions about the process are in the DLC protocol [26]). The training error is a classification error which is based on the manual placed markers as well as the DLC placed markers. So, the model is applied to the data after the training phase. On the other hand, the test error uses the dataset twice to calculate the classification error besides the training of the model which means there is a calculation on the data before the training phase (on the assessor placed markers). The error sizes for each network are displayed in table 2 and shows the errors for all markers, and markers above a certain threshold (p cut-off of 0.6). Overall, the distance between the assessor placed markers and the DLC placed markers are more minimised in the separated networks than in the combined 3D neural network, which concludes that building separate networks for each camera would do better marker detection in DLC. So, for this reason we applied the separate neural networks approach to the GMA performances of fifteen infants (ten congenital heart patients (GA range 37 – 40 weeks), three prematures (GA range 25 – 31 weeks) and two near-prematures defined as near-term (GA range 33 – 34 weeks)).

Table 2. Train and test errors of the trained HAI neural networks

Network	Train error (pixels)	Test error (pixels)
3D combined HAI	2.67	9.13
- with a p cut-off of 0.6	2.67	6.45
HAI Left camera	2.51	3.24
- with a p cut-off of 0.6	2.44	2.52
HAI Right camera	2.33	7.22
- with a p cut-off of 0.6	2.33	6.15
HAI Front camera	2.11	16.34
- with a p cut-off of 0.6	2.11	7.75

Table 3. Train and test errors of the trained GMA neural networks

Network	Train error (pixels)	Test error (pixels)
GMA Left camera	2.53	5.25
- with a p cut-off of 0.6	2.41	4.47
GMA Right camera	2.42	6.12
- with a p cut-off of 0.6	2.17	5.43
GMA Front camera	1.85	5.5
- with a p cut-off of 0.6	1.82	4.55

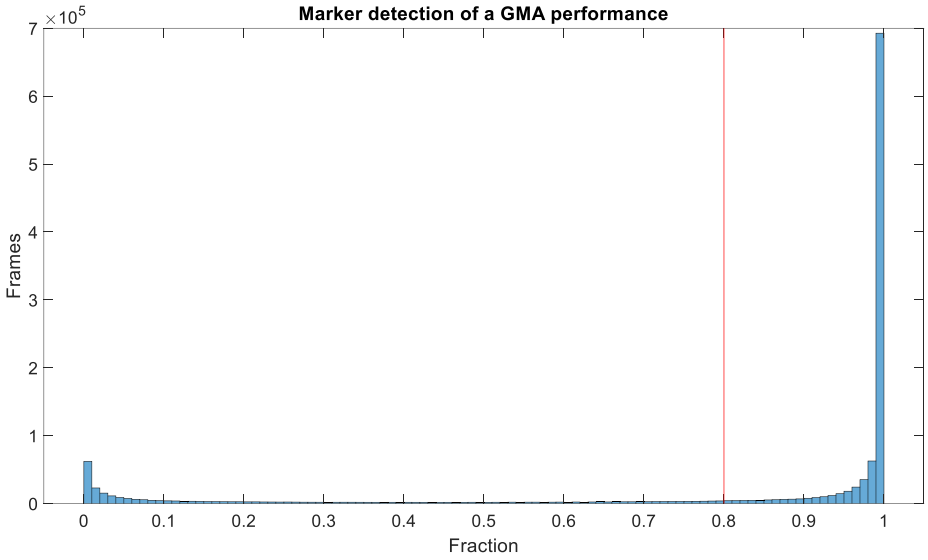


Figure 3. All frames and their likelihoods in marker detection by DLC. This figure illustrates the certainty of the neural network marker detection in one GMA measurement. A clear dichotomous result is shown, which means that the neural network has in 90% (0.8-1) of all frames a high certainty of marker detection and shows in 10% a completely uncertain outcome (0-0.1). A threshold of 0.8 (red line) is chosen for further analysis.

We next assessed a 2D-analysis on the GMA database. The GMA train and test errors are illustrated in table 3, which shows similar outcomes as the HAI-database (an average of 2–3-pixel distances). The output of the DLC contained CSV files includes information of each marker (x- and y- coordinates within all frames and the associated likelihoods). To make the movement kinematics more dependable, it is important to only include the frames which have a likelihood above a certain threshold. To make an informed decision about where to put the threshold, a determination of all likelihoods of all marker detections have been created separately for each GMA performance (figure 3). This revealed that most markers are placed, within frames, with a fraction of 0.8-1 (likelihood of 80-100%). Only a small fraction of markers was detected with a high uncertainty of 0-0.1 (0-10%). So, the certainty of labelling by the neural networks is dichotomous showing overall high certainty. We could therefore place a likelihood threshold at 0.8, which would still include around 90% of alle observations.

We then compared the marker detection between all three neural networks. A mean fraction of frames, above the 0.8 threshold, have been created for thirteen infants (infant 1-13 in table 1). Figure 4 illustrates seven markers with their mean fraction for each separate neural network (a distinct colour for each AOV) and their associated standard deviation (STD). Marker detection depends on multiple variables like the certainty of the neural network and the quality of the cameras. Figure 5 shows the high-quality of the cameras because the smaller joints are easy to distinguish (IDS UI-3060CP-C-HQ_Rev_2 cameras, with resolution of 1936 x 1216 Pixels). However, smaller joints like the digits do overall have a lower fraction in figure 4 (e.g., Right middle finger MCP and top). This indicates that the lower fractions are at least not due to error sizes or because of camera resolution. Furthermore, easily observed markers were detected by all the cameras at high fractions (e.g., nose), and detection of bigger joints were also at high fractions but do differ between the cameras because of often-occluded locations in their AOV (e.g., left- and right wrists and ankles). The left camera shows the lowest fractions because this camera is not optimally located, attached lower on the wall, which creates more occluded locations.

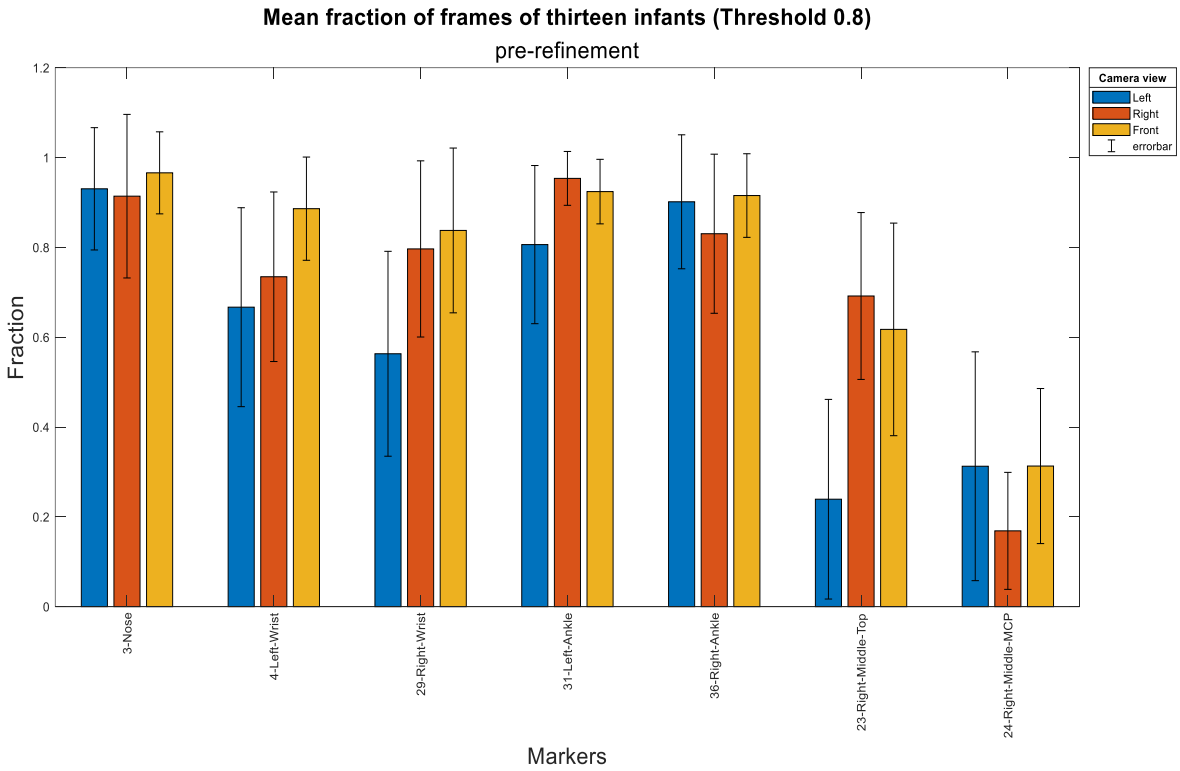


Figure 4. Mean fraction of marker detection of thirteen infants displayed for each DLC neural network. Markers that can be easily observed are detected by all the cameras at high fractions (e.g., nose). Bigger joints are also detected at high fractions but do differ between the AOV of the cameras because of often-occluded locations (e.g., left- and right wrists and ankles). However, smaller joints like the digits do overall have a lower fraction. Overall, the front (e.g., yellow) camera has a higher fraction of marker detection than the right- (e.g., red) or left (e.g., blue) camera.

In addition, we visualised the marker distributions to compare those with the body part orientations from the video. We obtained the graphs illustrating markers over time in space, by implementing x- and y- coordinates within each frame for each marker in MATLAB. The nose and bigger joints markers, as within figure 4, have been used to illustrate the orientation of the markers in space compared to the body parts orientation on the video. Figure 6 contains three graphs (obtained with MATLAB) which represent all three camera AOVs. Underneath the graphs are snapshots of the video to illustrate the markers and the orientation of the body parts in the video. It clearly shows that the MATLAB graphs of the markers are perfectly in line with the orientation of those body parts on the videos.

Figure 5. A frame of the infant's hand during a GMA performance. The digits can easily be distinguished showing a high-quality of the cameras, which is needed to receive a correct marker detection in DLC.

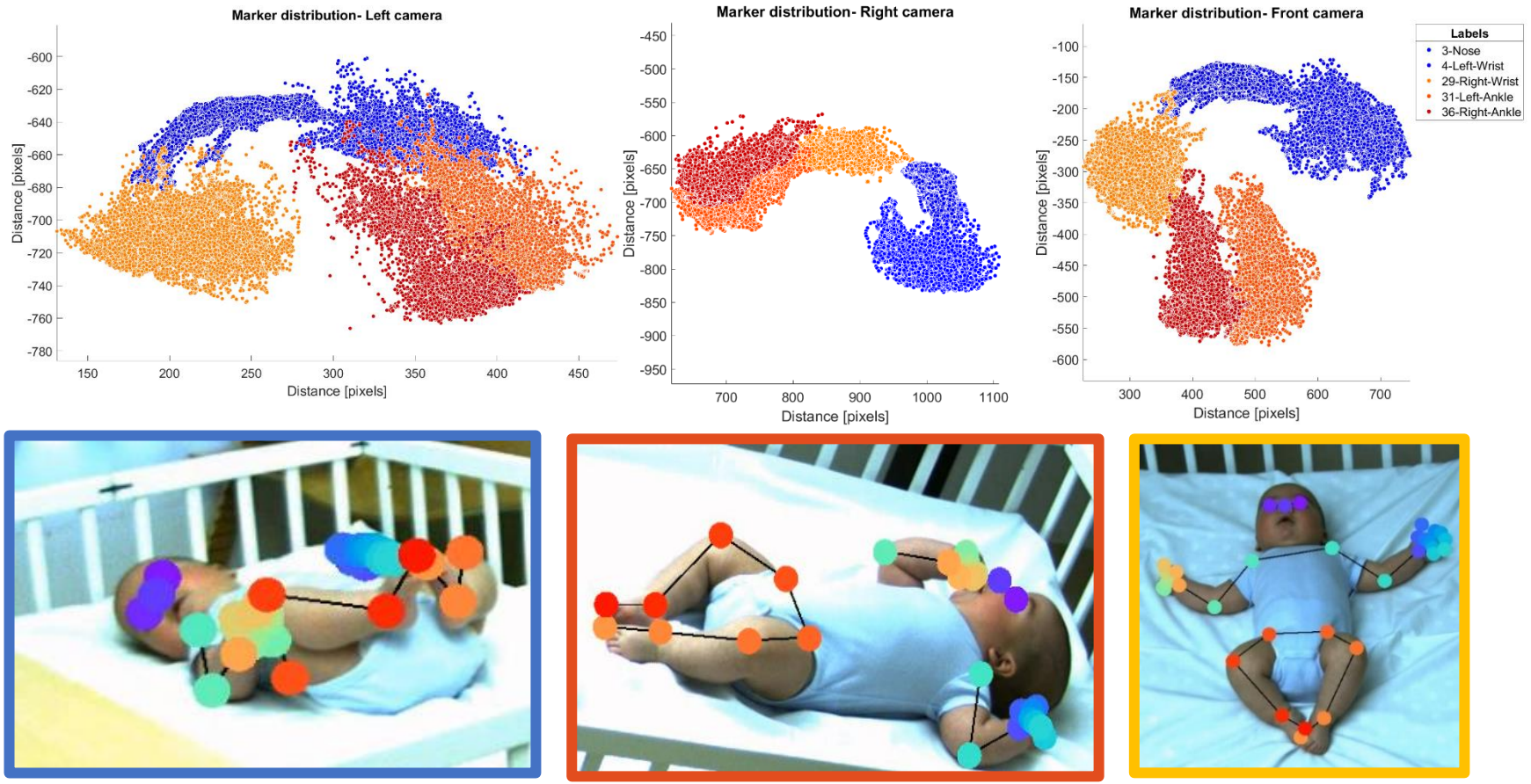


Figure 6. The nose left- and right wrist and the left- and right ankle are regenerated by MATLAB to show the body parts in space in relation to each other. This marker distribution is performed in all three camera neural networks of one infant. The blue stroked picture and the above associated MATLAB graph illustrate the output of the Left camera neural network. The red stroked picture and the above associated MATLAB graph illustrates the output of the right camera neural network, whereof the yellow stroked picture and MATLAB graph represent the output of the front camera AOV.

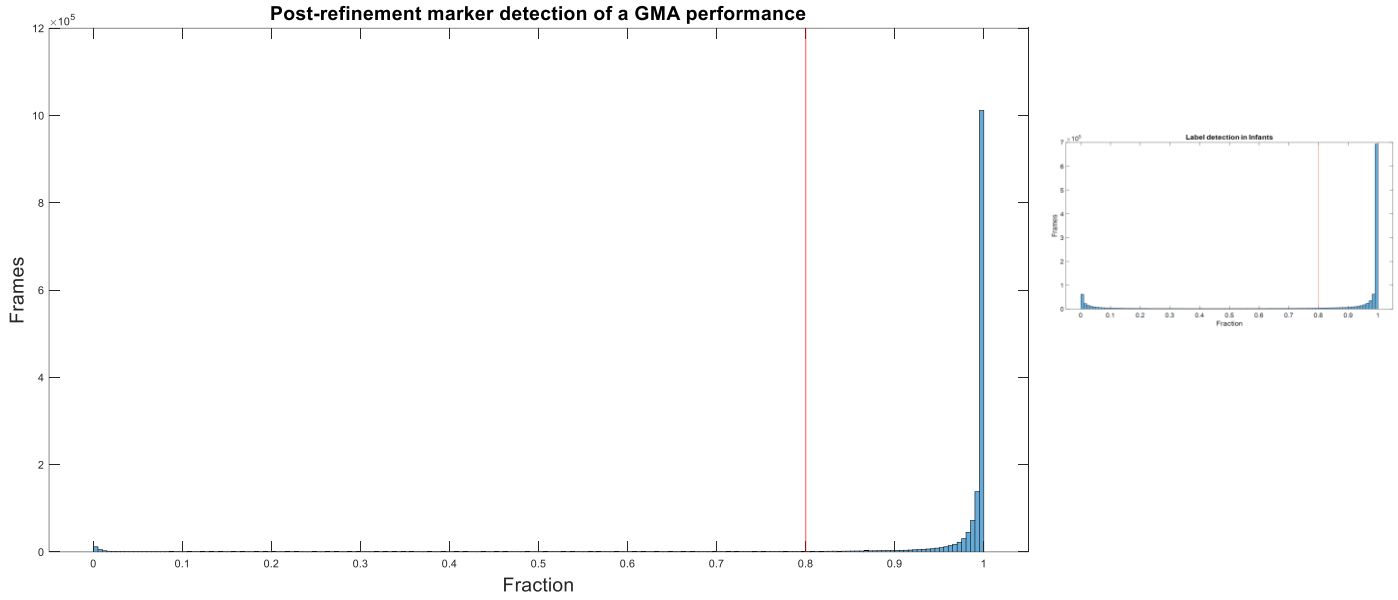


Figure 7. The post-refinement fractions of all frames and their likelihood in marker detection by DLC. The figure illustrates the certainty of the neural network marker detection in one GMA measurement. Compared to the pre-refinement GMA neural networks, illustrated on the right side of this figure, or in figure 3, the dichotomous distribution is shifted more to the high certainty side. The redline is the chosen threshold for marker detection.

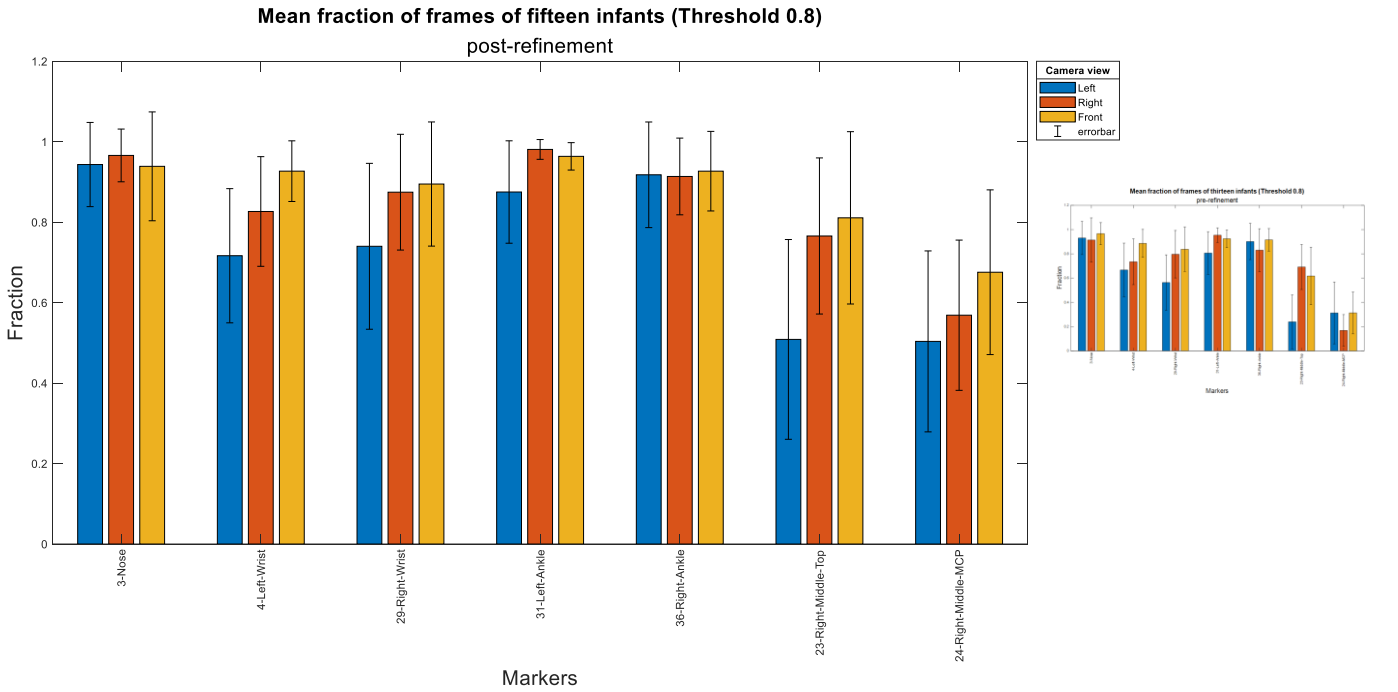


Figure 8. The post-refinement mean fraction of fifteen infants displayed for each network. The figure illustrates the fraction values for each GMA neural network for certain markers like nose, left- and right wrist, left- and right ankle and the MCP- and Top of the right middle digit. Compared to the pre-refinement GMA neural networks, illustrated on the right side of this figure, or in figure 4, the bigger joints do not show big fraction differences. However, the smaller joints, digits, show better marker detections.

In addition, we refined and retrained the neural networks to optimise the marker detection (detailed descriptions about the process are in DLC protocol [26]). Post-refinement figures have been compared to the pre-refinement figures to show the efficiency of refinement. Before retraining, two infants were initially added to the neural networks (infant 14 and 15 of table 1). Figure 7 displays the certainty of marker detection after refinement. The uncertain fractions at the post-refinement stage (figure 7) are significant less than in the pre-refinement stage (figure 3). For this outcome, a two-sample Kolmogorov Smirnov test has been performed ($D= 0.163$, $p<0.01$). Figure 8 displays the mean fraction of fifteen infants after refinement. Unfortunately, markers detection did not significantly improve according the one-way ANOVA test ($F(41)= 2.83$, $p= 0.1$). However, detection of smaller joints (e.g., digits) is significant improved ($t(10)= -2.43$, $p<0.05$) at the post-refinement stage (figure 8) in comparison to the pre-refinement stage (figure 4), according a paired t-test**.** To conclude, the figures perfectly show that refinement matter to improve marker detection in the neural networks, especially the smaller joints.

We then tested the fraction of occlusion of the smaller joints or digits. The neural networks cannot perform high certainty marker detection if body parts are not appearing in the AOV during video measurement. Therefore, the fraction of marker detection is also dependent on the AOV of the cameras and their occluded locations. The digits are important to critically analyse because earlier studies did not place markers on those body parts during the GMAs. We did place those markers to receive similar parameters as within the MOS-score[25] and to test whether those movements could also be a reliable predictive motor outcome in infants around 3 months of age. Additionally, the digits have the lowest fraction of marker detection that might be due to occlusion. Figure 9 displays the fraction of camera detection per finger (we combined the data of the MCP and top digit). The digits were detected around 80% of the time (Mean 0.82 ± SD 0.07) in two (yellow) or more cameras (purple), which concludes that partial (red) or entire occlusion (blue) of those digits did not appear often. This implies that low fraction in marker detection all depends on the neural network training. Fortunately, this can be improved by refining and retraining the neural networks.

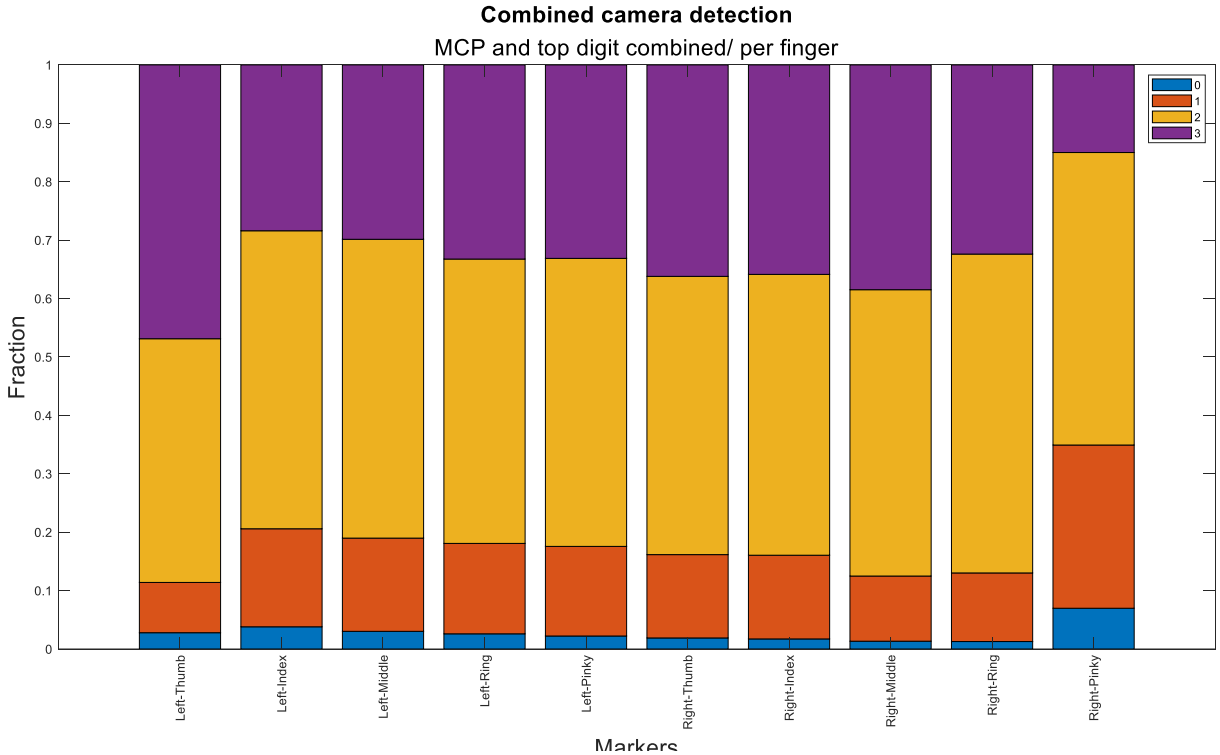


Figure 9. The fraction of the number of cameras detecting each finger (MCP and top digit combined). A small fraction around 5-8% represent the frames wherein none of the cameras could detect the digits (e.g., blue colour). Only one camera detecting each finger occurred in 10-20% of the frames (e.g., red). Two cameras detecting a finger occurred around 40% (e.g., yellow). All three cameras detecting the similar finger occurred in 15-50% of the frames (e.g., purple). Overall, each finger is detected most often by two cameras or more.

Before creating a 3D reconstruction, it is important to understand the motor onset and the involved trajectories within the brain. The CST is involved in the motor onset and is illustrated in figure 10. The areas within this CST are the motor cortex and the posterior limb of the internal capsule (PLIC). As already mentioned in the background, injury to the CST or basal ganglia could be seen as an early predictor of cerebral palsy. Neonatal MRI scans need to show asymmetrical development or injury of the PLIC, middle cerebral peduncle or white matter lesions of the CST. Abnormal myelination of the PLIC has a specificity and sensitivity of 29% and 98%. The negative predictive value of absent PLIC myelination and absent FMs is 90% for any and 98% for moderate or severe CP [6]. A description of the found brain abnormalities in all fifteen infants, according to the fellow-neonatologist, are in table 4. The characteristics of those infants are described in table 1. Three infants have a higher risk to develop CP because of their brain abnormalities (infant 4, 6 and 14 of table 4). Other infants did have brain abnormalities but those are not located in the region of interest (CST; figure 10). Figure 11 illustrates the abnormalities within the region of interest of the three infants-at-risk for CP. If we create a 3D reconstruction, we must focus on atypical movement patterns, especially on those three infants.

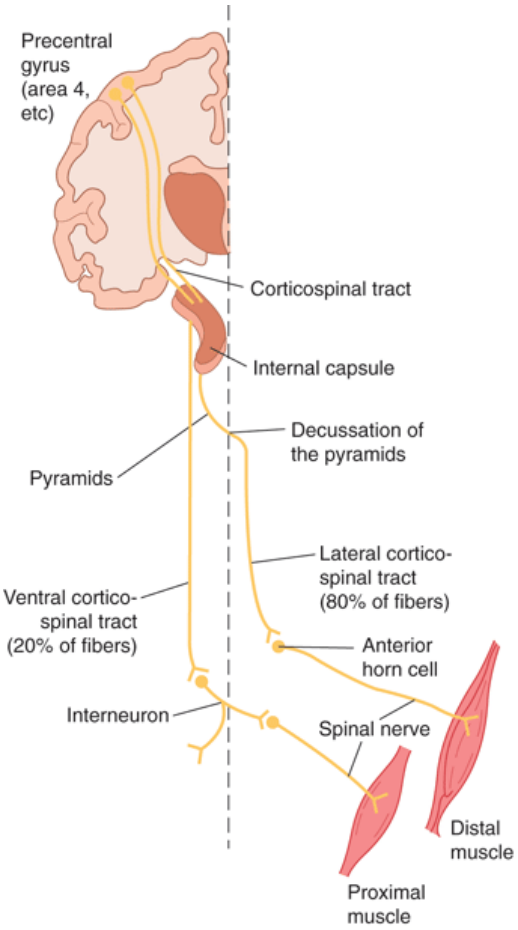
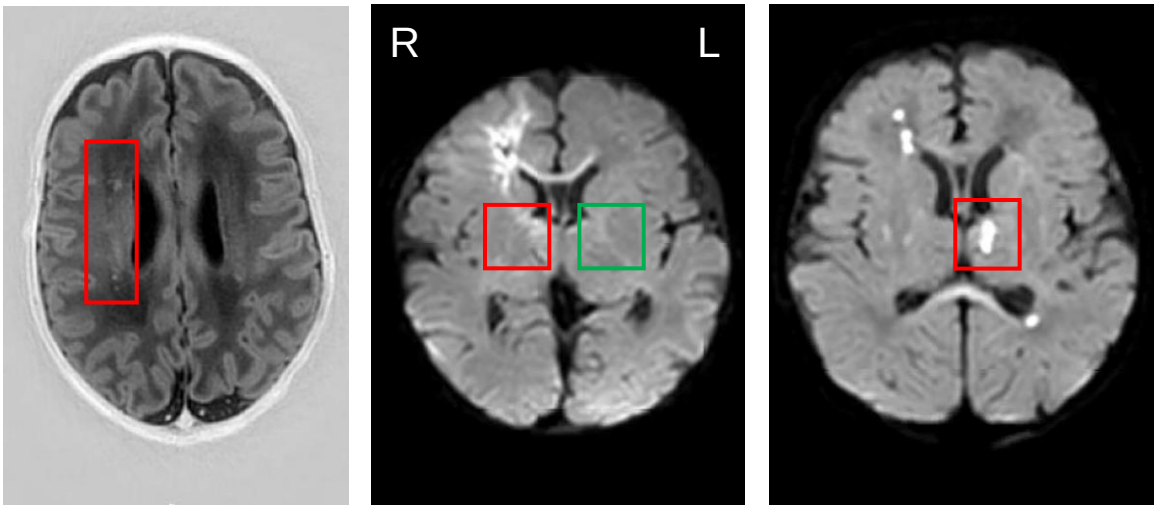


Figure 10. A clear overview of the corticospinal tract (e.g., yellow line) which is involved in the motor onset of the human body[34]. The brain areas, associated with motor onset, are the motor cortex and the posterior limb of the internal capsule (course past the basal ganglia).



Infant 3. Multiple ischemic lesions

Infant 6. Haemorrhagic infarction, asymmetrical myelinisation

Infant 14. Intraventricular bleeding

Figure 11. The MRI scans of the GMA performed infants-at-risk for CP. Infant number 3 shows multiple ischemic lesions within the CST. The ischemic white lesions appear hypo-intense within a FLAIR MRI scan (illustrated in the red box). Infant 6 has asymmetrical myelinisations in the PLIC, visible in a T1 weighted MRI. A clear representation of the PLIC is visible on the left hemisphere (green box), whereas the right hemisphere (red box) shows an unclear hyperintense white line. Infant 14 has the highest chance to develop CP, because of the intraventricular bleeding within the region of interest (red box). The red box shows scarring tissue which appear as hyperintense on a T1 weighted MRI scan. Infant 6 and 14 do have more multiple ischemic white lesions, however those are not mentioned because they are not located within the PLIC or CST.

Table 4. Clinical outcomes based on MRI data of the GMA infants

Infant	Category	Etiology	Location of Etiology	Risk of motor problems
1	Cardio	Small ischemic lesion	Right	None
2	Cardio	Ischemic lesion	Right> Left	None
3	Cardio	Two small ischemic lesions	Left and right	None
4	Cardio	Multiple ischemic lesions	Left and right	Small chance
5	Premature	No MRI performed	-	None
6	Cardio	Haemorrhagic infarction, asymmetrical myelinisation	Right	Small chance
7	Cardio	Subdural bleeding and ischemic lesion	Right	None
8	Near-term	No particularities	-	None
9	Near-term	Temporal bleeding	Right	None
10	Premature	No particularities	-	None
11	Cardio	No particularities	-	None
12	Cardio	No particularities	-	None
13	Cardio	Subdural bleeding and white matter oedema	Right	None
14	Premature	Intraventricular bleeding	Left and right	Higher chance
15	Cardio	No particularities	-	None

After the satisfactory 2D-analysis output, we attempted to create 3D reconstructions of those fifteen infants performing GMA. The software package Anipose used the calibration videos (made on the similar dates as the GMA performances) combined with the DLC output of the GMA videos to create such a 3D model (a detailed description of the process is given in the Anipose article[12]). However, the errors of fourteen calibration videos were impressively high, so those 3D reconstructions did not show an infant orientation as we aimed for. At least, Anipose created a suitable 3D reconstruction of only one infant which had a calibration error of 0.52 (appendix 5). Therefore, the 3D-analysis has only been performed on the 3D reconstruction of infant 10. According to the fellow-neonatologist, the infant showed no particularities on the MRI scan (table 4). Infant 10 was born before 28 weeks and low-weighted, which are two risk factors for CP (table 1). The manual MOS-score of GMA was 26 out of 28, which is almost a healthy motor performance. He did not receive all 28 points because he showed atypical jerky movements instead of a normal smooth and fluent motor representation. The next analysis will test whether those outcomes are in line with our 3D reconstruction calculations. Before analysis, a median filter has been applied on all 3D data.

We first assessed a scatter plot with MATLAB to check if the orientation of the markers is in line as expected from the orientation of the 2D output. The identical markers within figure 6 have been used to illustrate the 2D- versus 3D orientation. Figure 12 gives the 3D- scatterplot of the nose, left- and right wrists and left- and right ankles. The marker distribution is illustrated during a 10 minute period for the x-, y- and z-direction in pixels. The outcome showed a similar range of movements between the left and right side of the body. If measuring an infant at risk for neurodevelopmental diseases like CP, for instant infant 14, an asymmetrical range can occur between the left and right side of the body. For now, we can conclude that the orientation in the 3D-scatterplot is similar to the 2D input, which proves a suitable merging of the 2D output videos. However, if you want to compute movement kinematics out of the 3D reconstructions, the rigid bodies need to have consistent length over time.

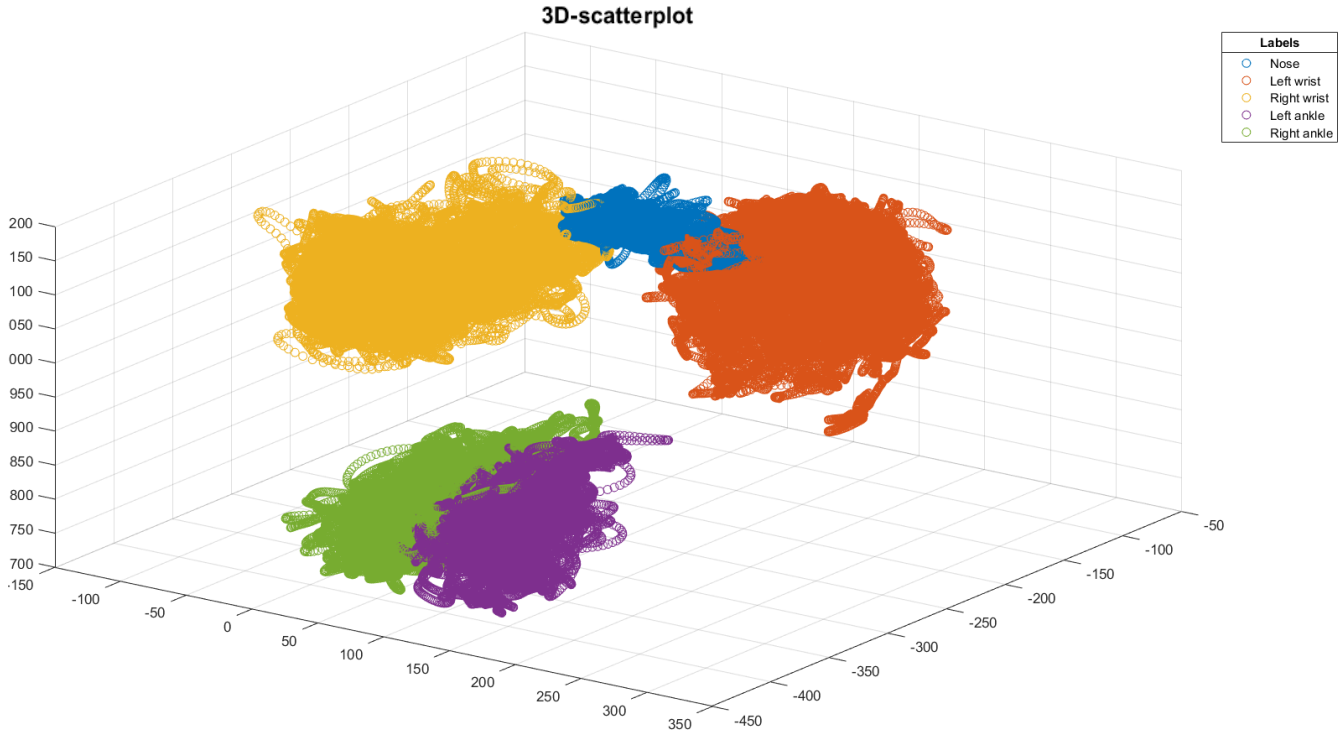


Figure 12. A MATLAB 3D-scatterplot of infant 10 showing the marker distribution of the nose, left- and right wrist and ankle. The x-, y- and z-direction are displayed in pixels. The left- and right side of the body show overall the same movement range over time.

We next assessed a calculation of the upper-arm and thigh of the left and right side of the body, to test whether those rigid bodies were consistent in time. The lengths of the rigid bodies (definition of rigid bodies is explained in subsection 2D-analysis of the method section) are obtained by measuring the length between the markers on both ends of the rigid body. For instance, the left-hip and left-knee are used to calculate the left-thigh length. And the right-shoulder and right-elbow are used to obtain the right upper-arm length. Figure 13 gives the length of the left and right upper-arm and both thighs (Left upper arm 78.7 ± 0.1 mm; right upper arm 74.0 ± 0.2 mm; left thigh 65.7 ± 0.0 mm; right thigh 67.3 ± 0.0 mm). Because infant 10 is born being low weighted, those values could be realistic but we did not measure those values at the beginning so we cannot prove it is the same length as in reality. Besides this, there is noise (high amplitude peaks). The upper-arm has more noise than the thighs (ranges: 74 – 81 mm left upper arm; 72 – 85 mm right upper arm; 65 – 66 mm left thigh; 67 – 68 mm right thigh) which can occur because of the chosen AOVs of the multi-camera approach. The front AOV was more located towards the legs than to the arms. However, overall, you can conclude that the length of the rigid bodies is consistent in time (maximal deviation of 0,3% total length). This outcome enables calculations of movement kinematics.

To compare the neurological and MOS-score outcomes with the 3D kinematics, the movement kinematics, such as position, velocity, and acceleration, were calculated. Figure 14 illustrates the kinematics of the left and right wrist of infant 10. The positions of the left and right wrist are illustrated in x-, y- and z-direction. The velocity is the first differentiation of the position, while the acceleration is obtained by calculating the second differentiation. The velocity has a maximum around 1.2 m/s while the acceleration has a maximum of 18 m/s^2. The correlation occurrence has been calculated for the velocity as well for the acceleration between left- and right wrists, illustrated in figure 15. The acceleration shows more uncorrelated than correlated movements in comparison to the velocity of the left- and right wrist (figure 15: acceleration has a higher peak around 0 than the velocity, which is more spread out). This high amount of speed peaks at the acceleration can be due to the jerky movement performance infant 10 has been showing during GMA[30]. One would expect less speed peaks with a healthy smooth and fluent motor representation. This figure illustrates the kinematic possibilities of the 3D reconstruction which enables detecting abnormalities in the infant's movement representation. In addition, the output of Anipose also includes joint kinematics calculations to dissect the contributions of specific joints to complex motor behaviours.

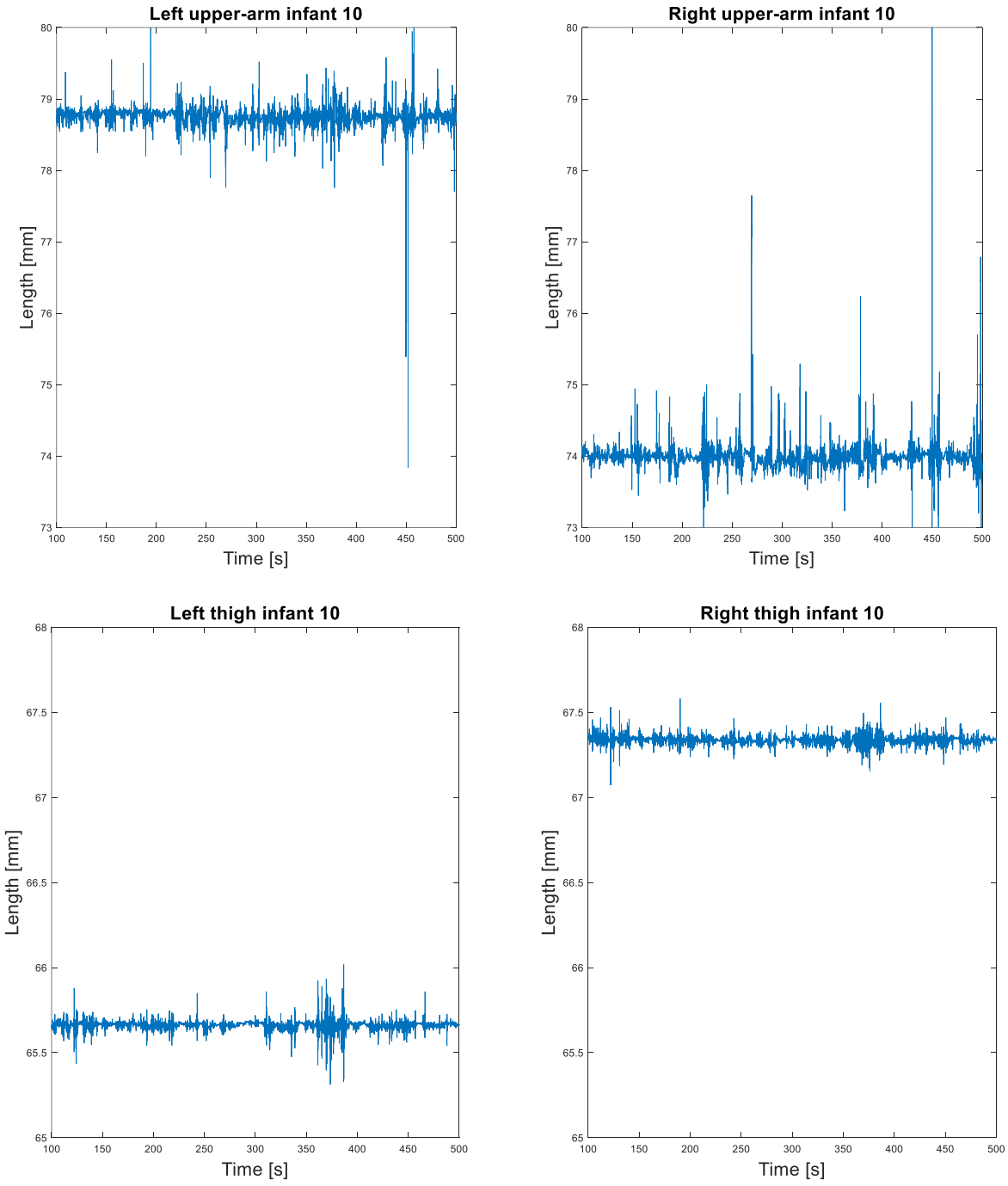


Figure 13. Length of the left- and right upper arms and thighs of infant 10 during GMA performance. The upper arm shows more noise than the thigh lengths, but overall, the lengths are quite consistent over time. The noise is due to the DLC labelling errors of the shoulders, elbows, hips, or knees. The mean length of the upper arm are around 7,5 centimetres while the length of the thighs are around 6,5 centimetres.

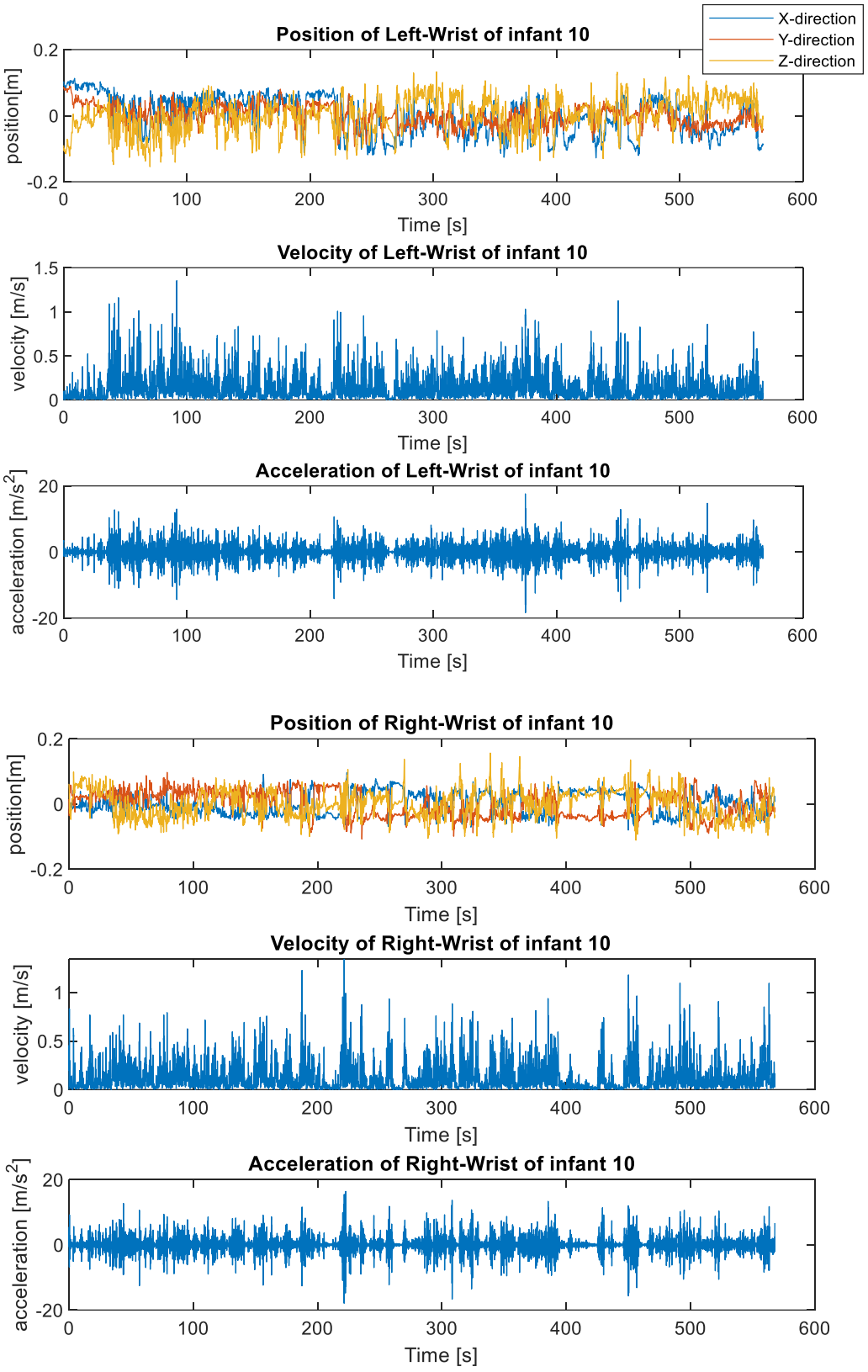


Figure 14. The kinematics of left- and right wrist of infant 10 during GMA performance. The position in x-, y- and z-direction. The velocity has in both wrists a maximum around 1.2 m/s, both at another timeframe. The acceleration differs more between left and right wrist but do both have a maximum of 18 m/s².

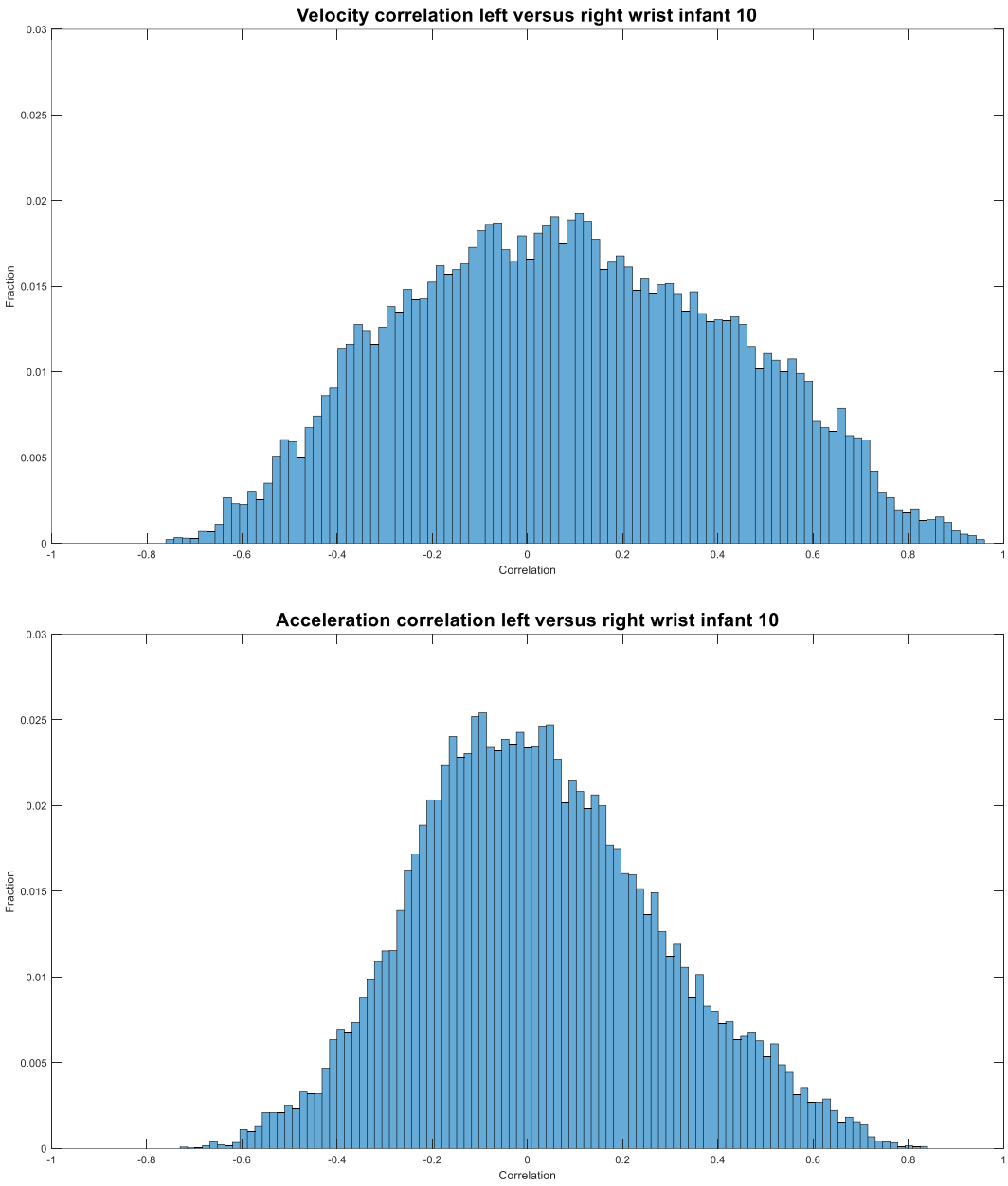


Figure 15. The kinematic correlation occurrence between left and right wrist of infant 10. The velocity, upper figure, shows a more widen curve than the acceleration, lower figure, does. So, we can conclude that the velocity shows more negative or positive correlations (correlations around -1 and 1) and the acceleration shows more uncorrelated values (correlations between -1 and 1).

Furthermore, we tested if the neurological and MOS-score outcome of infant 10 can be related to the joint kinematics of the 3D reconstruction of this infant. 3D tracking can be used to dissect the contributions of specific joints to complex motor behaviours[12]. For this matter, a comparison between left and right knee and between left and right elbow has been performed. Figure 16 illustrates the left versus right knee angles over time. Above this, a time-based correlation has been calculated to show the correlated and uncorrelated movements. Infant 10 shows more negative and positive correlations (knee: 13% negative correlation; 14% positive correlation) between left and right knee, rather than uncorrelated movements (knee: uncorrelated 73% of the time). Around 300 seconds the left knee makes a rotation, while the right knee remains stable. This is an example of a negative correlation. Uncorrelated movements are interesting to observe, for instance leg kicking. Positive correlations are expected to be more present within knee rotations than elbow rotations, because we biologically learn to synchronously

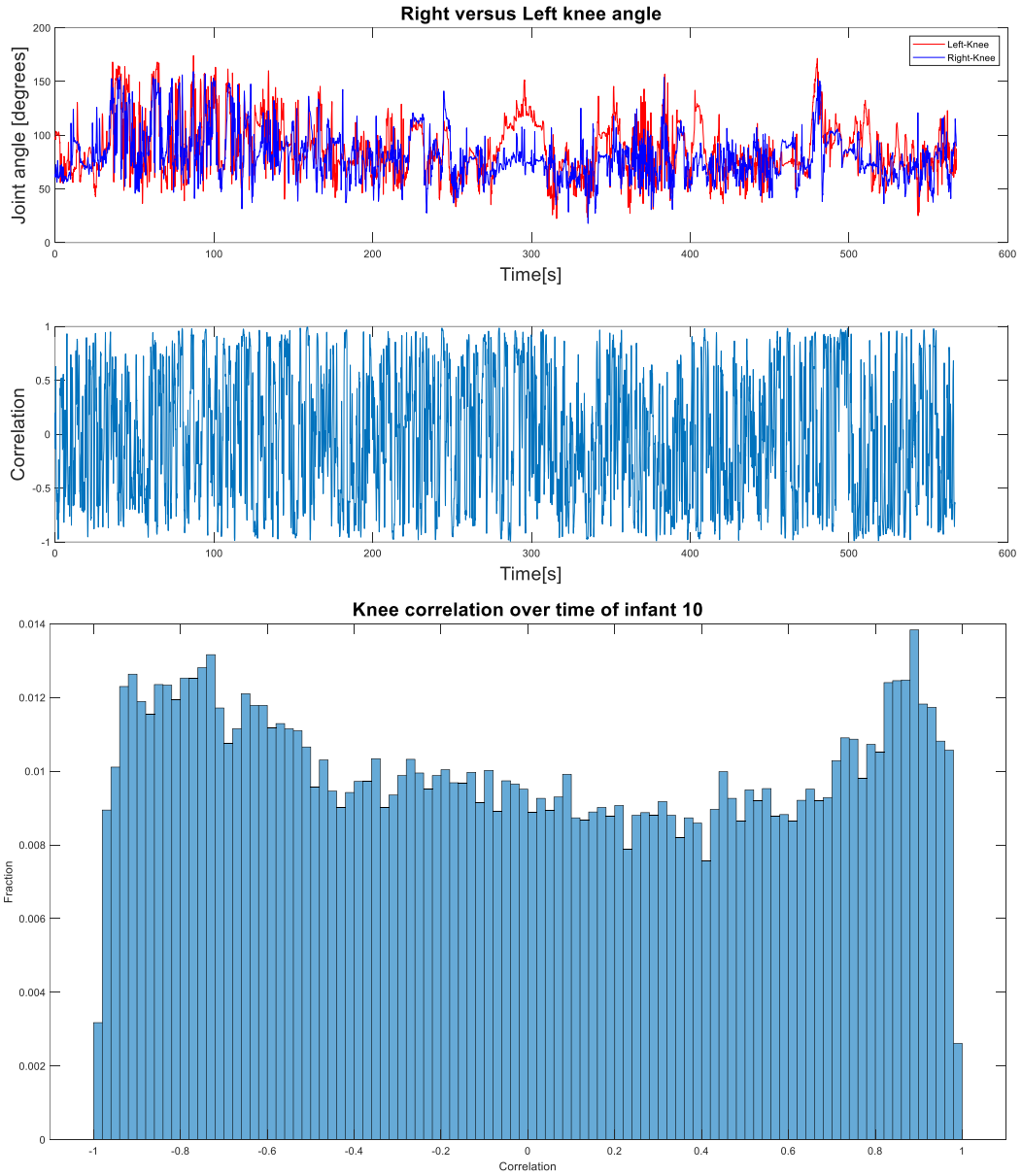


Figure 16. The rotation of the left (e.g. red line) versus right (e.g. blue) knee of infant 10 during a GMA performance. The middle figure displays a time-based correlation to show the positive or negative correlation of the left versus right knee over time. Those values are illustrated in a histogram, lower figure, showing the fraction of positive (1), negative (-1) or uncorrelated knee movements (between -1 and 1).

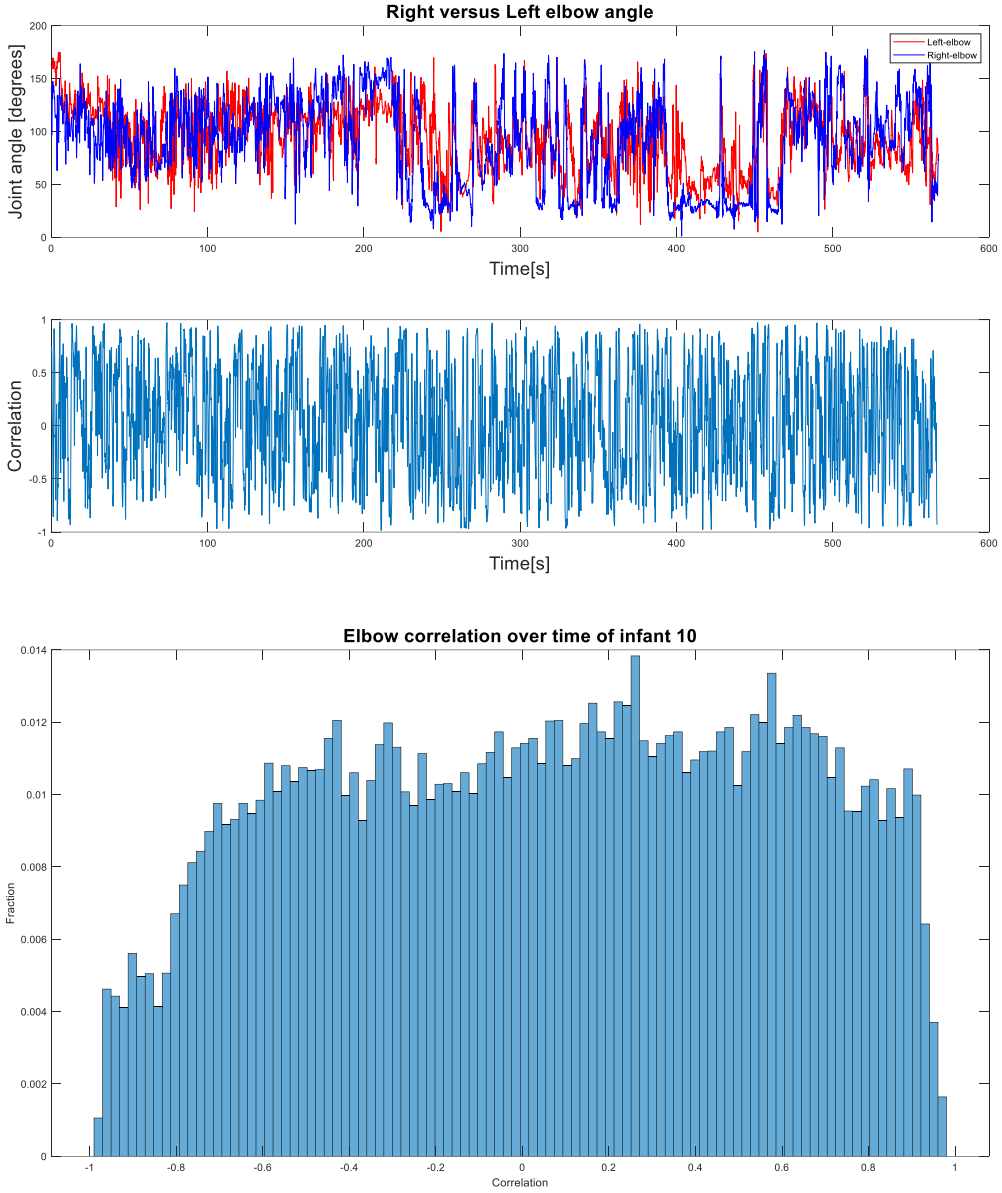


Figure 17. The rotation of the left (e.g. red line) versus right (e.g. blue) elbow of infant 10 during a GMA performance. The middle figure displays a time-based correlation to show the positive or negative correlation of the left versus right elbow over time. Those values are illustrated in a histogram, lower figure, showing the fraction of positive (1), negative (-1) or uncorrelated elbow movements (between -1 and 1).

move the legs to walk, while the arms are taught to grasp things and use them separately from each other. For this reason, we also created similar figures for left and right elbow rotations illustrated in figure 17. As expected, infant 10 shows more uncorrelated movements with the arms than with the legs (elbow: 2% of the time negative correlation; 1,5% of the time positive correlation; uncorrelated 96,5% of the time). The general Pearson coefficient is also calculated for as well the knees as the elbows (knee correlation coefficient of 0.47; elbow correlation coefficient 0.54). Both have a positive moderate correlation, which is in line with the hypothesis that the left versus right body are overall showing symmetrical movement patterns.

Conclusion

To evaluate the quality of a markerless multi-camera approach, we first assessed a 2D-analysis in MATLAB. We achieved markerless labelling of limbs on par with human labelling by using multiple cameras in combination with the multiple deep neural networks of DeepLabCut. The certainty of labelling by the neural networks was dichotomous, showing high certainty and low error of labelling when body parts were in view. Markers on digits or smaller joints are more challenging, but become increasingly better whenever the neural networks are repeatedly refined. We next assessed a 3D-analysis in MATLAB upon the 3D reconstruction created with Anipose. The 3D-analysis was performed to test whether precise kinematic parameters could be detected during the fidgety movements stage in approximately 3 months old infants-at-risk for CP. Our 3D reconstruction shows good tracking as indicated by a constant length of rigid bodies. Furthermore, preliminary results show the possibility to analyse positional data in 3D and kinematics of limbs and joints. We compared those aspects with neurological and MOS-score outcomes. Infant 10 shows a symmetric movement pattern as expected from the MRI scan. Interestingly, this patient shows frequent speed peaks in acceleration, which indicates a jerky instead of smooth movement pattern[30]. This atypical jerky motor presentation is also observed by manual GMA observation (26/28 MOS-score). Overall my results show that 3D reconstruction can reveal different precise kinematic parameters, but more 3D reconstructions of infants at high-risk are necessary. Furthermore, we need to determine precise kinematic parameters indicative of disrupted development to optimise each infant's functional potential.

Discussion

Our 3D-analysis findings showed overall interesting opportunities for future studies to pose estimation of infants-at-risk for CP. However, it was disappointing that we were able to build a 3D reconstruction of only one infant. The errors of most calibration videos were high, assuming something went wrong during the calibration step in Anipose. It could be that the Charuco board we used is not suitable for those measurements, or environmental factors like daylight were influencing the board detection. In future, it would be a better approach to test which aspects do have an influence on the calibration of the multi-camera approach. We recommend testing whether different Charuco boards or lighting conditions would make differences in error size of the calibration video, to enable more 3D reconstructions. Besides that, we could not verify if the values we calculated in the real world coordinates were realistic and true. The 3D-analysis did show a constant length of the rigid bodies, but this can only be checked to be true if the real length of those rigid bodies are measured. In the future, a valuable tool like a measuring tape needs to be on the video during GMA, or the limbs need to be measured beforehand.

The 2D-analysis of DLC output showed overall satisfactory outcomes and a low error of labelling. However, this only occurred when body parts were in view. For now, refinement and retraining the neural network would optimise the marker detection, but we recommend adding more cameras to the set-up to avoid those occluded locations. In this matter, there will still be complete data in case one of the body parts ever get blocked from one AOV. Besides the occluded locations, the human labelling pre-train step in DLC may also manipulate the DLC errors-sizes. In future studies, we could manipulate those marker locations by adding different jitters. It would be important to test whether those noisy factors would have an impact on the neural network and the final 3D reconstruction.

Furthermore, human labelling can become more reliable if we define precise marker locations. For instance, the wrist defines a surface of multiple centimetres underneath the hand, but it would be better if we narrow this location to the pisiforme of the wrist as this bone is the easiest to detect. In addition, we only used one similar marker for the internal or external side of the wrist. In future, it might therefore be helpful to test the usage of distinguished markers for the internal and external side of joints. This might also have a positive effect on the DLC error-sizes.

Focusing on time-consuming limitations, DLC training slows down the entire process. Our video durations are around 10 minutes, containing around 60.000 frames, which all need to be trained (acquisition speed of 100 Hz). Besides, the high quality of the IDS cameras (1936X1216 pixels) is time-consuming compared to the resolution used in earlier studies (640X480 pixels). Overall, we had to train eight times longer than the earlier studies. Only one neural network, containing all fifteen infant videos of one camera, lasted 96 hours while the entire DLC training of three neural networks lasted around 288 hours. The earlier recommendations to increase the number of cameras, and markers, would make this even a bigger issue. In future, we can prevent this limitation by making shorter videos and doing fewer refinements. In literature, a GMA performance lasts only 3 to 5 minutes while most of our infants had a performance of 10 minutes. This implies that our videos have a frame abundance which is time-consuming for DLC because each frame needs to be trained. Therefore, a decrease in video duration diminishes the time of DLC training. On top of that, adding data becomes less important because training makes the neural network stable, which is another time-saving factor.

On the other hand, I did not have the time to determine precise kinematic parameters. My internship lasted 9 months where I had to perform all the assessments, had to train all the videos in DLC, had to perform a 2D analysis on the trained neural networks, had to create 3D reconstructions, and finally had to perform a 3D analysis. Because I was the first intern on this project, I had to deal with multiple hitches. Overall, I did not have enough time to determine precise kinematic parameters indicative of disrupted development. The challenge is to translate the human terms of the manual MOS-scoring to computer logical terms. In this study, only a few general parameters are illustrated, but those parameters should become more specific and detailed in future, to really draw any conclusion on optimising individual therapies.

In addition to this, there were also limitations on the assessments. For instance, the HAI markers were only placed on the infant's body. However, to make a clear distinction between toy-targeted and free movements, it would be more effective to also place markers upon the toys to assess movements with or without toys. Another clinical challenge, for both assessments, is that the state of arousal between individuals could be widely different. In our study, this was partly accounted for by the Brazelton model to detect the infant's arousal state during measurement. However, the GMAs were performed after the outpatient clinic appointment, which already had an enormous impact on their arousal state. It would be more optimal when the assessment is performed beforehand or separately from an outpatient clinic appointment. The state of arousal would be better if the video is taken on the baby's optimal state, from home. Researchers are focusing on home-video methodologies to optimise the infants state of arousal and to make it easier to participate for parents [31]. However, parents could have an impact on the video. For instance, their smartphones could not be stable or they are not including the whole body within the video-frame throughout the measurement. Besides, the smartphone camera lacks high-quality and it only provides 2D analysis. For this reason, home assignment cannot analyse kinematics of body parts moving in three directions, which is an important factor of human motion.

Furthermore, an important challenge for novel tools is validation. All earlier studies with DLC do not provide a retrospective study because they perform 2D kinematic analysis instead of a 3D analysis. To resolve this, we compared the neurological outcome and the manual scoring of GMA with the Anipose output. The MOS-score lacks 2 points because of a jerky movement performance, which can be related back to the analysis of the 3D reconstruction. So far, those tools are in line with the 3D reconstruction output. However to validate even more, our 3D output should be compared with other golden standards detecting human motion. One study already tried to validate DLC by comparing it with Fastrak (high accurate electromagnetic 3D motion tracking system), which showed similar outcomes[32]. At our department, accelerometers are attached at the wrists of infants during several HAI performances. In future studies, accelerometer data can be used to validate the kinematic parameters of Anipose.

To make the DLC and Anipose more dependable, our dataset should be open-source so worldwide usage enables focus upon, for instance, infant-, gender- and disease-differences. If you widen the target audience you can compare different neurodevelopmental diseases with

each other, such as differences in motor outcome and development between CP, and for instance, spinal muscular atrophy (SMA) patients. SMA is caused by autosomal recessive mutations and results in the loss of motor neurons and progressive muscle weakness [33]. There are multiple treatments available, but it also lacks early detection within childhood. In addition to this, enlarging the dataset would also give more opportunities to measure parameters such as movement fluency and smoothness. The more infants being measured, the more information on average acceleration while reaching toys during HAI performance. If an infant would not show acceleration values within the average range, you can conclude that this infant has an atypical performance (more jerky or absurd movements). Therefore, our ultimate goal is to create a worldwide open-source database for paediatric movements.

Acknowledgements

The author thank the families of the infants participating the study, and the staff of the DDOD lab, Center for Child Development, Exercise and Physical Literacy and the UMCU neonatal intensive care unit for their support.

The 3D-video analysis interdisciplinary research group consists out of (from left to right):

- Chelsey Linnenbank, another research intern who has been focusing on 3D video analysis of mice.
- Dr. Laurens Witter, who is my daily supervisor and DDOD staff member.
- Roos Bos, the author of this paper
- Dr. Lianne Verhage, who is an occupational therapist of the Child Development, Exercise and Physical Literacy department.
- Dr. Niek van der Aa is missing on the picture, but he is a fellow neonatologist who works at the neonatal intensive care unit of the UMCU.

References

1. Ju YH, Hwang IS, Cherng RJ. Postural adjustment of children with spastic diplegic cerebral palsy during seated hand reaching in different directions. *Arch Phys Med Rehabil*. 2012;93(3):471-479. doi:10.1016/j.apmr.2011.10.004

2. Beckers LWME, Geijen MME, Kleijnen J, et al. Feasibility and effectiveness of home-based therapy programmes for children with cerebral palsy: A systematic review. *BMJ Open*. 2020;10(10). doi:10.1136/bmjopen-2019-035454

3. Støen R, Songstad NT, Silberg IE, Fjørtoft T, Jensenius AR, Adde L. Computer-based video analysis identifies infants with absence of fidgety movements. *Pediatr Res*. 2017;82(4):665-670. doi:10.1038/pr.2017.121

4. Raghuram K, Orlandi S, Church P, et al. Automated movement recognition to predict motor impairment in high-risk infants: a systematic review of diagnostic test accuracy and meta-analysis. *Dev Med Child Neurol*. 2021;63(6):637-648. doi:10.1111/dmcn.14800

5. Schroeder AS, Hesse N, Weinberger R, et al. General Movement Assessment from videos of computed 3D infant body models is equally effective compared to conventional RGB video rating. *Early Hum Dev*. 2020;144. doi:10.1016/j.earlhumdev.2020.104967

6. Glass HC, Li Y, Gardner M, et al. Early Identification of Cerebral Palsy Using Neonatal MRI and General Movements Assessment in a Cohort of High-Risk Term Neonates. *Pediatr Neurol*. 2021;118:20-25. doi:10.1016/j.pediatrneurol.2021.02.003

7. Einspieler C, Marschik PB, Pansy J, et al. The general movement optimality score: A detailed assessment of general movements during preterm and term age. *Dev Med Child Neurol*. 2016;58(4):361-368. doi:10.1111/dmcn.12923

8. Crichton A, Ditchfield M, Gwini S, et al. Brain magnetic resonance imaging is a predictor of bimanual performance and executive function in children with unilateral cerebral palsy. *Dev Med Child Neurol*. 2020;62(5):615-624. doi:10.1111/dmcn.14462

9. Silva N, Zhang D, Kulvicius T, et al. The future of General Movement Assessment: The role of computer vision and machine learning – A scoping review. *Res Dev Disabil*. 2021;110. doi:10.1016/j.ridd.2021.103854

10. Einspieler C, Bos AF, Krieber-Tomantschger M, et al. Cerebral palsy: Early markers of clinical phenotype and functional outcome. *J Clin Med*. 2019;8(10). doi:10.3390/jcm8101616

11. Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. 2018;21(9):1281-1289. doi:10.1038/s41593-018-0209-y

12. Karashchuk P, Rupp KL, Dickinson ES, et al. Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Rep*. 2021;36(13). doi:10.1016/j.celrep.2021.109730

13. Philippi H, Karch D, Kang KS, et al. Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy. *Dev Med Child Neurol*. 2014;56(10):960-967. doi:10.1111/dmcn.12477

14. Arac A, Zhao P, Dobkin BH, Carmichael ST, Golshani P. Deepbehavior: A deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front Syst Neurosci*. 2019;13. doi:10.3389/fnsys.2019.00020

15. McCay KD, Ho ESL, Shum HPH, Fehringer G, Marcroft C, Embleton ND. Abnormal Infant Movements Classification with Deep Learning on Pose-Based Features. *IEEE Access*. 2020;8:51582-51592. doi:10.1109/ACCESS.2020.2980269

16. Irshad MT, Nisar MA, Gouverneur P, Rapp M, Grzegorzek M. AI approaches towards prechtl's assessment of general movements: A systematic literature review. *Sensors (Switzerland)*. 2020;20(18):1-32. doi:10.3390/s20185321

17. Chambers C, Seethapathi N, Saluja R, et al. Computer vision to automatically assess infant neuromotor risk and directs the Rehabilitation Robotics Lab, GRASP. HHS Public Access. doi:10.6084/m9.figshare.8161430

18. Ihlen EAF, Støen R, Boswell L, et al. Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. *J Clin Med*. 2020;9(1). doi:10.3390/jcm9010005

19. Adde L, Helbostad JL, Jensenius AR, Taraldsen G, Grunewaldt KH, StØen R. Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. *Dev Med Child Neurol*. 2010;52(8):773-778. doi:10.1111/j.1469-8749.2010.03629.x

20. Raghuram K, Orlandi S, Shah V, et al. Automated movement analysis to predict motor impairment in preterm infants: a retrospective study. *Journal of Perinatology*. 2019;39(10):1362-1369. doi:10.1038/s41372-019-0464-0

21. Orlandi S, Raghuram K, Smith C, et al. *Detection of Atypical and Typical Infant Movements Using Computer-Based Video Analysis*.; 2018. doi:10.0/Linux-x86_64

22. Nakano N, Sakura T, Ueda K, et al. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Front Sports Act Living*. 2020;2. doi:10.3389/fspor.2020.00050

23. Mathis MW, Mathis A. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr Opin Neurobiol*. 2020;60:1-11. doi:10.1016/j.conb.2019.10.008

24. T. Berry Brazelton, J. Kevin Nugent. *Neonatal Behavioral Assessment Scale*. No. 137. Cambridge University Press; 1955.

25. Einspieler C, Bos AF, Krieber-Tomantschger M, et al. Cerebral palsy: Early markers of clinical phenotype and functional outcome. *J Clin Med*. 2019;8(10). doi:10.3390/jcm8101616

26. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc*. 2019;14(7):2152-2176. doi:10.1038/s41596-019-0176-0

27. Shin HI, Shin HI, Bang MS, et al. Deep learning-based quantitative analyses of spontaneous movements and their association with early neurological development in preterm infants. *Sci Rep*. 2022;12(1):3138. doi:10.1038/s41598-022-07139-x

28. Marchi V, Hakala A, Knight A, et al. Automated pose estimation captures key aspects of General Movements at eight to 17 weeks from conventional videos. *Acta Paediatrica, International Journal of Paediatrics*. 2019;108(10):1817-1824. doi:10.1111/apa.14781

29. Guzzetta A, Pizzardi A, Belmonti V, et al. Hand movements at 3 months predict later hemiplegia in term infants with neonatal cerebral infarction. *Dev Med Child Neurol*. 2010;52(8):767-772. doi:10.1111/j.1469-8749.2009.03497.x

30. Hogan N, Sternad D. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J Mot Behav*. 2009;41(6):529-534. doi:10.3200/35-09-004-RC

31. Boonzaaijer M, van Wesel F, Nuysink J, Volman MJM, Jongmans MJ. A home-video method to assess infant gross motor development: Parent perspectives on feasibility. *BMC Pediatr*. 2019;19(1). doi:10.1186/s12887-019-1779-x

32. Kosourikhina V, Kavanagh D, Richardson MJ, Kaplan DM. Validation of DeepLabCut as a tool for markerless 3D pose estimation. doi:10.1101/2022.03.29.486170

33. Schorling DC, Pechmann A, Kirschner J. Advances in Treatment of Spinal Muscular Atrophy - New Phenotypes, New Challenges, New Implications for Care. *J Neuromuscul Dis*. 2020;7(1):1-13. doi:10.3233/JND-190424