Annals of Biomedical Engineering
https://doi.org/10.1007/s10439-024-03658-y

BMES BIOMEDICAL ENGINEERING SOCIETY

ORIGINAL ARTICLE

Estimating Ground Reaction Forces from Gait Kinematics in Cerebral Palsy: A Convolutional Neural Network Approach

Mustafa Erkam Ozates[1] · Firooz Salami[2] · Sebastian Immanuel Wolf[2] · Yunus Ziya Arslan[3]

Received: 2 September 2024 / Accepted: 21 November 2024

Abstract

Purpose While gait analysis is essential for assessing neuromotor disorders like cerebral palsy (CP), capturing accurate ground reaction force (GRF) measurements during natural walking presents challenges, particularly due to variations in gait patterns. Previous studies have explored GRF prediction using machine learning, but specific focus on patients with CP is lacking. This research aims to address this gap by predicting GRF using joint angles derived from marker data during gait in patients with CP, thereby suggesting a protocol for gait analysis without the need for force plates.

Methods The study employed an extensive dataset comprising both typically developed (TD) subjects ($n=132$) and patients with CP ($n=622$), captured using motion capture systems and force plates. Kinematic data included lower limb angles in three planes of motion, while GRF data encompassed three axes. A one-dimensional convolutional neural network model was designed to extract features from kinematic time series, followed by densely connected layers for GRF prediction. Evaluation metrics included normalized root mean squared error (nRMSE) and Pearson correlation coefficient (PCC).

Results GRFs of patients with CP were predicted with nRMSE values consistently below 20.13% and PCC scores surpassing 0.84. In the TD group, all GRFs were predicted with higher accuracy, showing nRMSE values lower than 12.65% and PCC scores exceeding 0.94.

Conclusion The predictions considerably captured the patterns observed in the experimentally obtained GRFs. Despite limitations, including the absence of upper extremity kinematics data and the need for continuous model evolution, the study demonstrates the potential of machine learning in predicting GRFs in patients with CP, albeit with current prediction errors constraining immediate clinical applicability.

Keywords Gait analysis · Machine learning · Cerebral palsy · Ground reaction force

Introduction

Cerebral palsy (CP) is characterized by a range of neurological and motor impairments, encompassing a diverse array of disorders that profoundly affect an individual's neuromotor functions. Among patients with CP, a notable impairment is evident in their natural walking pattern [1, 2].

Gait analysis, a recognized and valuable tool in clinical decision-making, plays a vital role in evaluating and managing various neurological and musculoskeletal conditions, especially for patients with CP [3, 4]. In the clinical settings, the assessment of ground reaction force (GRF), which is a direct kinetic parameter measured using force plates that subjects step on during their natural walking, stands as a cornerstone of gait analysis.

The measurement of GRF encounters inherent challenges when capturing it during natural walking [5]. These

Associate Editor Elisabetta Zanetti oversaw the review of this article.

✉ Yunus Ziya Arslan
yunus.arslan@tau.edu.tr

[1] Department of Electrical Electronics Engineering, Faculty of Engineering, Turkish-German University, Istanbul, Turkey

[2] Clinic for Orthopaedics and Trauma Surgery, Heidelberg University Hospital, Heidelberg, Germany

[3] Department of Robotics and Intelligent Systems, Institute of Graduate Studies in Science and Engineering, Turkish-German University, Istanbul, Turkey

Published online: 30 November 2024

challenges are further compounded when dealing with abnormal gaits, especially within the context of CP [6]. The complexity lies in accurately quantifying GRF in real-time while individuals exhibit variations in their gait patterns, such as uneven foot placements and asymmetrical weight distribution. These deviations present significant hurdles to achieving precise and dependable GRF measurements. Providing GRF information that requires only kinematic data can aid in the clinical evaluation of patients with CP, particularly for those whose natural gait patterns make it difficult to capture GRF data with force plates or for whom no force plate-equipped laboratory is available.

Machine learning (ML) emerges as a potent tool for addressing tasks with incomplete measurements or the absence of physical models. Exploring the prediction of GRF during gait using ML techniques has been a vibrant area of research [7]. A notable inception in this domain dates back to 2013, where conventional ML algorithms were employed to predict GRF based on kinematic data [8]. Subsequent studies have extended this work, delving into statistical methodologies [9], as well as deep learning algorithms utilizing motion capture data [10] and incorporating spatio-temporal information [11].

ML algorithms have also found success in the realm of patients with CP [12], who exhibit non-uniform gait characteristics. Researchers have harnessed ML techniques for various purposes, including CP disease detection using video recordings [13] or gait kinematics [14]. Additionally, ML has been employed for classifying gait phases in patients with CP using electromyography [15] or marker data [16]. These applications underscore the potential of ML in aiding the diagnosis, evaluation, and categorization of gait anomalies in patients with CP, furnishing valuable insights for clinical decision-making and personalized treatment strategies.

Predicting GRF would simplify the application of gait analysis, eliminating the need for force plates and standard stepping protocols for subjects. However, to the best of our knowledge, no study has specifically focused on predicting GRF in patients with CP using ML algorithms. Accordingly, we aimed to predict GRF using joint angles derived from marker data during gait in patients with CP.

Materials and Methods

Subjects

The studies conducted within this research were granted ethical approval by the Local Ethical Committee of the University Hospital of Heidelberg (S-227/2021), ensuring strict adherence to essential ethical guidelines and considerations. For the research, an extensive dataset of gait information was employed, comprising anonymized retrospective data (Fig. 1). This dataset encompassed 329 typically developed (TD) subjects, exhibiting characteristic gait patterns, as well as 917 patients diagnosed with CP. The TD group had an average age of 26 years (with a standard deviation of ± 14), a mass of 70 kg (±15), and a height of 162 cm (±21). The patients with CP had an average age of 17 years (±9), a mass of 47 kg (±19), and a height of 152 cm (±20). This dataset was sourced from routine patient care, ensuring a pragmatic and clinically relevant context.

To capture kinematic data, the Plug-in Gait Model (Oxford Metrics, Oxford, UK) was employed, involving the placement of 19 markers on the subjects. Data capture took place using a 12-camera motion capture system (Vicon Motion Systems Ltd., Oxfordshire, UK) while subjects walked at their self-selected speed. Simultaneously, GRF data was gathered using two force plates (Kistler Instruments, Winterthur, Switzerland) and normalized by body mass.

The inclusion criteria for both TD and CP subjects did not impose specific age or gender constraints. In the case of TD subjects, those who walked barefoot and possessed complete measurements were incorporated into the study (Fig. 2).

Regarding the CP participants, their initial appointments at the clinic were taken into account, and only individuals walking barefoot, capable of walking unaided and possessing complete measurements were encompassed. This selection process is exemplified in Fig. 3 of the patient flowchart. The included patients with CP were confined to levels I and II of the Gross Motor Function Classification System (GMFCS).

Following the application of the inclusion and exclusion criteria, a total of 132 typically developing individuals and 622 patients with CP with spastic diplegia were chosen for further analysis.

Dataset

The kinematic gait data comprised measurements taken from various body segments, encompassing the trunk, pelvis, hip, knee, and ankle, in three planes of motion—namely, sagittal, coronal, and transverse. This compilation resulted in a total of 15 angles under consideration. The kinetic data included GRF along three axes: vertical, mediolateral, and anteroposterior (gait direction) axes.

Temporal foot-off values guided the segmentation of the time series into stance and swing phases. As GRF data were only accessible during the stance phase, this phase was only utilized in the study. Given that the duration of the stance time varies across subjects, achieving a consistent dataset size for ML algorithms training necessitated interpolation of the stance segments of time series to a standardized length of 60 data points. All values within the time series were normalized to a range of 0 to 1, regardless of their original units. This normalization ensured that time series with high or low

Fig. 1 A subject performing the walking test in the Heidelberg University Hospital, Gait Analysis Laboratory

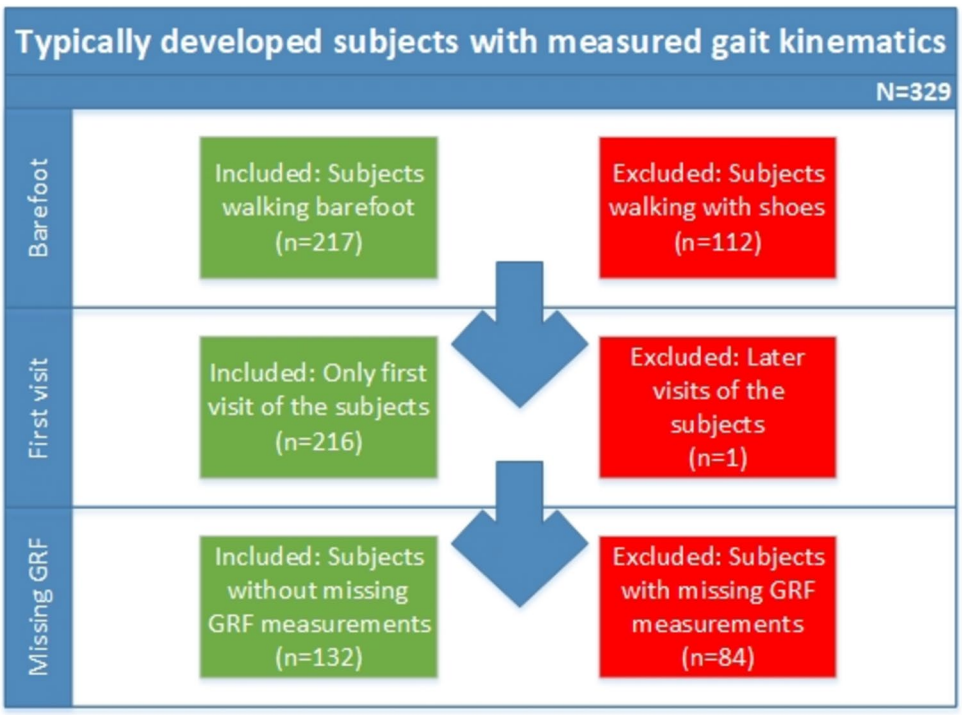

Fig. 2 Inclusion/exclusion flow of the typically developed subjects. *GRF* ground reaction force

magnitudes were represented on a common scale, balancing their contribution to the model. Since input features with different scales can hinder model training by causing slower convergence or instability, rescaling the time series to the same range facilitated successful training without altering the underlying patterns in the data.

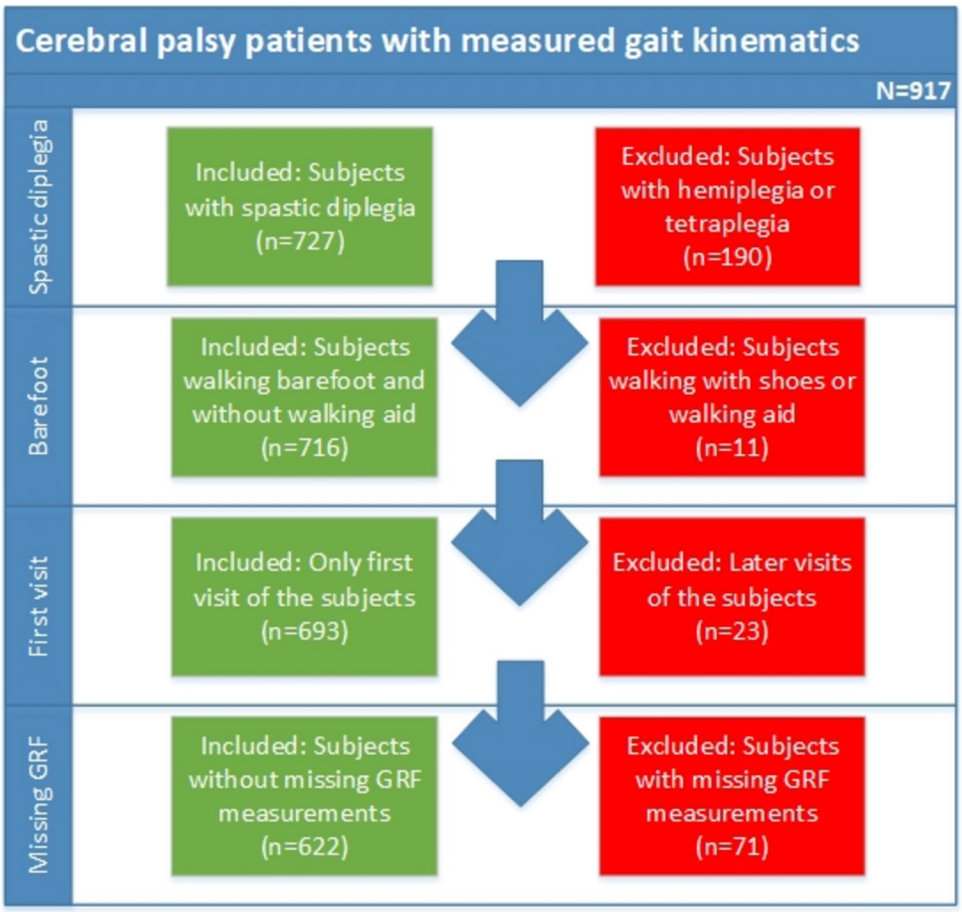

Fig. 3 Inclusion–exclusion flow of the subjects with cerebral palsy. *GRF* ground reaction force

Following normalization, the 15 kinematic time series, along with their standard deviations, were combined into a matrix measuring 30 rows and 60 columns. This matrix structure was repeated for each subject, leading to the creation of a total of 132 matrices for TD individuals and 622 matrices for patients with CP, which formed the basis for ML training and testing phases.

Machine Learning

To process the rows of the time series and extract distinct features from each, a one-dimensional (1D) convolution was implemented within the convolutional neural network (CNN). One-dimensional convolution layers extracted signal features from the time series data (joint angles) individually, encompassing diverse temporal ranges, thereby preserving essential information crucial for predicting another time series data (GRF).

The designated 1D-CNN model comprised five convolutional layers with filter quantities of [128, 128, 512, 1024, 2048] and corresponding 1D kernel sizes of [30, 15, 10, 5, 3]. Following the flattening of the convolutional layers' output, ten densely connected layers were employed, housing neuron quantities of [10000, 8000, 6000, 4000, 3000, 2000, 1000, 500, 250, 100]. All layers utilized the rectified linear unit (ReLU) activation function and featured a dropout layer with a dropout fraction of 1% attached to their outputs. Lastly, a densely connected output layer with a linear activation function was utilized, equipped with a neuron count of 60-matching the number of time points in the stance phase of the interpolated GRF time series.

The optimization algorithm employed for the learning process was the stochastic gradient descent algorithm, set with a learning rate of 0.01. The loss criterion was established based on the error metrics used in our study, namely normalized root mean squared error (nRMSE) and Pearson correlation coefficient (PCC). The described 1D-CNN algorithm was implemented using Keras on the TensorFlow platform [17].

In line with a common testing approach, a 10-fold cross-validation algorithm was adopted [18]. The dataset was divided into 10 equal segments, with 9 utilized for training and 1 for testing. Range normalization was independently applied to both training and testing sets to avert data leakage. Each subject was exclusively included in 1 of these 10 subsets to prevent the model from overfitting to specific walking patterns of subjects. Learning curves were plotted during training to monitor potential overfitting on the training data,

gauging the loss decrease on the training set in comparison to an isolated test set. These learning curves exhibited consistent behavior on both the training and isolated test sets. The training process lasted for 500 epochs for each split, with batch sizes set at 32.

Evaluation Metrics

The assessment of the predicted GRF time series involved the examination of normalized root mean square error [nRMSE (%)] and Pearson correlation coefficient (PCC) between the experimental and predicted time series. These metrics are widely recognized as reliable measures for evaluating GRF predictions utilizing ML algorithms [10, 19–21]. The nRMSE metric quantifies the normalized and point-to-point magnitude difference between the experimental and predicted GRF time series. Its computation entails dividing the RMSE value by the average peak to peak value of the experimental GRF (μPP) across all subjects within the same group, as described in Eq. (1). In the equation, GRF_P and GRF_E denote predicted and experimental GRFs, respectively. Sub-indices P and E denote predicted and experimental quantity, respectively.

$$\text{nRMSE} = \sqrt{\frac{\Sigma_n(\text{GRF}_\text{P} - \text{GRF}_\text{E})2}{n}}/\mu\text{PP}. \quad (1)$$

The PCC metric calculates the pattern similarity between the experimental and predicted GRFs [22], in which cross-covariance (cov(E, P)) and variance ($\boldsymbol{\sigma}_\text{E}, \boldsymbol{\sigma}_\text{P}$) of each of them were used [Eq. (2)].

$$\text{PCC} = \frac{\text{cov}(\text{E}, \text{P})}{\boldsymbol{\sigma}_\text{E}\boldsymbol{\sigma}_\text{P}}. \quad (2)$$

To mitigate the strong skewness in the PCC value distribution, we implemented Fisher's *Z* transformation on the PCC values. Subsequently, the mean was calculated using the transformed *Z* values, and we reverted this mean back to the original PCC scale. This procedure adhered to the methodology outlined in the reference of Silver and Dunlap [23].

Statistical analysis was carried out using SPSS software (Version 21.0; SPSS; Chicago, IL, USA). We conducted both inter-group comparisons (between patients with CP and TD subjects) and intra-group comparisons (within each subject group). In the inter-group comparisons, our hypothesis was defined regarding the distinctiveness of prediction success rates for GRF between the two groups—patients with CP and TD subjects. Within the intra-group comparisons, our hypothesis was defined regarding the differentiation of prediction success rates for GRF in all three dimensions within each subject group. A significance level of 0.05 was established.

To assess data normality, the Kolmogorov–Smirnov test was applied, indicating non-normal distribution. The predicted GRF values of TD subjects and patients with CP underwent statistical analysis using the Mann–Whitney *U*-test. For intra-group comparisons, Friedman's ANOVA test was employed. The Mann–Whitney *U*-test was utilized to identify noteworthy differences between the methods. To account for multiple comparisons, a Bonferroni correction was applied to adjust the *p*-value, setting the threshold at $p < 0.016$.

Results

For TD subjects, all GRFs were predicted with mean nRMSE values less than 12.65% ± 4.83 (Fig. 4). The mediolateral GRF is the least successfully predicted one in terms of nRMSE score (12.65% ± 4.83). The GRF along the anteroposterior axis is the most successfully predicted (5.70% ± 2.06). The vertical GRF is predicted with an nRMSE value of 7.47% ± 3.53 for TD subjects.

For patients with CP, all GRFs were predicted with mean nRMSE values less than 20.13% ± 9.63 (Fig. 4). The mediolateral GRF is the least successfully predicted one (20.13% ± 9.63). The GRF along the anteroposterior axis is the most successfully predicted (11.63% ± 5.14). The vertical GRF is predicted with an nRMSE value of 11.75% ± 6.88 for patients with CP.

For TD subjects, all GRFs were predicted with mean PCC scores higher than 0.94 (Fig. 5). The mediolateral, vertical, and anteroposterior GRFs have the mean PCC scores of 0.94, 0.98 and 0.99, respectively.

For the patient group, all GRFs were predicted with mean PCC scores higher than 0.84 (Fig. 5). The mediolateral, vertical, and anteroposterior GRFs have the mean PCC scores of 0.84, 0.94 and 0.96, respectively.

Tables 1 and 2 present the statistical significance of the nRMSE and PCC scores obtained for the GRFs of TD and CP groups, respectively. Within the TD group, the anteroposterior GRF exhibited significantly better predictions than the mediolateral and vertical GRFs in terms of nRMSE (Table 1). When examining the PCC scores, the observed significant differences in prediction rates are consistent with those observed for nRMSE, except for the relationship between predictions of anteroposterior and vertical GRFs.

Within the patient group, the significance of the performance differences between the GRFs in three dimensions were similar to those for the TD group, whereas for vertical vs. anteroposterior performance comparison, the significance did not exist not only regarding PCC but also regarding nRMSE (Table 2).

Table 3 demonstrates the statistical significance of the scores between the TD individuals and patient groups. In

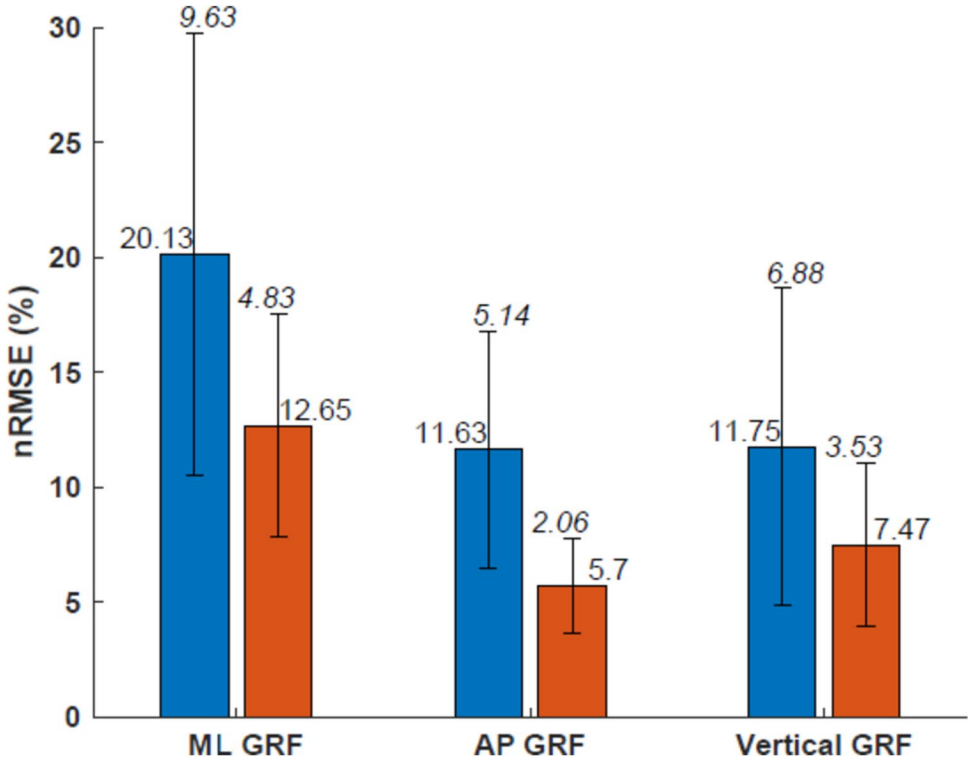

Fig. 4 Normalized root mean square error (nRMSE) scores for ground reaction force (GRF) predictions of typically developed subjects (red) and patients with CP (blue). *ML* mediolateral, *AP* anteroposterior

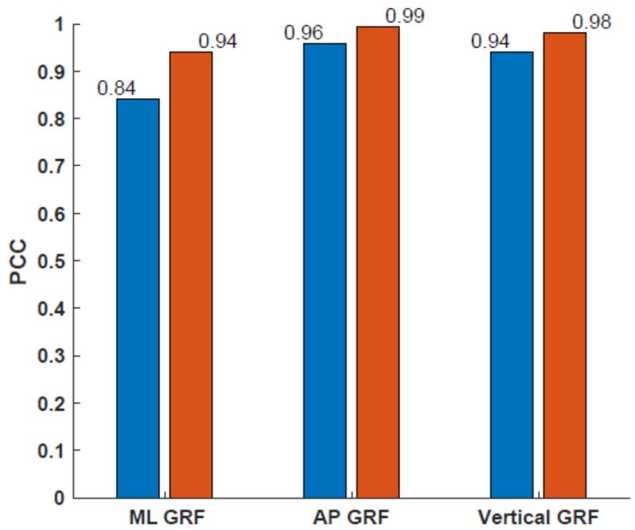

Fig. 5 Pearson correlation coefficient (PCC) scores for ground reaction force (GRF) predictions of typically developed subjects (red) and patients with CP (blue). *GRF* gait reaction force, *ML* mediolateral, *AP* anteroposterior

terms of both nRMSE and PCC, all GRFs were predicted significantly higher in the TD group than in the patient group.

Table 1 *p*-Values obtained for the nRMSE and PCC values of ground reaction force predictions for typically developed subjects

	nRMSE	PCC
Anteroposterior vs.		
Mediolateral	**0.004**	**0.012**
Vertical	**0.012**	0.029
Mediolateral vs.		
Anteroposterior	**0.004**	**0.012**
Vertical	**0.009**	**0.012**
Vertical vs.		
Anteroposterior	**0.012**	0.029
Mediolateral	**0.009**	**0.012**

Significant differences were marked bold

Figures 6 and 7 show some representative predicted and experimental GRFs of TD subjects and patients with CP, respectively. These figures are provided for a better understanding of the trained models' capability of predicting GRFs with varying patterns. To ensure the representativeness of the models' capability in predicting GRFs, 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).

Table 2 *p*-Values obtained for the nRMSE and PCC values of ground reaction force predictions for the patients with CP

	nRMSE	PCC
Anteroposterior vs.		
Mediolateral	**0.003**	**0.006**
Vertical	**0**.018	0.032
Mediolateral vs.		
Anteroposterior	**0.003**	**0.006**
Vertical	**0.007**	**0.004**
Vertical vs.		
Anteroposterior	0.018	0.032
Mediolateral	**0.007**	**0.004**

Significant differences were marked bold

Table 3 *p*-Values obtained for the comparison of the nRMSE and PCC values of ground reaction force predictions for the healthy subjects and patients with CP

	nRMSE	PCC
TD individuals vs. patients with CP		
Anteroposterior	**0.002**	**0.011**
Mediolateral	**0.011**	**0.013**
Vertical	**0.008**	**0.014**

Significant differences were marked bold

Discussion

Given the significant role of GRF as a pivotal assessment parameter in the management of CP, and considering that it represents a kinetic parameter that is both directly measurable however challenging to capture experimentally, our study focused on predicting the GRFs of patients with CP in vertical, mediolateral, and anteroposterior axes during gait. This prediction was achieved using a 1D CNN model based on joint angles. To evaluate the models' performance, blind testing was conducted using randomly selected test splits across all participants, demonstrating their potential for prediction of gait kinetics across a spectrum of gait patterns.

Our findings indicated that the GRFs of patients with CP were predicted with nRMSE values consistently below 20.13% and PCC scores surpassing 0.84. In the TD group, all GRFs were predicted with higher accuracy, showing nRMSE values lower than 12.65% and PCC scores exceeding 0.94. These predictions considerably captured the patterns observed in the experimentally obtained GRFs.

In a study predicting the GRFs for TD, Mundt et al. predicted GRFs from joint angles using a densely connected feed-forward and an LSTM neural network achieved nRMSE scores between 12.14 and 15.00% and PCC scores between 0.93 and 0.97 on cross-validation splits, whereas in our study the 1D-CNN model achieved nRMSE scores between 5.7 and 12.65% and PCC scores between 0.94 and 0.99 for TD subjects (Figs. 4, 5) [10].

When comparing the prediction of GRFs between TD and CP individuals, it became evident that TD subjects achieved significantly higher success in terms of both nRMSE and PCC (Table 3). The diverse gait abnormalities seen in CP cases present challenges for the learning process of CNN models, leading to inferior performance in predicting GRFs for patients with CP when compared to TD subjects. This outcome was to be expected due to the heightened complexity of the interconnected relationship between joint angles and GRFs in patients with CP. Despite the TD group having a relatively smaller subject count in contrast to the CP group, the models designed for TD subjects showcased superior success rates in predicting GRFs. This observation is of significance, particularly considering the recognized drawback of training machine learning models with a restricted sample size. It illustrates that the widely varying gait characteristics observed in patients with CP hinder ML models from benefiting from a larger sample size, unlike the success seen in TD group predictions. The patterns provide compelling evidence of the models' efficacy in predicting GRFs with diverse characteristics (Figs. 6, 7).

One limitation of this study is the use of the Plug-in Gait model, which, while widely utilized in gait laboratories, has certain limitations in marker placement and segment assumptions. This simplified model may restrict the detailed analysis of foot and ankle kinematics, particularly in populations where subtle joint mechanics are crucial for understanding movement, such as individuals with gait impairments. Utilizing a more detailed model could address these limitations by capturing additional joint segments and movements, allowing for a more comprehensive assessment of specific joint mechanics. Nonetheless, as the same gait model and associated limitations were consistently applied across all subjects, these model-specific limitations are unlikely to impact the accuracy of our machine learning model's predictions. Moreover, the kinematics data in our study was limited to the trunk within the upper body. Nevertheless, integrating additional kinematics data from upper extremities, such as the arms, could furnish valuable insights, potentially enhancing the predictive prowess of the ML models. Another aspect to consider is the model's capability to predict GRF variations for patients with CP who demonstrate deviations that have never been observed in our existing subject dataset. While our dataset is substantial

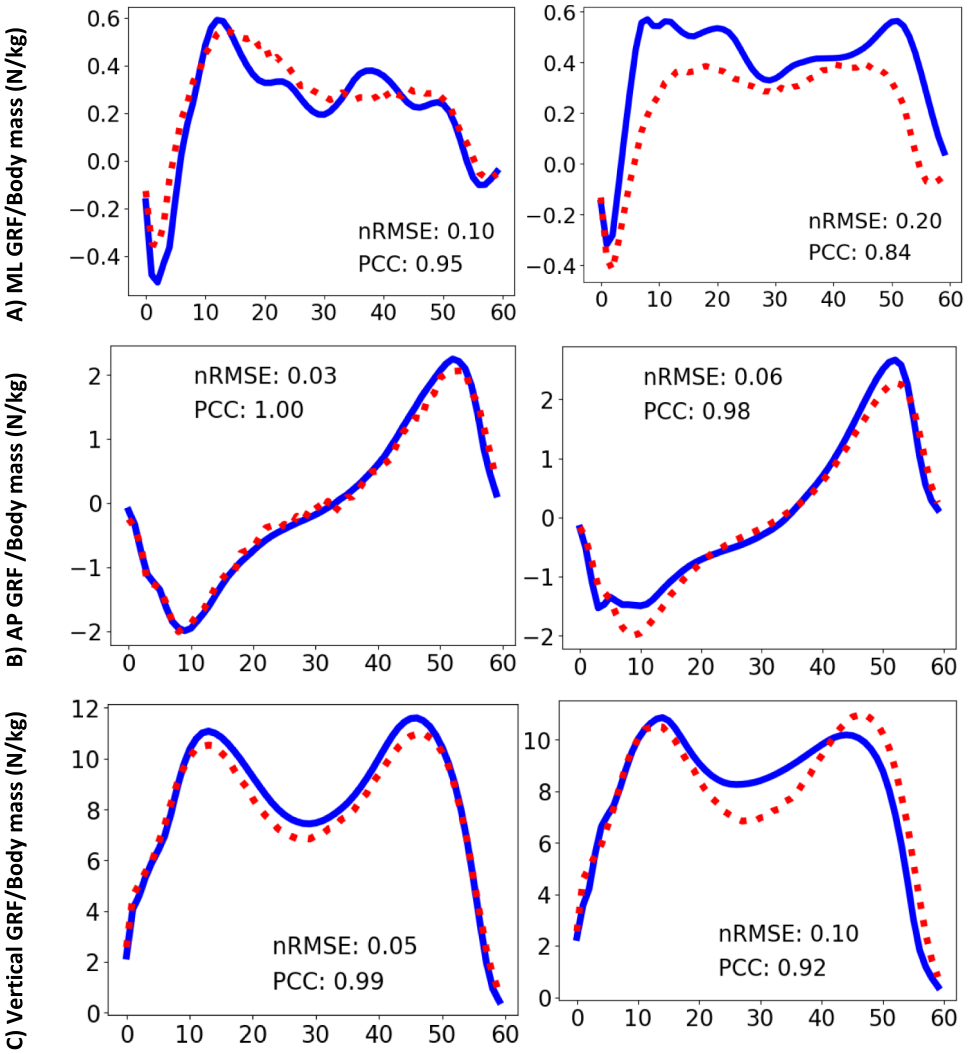

Fig. 6 Representative GRF prediction results of typically developed subjects. GRFs along **A** mediolateral (ML), **B** anteroposterior (AP), and **C** vertical directions. The predictions on the left column correspond to above-average success rates, while those on the right column correspond to below-average success rates. The blue line represents the experimental GRF, while the red points line represents the predicted GRF. *nRMSE* normalized root mean square error, *PCC* Pearson correlation coefficient

and spans a period of over two decades, the ML algorithm should continuously evolve by incorporating potential data from new cases. Another important point to note is that our study did not encompass subjects with hemiplegia. Introducing ML algorithms to such patients would undoubtedly enhance the practicality of the proposed approach for predicting joint kinetics.

In conclusion, utilizing machine learning for predicting GRF from kinematic data holds promise as a potential alternative to traditional methods in gait analysis for patients with CP in the future. This approach has the potential to streamline the clinical evaluation of patients with CP by reducing the number of measurements and the equipment needed in gait laboratories. Nevertheless, the current presence of prediction errors constrains the immediate clinical applicability of this machine learning-based approach.

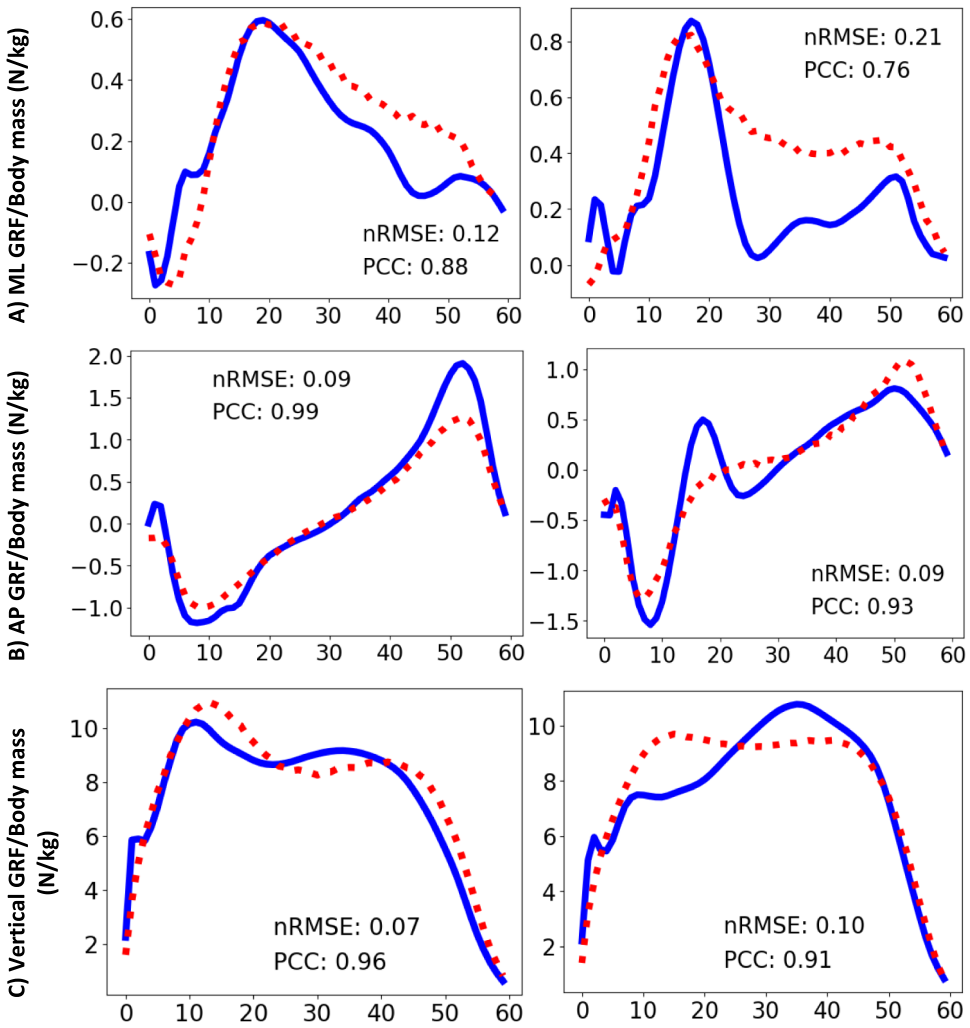

Fig. 7 Representative GRF prediction results of patients with CP. GRFs along **A** mediolateral (ML), **B** anteroposterior (AP), and **C** vertical directions. The predictions on the left column correspond to above-average success rates, while those on the right column correspond to below-average success rates. The blue line represents the experimental GRF, while the red points line represents the predicted GRF. *nRMSE* normalized root mean square error, *PCC* Pearson correlation coefficient

Funding No funding was received for conducting this study.Financial interests The authors declare they have no financial interests.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Lin, C. J., L. Y. Guo, F. C. Su, Y. L. Chou, and R. J. Cherng. Common abnormal kinetic patterns of the knee in gait in spastic diplegia of cerebral palsy. *Gait Posture*. 11(3):224–232, 2000.
2. Lenhart, R. L., S. C. Brandon, C. R. Smith, T. F. Novacheck, M. H. Schwartz, and D. G. Thelen. Influence of patellar position on the knee extensor mechanism in normal and crouched walking. *J. Biomech.* 51:1–7, 2017.
3. Gage, J. R. The clinical use of kinetics for evaluation of pathological gait in cerebral palsy. *J. Bone Jt Surg.* 76(4):622–631, 1994.
4. Novacheck, T. F., and J. R. Gage. Orthopedic management of spasticity in cerebral palsy. *Child's Nerv. Syst.* 23(9):1015–1031, 2007.
5. Caldas, R., T. Fadel, F. Buarque, and B. Markert. Adaptive predictive systems applied to gait analysis: a systematic review. *Gait Posture*. 77:75–82, 2020.
6. White, R., I. Agouris, R. D. Selbie, and M. Kirkpatrick. The variability of force platform data in normal and cerebral palsy gait. *Clin. Biomech.* 14(3):185–192, 1999.
7. Mouloodi, S., H. Rahmanpanah, S. Gohari, C. Burvill, K. M. Tse, and H. M. Davies. What can artificial intelligence and machine learning tell us? A review of applications to

equine biomechanical research. *J. Mech. Behav. Biomed. Mater.*123:104728, 2021.
8. Oh, S. E., A. Choi, and J. H. Mun. Prediction of ground reaction forces during gait based on kinematics and a neural network model. *J. Biomech.* 46(14):2372–2380, 2013.
9. Johnson, W. R., A. Mian, C. J. Donnelly, D. Lloyd, and J. Alderson. Predicting athlete ground reaction forces and moments from motion capture. *Med. Biol. Eng. Comput.* 56:1781–1792, 2018.
10. Mundt, M., A. Koeppe, S. David, F. Bamer, W. Potthast, and B. Markert. Prediction of ground reaction force and joint moments based on optical motion capture data during gait. *Med. Eng. Phys.* 86:29–34, 2020.
11. Johnson, W. R., J. Alderson, D. Lloyd, and A. Mian. Predicting athlete ground reaction forces and moments from spatio-temporal driven CNN models. *IEEE Trans. Biomed. Eng.* 66(3):689–694, 2018.
12. Ozates, M. E., D. Karabulut, F. Salami, S. I. Wolf, and Y. Z. Arslan. Machine learning-based prediction of joint moments based on kinematics in patients with cerebral palsy. *J. Biomech.*155:111668, 2023.
13. Ihlen, E. A., R. Støen, L. Boswell, R. A. de Regnier, T. Fjørtoft, D. Gaebler-Spira, C. Labori, M. C. Loennecken, M. E. Msall, U. I. Möinichen, C. Peyton, M. D. Schreiber, I. E. Silberg, N. T. Songstad, R. T. Vågen, G. K. Øberg, and L. Adde. Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: a multi-site cohort study. *J. Clin. Med.* 9(1):5, 2019.
14. Zhang, Y., and M. Ye. 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, 2019.
15. Morbidoni, C., A. Cucchiarelli, V. Agostini, M. Knaflitz, S. Fioretti, and F. Di Nardo. Machine-learning-based prediction of gait events from EMG in cerebral palsy children. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29:819–830, 2021.
16. Kim, Y. K., R. M. Visscher, E. Viehweger, N. B. Singh, W. R. Taylor, and F. Vogl. 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, 2022.
17. Chollet, F. 2015. Keras. GitHub.
18. Refaeilzadeh, P., L. Tang, and H. Liu. Cross-validation. In: Encyclopedia of Database Systems, vol. 5. New York: Springer, 2009, pp. 532–538.
19. Ardestani, M. M., X. Zhang, L. Wang, Q. Lian, Y. Liu, J. He, D. Li, and Z. Jin. Human lower extremity joint moment prediction: a wavelet neural network approach. *Expert Syst. Appl.* 41(9):4422–4433, 2014.
20. Mundt, M., A. Koeppe, S. David, T. Witter, F. Bamer, W. Potthast, and B. Markert. Estimation of gait mechanics based on simulated and measured IMU data using an artificial neural network. *Front. Bioeng. Biotechnol.* 8:41, 2020.
21. Ripic, Z., C. Kuenze, M. S. Andersen, I. Theodorakos, J. Signorile, and M. Eltoukhy. Ground reaction force and joint moment estimation during gait using an Azure Kinect-driven musculoskeletal modeling approach. *Gait Posture*. 95:49–55, 2022.
22. Savelberg, H., and H. Walter. Prediction of dynamic tendon forces from electromyographic signals: an artificial neural network approach. *J. Neurosci. Methods*. 78(1–2):65–74, 1997.
23. Silver, N. C., and W. P. Dunlap. Averaging correlation coefficients: should Fisher's z transformation be used? *J. Appl. Psychol.* 72(1):146, 1987.