Journal of Biomechanics 155 (2023) 111668

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech

Machine learning-based prediction of joint moments based on kinematics in patients with cerebral palsy

Mustafa Erkam Ozates [a], Derya Karabulut [b], Firooz Salami [c], Sebastian Immanuel Wolf [c], Yunus Ziya Arslan [a,*]

[a] Department of Robotics and Intelligent Systems, Institute of Graduate Studies in Science and Engineering, Turkish-German University, Istanbul, Turkey
[b] Department of Mechanical Engineering, Faculty of Engineering, Istanbul University-Cerrahpaşa, Istanbul, Turkey
[c] Clinic for Orthopaedics and Trauma Surgery, Heidelberg University Hospital, Heidelberg, Germany

ARTICLE INFO

Keywords:
Convolutional neural networks
Joint moments
Gait kinematics
Computational gait analysis
Cerebral palsy
Machine Learning

ABSTRACT

Joint moments during gait provide valuable information for clinical decision-making in patients with cerebral palsy (CP). Joint moments are calculated based on ground reaction forces (GRF) using inverse dynamics models. Obtaining GRF from patients with CP is challenging. Typically developed (TD) individuals' joint moments were predicted from joint angles using machine learning, but no such study has been conducted on patients with CP. Accordingly, we aimed to predict the dorsi-plantar flexion, knee flexion–extension, hip flexion–extension, and hip adduction-abduction moments based on the trunk, pelvis, hip, knee, and ankle kinematics during gait in patients with CP and TD individuals using one-dimensional convolutional neural networks (CNN). The anonymized retrospective gait data of 329 TD (26 years ± 14, mass: 70 kg ± 15, height: 167 cm ± 89) and 917 CP (17 years ± 9, mass:47 kg ± 19, height:153 cm ± 36) individuals were evaluated and after applying inclusion–exclusion criteria, 132 TD and 622 CP patients with spastic diplegia were selected. We trained specific CNN models and evaluated their performance using isolated test subject groups based on normalized root mean square error (nRMSE) and Pearson correlation coefficient (PCC). Joint moments were predicted with nRMSE between 18.02% and 13.58% for the CP and between 12.55% and 8.58% for the TD groups, whereas with PCC between 0.85 and 0.93 for the CP and between 0.94 and 0.98 for the TD groups. Machine learning-based joint moment prediction from kinematics could replace conventional moment calculation in CP patients in the future, but the current level of prediction errors restricts its use for clinical decision-making today.

1. Introduction

Cerebral Palsy (CP) is a group of disorders that affects a person's neuromotor functions. Lower limb joint moments are of great importance in the assessment, monitoring, and treatment of CP (Lai et al., 1988; Gage, 1994; Ounpuu et al., 1996; Lin et al., 2000; Novacheck and Gage, 2007). Joint moments have also great potential to provide insights about muscle behaviors during exerting joint motion. Research showed that pre- and postoperative analyses of joint moments, particularly those of the lower extremity in the sagittal plane and hip joint moment in the frontal plane, have an impact on the decision-making process of the treatment of CP (Ounpuu et al., 1996; DeLuca et al., 1997; Kay et al., 2000, Rhodes et al., 2023). For example, the insufficiency of quadriceps strength is one of the primary factors that contribute to crouch gait in patients with CP (Lenhart et al., 2017). Therefore, the magnitude and pattern of the knee extension moment could reflect the impact of quadriceps weakness on crouch gait, making it crucial for surgical decision-making in such cases (Lenhart et al., 2017; Karabulut et al., 2021).

Obtaining joint moments in clinical gait analysis leads to challenges. It requires ground reaction force (GRF) measurements to be calculated using inverse dynamics models (Winter, 2009; Whittle, 2014). However, the measurement of the GRF can be hard to capture during natural walking (Caldas et al., 2020) and even more problematic for deviated gaits as in cases of CP (White et al., 1999). Therefore, some attempts on predicting joint moments have been made in recent years, which can eliminate clinical measurement challenges in patients with deviated gait by using either machine learning (ML) techniques (De Brabandere et al.,

* Corresponding author at: Institute of Graduate Studies in Science and Engineering, Turkish-German University, Sahinkaya Cad. No, 94 - 34820 Beykoz / Istanbul, Turkey.
E-mail address: yunus.arslan@tau.edu.tr (Y.Z. Arslan).

https://doi.org/10.1016/j.jbiomech.2023.111668
Accepted 26 May 2023
Available online 27 May 2023

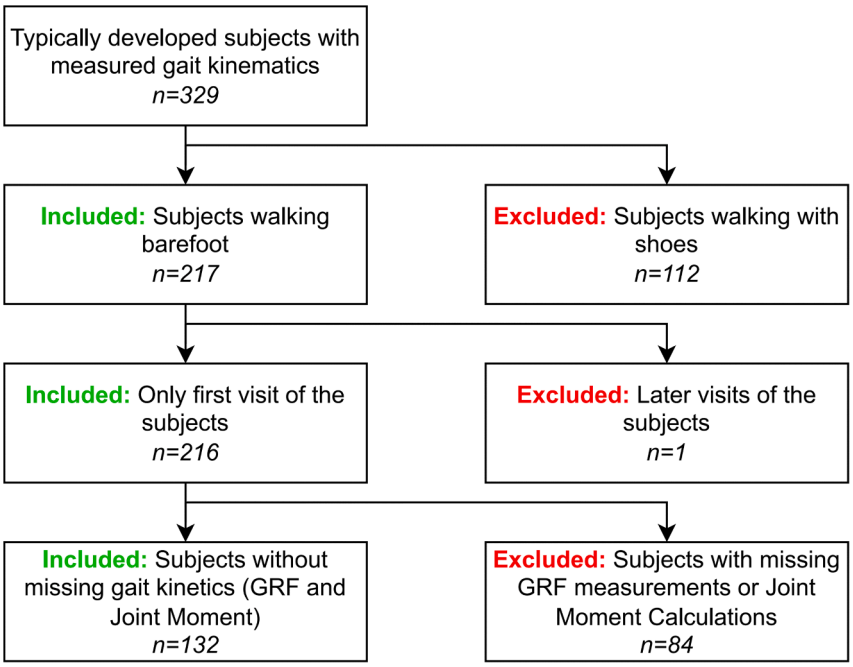

Fig. 1. Inclusion-exclusion flow of the typically developed subjects. GRF: Ground reaction force.

2020; Giarmatzis et al., 2020) or musculoskeletal based predictive models (Oh et al., 2020; Richards et al., 2018; Shao et al., 2009; Kloeckner et al., 2023).

ML is a powerful tool for solving tasks with missing measurements or lacking physical models. ML algorithms have already been applied to patients with CP, which have non-uniform gait characteristics (Arslan and Karabulut, 2021), for different tasks such as the detection of CP disease using video recording (Ihlen et al., 2019) or gait kinematics (Zhang and Ye, 2019) and the classification of the gait phase of CP patients using electromyography (EMG) (Morbidoni et al., 2021) or using marker data (Kim et al., 2022). Among the ML techniques, the one-dimensional (1D) convolutional neural network (CNN) is particularly

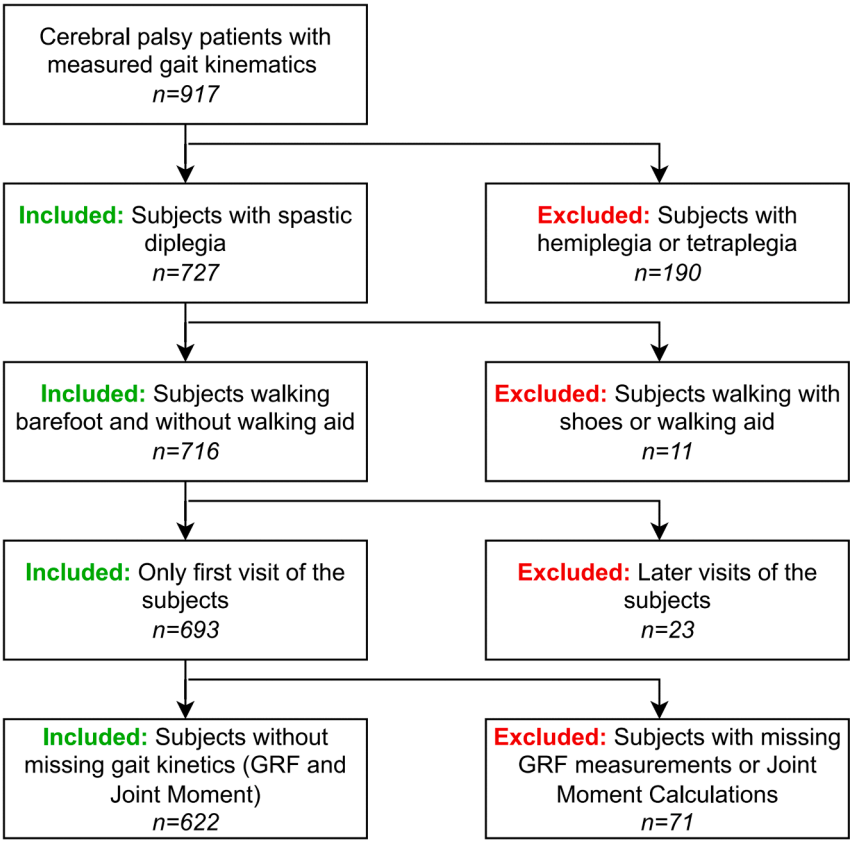

Fig. 2. Inclusion-exclusion flow of the subjects with cerebral palsy. GRF: Ground reaction force.

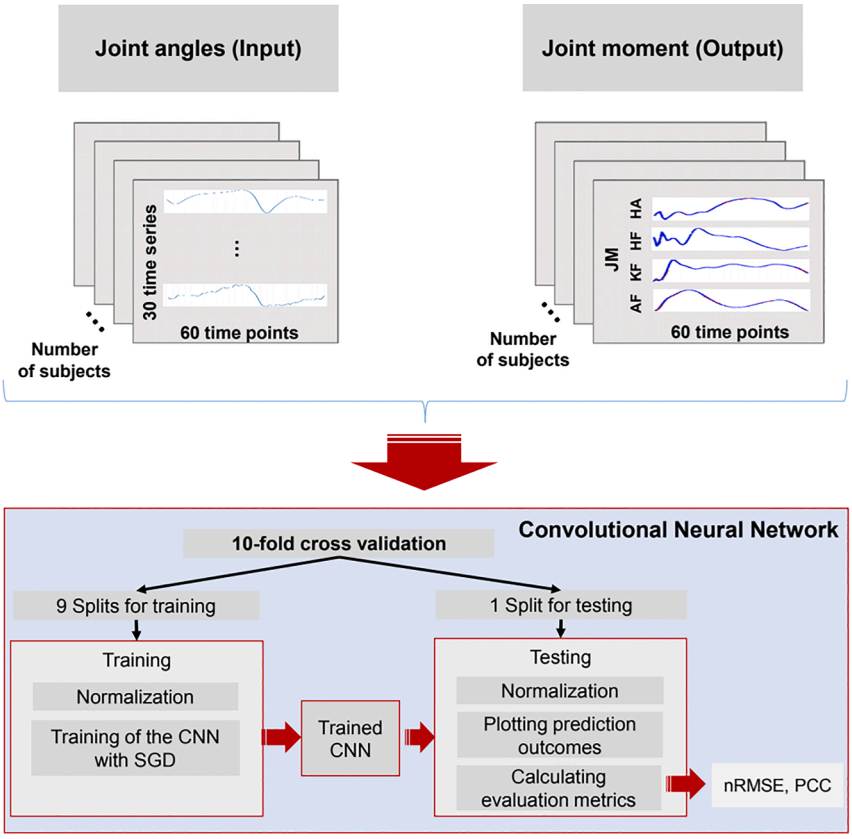

Fig. 3. Data processing and machine learning pipeline. JM: Joint moment, AF: ankle dorsi-plantar flexion, KF: knee flexion–extension, HF: hip flexion–extension, HA: hip adduction-abduction, CNN: convolutional neural network, SGD: stochastic gradient descent, nRMSE: normalized root mean square error, PCC: Pearson correlation coefficient.

suitable for processing time series information.

A few attempts were made to predict joint moments in typically developed (TD) subjects using ML during gait. For example, a feed-forward neural network and a long short-term memory neural network were able to predict joint moments of TD subjects using kinematic data from three-dimensional motion capturing successfully (Mundt et al., 2020a). A wavelet neural network, taking frequency information into account, was successful in predicting joint moments of TD subjects using both kinematic and EMG data (Ardestani et al., 2014). However, no study has yet been conducted on predicting joint moments in patients with CP.

Accordingly, we aimed to predict the dorsi-plantar flexion, knee flexion–extension, hip flexion–extension, and hip adduction-abduction moments of patients with CP during gait from joint angles based on marker data in this study. To accomplish this, the 1D CNNs were trained using the kinematic data to predict joint kinetics.

2. Methods

2.1. Subjects

The study was approved by the local ethical committee of the University Hospital (S-227/2021). The anonymized retrospective gait data of 329 TD subjects (age: 26 years ± 14, mass: 70 kg ± 15, height: 167 cm ± 89) with typical gait characteristics and 917 CP patients (age: 17 years ± 9, mass: 47 kg ± 19, height: 153 cm ± 36), which had been collected in the course of patient care, were used. Kinematic data were gathered according to the Plugin Gait Model (Oxford Metrics, Oxford, UK) with 19 markers, using a 12-camera motion capture system (Vicon Motion Systems Ltd. Oxfordshire UK) while the subjects walked at a self-selected speed. GRF data were simultaneously collected using force plates (Kistler Instruments, Winterthur, Switzerland) and joint moments, normalized by the body mass, were calculated using an inverse kinematics model (Harrison et al., 2012).

No age or gender criteria were set for including TD and CP subjects. TD subjects who walked barefoot and without missing measurements were included in the study as shown in the flowchart (Fig. 1).

The first visits of the spastic diplegia subjects who can walk without assistive devices and missing measurements were included as shown in the patient flowchart (Fig. 2). Gross Motor Function Classification System (GMFCS) levels of the patients were I and II.

2.2. Dataset

Kinematic data from the trunk, pelvis, hip, knee, and ankle in sagittal, coronal, and transverse planes (15 angles in total) and kinetic data of the flexion–extension moments of the ankle, knee, hip, and adduction-abduction moment of the hip were taken into account. All data were averaged across 7–10 strides for each subject and normalized to a percentage gait cycle. In each time series, there are 101-time points representing the gait cycle from 0% to 100%. In addition to the averaged time series, standard deviations of the time series throughout the strides were contained in the dataset as well.

The time series were segmented into stance and swing phases regarding their temporal foot-off values. Since GRF data were not available during the swing phase, only the stance phases of the time series were used to include only directly calculated moment data. Each subject has naturally a different duration of stance time, whereas the ML algorithms require the same size of data for being trained with subjects. To satisfy this, stance segments of the time series were interpolated to a standard length of 60-time points. All of the time series values were normalized in the range between 0 and 1, regardless of their unit to

avoid dominance of the time series with higher magnitudes in the learning process.

After normalization, all of the aforementioned 15 kinematic time series were stacked in a matrix with a size of 30 rows (for 15-time series and their standard deviations) and 60 columns. In total, 132 (for TD) and 622 (for patients with CP) matrices were created for training and testing in the ML process.

2.3. Machine learning

To process the time series rows and extract the features from them distinctly per time series, 1D convolution was applied in the CNN model. One-dimensional convolution layers extract features from the time series' data (in our case joint angles) separately with different temporal ranges, which sustain valuable information for predicting another time series' data (in our case joint moments) (Hua et al., 2020; Malek et al., 2018). The size of the 1D-CNN model was manually defined by increasing its complexity until no significant decrease in loss was observed on a separate development set consisting of 42 patients with CP. The designated 1D-CNN model consisted of five convolutional layers with the following number of filters [128, 128, 512, 1024, 2048] having the following 1D sizes of kernels respectively [30, 15, 10, 5, 3]. The ascending number of filters over layers with decreasing filter sizes works for extracting features increasingly for decreasing time intervals. After flattening the output of the convolutional layers, ten densely connected layers were used with the following number of neurons: [10000, 8000, 6000, 4000, 3000, 2000, 1000, 500, 250, 100]. As a general approach for forward propagation in densely connected layers, the descending number of neurons works for transforming the information to the desired output size throughout the learning process. Based on the results of the preliminary trials on the development set, the number of filters selected for the convolutional layers was sufficient for feature extraction during the learning process, and the number of neurons selected for the densely connected layers was adequate for learning using the extracted features. All of the layers used a rectified linear unit activation function and had a dropout layer attached to their outputs with a 1% dropout fraction. Finally, a densely connected output layer with a linear activation function was used, which has a neuron size of 60 equaling the number of time points in the stance phase of the interpolated joint moment time series.

An optimization algorithm for the learning process, namely the stochastic gradient descent (SGD) algorithm, was used with a learning rate of 0.01. The loss criterion was set based on root mean squared error (RMSE) and Pearson correlation coefficient (PCC) between the experimental and predicted time series. The explained 1D-CNN algorithm was implemented with Keras on Tensorflow (Chollet et al., 2015). The data processing and ML pipelines are given in Fig. 3.

As a common approach for testing, a 10-fold cross-validation algorithm was used (Refaeilzadeh et al., 2009). The dataset was divided into ten equal parts, with nine parts used for training and one part for testing. Range normalization was applied separately to both the training and testing sets to prevent information leakage between them (Fig. 3). Each subject was only included in one of these 10 subsets to avoid over-fitting the model to the subject-specific walking patterns. Learning curves were plotted during the training process to check whether the models were overfitting on the training data. The decrease of loss on the training set was compared to that of an isolated test set. The learning curves did show concurrent behaviour on the training and isolated test sets. The training was limited to 500 epochs for each split with batches of size 32.

2.4. Evaluation metrics

The predicted joint moment time series were evaluated regarding the normalized root mean square error (nRMSE (%)) and Pearson correlation coefficient (PCC) calculated between the experimental and predicted joint moments. These metrics are well-accepted measures for the evaluation of joint moment predictions using ML algorithms (Ardestani et al., 2014; Mundt et al., 2020a; Mundt et al., 2020b; Ripic et al., 2022). The nRMSE metric gives the normalized and time point-wise magnitude difference of the predicted joint moment time series from the experimental joint moment time series. nRMSE was calculated by dividing the RMSE value by the mean range of the experimental joint moment (μRoM) across all subjects of the same group as stated in Eq. (1). In the equation, JM_P and JM_E denote predicted and experimental joint moments, respectively. Subindices P and E denote predicted and experimental quantity, respectively.

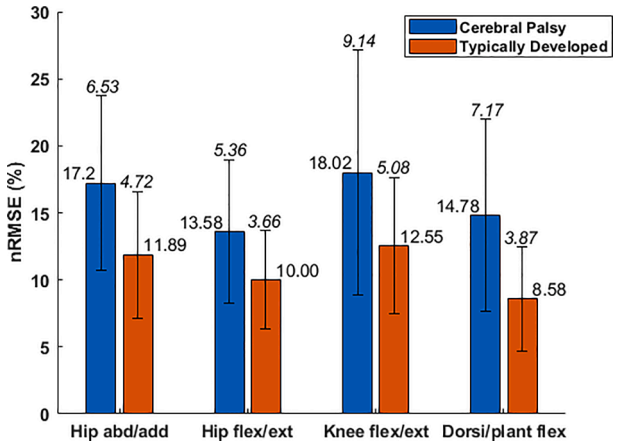

Fig. 4. Normalized root mean square error (nRMSE) scores for joint moment predictions of TD subjects (red) and patients with CP (blue). Hip abd/add: hip adduction-abduction, Hip flex/ext: hip flexion–extension, Knee flex/ext: knee flexion–extension, Dorsi/plant flex: dorsi-plantar flexion.

$$nRMSE = \sqrt{\frac{\Sigma_n (JM_P - JM_E)2}{n}} / \mu RoM \tag{1}$$

The PCC metric calculates the pattern similarity between the experimental and predicted joint moments (Savelberg and Herzog, 1997), in which cross-covariance ($cov(E, P)$) of them and variance of each of them ($\boldsymbol{\sigma}_E, \boldsymbol{\sigma}_P$) respectively were used (Eq. (2)).

$$PCC = \frac{cov(E, P)}{\boldsymbol{\sigma}_E \boldsymbol{\sigma}_P} \tag{2}$$

In order to avoid strong skewness in the distribution of the PCC values, we applied Fisher's Z transformation to the PCC values. The mean was computed using the resulting Z values and we then transformed the mean back to the PCC scale, following the approach described in Silver and Dunlap (1987).

The statistical analysis was conducted using SPSS software (Version 21.0; SPSS; Chicago, IL, USA). Both inter-group (between patients with CP and TD subjects) and intra-group (within each group of subjects) comparisons were performed. For the inter-group comparison, we hypothesized that the prediction results for the joint moments would differ significantly between the two groups. In the intra-group comparison, we hypothesized that the prediction results for the joint moments would differ significantly within each group of subjects. The level of significance was set at 0.05. The Kolmogorov-Smirnov test was used to test the normality of the data, which was found not to be normally distributed. The predicted joint moments of the TD subjects and patients with CP were statistically analyzed using the Mann–Whitney U test. For the intra-comparisons, Friedman's ANOVA test was used. A Bonferroni correction was applied to adjust the p-value for multiple comparisons ($p < 0.016$). Please refer to Appendix A for details on the intra- and inter-comparison groups and the types of analysis.

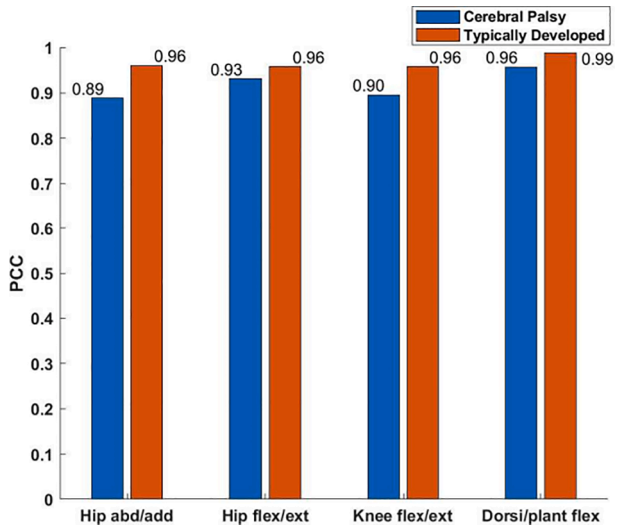

Fig. 5. Pearson correlation coefficient (PCC) scores for joint moment predictions of TD subjects (red) and patients with CP (blue). Hip abd/add: hip adduction-abduction, Hip flex/ext: hip flexion–extension, Knee flex/ext: knee flexion–extension, Dorsi/plant flex: dorsi plantar-flexion.

Table 1
P-values obtained for the nRMSE and PCC values of joint moment predictions for healthy subjects. Significant differences were marked bold. Hip add/abd: hip adduction-abduction, hip flex/ext: hip flexion–extension, knee flex/ext: knee flexion–extension, dorsi/plan flex: dorsi-plantar flexion.

		nRMSE	PCC
Hip add/abd vs.	Hip flex/ext	**0.014**	0.021
	Knee flex/ext	0.019	0.020
	Dorsi/planflex	**0.015**	0.018
Hip flex/ext vs.	Hip add/abd	**0.014**	0.021
	Knee flex/ext	**0.012**	0.019
	Dorsi/planflex	0.020	0.016
Knee flex/ext vs.	Hip add/abd	0.019	0.020
	Hip flex/ext	**0.012**	0.019
	Dorsi/planflex	**0.014**	**0.014**
Dorsi/plan flex vs.	Hip add/abd	**0.015**	0.018
	Hip flex/ext	0.020	0.016
	Knee flex/ext	**0.014**	**0.014**

3. Results

For TD subjects, all joint moments were predicted with mean nRMSE values less than 12.55%±5.08 (Fig. 4). The knee flexion–extension moment is the least successfully predicted joint moment in terms of nRMSE score (12.55%±5.08). The dorsi-plantar flexion is the most successfully predicted joint moment (8.58%±3.87). The hip adduction-abduction and hip flexion–extension moments were predicted with an nRMSE value of 11.89%±4.72 and 10%±3.66 for TD subjects, respectively.

For patients with CP, all joint moments were predicted with mean nRMSE values less than 18.02%±9.14 (Fig. 4). The knee flexion–extension moment is the least successfully predicted joint moment in terms of nRMSE (18.02%±9.14), while the hip flexion–extension is the most successfully predicted joint moment (13.58%±5.36). The hip adduction-abduction and dorsi-plantar flexion moments were predicted with an nRMSE value of 17.2%±6.53 and 14.78%±7.17 for the CP group, respectively.

For TD subjects, all joint moments were predicted with mean PCC scores higher than 0.96 (Fig. 5). The dorsi-plantar flexion is the most successfully predicted joint moment in terms of PCC score (0.99), while the others have the same PCC (0.96).

For the patient group, all joint moments were predicted with mean PCC scores higher than 0.89 (Fig. 5). The hip adduction-abduction moment is the least successfully predicted moment in terms of PCC score (0.89), while the dorsi-plantar flexion is the most successfully predicted one (0.96).

Table 2
P-values obtained for the nRMSE and PCC values of joint moment predictions for the patients with CP. Significant differences were marked bold. Hip add/abd: hip adduction-abduction, hip flex/ext: hip flexion–extension, knee flex/ext: knee flexion–extension, dorsi/plan flex: dorsi-plantar flexion.

		nRMSE	PCC
Hip add/abd vs.	Hip flex/ext	0.016	**0.015**
	Knee flex/ext	0.018	0.032
	Dorsi/planflex	**0.014**	**0.011**
Hip flex/ext vs.	Hip add/abd	0.016	**0.015**
	Knee flex/ext	**0.014**	**0.013**
	Dorsi/planflex	0.018	0.028
Knee flex/ext vs.	Hip add/abd	0.018	0.032
	Hip flex/ext	**0.014**	**0.013**
	Dorsi/planflex	**0.012**	**0.011**
Dorsi/plan flex vs.	Hip add/abd	**0.014**	**0.011**
	Hip flex/ext	0.018	0.028
	Knee flex/ext	**0.012**	**0.011**

Table 3
P-values obtained for the comparison of the nRMSE and PCC values of joint moment predictions for the healthy subjects and patients with CP. Significant differences were marked in bold. Hip add/abd: hip adduction-abduction, hip flex/ext: hip flexion–extension, knee flex/ext: knee flexion–extension, dorsi/plan flex: dorsi-plantar flexion.

		nRMSE	PCC
Healthy vs. Patients with CP	Hip add/abd	**0.041**	**0.033**
	Hip flex/ext	**0.047**	0.051
	Knee flex/ext	**0.038**	**0.037**
	Dorsi/plan flex	**0.034**	0.055

Tables 1 and 2 present the statistical significance of the nRMSE and PCC scores obtained for the joints of TD and patient groups, respectively. Within the TD group, the dorsi-plantar flexion and hip flexion–extension moments exhibited significantly better predictions than the hip abduction–adduction and knee flexion–extension moments in terms of nRMSE (Table 1). When considering the PCC scores, the prediction results for the dorsi-plantar flexion moment was significantly higher than that for the knee flexion–extension joint moment (Table 1).

Within the patient group, the dorsi-plantar flexion moment was significantly better predicted than the hip abduction–adduction and knee flexion–extension moments in terms of nRMSE (Table 2). Furthermore, when taking the PCC scores into account, the prediction results for the dorsi-plantar flexion and hip flexion–extension moments were significantly higher than those for the knee flexion–extension and hip abduction–adduction moments (Table 2).

Table 3 demonstrates the statistical significance of the scores between the TD individuals and patient groups. In terms of nRMSE, all four joint moments were predicted significantly higher in the TD group than in the CP group.

Figs. 6 and 7 show some representative predicted and experimental joint moments of TD subjects and patients with CP, respectively. These figures are provided for a better understanding of the trained models' capability of predicting joint moments with varying patterns. To ensure the representativeness of the models' capability in predicting joint moments, the figures in the left column show relatively successful predictions (with lower nRMSE and higher PCC values than the average), while the figures in the right column show relatively less successful predictions (with higher nRMSE and lower PCC values than the average).

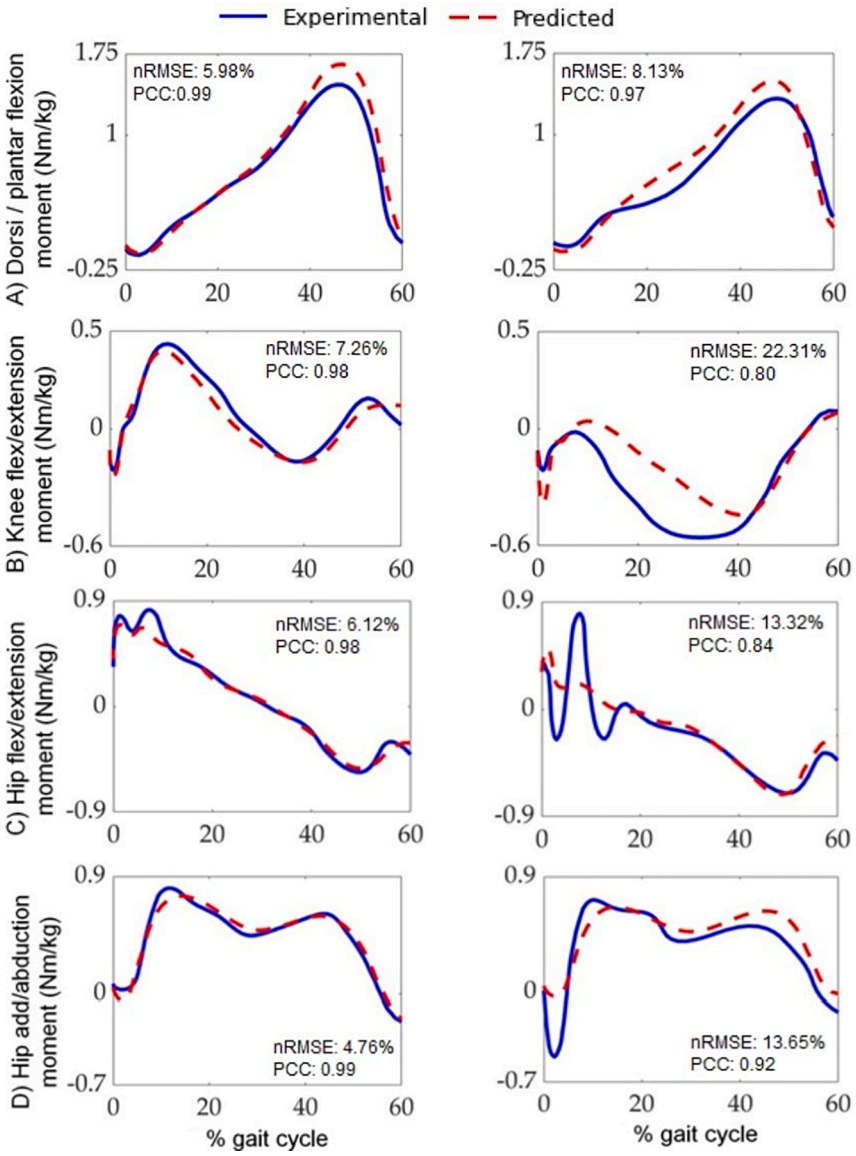

Fig. 6. Secondary, representative results for aiding interpretations of Fig. 4 and Fig. 5. Joint moments of a) dorsi-plantar flexion, b) knee flexion–extension, c) hip flexion–extension, d) hip adduction-abduction for representative typically developed subjects. The predictions in the left column correspond to results indicating above-average success, while those in the right column correspond to below-average success. The blue line represents the experimental joint moment, while the red dashed line represents the predicted joint moment.

4. Discussion

Since joint moments are valuable assessment parameters in the management of CP (Lai et al., 1988) and hard to capture experimentally, we predicted the dorsi-plantar flexion, knee flexion–extension, hip flexion–extension, and the hip adduction-abduction moments of patients with CP during gait from joint angles using 1D CNN in our study. We found that the joint moments of patients could be predicted with nRMSE values less than 18.02% and PCC scores higher than 0.85. In the TD group, all joint moments were predicted with nRMSE values less than 12.55% and PCC scores higher than 0.94. The predictions mostly captured the patterns and magnitudes of the experimentally obtained joint moments.

Mundt et al. predicted joint moments from joint angles of TD subjects using a densely connected feed-forward and an LSTM neural network achieved nRMSE scores between 12.14% and 15.00% and PCC scores between 0.92 and 0.97 on cross validation splits (Mundt et al., 2020a), whereas in our study the CNN model achieved nRMSE scores between 8.58% and 12.55% and PCC scores between 0.94 and 0.98 for TD subjects (Figs. 4 and 5). There is another study predicting joint moments of TD subjects based on EMG and GRF components using wavelet neural networks, which achieved higher success in terms of nRMSE (lower than 5.69%) and PCC (above 0.99) (Ardestani et al., 2014). Using GRF as input information would increase the prediction success since GRF and joint moments are biomechanically coupled. Thus, GRF that was leaked into the calculated joint moments was considered the golden standard in this study. However, in our research, only the joint angles were used as input which were separately measured and easily accessible information in routine gait analysis, hence there is no further need for costly equipment like force plates.

The prediction of joint moments for TD subjects was achieved with a significantly higher success regarding nRMSE within all considered joint moments and with a significantly higher PCC within hip adduction-abduction and knee flexion–extension moments (Table 3) when compared to those for subjects with CP. The varying deviation of gait in CP cases makes the learning process of the CNN models harder, which caused less moment prediction performance in the patient group compared to TD subjects. This was expected due to the coupled relation between joint angles and joint moments becoming more complex in patients with CP. The models for TD subjects have achieved higher prediction results despite having a relatively smaller number of subjects than the CP group, which is a commonly recognized disadvantage when training ML models. The sub-classification of CP groups based on altered gait patterns, such as crouch gait and tip-toe, and training separate ML

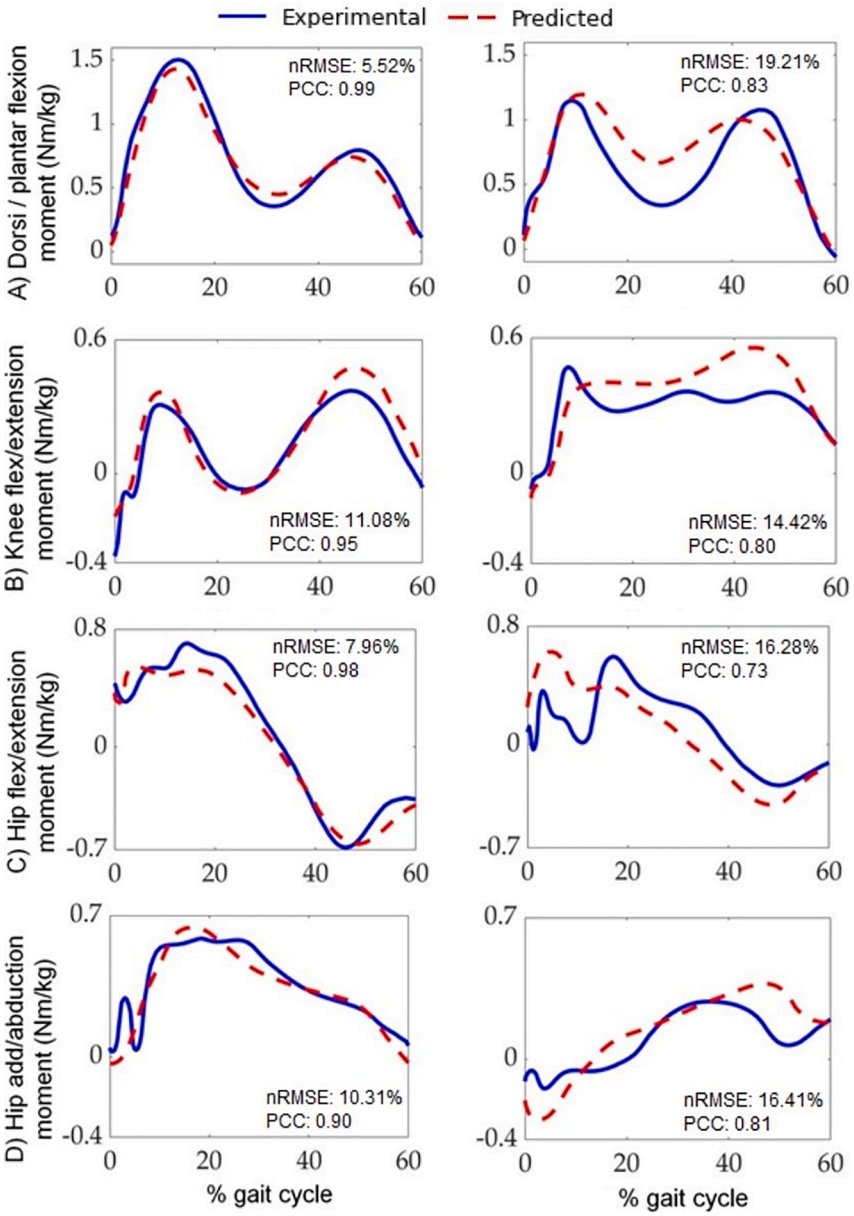

Fig. 7. Secondary, representative results for aiding interpretations of Fig. 4 and Fig. 5. Joint moment predictions of a) dorsi-plantar flexion, b) knee flexion–extension, c) hip flexion–extension, d) hip adduction-abduction for representative patients with cerebral palsy. The predictions in the left column correspond to results indicating above-average success, while those in the right column correspond to below-average success. The blue line represents the experimental joint moment, while the red dashed line represents the predicted joint moment.

models for each subgroup could improve the prediction accuracy. We consider this attempt as the next step in joint moment prediction studies for CP patients.

One could argue that the prediction of moments in the joints that are closer to the ground (distal joints) would be more successful than those that are further from the ground (proximal joints) because the calculation of joint moments that is based on inverse dynamics is performed in a stepwise fashion from bottom to the top resulting in accumulating errors in calculations (Whittle, 2014). This was not totally observable in our results however, the joint moment with the highest prediction success for TD subjects was the ankle dorsi-plantar flexion, which fits that expectation.

The representative joint moments presented in Fig. 6 and Fig. 7 indicate that the models were able to predict joint moments with different profiles. The models were blindly tested with randomly selected test splits across all included subjects, hence the performance of the models is promising for predicting joint moments of varying gaits. Although the results are promising, the fact that the gait analysis is used for surgical decision-making in CP makes the use of ML-based joint moment predictions limited, since the obtained error values might still be critical for surgical decision-making. The accuracy of obtaining kinematics data from markers directly affects the correctness of joint moment prediction. Moreover, inaccurate recording of kinematics data, caused by marker misplacement or soft tissue artifacts, can result in biomechanically inaccurate joint moments (Fonseca et al, 2020). However, the successful application of this workflow would facilitate the gait analysis of patients with CP by reducing laboratory effort and eliminating the need for complex musculoskeletal models for calculating joint moments. Furthermore, this workflow can help clinicians with the treatment protocol by providing joint moments of the patients with CP, whose GRFs could not be correctly measured at all due to using assistive devices or very short stride length.

Limitations of this study should be considered. Firstly, the models were limited to the aforementioned four joint moments, which are major kinetic parameters for the management of CP, however additional joint moments like hip internal/external rotation and ankle inversion/eversion may also be taken into account in monitoring CP. Secondly, the kinematics data included only the trunk from the upper body, however further kinematics data from upper extremities like arms may provide valuable information, thereby improving the ML models' prediction

Table A1
Details of the statistical analysis.

Type of analysis	• Friedman's ANOVA (for the intra-comparison) • Mann-Whitney *U* test (for the inter-comparison)
Intra-comparison	• nRMSE values calculated between the predicted and experimental hip abduction–adduction moment vs. nRMSE values calculated between the predicted and experimental hip flexion–extension moment, • nRMSE values calculated between the predicted and experimental hip abduction–adduction moment vs. nRMSE values calculated between the predicted and experimental knee flexion–extension moment, • nRMSE values calculated between the predicted and experimental hip abduction–adduction moment vs. nRMSE values calculated between the predicted and experimental dorsi-plantar flexion moment, • nRMSE values calculated between the predicted and experimental hip flexion–extension moment vs. nRMSE values calculated between the predicted and experimental knee flexion–extension moment, • nRMSE values calculated between the predicted and experimental hip flexion–extension moment vs. nRMSE values calculated between the predicted and experimental dorsi-plantar flexion moment, • nRMSE values calculated between the predicted and experimental knee flexion–extension moment vs. nRMSE values calculated between the predicted and experimental dorsi-plantar flexion moment, • PCC values calculated between the predicted and experimental hip abduction–adduction moment vs. PCC values calculated between the predicted and experimental hip flexion–extension moment, • PCC values calculated between the predicted and experimental hip abduction–adduction moment vs. PCC values calculated between the predicted and experimental knee flexion–extension moment, • PCC values calculated between the predicted and experimental hip abduction–adduction moment vs. PCC values calculated between the predicted and experimental dorsi-plantar flexion moment, • PCC values calculated between the predicted and experimental hip flexion–extension moment vs. PCC values calculated between the predicted and experimental knee flexion–extension moment, • PCC values calculated between the predicted and experimental hip flexion–extension moment vs. PCC values calculated between the predicted and experimental dorsi plantar flexion moment, • PCC values calculated between the predicted and experimental knee flexion–extension moment vs. PCC values calculated between the predicted and experimental dorsi-plantar flexion moment,
Inter-comparison	• nRMSE values calculated between the predicted and experimental hip abduction–adduction moment of the CP patients vs. those calculated healthy subjects, • nRMSE values calculated between the predicted and experimental hip flexion–extension moment of the CP patients vs. those calculated healthy subjects, • nRMSE values calculated between the predicted and experimental knee flexion–extension moment of the CP patients vs. those calculated healthy subjects, • nRMSE values calculated between the predicted and experimental dorsi-plantar flexion moment of the CP patients vs. those calculated healthy subjects, • PCC values calculated between the predicted and experimental hip abduction–adduction moment of the CP patients vs. those calculated healthy subjects, • PCC values calculated between the predicted and experimental hip flexion–extension moment of the CP patients vs. those calculated healthy subjects, • PCC values calculated between the predicted and experimental knee flexion–extension moment of the CP patients vs. those calculated healthy subjects, • PCC values calculated between the predicted and experimental dorsi -plantar flexion moment of the CP patients vs. those calculated healthy subjects.

CP: Cerebral palsy, nRMSE: normalized root-mean-square error, PCC: Pearson cross-correlation coefficient.

performance. Thirdly, it is ambiguous if the model would be able to predict a marginal joint moment from a CP patient with a novel form of deviation, which did not show up in our subject dataset. Although the used dataset is large and has been collected over two decades, the ML algorithm should always be further developed with potential new cases' data. For example, we did not include hemiplegic and tetraplegic subjects in the study. The implementation of ML algorithms on such patients would improve the applicability of the proposed joint kinetics prediction procedure. Lastly, the developed models are based on the CNN algorithms, however other ML algorithms like long-short term memory neural networks can also be used for the same task and compared in terms of their prediction performance.

In conclusion, the results of the study showed that machine learning-based prediction of joint moments based on kinematics could be an alternative technique to conventional joint moment calculation in gait analysis of patients with CP in the near future, however, the level of prediction errors limits the use of machine learning-based technique for clinical decision making today.

CRediT authorship contribution statement

Mustafa Erkam Ozates: Writing – original draft, Writing – review & editing, Investigation, Conceptualization, Methodology, Software, Validation, Visualization. **Derya Karabulut:** Validation, Methodology, Writing - original draft, Writing – review & editing. **Firooz Salami:** Data curation, Methodology. **Sebastian Immanuel Wolf:** Writing – review & editing, Supervision, Conceptualization. **Yunus Ziya Arslan:** Writing – review & editing, Supervision, Conceptualization, Methodology, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Mustafa Erkam Özates and Yunus Ziya Arslan visited Heidelberg University Hospital, Clinic for Orthopaedics and Trauma Surgery as part of the DAAD-Forschungsaufenthalt Program.

Appendix A. Details of the statistical analysis

See Table A1

References

Ardestani, M.M., Zhang, X., Wang, L., Lian, Q., Liu, Y., He, J., Li, D., Jin, Z., 2014. Human lower extremity joint moment prediction: A wavelet neural network approach. Expert Syst. Appl. 41 (9), 4422–4433.
Arslan, Y.Z., Karabulut, D., 2021. Sensitivity of model-predicted muscle forces of patients with cerebral palsy to variations in muscle-tendon parameters. J. Mech. Med. Biol. 21, 2150008.
Caldas, R., Fadel, T., Buarque, F., Markert, B., 2020. Adaptive predictive systems applied to gait analysis: A systematic review. Gait Posture 77, 75–82.
De Brabandere, A., Emmerzaal, J., Timmermans, A., Jonkers, I., Vanwanseele, B., Davis, J., 2020. A machine learning approach to estimate hip and knee joint loading using a mobile phone-embedded IMU. Front. Bioeng. Biotechnol. 8, 320.
DeLuca, P., Davis, R., Õunpuu, S., Rose, S., Sirkin, R., 1997. Alterations in surgical decision making in patients with cerebral palsy based on three-dimensional gait analysis. J. Pediatr. Orthop. 17 (5), 608–614.
Fonseca, M., Gasparutto, X., Leboeuf, F., Dumas, R., Armand, S., 2020. Impact of knee marker misplacement on gait kinematics of children with cerebral palsy using the Conventional Gait Model—A sensitivity study. PLoS One 15 (4), e0232064.
Gage, J.R., 1994. The clinical use of kinetics for evaluation of pathological gait in cerebral palsy. J. Bone Joint Surg. 76 (4), 622–631.
Giarmatzis, G., Zacharaki, E.I., Moustakas, K., 2020. Real-time prediction of joint forces by motion capture and machine learning. Sensors 20 (23), 6933.

Harrison, S.M., Whitton, R.C., King, M., Haussler, K.K., Kawcak, C.E., Stover, S.M., Pandy, M.G., 2012. Forelimb muscle activity during equine locomotion. J. Exper. Biol. 215, 2980–2991.
Hua, X., Han, J., Zhao, C., Tang, H., He, Z., Chen, Q., Tang, S., Tang, J., Zhou, W., 2020. A novel method for ECG signal classification via one-dimensional convolutional neural network. Multimedia Syst. 1–13.
Ihlen, E.A., Støen, R., de Boswell, L., Regnier, R.A., Fjørtoft, T., Gaebler-Spira, D., Labori, C., Loennecken, M.C., Msall, M.E., Möinichen, U.I., Peyton, C., Schreiber, M. D., Silberg, I.E., Songstad, N.T., Vågen, R.T., Øberg, G.K., Adde, L., 2019. Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: A multi-site cohort study. J. Clin. Med. 9 (1), 5.
Karabulut, D., Arslan, Y.Z., Götze, M., Wolf, S.I., 2021. The Impact of Patellar Tendon Advancement on Knee Joint Moment and Muscle Forces in Patients with Cerebral Palsy. Life 11 (9), 944.
Kay, R.M., Dennis, S., Rethlefsen, S., Reynolds, R.A., Skaggs, D.L., Tolo, V.T., 2000. The effect of preoperative gait analysis on orthopaedic decision making. Clin. Orthop. Relat. Res. 1976–2007 (372), 217–222.
Chollet, F., & Others, 2015. Keras. Retrieved from https://keras.io.
Kim, Y.K., Visscher, R.M., Viehweger, E., Singh, N.B., Taylor, W.R., Vogl, F., 2022. A deep-learning approach for automatically detecting gait-events based on foot-marker kinematics in children with cerebral palsy—Which markers work best for which gait patterns? PLoS One 17 (10), e0275878.
Kloeckner, J., Visscher, R.M., Taylor, W.R., Viehweger, E.D., Pieri, E., 2023. Prediction of ground reaction forces and moments during walking in children with cerebral palsy. Front. Hum. Neurosci.
Lai, K.A., Kuo, K.N., Andriacchi, T.P., 1988. Relationship between dynamic deformities and joint moments in children with cerebral palsy. J. Pediatr. Orthop. 8 (6), 690–695.
Lenhart, R.L., Brandon, S.C., Smith, C.R., Novacheck, T.F., Schwartz, M.H., Thelen, D.G., 2017. Influence of patellar position on the knee extensor mechanism in normal and crouched walking. J. Biomech. 51, 1–7.
Lin, C.J., Guo, L.Y.S., Chou, F.C., Cherng, Y.L., R. j., 2000. Common abnormal kinetic patterns of the knee in gait in spastic diplegia of cerebral palsy. Gait Posture 11 (3), 224–232.
Malek, S., Melgani, F., Bazi, Y., 2018. One-dimensional convolutional neural networks for spectroscopic signal regression. J. Chemom. 32 (5), e2977.
Morbidoni, C., Cucchiarelli, A., Agostini, V., Knaflitz, M., Fioretti, S.D., Nardo, F., 2021. Machine-learning-based prediction of gait events from EMG in cerebral palsy children. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 819–830.
Mundt, M., Koeppe, A., David, S., Bamer, F., Potthast, W., Markert, B., 2020a. Prediction of ground reaction force and joint moments based on optical motion capture data during gait. Med. Eng. Phys. 86, 29–34.
Mundt, M., Koeppe, A., David, S., Witter, T., Bamer, F., Potthast, W., Markert, B., 2020b. Estimation of gait mechanics based on simulated and measured IMU data using an artificial neural network. Front. Bioeng. Biotechnol. 8, 41.
Novacheck, T.F., Gage, J.R., 2007. Orthopedic management of spasticity in cerebral palsy. Childs Nerv. Syst. 23 (9), 1015–1031.
Oh, J., Eltoukhy, M., Kuenze, C., Andersen, M.S., Signorile, J.F., 2020. Comparison of predicted kinetic variables between Parkinson's disease patients and healthy age-matched control using a depth sensor-driven full-body musculoskeletal model. Gait Posture 76, 151–156.
Ounpuu, S., Davis, R.B., Deluca, P.A., 1996. Joint kinetics: methods, interpretation and treatment decision-making in children with cerebral palsy and myelomeningocele. Gait Posture 4 (1), 62–78.
Refaeilzadeh, P. Tang, L. Liu, H., 2009. Cross-validation. Encyclopedia of Database Systems 5, 532–538.
Rhodes, J.T.T., Moore, A., Holcomb, L., Carry, A., Skinner, P., Miller, A., De, S., Carollo, S., J.,, 2023. Rectus femoris transfers with and without a hamstring lengthening will not change hip kinematics in children with cerebral palsy. Gait Posture 99, 119–123.
Richards, R.E., Andersen, M.S., Harlaar, J.V., Den Noort, J.C., 2018. Relationship between knee joint contact forces and external knee joint moments in patients with medial knee osteoarthritis: effects of gait modifications. Osteoarthr. Cartil. 26 (9), 1203–1214.
Ripic, Z., Kuenze, C., Andersen, M.S., Theodorakos, I., Signorile, J., Eltoukhy, M., 2022. Ground reaction force and joint moment estimation during gait using an Azure Kinect-driven musculoskeletal modeling approach. Gait Posture 95, 49–55.
Savelberg, H.H., Herzog, W., 1997. Prediction of dynamic tendon forces from electromyographic signals: An artificial neural network approach. J. Neurosci. Methods 78 (1–2), 65–74.
Shao, Q., Bassett, D.N., Manal, K., Buchanan, T.S., 2009. An EMG-driven model to estimate muscle forces and joint moments in stroke patients. Comput. Biol. Med. 39 (12), 1083–1088.
Silver, N.C., Dunlap, W.P., 1987. Averaging correlation coefficients: should Fisher's z transformation be used? J. Appl. Psychol. 72 (1), 146.
White, R., Agouris, I., Selbie, R.D., Kirkpatrick, M., 1999. The variability of force platform data in normal and cerebral palsy gait. Clin. Biomech. 14 (3), 185–192.
Whittle, M.W., 2014. Gait analysis: an introduction. Butterworth-Heinemann.
Winter, D.A., 2009. Biomechanics and motor control of human movement. John Wiley & Sons.
Zhang, Y., Ye, M., 2019. Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. Comput. Biol. Med. 106, 33–39.