

1 **Rectus Femoris Electromyography Signal Clustering: Data-Driven Management of Crouch**
2 **Gait in Patients with Cerebral Palsy**

3 **Mehrdad Davoudi¹, Firooz Salami¹, Robert Reisig¹, Dimitrios A. Patikas², Sebastian I. Wolf^{1*}**

4 ¹ Clinic for Orthopaedics, Heidelberg University Hospital, Heidelberg, Germany.

5 ² Laboratory of Neuromechanics, School of Physical Education and Sports Science at Serres,
6 Aristotle University of Thessaloniki, Greece.

7 *** Correspondence:**

8 Sebastian I. Wolf

9 Sebastian.Wolf@med.uni-heidelberg.de

10 Clinic for Orthopedics, Schlierbacher Landstr. 200a, 69118 Heidelberg, Germany.

11

12

13

14

15

16

17

18

19

20

21 **Abstract**

22 This study aimed to investigate how electromyography (EMG) cluster analysis of the rectus femoris
23 (RF) could help to better interpret gait analysis in patients with cerebral palsy (CP). The retrospective
24 gait data of CP patients were categorized into two groups: initial examination (E1, 881 patients) and
25 subsequent examination (E2, 377 patients). Envelope-formatted EMG data of RF were collected. Using
26 PCA and a combined PSO-K-means algorithm, main clusters were identified. Patients were further
27 classified into crouch, jump, recurvatum, stiff and mild gait for detailed analysis. The clusters (labels)
28 were characterized by a significant peak EMG activity during mid-swing (L1), prolonged EMG activity
29 during stance (L2), and a peak EMG activity during loading response (L3). Notably, L2 contained 76%
30 and 92% of all crouch patients at E1 and E2, respectively. Comparing patients with a crouch gait pattern
31 in L2-E1 and L2-E2, two subgroups emerged: patients with persistent crouch (G1) and patients
32 showing improvement at E2 (G2). The minimum activity of RF during 20-45% of the gait was
33 significantly higher ($p= 0.025$) in G1 than in G2. A greater chance of improvement from crouch gait
34 might be associated with lower RF activity during the stance phase. Using our findings, we could
35 potentially establish an approach to improve clinical decision-making regarding treatment of patients
36 with CP.

37 **Keywords: EMG, Rectus femoris, cerebral palsy, cluster analysis, k-means.**

38 **1 Introduction**

39 Cerebral palsy (CP) is a nonprogressive movement and posture disorder that develops in a fetus during
40 pregnancy or infancy [1]. It is caused by an injury to the developing brain, which mostly happens
41 before birth. Spastic CP is the most common form of the disease, which arises from damage to the
42 motor cortex [1]. In this condition, muscles appear stiff and tight. Abnormal muscle tone and motor
43 deficit affect the walking ability of patients with CP [2].

44 Surface electromyography (EMG) can be used to measure muscle activity in a noninvasive and
45 clinically meaningful manner. In patients with CP, 3D gait analysis with simultaneous EMG
46 measurements is often conducted to gain insight into muscle function as part of prescribing treatment
47 and evaluating treatment effect [3]. Combining EMG data of the rectus femoris (RF) muscle with
48 kinematics, kinetics, and clinical data, Reinbold et al. demonstrated a method to predict the outcome
49 of RF transfer surgery [4]. They concluded that a diminished range of knee flexion and a high activity
50 of RF during swing phase, in addition to a positive RF spasticity (Ely) test, are the main factors for
51 deciding to perform RF transfer surgery [4]. Patikas et al. [5] also suggested that EMG could be used
52 to better interpret gait in children with hemiplegic spastic CP. They reported a prolonged activation of
53 RF during the swing phase as a part of the underlying gait compensatory mechanisms in these patients.
54 Additionally, they stated that EMG signals can describe the clinical condition of patients before and
55 after surgery [6]. Particularly, several studies also reported a strong association between the EMG of
56 RF and gait impairments in these patients [7-9].

57 Although the literature confirms the importance of EMG for treatment decision-making in patients
58 with CP, there are still some limitations here, such as cross-talk, artifacts, and poor signal quality [10],
59 for using these data in clinical settings. In addition, interpreting the results derived from these signals
60 requires the expertise of a clinician and, currently, this is primarily conducted visually and qualitatively
61 [11]. Consequently, there is a need for an approach that can assist clinicians by providing an objective
62 analysis and interpretation of the EMG signals.

63 Some studies quantitatively classified CP patients into different groups based on kinematics. Sutherland
64 and Davids [12] classified common gait abnormalities of the knee in CP into four types: crouch, jump,
65 recurvatum, and stiff knee. Rodda et al. [13] proposed an algorithm using a combination of 3D gait
66 analysis, videos, and clinical examinations to classify the gait of patients with hemiplegia and diplegia.
67 While these studies [14] have shown that kinematics-based grouping is helpful for treatment

68 management and clinical decision-making in CP, a comprehensive EMG-based grouping system for
69 these patients remains a challenge. The lack of access to an extensive EMG database definitively poses
70 a significant problem for developing such systems.

71 Clustering analysis represents an analytical technique to group-unlabeled data for extracting
72 meaningful information [15]. In recent studies, researchers used a K-means clustering algorithm as an
73 unsupervised approach to classify the pathological gait patterns observed in patients with CP [16, 17].
74 Sangeux et al. [18] conducted a study using a large dataset of CP patients to compare sagittal gait
75 patterns and K-means clustering. They introduced the "Plantar flexor–Knee extension (PFKE) couple
76 index", which measures the distance of ankle and knee kinematics during 20 to 45% of the gait cycle
77 relative to normative data. Their findings revealed a significant association between the traditional CP
78 gait groups and the five clusters identified through the PFKE-based K-means algorithm. Additionally,
79 they observed a correlation between the clusters and spasticity in the gastrocnemius-soleus muscles.

80 K-means is a powerful clustering algorithm that is widely used for various clustering problems.
81 However, this method is associated with two significant limitations : 1) converging to a local minimum
82 and 2) sensitivity to selecting the initial cluster centroids, which converges on the local rather than the
83 global optima [19]. Therefore, the initial selection of cluster centroids plays a critical role in processing
84 the K-means algorithm. This challenge can also be considered as the optimization of an "objective
85 function" that effectively groups the points in the data space into clusters. To address this problem,
86 researchers have proposed several methods that employ global optimization search algorithms to
87 determine the initial points for the K-means algorithm [20]. Particle swarm optimization (PSO) is a
88 good, nature-inspired and population-based, effective global optimization algorithm [21].

89 To the best of our knowledge, no clustering research has been conducted regarding EMG in patients
90 with CP while walking. In the current study, we focused on EMG of the RF because of the importance
91 of this muscle in the gait of these patients. Developing a hybrid PSO- K-means clustering algorithm,

92 we aimed to analyze the EMG data of patients with CP and to evaluate the relationship between
93 common CP gait abnormalities, changes in gait over time, and the identified clusters. Therefore, we
94 hypothesize that the RF EMG patterns of patients with CP are not uniform and that they can be
95 classified into clusters that might be linked to the clinical picture. Furthermore, examining these EMG
96 patterns before and after treatment might help us identify EMG features that could aid in patient
97 prognosis.

98 **2 Materials and Methods**

99 **2-1 Participants**

100 The data analyzed in this study were part of a larger database established at the local University Clinics
101 in the years 2000-2022 and derived from more than 2000 hemiplegia/diplegia patients with CP and
102 about 350 typically developed individuals. This study was approved by the ethics committee of the
103 Medical Faculty of Heidelberg University, Germany, with the serial number S-243/2022. The day of
104 access to the data in the local clinical data base was May 9th 2022. Only author (SIW) had access to
105 the identity of patients at the day of data collection. After data collection the data set was anonymized.
106 In agreement the local ethics committee we refrained from obtaining patients'/subjects' consent since
107 a) the data to be evaluated are already available at the investigating institution and the original data
108 collection was carried out as part of routine medical care and the data are pseudonymized and further
109 processed for research purposes only; and b) for the purpose of research, only persons who have
110 previously been authorized to inspect patients' data for routine medical care and have access to
111 personal data may access the data again, and c) there is no disclosure of personal data to external
112 bodies.

113

114 The retrospective gait data used for this study were primarily divided into two groups. The group of
 115 patients who visited the gait lab for the first time (E1, first examination) and those who visited a second
 116 time, too (E2, second examination). Therefore, it should be noted that the E2 patients in our study
 117 represent the same individuals examined at two different time points. The inclusion criteria were
 118 availability of EMG, kinematics and kinetics data for each subject, walking only barefoot without any
 119 assistive device, and classified as GMFCS level I, II, or III [14]. For hemiplegia patients, only the
 120 affected side was considered. After applying these criteria, 881 and 377 patients were recruited as E1
 121 and E2, respectively, and 117 persons as typically developed (TD) (reference group). In addition, the
 122 patients with CP were classified as crouch, jump, recurvatum, stiff knee, and mild gait [12, 22]. The
 123 characteristics of the participants are shown in Table 1.

124

125

Table 1. Demographic and descriptive data of the participants at their first (E1) and second (E2) examinations and also healthy individuals.

	<i>E1 (n=881)</i>	<i>E2 (n=377)</i>	<i>TD (n=117)</i>
<i>Age (years)</i>	16.8 ± 9.6 (58.3-3.3)	16.9 ± 8.5 (54.2-3.7)	21.7 ± 12.3 (46 -6)
<i>Height (cm)</i>	151.9 ± 37.1 (197-94)	152.6 ± 18.2 (196-98)	161.9 ± 19.8 (195-108)
<i>Body mass (kg)</i>	46 ± 19.5 (125.1-13.6)	46.8 ± 17.3 (101.8-14.5)	66.4 ± 16.8 (91-19)
<i>Sex (male/female)</i>	510/371	212/165	58/59
<i>CP type (diplegia/ hemiplegia)</i>	700/181	332/45	
<i>Subgroups (crouch/ jump/ recurvatum/ stiff/ mild)</i>	72/65/109/42/593	38/28/24/19/268	
<i>GMFCS level (I/ II/ III)</i>	140/130/15	59/67/3	
<i>Interval between examinations (years)</i>	2.4 ± 1.9 (13.5-0.1)		

126

127 **2-2 Data acquisition**

128 All subjects walked barefoot at a self-selected speed along a lane 15 m in length during data acquisition.
129 Kinematics and kinetics were recorded using a twelve-camera 3D motion analysis system (VICON,
130 Oxford Metrics Limited, UK) operating at 120 Hz and using three force plates (Kistler Instruments
131 Co.), respectively. The skin-mounted markers were applied according to the protocol of Kadaba et al.
132 [23] and the plug-in-gait model was chosen for analysis. Subsequently, gait parameters of at least seven
133 strides were determined.

134 The EMG data were recorded from eight lower extremity major muscles including the RF, right and
135 left legs, using myon 320 (Myon AG, Schwarzenberg, CH). Bipolar surface adhesive electrodes (Blue
136 Sensor, Ambu Inc., Glen Burnie, MD, USA) were placed on the targeted muscles, following the
137 guidelines provided by SENIAM [24]. The distance between the electrodes was set at 2 cm [6]. To
138 amplify the EMG signal, the Biovision EMG apparatus (Biovision Inc., Wehrheim, Germany) before
139 2013/2014 and via Delsys (Delsys Inc., Natick, MA, USA) after 2013/14, was utilized with a
140 preamplification factor of $\times 5000$.

141 Clinical examination was exclusively performed by two physiotherapists controlling each other. Knee
142 extensor muscle strength was assessed according to the Medical Research Council (MRC) [25]. The
143 spasticity of the RF was tested both by the Duncan-Ely test [26] and by the Tardieu test [27]. According
144 to the MRC, muscle strength ranged from 5 (the strongest) to 1 (the weakest) The scale for spasticity
145 ranged from 0 (no spasticity) to 4 (severe spasticity). More information about these (strength and
146 spasticity) grading systems is available in our previous work [28].

147 **2-3 Signal processing**

148 The raw EMG data were then band-pass filtered (Butterworth filter with a cutoff frequency of 20-350
149 Hz), rectified, and the signal smoothed (Butterworth low-pass filter with a cutoff frequency of 9 Hz),
150 amplitude-normalized to the mean of signal, time-normalized within one gait cycle (101 datapoint),
151 and eventually averaged across valid strides in MatLab (The MathWorks, Inc. USA) [6]. Six main

152 features (mean, range, max, min, and their timing) during ten gait phases [6], including whole-gait-
153 cycle stance, swing, loading response, mid-stance, terminal stance, preswing, initial swing, mid-swing,
154 and terminal swing phase, were extracted for clustering.

155 To determine a magnitude that describes the deviation of a patient's EMG feature from the reference
156 group, the norm-distance (NDi) was calculated according to [6] by Equation (1). ND was defined as
157 the absolute difference between the i _th feature of the EMG of the patient p (F_{pi}) and the mean value
158 of the same feature in the reference group (\bar{F}_{ni}), divided by the corresponding standard deviation within
159 the reference group (SD_{ni}). This standardization process served as the initial input for subsequent steps
160 and can potentially help as a data transformation method to identify meaningful clusters.

$$\text{Eq. 1.} \quad NDi = \frac{|F_{pi} - \bar{F}_{ni}|}{SD_{ni}}$$

161 **2-4 Cluster analysis**

162 We applied cluster analysis as an unsupervised stand-alone tool to gain insight into the data
163 distribution, examine the distinct characteristics of each cluster, and prioritize specific clusters for
164 subsequent analysis. Therefore, we applied two feature matrices with dimension of 881×60 (number
165 of feature (6) \times number of gait phases (10)) and 377×60 for E1 and E2 conditions, respectively. Prior
166 to performing clustering analysis, we employed principal component analysis (PCA) to reduce the
167 dimensionality of the input matrices [29]. The first principal components (PCs), which explained more
168 than 96% of the total variance, were used for the clustering. The hybrid clustering algorithm was
169 developed using MatLab software based on the details described in Supplementary material (Appendix
170 1). In the initial stage, the PSO algorithm was employed for a global search to explore the possible
171 optimal solutions to predefine the number of clusters. The output of PSO served as the initial centroids
172 for the K-means algorithm, which was then utilized to refine and generate the final result. Using the
173 elbow method [30], the number of 'K' was determined. In this method, the changes in the sum of

174 squared differences between the observations and their cluster (SSE) were analyzed by adding the
175 number of clusters. The point at which there is a sharp change in the elbow curve indicates the K. In
176 this study, we applied the elbow method on the E1 dataset to determine the K; then we set this for the
177 E2 clustering as well.

178 **2-5 Comparison between the clusters**

179 The Pearson correlation (r) was applied to compare the averaged RF EMG of patients in the different
180 clusters with that of the typically developed group. A more comprehensive investigation of the clusters
181 (labels)' characteristics conducted through 1) extracting the average hip, knee, and ankle joint angles
182 and moments of each cluster at E1 and E2 in the sagittal plane and 2) examining the population of the
183 gait subgroups (crouch, jump, recurvatum, stiff knee, and mild) in each cluster. Descriptive statistics
184 (mean and standard deviation) were used to compare the clinical examination data between the
185 conditions.

186 **2-6 Comparison between the groups**

187 Considering the changes in the EMG cluster and gait kinematics of the patients from their second
188 examination (E2), we set two cohorts of patients in E1 in an identified cluster (section 3) who were
189 determined as crouch and who did not show any changes (G1) or showed improvement (G2) in their
190 gait later according to E2. All of these patients underwent single-event multilevel surgery (SEMLS)
191 between the examinations. Gait profile score (GPS) [31] was computed for the groups in both E1 and
192 E2 to assess gait improvement.

193 To investigate the gait factors resulting in these different responses to the treatment, the EMG data of
194 the RF muscle from G1 and G2 patients at E1 were compared by extracting six main features (section
195 2-3) during 20 to 45% of their gait cycle, as described by Sangeux et al. [18]. Applying the
196 nonparametric Kruskal-Wallis test, we compared the features between G1 and G2 at E1 (p-Value =
197 0.05). Furthermore, we used statistical parametric mapping (SPM, www.spm1d.org) implemented in

198 MatLab [32] to compare joint patterns. Fig 1 illustrates the procedure we used in this study in a flow
199 chart.

200 **Please insert Fig 1**

201 **Fig 1. Flowchart of the methodology used in this study. EMG, electromyography; PSO, particle**
202 **swarm optimization**

203 **3 Results**

204 Fig 2 shows an elbow plot with $K=3$, representing the number of clusters (L1, L2, and L3). The first
205 25 PCs accounted for approximately 96.1% of the variance in the primary dataset and were utilized for
206 the clustering analysis. Comparing the EMG of RF between the three identified clusters in two
207 examinations (Fig 3A), they were characterized by a peak during mid-swing (L1), prolonged activity
208 during (mid- and terminal) stance (L2), and a significant peak during loading response (L3). Notably,
209 L2 contained 76% (55 out of 72) and 92% (35 out of 38) of all crouch patients at E1 and E2, respectively
210 (Table 2). An excessive knee flexion during the stance phase of gait (Fig 3B), along with lack of an
211 extension moment (Fig 3C), for patients detected as cluster 2, supported the sensitivity of our clustering
212 to the crouch gait.

213 **Please insert Fig 2**

214 **Fig 2. Determination of the number of clusters using the elbow method. The scatter plots show**
215 **the distribution of input data (E1) in different clusters in which the axes are PC1 and PC2.**

216 **Each color represents a cluster.**

217 Correlation analysis (Table 2) revealed a similarity of more than 0.9 between the average EMG pattern
218 of patients placed in L3 and the normal population, while L1 patients had the lowest coefficient by 39
219 and 53% at E1 and E2, respectively.

220 In addition, averaged clinical assessment of the patients (as presented in Table 2 descriptively) revealed
 221 that the L2 group had lower strength and higher levels of spasticity in their knee extensors than did the
 222 other two groups. Moreover, the number of patients with more severe motor impairment (GMFCS level
 223 III) was also higher in the L2 than in the L1 and L3 groups. Comparing G1 (16 patients, 32 limbs) and
 224 G2 (8 patients, 15 limbs), there was a significant improvement in the GPS of G2 at E2 (Table 3).

225

226

227

228

229

Table 2- Number of patients with a gait abnormality, clinical examination data (mean \pm SD), and number of patients with different GMFCS level in each cluster examination.

<i>Conditions (Clusters- Examinations)</i>	Gait subgroups count per condition					r- correlation to normal EMG	Clinical Examination Parameters			GMFCS level (I/ II/ III)
	Stiff	Recurvatum	Mild	Jump	Crouch		Knee extensors strength	Rectus spasticity Tardieu	Rectus spasticity Duncan-Ely	
<i>E1-L1</i>	11	63	224	25	7	0.39	4.5 \pm 0.6	0.66 \pm 1	0.64 \pm 0.5	57/40/2
<i>E1-L2</i>	21	26	178	23	55	0.67	4.3 \pm 0.6	1.1 \pm 1	0.71 \pm 0.4	19/45/11
<i>E1-L3</i>	10	21	190	17	10	0.91	4.6 \pm 0.5	0.44 \pm 0.8	0.49 \pm 0.5	64/45/2
<i>E2-L1</i>	6	6	107	10	3	0.53	4.3 \pm 0.6	0.68 \pm 1	0.72 \pm 0.5	30/20/0
<i>E2-L2</i>	5	9	79	12	35	0.7	4.2 \pm 0.6	0.96 \pm 1	0.69 \pm 0.5	10/29/2
<i>E2-L3</i>	8	8	83	6	0	0.92	4.4 \pm 0.6	0.44 \pm 0.9	0.59 \pm 0.5	19/18/1

230

Please insert Fig 3

231 **Fig 3. Average rectus femoris electromyography (EMG) (A), knee kinematic (B) and knee**
 232 **kinetic (C) patterns for different clusters and examinations. Red: L1, Blue: L2, Green: L3,**
 233 **Black: reference group, Solid lines: E1, Dashed lines: E2.**

Table 3- Mean \pm SD and statistical comparison for gait profile score (GPS) and rectus femoris electromyography (EMG) features at 20-45% gait between patients with crouch gait (cluster 2) measured before and after surgery (E1 and E2) and demonstrated no significant improvement (G1) or significant improvement during the second examination (G2).

	GPS			
	G1	G2	Normal	
<i>E1</i>	16.25 \pm 2.6	18.4 \pm 5.9	4.87 \pm 1.09	
<i>E2</i>	15.16 \pm 3.9	14.3 \pm 6.3		
<i>p-value between E1 and E2 within each Group</i>	0.167	0.033*		
	Rectus EMG features in 20-45% gait cycle during E1			
	G1	G2	<i>p-value between G1 and G2 for each feature</i>	Normal
<i>Min 20-45%</i>	85.1 \pm 20.2	69.2 \pm 25.7	0.025*	38.3 \pm 14.5
<i>Mean 20-45%</i>	110.8 \pm 16.3	100.8 \pm 23.4	0.144	56.7 \pm 21
<i>Max 20-45%</i>	137.8 \pm 19.1	128.1 \pm 27.4	0.121	92.2 \pm 36.8

*p-value<0.05

234

Please insert Fig 4

235 **Fig 4. Average rectus femoris electromyography (EMG) (A), knee kinematic (B) and knee**
236 **kinetic (C) patterns for cluster 2 patients that did not improve (G1) or improved (G2) after**
237 **surgery (E2) compared to their condition before surgery (E1). Red: G1, Blue: G2, Solid line:**
238 **E1, Dashed line: E2.**

239 The SPM results did not show any systematic difference ($p < 0.05$) between knee kinematics of G1 and
240 G2 individuals at E1 (Fig 4A and Fig 4B). Visually, however, the stance peak extension moment for
241 both groups is at the same level (Fig 4C). As increased RF activity during stance was the main
242 characteristic of cluster 2, a subjective comparison between the two groups in this phase (Fig 4A)
243 showed a higher average EMG in G1-E1 than G2-E1. Statistically, as shown in Table 2, there is a
244 systematic difference between the minimum activity of RF during 20-45% gait between the two groups
245 ($p = 0.025$). Mean and maximum (features) were also lower for G2. We only reported the main three
246 features with the lowest p-value.

247 **4 Discussion**

248 Using an unsupervised hybrid PSO- K-means cluster analysis, three main groups were identified from
249 EMG data of the RF in patients with CP. These clusters differed from each other in level of activity in
250 swing (L1), stance (L2), and loading response (L3) (Fig 3A). Applying a pre-clustering standardization
251 technique in combination with PCA, our clustering system could categorize the patients in relation to
252 the deviation of their EMG results from those of a healthy population. Subsequently, a correlation of
253 more than 90% was observed between one of the clusters (L3) and a normal EMG (Table 2).
254 Descriptively, L3 patients showed a stronger knee extensor and a lower RF spasticity. On the other
255 hand, patterns in patients identified as L1 correlated least with normal patterns (Table 2). It has been
256 reported that in the majority of children with CP the RF is active during the mid-swing phase, when
257 this muscle is normally inactive [33]. A high swing peak observed in the mean EMG of this cluster
258 (L1) aligns with this typical feature of CP.

259 Furthermore, identifying a significant number of patients with a crouch gait as L2 (Table 2) is
260 consistent with the prolonged activation of RF during stance phase, which is supported by the current
261 literature [12]. In healthy individuals, quadriceps muscles are typically active for a small portion of the
262 stance phase. However, in patients with crouch gait, the positioning of the ground reaction force behind
263 the center of the knee joint requires the quadriceps to be engaged throughout the entire stance phase in
264 order to maintain stability of the knee joint [34]. Table 1 demonstrates higher levels of RF and weaker
265 knee extensor strength among individuals classified as L2 at E1 compared to L1 and L3. This cluster
266 also exhibited a greater proportion of patients with a higher GMFCS level, which may support the
267 notion that this EMG pattern might be linked to severity of the disability.

268 Investigating the relationship between EMG activity and altered kinematics in G1 patients with
269 persistent crouch condition and G2 patients with significant GPS improvement, we observed (Table 3)
270 that a lower minimum activity of RF during 20-45% of the gait cycle was a significant ($p=0.025$)
271 indicator of a better GPS at E2. However, the changes from E1 to E2 for sagittal knee kinematics of
272 G1 and G2 were not significant. These findings suggest that EMG patterns are important for treatment
273 decision-making, but that the effectiveness of kinematics, as a widely used clinical measure, is limited.
274 The G1-G2 EMG comparison was made for the period 20-45% of the gait cycle. Sangeux et al. [18]
275 developed an index to categorize CP subgroups by considering only this 25% of the gait cycle, mainly
276 to avoid the loading response effects on the stance phase features. As the relationship between muscle
277 EMGs and joint kinetics is nonlinear [35], it is difficult to find a direct explanation for the differences
278 in the kinetics patterns of G1-G2 at E1 based on the RF activity.

279 Several algorithms are available in the literature to determine the optimal number of clusters [36];
280 however, in this study, the elbow curve at $K=3$ was obvious and an exact biomechanical meaning for
281 the clusters supports the algorithm results. For instance, each cluster showed a prominent activation in
282 different gait phases: there was a cluster with a significant number of patients with crouch and a weaker

283 RF, in addition to a cluster with patients with (correlated) normal EMG patterns. In this research, we
284 used blind, unsupervised clustering without any prior information about the available EMG data. The
285 only inclusion criteria applied were the availability of data from patients who walked barefoot and
286 without any assistance. Applying such an algorithm on a large database (collected in our center) with
287 more than 1000 examinations aided us in identifying general trends in the EMG data of CP patients.
288 Furthermore, the term E2 in our study addressed the second examination of patients in whom a gait
289 test (E1) had previously been performed in the lab. As a result, this does not apply to a before-after
290 surgery scenario. However, it is important to note that these patients are typically engaged in an
291 everyday training program. In the present study, the inclusion or exclusion criteria did not specifically
292 consider the aspect of "treatment" initially. Moreover, the crosstalk from the surrounding vastus
293 lateralis muscle on the activity of RF has recently been extensively reported as a common issue when
294 employing surface EMG, especially in the presence of crouch gait observed in children with CP [37].
295 However, utilizing wire fine EMG, true activation of RF was noted in 30-45% of crouched gait cycles
296 [37]. Our study specifically focuses on the 20-45% phase of gait in G1 and G2, addressing this actual
297 part of RF activity. Nevertheless, despite lacking access to an extensive database of wire fine EMG
298 data from CP patients, we believe that surface EMG, as a non-invasive and clinically relevant measure
299 of gait in individuals with CP, has the capacity to capture the primary characteristics of muscle activity
300 in these patients.

301 To overcome the K-mean algorithm initialization problem, we developed a hybrid K-mean and PSO
302 optimization approach. As no similar study on the clustering of EMG data in CP has been conducted,
303 our research focused only on one muscle (RF) and, further, the crouch condition to investigate the
304 possibility of developing CP-EMG assessment approaches. These findings seem promising and suggest
305 that clustering analyses should be applied on datasets with more muscles. In addition, we mainly

306 evaluated knee movement while other joints and gait abnormalities could also be included in future
307 studies.

308 **5 Conflict of Interest**

309 ‘The authors declare that the research was conducted in the absence of any commercial or financial
310 relationships that could be construed as a potential conflict of interest.’

311 There is no conflict of interest.

312 **6 Author Contributions**

313 MD: Writing original draft, Data analysis; FS: Review & editing, Methodology; RR: Data analysis;
314 DAP: Review & editing; SIW: Review & editing, Conceptualization, Methodology, Project
315 administration.

316 **7 Funding**

317 This research was funded by German Research Foundation (DFG) (no: WO 1624/ 8-1).

318 **8 Supplementary material**

319 “S1 Appendix”.

320 **References**

- 321 1. Carr LJ, Reddy SK, Stevens S, Blair E, Love S. Definition and classification of cerebral
322 palsy. *Developmental medicine and child neurology*. 2005;47(8):508-10.
- 323 2. Patel DR, Neelakantan M, Pandher K, Merrick J. Cerebral palsy in children: a clinical
324 overview. *Translational pediatrics*. 2020;9(Suppl 1):S125-s35.
- 325 3. Dreher T, Wolf SI, Maier M, Haggmann S, Vegvari D, Gantz S, et al. Long-term results after
326 distal rectus femoris transfer as a part of multilevel surgery for the correction of stiff-knee gait in
327 spastic diplegic cerebral palsy. *The Journal of bone and joint surgery American volume*.
328 2012;94(19):e142(1-10).
- 329 4. Reinbolt JA, Fox MD, Schwartz MH, Delp SL. Predicting outcomes of rectus femoris transfer
330 surgery. *Gait & posture*. 2009;30(1):100-5.
- 331 5. Patikas D, Wolf S, Döderlein L. Electromyographic evaluation of the sound and involved side
332 during gait of spastic hemiplegic children with cerebral palsy. *European journal of neurology*.
333 2005;12(9):691-9.

- 334 6. Patikas D, Wolf SI, Schuster W, Armbrust P, Dreher T, Döderlein L. Electromyographic
335 patterns in children with cerebral palsy: do they change after surgery? *Gait & posture*.
336 2007;26(3):362-71.
- 337 7. Kay RM, Pierz K, McCarthy J, Graham HK, Chambers H, Davids JR, et al. Distal rectus
338 femoris surgery in children with cerebral palsy: results of a Delphi consensus project. *Journal of*
339 *children's orthopaedics*. 2021;15(3):270-8.
- 340 8. Gage JR, Perry J, Hicks RR, Koop S, Werntz JR. Rectus femoris transfer to improve knee
341 function of children with cerebral palsy. *Developmental medicine and child neurology*.
342 1987;29(2):159-66.
- 343 9. Josse A, Pons C, Printemps C, Chan-Wai-Nam J, Affes H, Brochard S, et al. Rectus femoris
344 surgery for the correction of stiff knee gait in cerebral palsy: A systematic review and meta-analysis.
345 *Orthopaedics & traumatology, surgery & research : OTSR*. 2023;109(3):103022.
- 346 10. Türker KS. Electromyography: some methodological problems and issues. *Physical therapy*.
347 1993;73(10):698-710.
- 348 11. Schmidt-Rohlfing B, Bergamo F, Williams S, Erli HJ, Rau G, Niethard FU, et al.
349 Interpretation of surface EMGs in children with cerebral palsy: An initial study using a fuzzy expert
350 system. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*.
351 2006;24(3):438-47.
- 352 12. Sutherland DH, Davids JR. Common gait abnormalities of the knee in cerebral palsy. *Clinical*
353 *orthopaedics and related research*. 1993(288):139-47.
- 354 13. Rodda JM, Graham HK, Carson L, Galea MP, Wolfe R. Sagittal gait patterns in spastic
355 diplegia. *The Journal of bone and joint surgery British volume*. 2004;86(2):251-8.
- 356 14. Papageorgiou E, Nieuwenhuys A, Vandekerckhove I, Van Campenhout A, Ortibus E,
357 Desloovere KJG, et al. Systematic review on gait classifications in children with cerebral palsy: an
358 update. 2019;69:209-23.
- 359 15. Oyewole GJ, Thopil GA. Data clustering: application and trends. *Artificial intelligence*
360 *review*. 2023;56(7):6439-75.
- 361 16. Kuntze G, Nettel-Aguirre A, Ursulak G, Robu I, Bowal N, Goldstein S, et al. Multi-joint gait
362 clustering for children and youth with diplegic cerebral palsy. *PloS one*. 2018;13(10):e0205174.
- 363 17. Choisne J, Fourrier N, Handsfield G, Signal N, Taylor D, Wilson N, et al. An Unsupervised
364 Data-Driven Model to Classify Gait Patterns in Children with Cerebral Palsy. *Journal of clinical*
365 *medicine*. 2020;9(5).
- 366 18. Sangeux M, Rodda J, Graham HK. Sagittal gait patterns in cerebral palsy: the plantarflexor-
367 knee extension couple index. *Gait & posture*. 2015;41(2):586-91.
- 368 19. Prakash C, Kumar R, Mittal N. Optimized Clustering Techniques for Gait Profiling in
369 Children with Cerebral Palsy for Rehabilitation. *The Computer Journal*. 2018;61(11):1683-94.
- 370 20. Rana S, Jasola S, Kumar RJIJoE, Science, Technology. A hybrid sequential approach for data
371 clustering using K-Means and particle swarm optimization algorithm. 2010;2(6).
- 372 21. Van Der Merwe D W. Data clustering using particle swarm optimization. *The 2003 Congress*
373 *on Evolutionary Computation (CEC03)*. 2003;1:215-20.
- 374 22. Guo L-Y, Su F-C, Lin C-J, Cherng R-J, Chou Y-LJG, Posture. Common abnormal kinetic
375 patterns of the knee in gait in cerebral palsy. 1997;2(5):151-2.
- 376 23. Kadaba MP, Ramakrishnan H, Wootten MJJoor. Measurement of lower extremity kinematics
377 during level walking. 1990;8(3):383-92.
- 378 24. Hermens HJ, Freriks B, Merletti R, Stegeman D, Blok J, Rau G, et al. European
379 recommendations for surface electromyography. 1999;8(2):13-54.
- 380 25. John J, John JIJorr. Grading of muscle power: comparison of MRC and analogue scales by
381 physiotherapists. 1984;7(2):173-82.

- 382 26. Marks M, Alexander J, Sutherland D, Chambers HJ, neurology c. Clinical utility of the
383 Duncan-Ely test for rectus femoris dysfunction during the swing phase of gait. 2003;45(11):763-8.
- 384 27. Gracies J-M, Burke K, Clegg NJ, Browne R, Rushing C, Fehlings D, et al. Reliability of the
385 Tardieu Scale for assessing spasticity in children with cerebral palsy. 2010;91(3):421-8.
- 386 28. Wolf SI, Mikut R, Kranzl A, Dreher TJG, Posture. Which functional impairments are the
387 main contributors to pelvic anterior tilt during gait in individuals with cerebral palsy?
388 2014;39(1):359-64.
- 389 29. Amene J, Krzak JJ, Kruger KM, Killen L, Graf A, Altiok H, et al. Kinematic foot types in
390 youth with pes planovalgus secondary to cerebral palsy. 2019;68:430-6.
- 391 30. Joshi KD, Nalwade PS, JoCS, Computing M. Modified k-means for better initial cluster
392 centres. 2013;2(7):219-23.
- 393 31. Baker R, McGinley JL, Schwartz MH, Beynon S, Rozumalski A, Graham HK, et al. The gait
394 profile score and movement analysis profile. 2009;30(3):265-9.
- 395 32. Pataky TC. Generalized n-dimensional biomechanical field analysis using statistical
396 parametric mapping. 2010;43(10):1976-82.
- 397 33. DeLuca P, Bell K, Davis RJG, Posture. Using surface electrodes for the evaluation of the
398 rectus femoris, vastus medialis and vastus lateralis muscles in children with cerebral palsy.
399 1997;5(3):211-6.
- 400 34. Ganjwala D, Shah H. Management of the knee problems in spastic cerebral palsy.
401 2019;53:53-62.
- 402 35. Lloyd DG, Besier TF. An EMG-driven musculoskeletal model to estimate muscle forces
403 and knee joint moments in vivo. 2003;36(6):765-76.
- 404 36. Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-Means
405 Clustering. 2013;1(6):90-5.
- 406 37. Barr KM, Miller AL, Chapin KBJG, posture. Surface electromyography does not accurately
407 reflect rectus femoris activity during gait: impact of speed and crouch on vasti-to-rectus crosstalk.
408 2010;32(3):363-8.

409