scientific reports

Check for updates

OPEN

Deep learning-based classification of hemiplegia and diplegia in cerebral palsy using postural control analysis

Javiera T. Arias Valdivia[1✉], Valeska Gatica Rojas[2] & César A. Astudillo[3]

Cerebral palsy (CP) is a neurological condition that affects mobility and motor control, presenting significant challenges for accurate diagnosis, particularly in cases of hemiplegia and diplegia. This study proposes a method of classification utilizing Recurrent Neural Networks (RNNs) to analyze time series force data obtained via an AMTI platform. The proposed research focuses on optimizing these models through advanced techniques such as automatic parameter optimization and data augmentation, improving the accuracy and reliability in classifying these conditions. The results demonstrate the effectiveness of the proposed models in capturing complex temporal dynamics, with the Bidirectional Gated Recurrent Unit (BiGRU) and Long Short-Term Memory (LSTM) model achieving the highest performance, reaching an accuracy of 76.43%. These results outperform traditional approaches and offer a valuable tool for implementation in clinical settings. Moreover, significant differences in postural stability were observed among patients under different visual conditions, underscoring the importance of tailoring therapeutic interventions to each patient's specific needs.

Cerebral palsy (CP) is a chronic neurological condition identified by altered body movements, psychomotor retardation, and, if not treated promptly, can lead to deformities of the limbs and trunk. Depending on the level of the lesion, four types of paralysis can be distinguished: monoparesis, diplegia, hemiplegia, and tetraparesis. In this study, a total of 57 pediatric patients diagnosed with hemiplegia and diplegia were included, classified as follows: 25 patients with right-sided hemiplegia, 10 patients with left-sided hemiplegia, and 22 patients with spastic diplegia. In hemiplegia (unilateral cerebral palsy), children exhibit greater motor impairments on one side of the body compared to the contralateral side. In contrast, in spastic diplegia (bilateral cerebral palsy), both sides of the body are affected, and the lower extremities are commonly more involved than the upper extremities.

These conditions are caused by a disturbance in the brain when the nervous system has not finished maturing (before the age of five). Causes can occur during pregnancy, at birth, after birth, and up to the age of five. Although the condition does not worsen over time, sequelae can worsen if left untreated. CP affects approximately 2-3 of every 1,000 live births worldwide[1].

Early and accurate diagnosis of these conditions is crucial for timely intervention and significant improvement in quality of life[2]. However, the complexity of data and identification presents challenges for treatment specialists.

Advanced artificial intelligence (AI) techniques and time series analysis have shown the potential to improve diagnostic accuracy and support clinical diagnosis. Recent studies have demonstrated the effectiveness of computer vision and deep learning in the analysis of infant movements, crucial for the early detection of CP[2–5]. Pain in children with CP has been classified[6], and robotic-assisted treatment has also been implemented[7]. Moreover, using RGB videos to extract pose sequences and analyze motion patterns has shown promising results in motion pattern recognition[2,8,9]. In addition, AI models such as long-short-term memory and gated recurrent units are effective in finding complex temporal patterns, thus improving the accuracy of various diagnostic and recognition[3,10,11], but not in diagnosing infants with hemiplegia or diplegia.

[1]Doctorado en Sistemas de Ingeniería, Faculty of Engineering, Universidad de Talca, Curicó 3340000, Chile. [2]Faculty of Health Sciences, University of Talca, Talca 3460000, Chile. [3]Department of Computer Science, Faculty of Engineering, Universidad de Talca, Curicó 3340000, Chile. ✉email: javiera.arias@utalca.cl

Several studies have addressed the use of AI in diagnosing and treating CP. For example, Morbidoni et al.[12] used electromyography (EMG) signals and machine learning (ML) to predict gait events in children with CP, obtaining high accuracy in predicting gait events (heel strike and toe off) even under conditions of high signal variability. Silva et al.[13] reviewed computer vision and machine learning-based approaches for the assessment of general infant movement, highlighting the importance of large annotated datasets to improve automated solutions.

Studies report that gait assessment analyzes a patient's movement pattern on the walk to identify and diagnose conditions such as cerebral palsy. They use sensors, cameras, or pressure platforms to capture data and identify gait abnormalities. Xiong et al.[14] developed a neural interface for gait tracking based on muscle synergies and deep neural networks. Similarly, Jung et al.[15] proposed a novel approach for multiple classifications of human gait using time-frequency representations and deep convolutional neural networks. However, Rueangsirarak et al.[16] proposed a framework for classifying neurological disorders in older adults using 3D movement data.

In addition, Dolatabadi et al.[17] explored using non-invasive sensor technology and machine learning methods to classify healthy and pathological gait patterns. Donahue and Hahn[18] compared a heuristic feature identification algorithm with a hidden Markov model for the identification of gait events, showing that both approaches can accurately identify gait events in different locomotion modes. Ricardo et al.[10] reviewed the impact of ankle-foot prosthesis on gait patterns in children with bilateral spastic CP, finding significant improvements in gait speed and stride length.

On the other hand, Xiong et al.[14] demonstrated that the innovative neural interface based on muscle synergies can accurately estimate joint angles during gait. Ren et al.[19] demonstrated the effectiveness of AI to evaluate different balance control subsystems using the center of pressure data. In addition, Agostini et al.[20] presented an algorithm for automatic segmentation and classification of gait cycles using foot switch signals, demonstrating an accuracy of 100% for healthy subjects and 98% for pathological subjects. Pérez-Ibarra et al.[21] developed unsupervised adaptive algorithms to identify gait events in healthy subjects with Parkinson's using inertial sensors. In addition, Zhang et al.[11] analyzed the accuracy of ambulatory gait analysis using machine learning models and instrumented shoe insoles. Duong et al.[22] demonstrated the efficacy of recurrent deep learning models for estimating Center of pressure (COP) trajectories using multimodal instrumented shoe insoles, providing a valuable tool for gait monitoring in real-world settings and significantly correlating with clinical measures of ambulatory function and lower extremity muscle strength.

Recent studies have demonstrated significant advances in the assessment of motor function using monocular video and deep learning techniques. Zhao et al.[23] proposed a spatio-temporal graph convolution network (STGCN) to extract motion features from pose data obtained from monocular videos, achieving an accuracy of 76.6% in the assessment of children with CP using the Gross Motor Function Classification System (GMFCS), improving accuracy by 5% compared to current methods. Rahmati et al.[3] developed a frequency analysis and feature reduction method to predict early infant CP. In addition, Skaramagkas et al.[24] conducted a systematic review of deep learning techniques applied to the diagnosis of Parkinson's disease (PD), highlighting the importance of data availability and model interpretability.

However, although gait assessment has been extensively studied, this approach has limitations regarding the complexity and variability of data collected in uncontrolled settings. In contrast, postural control measures stability and balance in a static position, which is crucial for patients with severe movement limitations, reducing physical strain for patients. This test is less invasive for the patient because it does not require the individual to perform complex movements or move around during the test. It benefits patients with cerebral palsy who may find it tiring or painful to perform a gait. In addition, the duration of a postural control test is generally shorter and less demanding than a gait assessment, reducing the overall assessment time and decreasing patient fatigue, improving the quality of the data collected. This system provides quantitative data that accurately reflect motor skills and postural control[25]. Studies show that posturographic measures, such as the displacement area of COP and the velocities in the mediolateral and anteroposterior directions, are sensitive to changes in postural control and can provide a solid basis for diagnosis and rehabilitation in diverse clinical populations[25–29].

However, the literature has continued to explore the potential of AI in clinical practice. Ullrich et al.[30] developed an algorithm to automatically detect unsupervised standardized gait tests from data from the Inertial Measurement Unit (IMU) in Parkinson's patients. Eguchi and Takahashi[31] proposed an efficient method to estimate the vertical ground reaction force (vGRF) during gait using instrumented insoles, showing high accuracy and providing a viable alternative to force plates. Pham et al.[32] used texture analysis for the classification of gait patterns in patients with neurodegenerative diseases, achieving high accuracy. Iosa et al.[33] reviewed the use of gaming technologies in pediatric neurorehabilitation, showing promising results in improving children's motivation and participation in therapies. Chakraborty and Nandy[34] investigated the use of computational intelligence techniques for the automatic diagnosis of gait in CP using a low-cost multi-sensor approach, demonstrating high accuracy in detecting gait abnormalities. Gombolay et al.[35] explored the application of AI in pediatric neurology, highlighting its potential to improve the accuracy and personalization of treatments. Khanna et al.[36] reviewed neurodevelopmental treatment (NDT) in children with CP, showing positive results in improving motor function, balance and postural control.

Integrating advanced AI techniques and time series analysis offers a promising approach for an early and accurate diagnosis of cerebral palsy. AI models such as LSTM and GRU have shown improved diagnostic accuracy and identification of movement patterns, facilitating the adoption of these technologies in clinical practice and improving long-term outcomes for CP patients.

The diagnosis of hemiplegia and diplegia in pediatric patients using COP data presents significant challenges. The diverse nature of the data, including variations in collection methods, demographic differences, and testing conditions, makes it difficult to apply a single algorithm for all tasks. These unique characteristics require a customized approach. The absence of previous studies addressing these factors further underscores the need

for AI models tailored to the specific conditions of the data. Developing and evaluating such models will enable more accurate diagnoses and more effective treatments, ultimately enhancing clinical outcomes for pediatric patients.

The present study focuses on the development of a decision support system that automatically classifies hemiplegia or diplegia in pediatric patients. The system improves diagnostic accuracy and facilitates early and personalized intervention using LSTM and GRU models to capture complex temporal dynamics in force time series data.

This article develops a deep learning-based classifier to classify hemiplegia or diplegia in pediatric patients. It uses long short-term memory (LSTM), gated recurrent units (GRU), bidirectional long short-term memory (BiLSTM), bidirectional gated recurrent units (BiGRU), and the autoregressive integrated moving average (ARIMA) models to analyze temporal patterns and postural control in various visual conditions, thereby enhancing clinical diagnosis and supporting early intervention in cerebral palsy.

Methodology

The present study aims to automatically classify hemiplegia or diplegia in pediatric patients, supporting clinical diagnosis using advanced time series analysis techniques and explainable artificial intelligence. To this end, force data were collected via an AMTI force platform during tests conducted with eyes open. Long Short-Term Memory (LSTM), Gated Recurrent Units neural network models, and ARIMA were used to capture the complex temporal dynamics present in the biomechanical data.

Data collection and preprocessing

This study collected force data, referring to the measurement of forces exerted by the patients on the force platform during postural assessments, from 57 patients, aged 7 to 14 years (age: 9.2 ± 1.8 years; 29 males, 28 females), all diagnosed with hemiplegia or diplegia. The data were collected during the patient evaluation stage, prior to the start of any rehabilitation process, using clinical evaluations that comply with international standards. This study was approved by the Scientific Ethics Committee of the University of Talca, Chile, under protocol number 24-2018 on 26/09/2018. Furthermore, the study is registered with the Australian New Zealand Clinical Trials Registry (ANZCTR) under the number ACTRN12621000117819 with date of trial registration 05/02/2021. The data was collected using an AMTI force platform by a team of researchers specializing in postural control for cerebral palsy in Chile[25–28]. The AMTI force platform is a highly precise device that measures the forces and moments generated by the subjects during the tests. In detail, the structure of the clinical trial study (Fig. 1) consists of 3 trials (t = 3) with a total of 103 samples. The trials have different sample sizes: trial 1 (n = 34), trial 2 (n = 5), and trial 3 (n = 18), each with different repetitions (rep). Trial 1 is performed with a single repetition of 34 samples, while trials 2 and 3 are repeated three times with 15 and 54 samples, respectively. The samples are classified into two clinical conditions: hemiplegia (Hem) and diplegia (Di), with 64 cases of hemiplegia and 39 cases of diplegia. In clinical neurology, particularly in postural control research within pediatric populations with CP, smaller sample sizes are common due to the specificity and complexity of the condition. In Chile, CP represents a significant proportion of pediatric neurorehabilitation cases, with approximately 33% of all neurorehabilitation services corresponding to children with CP, according to the Teletón Foundation. National disability studies further confirm the high prevalence and clinical relevance of CP in the country[37]. Given these factors, the data utilized in this study are representative and sufficient to provide meaningful insights into supporting clinical diagnosis in CP subtypes.
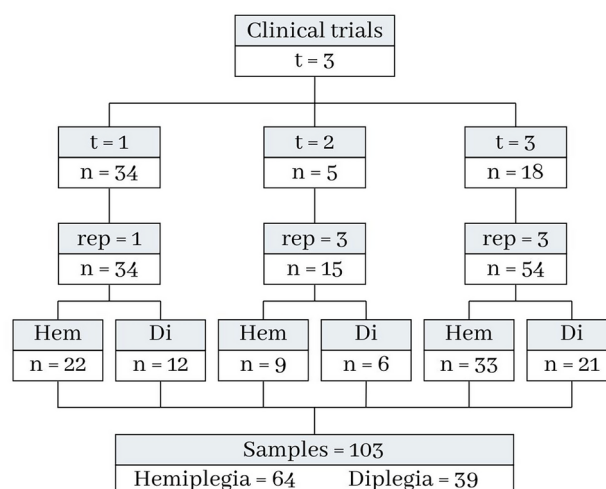

Fig. 1. Structure of the clinical trial study with three trials (t = 3) and 103 samples from 57 participants, categorized into hemiplegia (Hem, n = 64) and diplegia (Di, n = 39). Each trial varies in repetitions and sample size, enabling comparative analysis of outcomes between the two conditions

Each child participated in a test where they stood barefoot on the force platform in a natural posture, looking at an 'X' positioned at eye level on a wall in front of them. They maintained this position for 30 seconds while force data were recorded at a frequency of 200 Hz. This high sampling rate allowed for a detailed capture of the child's postural control.

The AMTI platform measures the forces applied on its axes (F_x, F_y, F_z) and the moments generated (M_x, M_y, M_z). These data are essential for analyzing postural control in different visual conditions, providing a solid basis for diagnostic and rehabilitation research. The platform was calibrated before each testing session to ensure the accuracy of the measurements.

Unlike other studies, the collected data were not pre-processed to remove outliers; instead, it was processed entirely, allowing for a more representative analysis of movement dynamics and postural control. This methodology provides a detailed assessment of patient conditions, contributing to a more reliable evaluation of postural impairments.

Centre of pressure
Centre of Pressure (COP) is a key measure used to represent the point where the resultant of the ground reaction forces acts on the body's supporting surface. The position is quantified through coordinates on the x- and y-axes, which are determined using specific equations (1 and 2) that account for the distribution of forces across the platform.

$$COP_x = -\frac{My + Fx \cdot dz}{Fz} \tag{1}$$

$$COP_y = \frac{Mx - Fy \cdot dz}{Fz} \tag{2}$$

where $dz = 41.3$ mm. This value represents a calibration constant used to calculate the COP on the AMTI force platform.

Velocity
Measures how quickly a patient's position changes over time, making it essential for assessing posture. It helps assess motor function, detect asymmetries, and follow rehabilitation progress. Equations 3 and 4 are used for this variable.

$$Velocidad_x = \frac{\Delta COP_x}{\Delta t} \tag{3}$$

$$Velocidad_y = \frac{\Delta COP_y}{\Delta t} \tag{4}$$

where $\Delta t = 1/200\ (seconds)$ (5 milliseconds) is the time interval between consecutive measurements, corresponding to the sampling frequency of 200 Hz used by the force platform. This interval is the time between each data point recorded during the test.

Standard deviation
Standard Deviation (STD) quantifies the variability or dispersion of values, providing a measure of stability in COP positions. In this study, the standard deviation of COP positions along the x and y directions serves as an indicator of postural stability. A higher standard deviation suggests increased variability and potential instability in maintaining balance. Specifically, the standard deviations STD_x and STD_y capture the average deviation of COP positions from their respective mean values, calculated using the root mean square approach (Equations 5 and 6).

$$STD_x = \sqrt{\frac{\sum (COP_x - C\bar{O}P_x)^2}{N}} \tag{5}$$

$$STD_y = \sqrt{\frac{\sum (COP_y - C\bar{O}P_y)^2}{N}} \tag{6}$$

where N refers to the total number of data points used to calculate the standard deviation, based on measurements taken over 30 seconds at 200 Hz, resulting in 6000 data points per trial.

Area of the ellipse
that encompasses the COP positions provides a measure of the extent of variability of the COP position in the plane. This area is calculated using the standard deviations in the x and y directions and is expressed as:

$$Area\ elipse = \pi \cdot STDx \cdot STDy \tag{7}$$

Given the 200 Hz sampling frequency and the 30-second recording, the COPs form a 6000-point time series, while the velocities are a 30-point time series. Initially, STDS and Area elipse are static values. However, to use these variables as input for the model, they were transformed into time series through a sliding-window technique. Specifically, six sliding windows were applied, each 5 seconds long and containing 1000 points. This

approach effectively converted static variables into dynamic time series, capturing the evolution of postural stability over time.

Several COP variables were calculated from these data, which are necessary to assess the subject's postural stability. These variables quantitatively measure oscillations and balance control in response to different visual conditions.

Finally, all tests and procedures were conducted following established ethical considerations, ensuring informed consent of the participants and the confidentiality of the data.

Data augmentation
Data augmentation techniques were used to address the class imbalance in the data set. These methods expand the data set and improve model robustness by generating or modifying data, helping the model handle imbalanced distributions.

SMOTE
Synthetic Minority Over-sampling Technique (SMOTE) generates synthetic samples by selecting instances of the minority class and interpolating between the selected instance and its closest neighbors. This increases the representation of the minority class, balancing the dataset and reducing classifier bias towards the majority class, thereby improving the model's ability to classify minority class instances correctly.

TSAUG
Time Series Augmentation (TSAUG) is designed to augment time series data. Modifies the temporal structure of the data while preserving underlying patterns, introducing slight variations through methods such as time warping, shifting, and scaling. These transformations enhance the model's ability to capture temporal dependencies by simulating real-world variability, such as shifts or fluctuations in time-dependent signals.

Jittering
This technique is a data augmentation method in which slight random noise is added to the data. This noise enhances model robustness by encouraging it to focus on underlying patterns rather than memorizing specific examples, thereby reducing the risk of overfitting, particularly in small datasets.

Models and validation
Recurrent Neural Networks (RNNs) specialize in processing data sequences, such as time series and text. Unlike traditional neural networks, RNNs have feedforward connections that allow them to maintain a short-term memory of previous inputs in the sequence. This is especially useful for tasks where context and order of data are crucial. While RNNs are useful for processing sequential data, they have limitations, especially in learning long-term dependencies. This is due to the vanishing and exploding gradient problems that occur during training.

Long short-term memory
Long Short-Term Memory (LSTM)[38] is an improvement over standard RNN, designed to address the difficulty of learning long-term dependencies. The LSTM architecture includes a cell memory and three gates that control the flow of information: the input gate, the forgetting gate, and the output gate. This allows LSTMs to retain information for long periods and better manage temporal dependencies. The LSTM equations can be expressed as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{9}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{10}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{11}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{12}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{13}$$

Where:

- f_t is the forgetting gate that decides which information to forget from the previous cell state.
- i_t is the input gate that decides what new information to store.
- $\tilde{C}_t$ is the candidate cell state, which is the new candidate information to be added to the cell state.
- C_t is the updated cell state.
- o_t is the output gate that decides which part of the cell state will be used for the output.
- σ is the sigmoid function, which limits the values between 0 and 1.
- $\tanh$ the hyperbolic tangent function, which produces values between -1 and 1.

Gated recurrent unit
Gated Recurrent Unit (GRU)[39] is a simpler variant of LSTM that was introduced to improve the efficiency of recurrent neural networks. GRUs combine the input and forgetting gates into a single update gate and eliminate

the separate cell state, simplifying the architecture, and reducing the number of parameters. This simplicity allows for faster and more efficient training. The GRU equations are the following.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{14}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{15}$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t] + b_h) \tag{16}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{17}$$

where:

- z_t is the update gate that decides how much of the previous state h_{t-1} and the new candidate state $\tilde{h}_t$ should be combined.
- r_t is the reset gate that decides how much of the past memory should be forgotten.
- $\tilde{h}_t$ is the candidate hidden state, which is the new value proposed for the hidden state.

In the GRU cell, the two vectors in the LSTM cell are combined into a single vector h_t. A gate driver handles the entry and forgets the gates. When z_t is 1, the forgetting gate opens, the input gate closes, and vice versa. This simplified architecture allows for more efficient processing in terms of computation and memory, although in some cases it may be less effective in capturing very long-term dependencies.

Bidirectional
Bidirectional models[40], such as bidirectional LSTM (BiLSTM) and bidirectional GRU (BiGRU), allow information to flow in both directions through the data stream. Unlike unidirectional LSTM and GRU models, which process information in one temporal direction (past to future), bidirectional models process sequences in two directions (past to future and future to past). This ability to capture temporal dependencies in both directions improves performance in tasks where both past and future contexts are essential for classification, such as in identifying patterns in complex time series.

AutoRegressive integrated moving average
AutoRegressive Integrated Moving Average (ARIMA) A statistical model widely used to analyze and predict linear time series. This model combines three main components: an autoregressive (AR) component, a differencing (I) component, and a moving average (MA) component[41]. These components were not optimized. Instead, the default values for the parameters were used during the analysis. The autoregressive component uses past values of the time series to predict the current value. The moving average component captures the relationship between the current value and past prediction errors. Finally, integration is used to make the series stationary if necessary. The general equation of ARIMA (p,d,q) is expressed as:

$$\begin{aligned} y_t \quad &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\ &+ \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \end{aligned} \tag{18}$$

Where:

- y_t the value of the time series over time t.
- c is the constant.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients that measure the relation between the current value and the past values $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients that measure the relationship between past errors $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$.
- ϵ_t is the error or term of white noise in time t.
- p is the order of the autoregressive component (AR).
- d is the order of differentiation (I).
- q is the order of the moving average (MA) component.

Validation
To assess the effectiveness of the models, the data set was split into training and test subsets. The model was trained using the training data and then evaluated on the test data to assess its performance. This approach helps to ensure that the model generalizes well to unseen data and is not overfitted to a specific subset.

In addition, performance metrics such as precision, accuracy, sensitivity, and F1 score were implemented to compare the LSTM and GRU models. These metrics provide a comprehensive assessment of model performance, considering the ability to correctly identify hemiplegia and diplegia conditions and minimization of false positives and negatives.

The validation results are presented in the results section, where the performance metrics of the LSTM and GRU models are compared using separate test data. The final model is chosen based on its performance in these metrics, ensuring the best combination of precision and accuracy for the clinical diagnosis of cerebral palsy.

Results

This section presents the results obtained by implementing GRU, LSTM, Bidirectional GRU, Bidirectional LSTM and ARIMA models to classify hemiplegia and diplegia using time series data from an AMTI force platform. The analysis involved extracting COP variables, such as COP coordinates, velocity, and standard deviations, essential to evaluate postural stability under different visual conditions. As illustrated in Fig. 2, the methodology included multiple steps, such as data collection from 103 samples, the application of various data augmentation techniques (SMOTE, TSAUG, and Jittering) to balance the dataset, and the optimization of the model architectures. The results were then analyzed to assess the effectiveness of each model in accurately distinguishing between hemiplegia and diplegia.

Input data analysis

Force data were collected from 103 instances of 57 patients diagnosed with hemiplegia and diplegia. Measurements were taken at a frequency of 200 Hz, allowing a detailed analysis of postural control. The variables calculated included the COP coordinates, the velocity of the COP, the standard deviation of the COP on the x and y axes and the ellipse area that encompasses the displacement of the COP. These variables are crucial in assessing postural stability under different visual conditions.

Data balance

The data imbalance was addressed using three algorithms: SMOTE, TSAUG, and Jittering. Five iterations were performed for each technique. A GRU model was used as the base classifier to evaluate the techniques. The GRU was configured with 50 units in the first layer and a dropout rate of 0.2. The second layer also had 50 units with a dropout rate of 0.4. The dense layer consisted of 10 units with an additional dropout of 0.2, and the learning rate was set to 0.1. This base architecture was chosen for its simplicity and computational efficiency, serving as a consistent benchmark to compare the performance of the data balancing techniques. The primary goal of this step was to identify the best data augmentation method. Once the optimal technique (TSAUG) was identified, the balanced dataset was used to evaluate and optimize the five distinct architectures for each model (GRU, LSTM, Bi-GRU, and Bi-LSTM). TSAUG achieved the best performance (Table 1) and improved metrics such as recall and F1 score due to controlled diversification of the data. Specifically, the TSAUG package applied the following methods for augmentation: TimeWarp(), Crop(), Quantize(), Drift(), and Reverse().
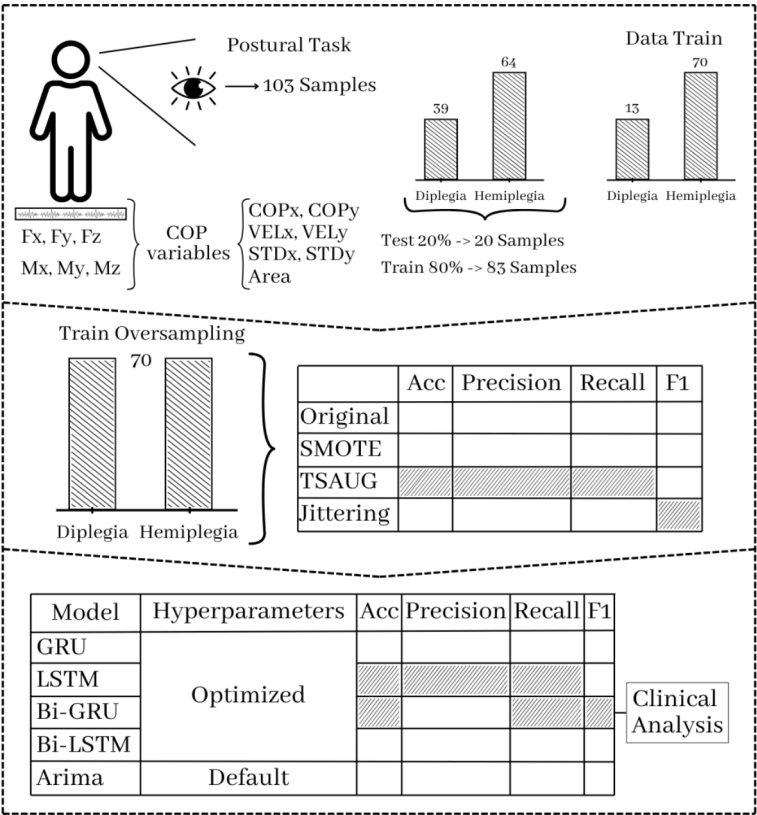

Fig. 2. Workflow for classifying hemiplegia and diplegia using time series data from an AMTI force platform. It includes data acquisition, extraction of COP variables, data augmentation (SMOTE, TSAUG, Jittering), and training of deep learning models (GRU, LSTM, BiGRU, BiLSTM) and ARIMA, optimized

Tecnic	Accuracy	Precision	Recall	F1 Score
Original	0.48	0.49	0.48	0.48
SMOTE	0.52	0.52	0.52	0.52
TSAUG	0.62	0.58	0.62	0.54
Jittering	0.57	0.56	0.57	0.56

Table 1. Comparative performance of data augmentation techniques applied to the training set, evaluated in terms of accuracy, precision, recall, and F1 Score over five iterations. The results were consistent across all iterations.

Model	Arch.	Accuracy	Precision	Recall	F1 Score
GRU	1	0.67	0.78	0.67	0.57
LSTM	1	0.71	0.80	0.71	0.66
Bi GRU	4	0.76	0.77	0.76	0.75
Bi LSTM	3	0.62	0.58	0.62	0.54
ARIMA	-	0.43	0.77	0.43	0.31

Table 2. Optimal performance achieved by each model using optimised architectures. Accuracy, precision, recall, and F1 score values are presented, and recurrent models (GRU, LSTM, BiGRU, BiLSTM) and the ARIMA model are compared.

Modelling evaluation
In this study, the available data were partitioned into training and test sets to evaluate the performance of the models in the classification task of hemiplegia and diplegia.

Data splitting
The data was split using the train-test split technique. Specifically, the data set was split into an 80% for training and a 20% for testing. This split ensures that a significant portion of the data is used to train the models. In contrast, a smaller portion is reserved for evaluating the generalizability of the models on data not seen during training.

Training set (80): This set was used to fit the GRU, LSTM, BiGRU, BiLSTM, and ARIMA models, applying different pre-processing and optimization techniques during training. During this process, an automatic hyperparameter optimization tool was used to find the optimal settings for each model, selecting those that improved the model's performance on the training set.

Test set (20): This test set evaluated trained and optimized models. Provides an estimate of the actual performance of the model on data not observed during training.

Parameter optimization
This study evaluates the performance of four deep learning algorithms: GRU, LSTM, BiGRU, and BiLSTM. For each algorithm, five distinct architectures were explored and optimized to maximize classification performance. The optimization process focused on tuning shared hyperparameters across all algorithms, including the number of units per layer, learning rate, and dropout rate.

Each model consisted of three layers of neurons. The number of units per layer was tested within the following ranges: [50, 200] for the first layer, [50, 200] for the second layer, and [10, 50] for the third layer. The dropout rate was varied within the range [0.2, 0.5], while the learning rate was tested in the interval $[10^{-5}, 10^{-3}]$. The models were trained using the Adam optimizer and the categorical cross-entropy loss function. Accuracy was used as the primary evaluation metric.

The models were trained for 50 epochs, meaning the model passed through the entire training dataset 50 times. The data was split into training and test sets using the *train_test_split* function from *sklearn*, with 80% allocated for training and 20% for testing. The training process used a batch size of 32, meaning that the data was processed in smaller chunks during each epoch. Furthermore, early stopping was applied to prevent overfitting, and training was halted if no improvement was observed in the validation set. To systematically identify the optimal hyperparameters, Bayesian optimization was employed, allowing efficient exploration of the hyperparameter space and ensuring robust model performance across the selected architectures.

The best performance achieved by each model is shown in Table 2. The architectures corresponding to these models that perform best are illustrated in Figure 3. For the GRU model, the highest performance was obtained with Architecture 1, which consisted of 100 units in the first layer, 150 units in the second layer and 50 units in the third layer, with dropout rates of 0.3, 0.4, and 0.3, respectively, and a learning rate of 0.001. The LSTM model also reached its best performance using Architecture 1, which comprised 50 units in the first layer, 50 units in the second layer, and ten units in the third layer, with dropout rates of 0.2, 0.4, and 0.2, respectively, and a learning rate of 0.001. Architecture 4 yielded the best results for the BiGRU model, characterized by 200 units in the first layer, 150 in the second layer, and 40 in the third layer, with dropout rates of 0.5, 0.3, and 0.2 and a
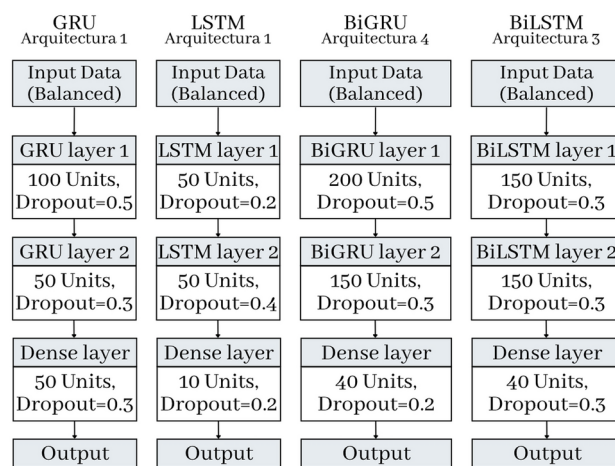
Fig. 3. Optimized Architectures of the Best-Performing Recurrent Neural Network Models for Postural Control Assessment in Cerebral Palsy Diagnosis

learning rate of 0.0006. Finally, the optimal performance of the BiLSTM model was achieved using Architecture 3, which included 150 units in the first layer, 150 in the second layer and 40 in the third layer, with dropout rates of 0.3 for all layers and a learning rate of 0.0003. The details of the parameters tested in these experiments are shown in Table S1.

Discussion

This study classifies hemiplegia and diplegia in children with cerebral palsy (CP) using force data (F_x, F_y, F_z) and moment (M_x, M_y, M_z) obtained from an AMTI force platform. Subsequently, we converted the data into Center of pressure (COP) variables. The decision to use COP variables instead of raw force or moment data was based on clinical expertise in the domain, as COP is a widely accepted indicator of postural control and balance in rehabilitation. The method enables the recognition of dynamic postural control patterns instead of static features. This approach ensures that our results are accepted within the clinical field. Motion-based studies often require patients to perform physically demanding tasks. In contrast, our study involves a brief 30-second static postural control test, making it a non-invasive alternative for children with CP, minimizing patient fatigue and discomfort.

We considered recurrent neural networks, including LSTM, GRU, BiLSTM, and BiGRU, to detect hemiplegia and diplegia based on time series. Bidirectional models, particularly BiGRU, shown superior performance by effectively capturing temporal dependencies in both directions. These findings are consistent with previous research stating that bidirectional architectures outperform unidirectional models in tasks that require detailed temporal analysis[42].

The relatively inferior performance of standard LSTM and GRU models could be attributed to their inability to capture bidirectional dependencies. However, their performance remains significant, suggesting that these models are still valuable for clinical applications where bidirectionality is not essential. The choice of the model will depend on the characteristics of the dataset and the clinical problem to be solved, highlighting the need for more comparative studies in different clinical settings.

Specifically, when comparing the results of LSTM and BiGRU, both models achieve a similar accuracy (0.71 - 0.76). However, BiGRU obtained a higher F1 score (0.75) compared to LSTM (0.66), suggesting that BiGRU is more effective in identifying the minority class in this clinical dataset. The superior precision of the LSTM model relative to BiGRU suggests that it produces fewer false positives, leading to a higher specificity in classification. This implies that when LSTM assigns a positive label, it is more likely to be correct. However, since recall remains the same for both models, LSTM does not identify more true positives than BiGRU but instead classifies fewer false positives. Consequently, this imbalance between precision and recall reduces the F1 score, given that this metric represents the harmonic mean of both. In contrast, BiGRU benefits from bidirectional processing, enhancing its ability to capture temporal dependencies in past and future contexts. This bidirectionality enables a more balanced trade-off between precision and recall, ultimately leading to a higher F1 score. LSTM did not achieve the same level of balance between precision and recall, which limits its classification ability in situations where classes are unbalanced. Moreover, BiGRU surpasses LSTM in performance and benefits from a simpler architecture, reinforcing its position as a more efficient and practical alternative. Healthcare specialists concurred that these accuracy levels are acceptable, considering the complexity of the process.

ARIMA was used as a traditional benchmark model, but it was unable to capture the dynamics present in the postural control data. This highlights the complexity of the problem and justifies the use of deep learning models, which are designed to handle high-dimensional sequential data. However, due to the focus of the research, the alternative of converting the dataset into static data and comparing it with a traditional machine learning model was not considered.

Although deep learning models exhibit strong performance, their limited interpretability poses a significant challenge to clinical adoption. Integrating explainability tools is crucial to building trust and enabling healthcare professionals to effectively use these methods. We anticipate that future research will prioritize addressing this critical need.

To tackle the class imbalance, we evaluated three data augmentation techniques: TSAUG, SMOTE, and Jittering, and compared them with the original unbalanced data. A GRU model with a simple architecture was employed as a baseline classifier to ensure consistent evaluation across techniques. This choice allowed us to focus on comparing the augmentation methods under controlled conditions without introducing variability from complex model architectures. TSAUG proved to be the most effective technique, as it preserves the temporality of the time series while generating synthetic data. In particular, TSAUG improved precision (0.58), accuracy (0.62), and recall (0.62). These values suggest that TSAUG is better at generating new instances while preserving the temporal dependencies of the data. However, it is important to note that TSAUG achieved a slightly lower F1 score (0.54) compared to Jittering (0.56), which could indicate a moderate false positive rate, limiting the overall accuracy of the model. As shown in Table 1, Jittering performed poorly in the other metrics.

The proposed system provides a noninvasive tool to assist in the clinical diagnosis of hemiplegia and diplegia. Traditionally, diagnosis has relied on clinical evaluations, visual analysis, and standardized tests performed by specialists, which can introduce variability in diagnostic outcomes. Early and accurate diagnosis can aid in the rehabilitation of children with cerebral palsy (CP), allowing timely interventions and potentially reducing the long-term impact of motor impairments. Additionally, our deep learning method provides reliable results through a brief and non-intrusive test, making it a practical option for clinical settings where efficiency and patient comfort are important.

Despite the results obtained with the deep learning models for classifying hemiplegia and diplegia, several limitations should be considered. The study used force data from a small cohort of 57 pediatric patients, which can limit the generalizability of the findings to more extensive and diverse populations. Furthermore, data were collected over a 30-second period, which may not capture the full variability in postural control, potentially affecting the model's performance in different clinical scenarios. Although data augmentation techniques, such as SMOTE, TSAUG, and Jittering, were applied to address class imbalance, their impact remains influenced by data distribution and may not fully represent the complexities of real-world clinical datasets. The black-box nature of deep learning models limits their interpretability, requiring improvements in explainability for clinical application.

Conclusion

The proposed method for classifying hemiplegia and diplegia offers a substantial advance in diagnostic precision compared to traditional approaches. The method captures complex temporal dynamics in force data using time series analysis with LSTM, GRU, BiLSTM, BiGRU, and ARIMA models, allowing more accurate differentiation of the characteristic variations of postural control of each condition. This capability is critical for refining the diagnostic process and addressing the challenges associated with traditional clinical assessments.

Transitioning from static diagnostic methods to a data-driven analysis framework allows a deeper exploration of temporal patterns in motor functions. Bidirectional models, which account for past and future dependencies in time series data, have proven especially effective in distinguishing between hemiplegia and diplegia, achieving accuracy metrics of 76% obtained by BiGRU. Health specialists agreed that these accuracies are acceptable given the difficulty of the process. This approach facilitates a more refined analysis of motor patterns in pediatric patients, significantly improving diagnostic accuracy.

The proposed method supports clinical diagnosis, providing a systematic and objective tool to distinguish between hemiplegia and diplegia. This advancement represents a significant leap forward in the field of pediatric neurology, offering new opportunities for the early and effective management of cerebral palsy.

Data availibility

References

1. Tranchida, G. V. & Van Heest, A. Cerebral Palsy: An Overview. *American Family Physician* **101**, 213–220. https://doi.org/10.1177/1753193419878973 (2020).
2. Stahl, A. et al. An optical flow-based method to predict infantile cerebral palsy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **20**, 605–614. https://doi.org/10.1109/TNSRE.2012.2195030 (2012).
3. Rahmati, H. et al. Frequency analysis and feature reduction method for prediction of cerebral palsy in young infants. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **24**, 1225–1234. https://doi.org/10.1109/TNSRE.2016.2539390 (2016).
4. Shanbhag, J. et al. Methods for integrating postural control into biomechanical human simulations: a systematic review. *Journal of NeuroEngineering and Rehabilitation* **20**, 111. https://doi.org/10.1186/s12984-023-01235-3 (2023).
5. McCay, K. D. et al. A Pose-Based Feature Fusion and Classification Framework for the Early Prediction of Cerebral Palsy in Infants. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **30**, 8–19. https://doi.org/10.1109/TNSRE.2021.3138185 (2022).

6. Vinkel, M. N., Rackauskaite, G. & Finnerup, N. B. *Classification of pain in children with cerebral palsy* https://doi.org/10.1111/dmcn.15102 (2022).
7. Santamaria, V. et al. Promoting Functional and Independent Sitting in Children with Cerebral Palsy Using the Robotic Trunk Support Trainer. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 2995–3004. https://doi.org/10.1109/TNSRE.2020.3031580 (2020).
8. Zhao, P. et al. Computer Vision for Gait Assessment in Cerebral Palsy: Metric Learning and Confidence Estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **32**, 2336–2345. https://doi.org/10.1109/TNSRE.2024.3416159 (2024).
9. Wu, Q. et al. A Training-Free Infant Spontaneous Movement Assessment Method for Cerebral Palsy Prediction Based on Videos. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **31**, 1670–1679. https://doi.org/10.1109/TNSRE.2023.3255639 (2023).
10. Ricardo, D. et al. *Effects of ankle foot orthoses on the gait patterns in children with spastic bilateral cerebral palsy: A scoping review* https://doi.org/10.3390/children8100903 (2021).
11. Zhang, H., Guo, Y. & Zanotto, D. Accurate Ambulatory Gait Analysis in Walking and Running Using Machine Learning Models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 191–202. https://doi.org/10.1109/TNSRE.2019.2958679 (2020).
12. Morbidoni, C. et al. Machine-Learning-Based Prediction of Gait Events from EMG in Cerebral Palsy Children. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 819–830. https://doi.org/10.1109/TNSRE.2021.3076366 (2021).
13. Silva, N. et al. The future of General Movement Assessment: The role of computer vision and machine learning - A scoping review. *Research in Developmental Disabilities* **110**, 103854. https://doi.org/10.1016/j.ridd.2021.103854 (2021).
14. Xiong, D., Zhang, D., Zhao, X., Chu, Y. & Zhao, Y. Synergy-Based Neural Interface for Human Gait Tracking with Deep Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 2271–2280. https://doi.org/10.1109/TNSRE.2021.3123630 (2021).
15. Jung, D., Nguyen, M. D., Park, M., Kim, J. & Mun, K. R. Multiple Classification of Gait Using Time-Frequency Representations and Deep Convolutional Neural Networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 997–1005. https://doi.org/10.1109/TNSRE.2020.2977049 (2020).
16. Rueangsirarak, W., Zhang, J., Aslam, N., Ho, E. S. & Shum, H. P. Automatic Musculoskeletal and Neurological Disorder Diagnosis with Relative Joint Displacement from Human Gait. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**, 2387–2396. https://doi.org/10.1109/TNSRE.2018.2880871 (2018).
17. Dolatabadi, E., Taati, B. & Mihailidis, A. An automated classification of pathological gait using unobtrusive sensing technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**, 2336–2346. https://doi.org/10.1109/TNSRE.2017.2736939 (2017).
18. Donahue, S. R. & Hahn, M. E. Feature Identification With a Heuristic Algorithm and an Unsupervised Machine Learning Algorithm for Prior Knowledge of Gait Events. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **30**, 108–114. https://doi.org/10.1109/TNSRE.2021.3131953 (2022).
19. Ren, P. et al. Assessment of Balance Control Subsystems by Artificial Intelligence. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 658–668. https://doi.org/10.1109/TNSRE.2020.2966784 (2020).
20. Agostini, V., Balestra, G. & Knaflitz, M. Segmentation and classification of gait cycles. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22**, 946–952. https://doi.org/10.1109/TNSRE.2013.2291907 (2014).
21. Perez-Ibarra, J. C., Siqueira, A. A. & Krebs, H. I. Identification of Gait Events in Healthy Subjects and with Parkinson's Disease Using Inertial Sensors: An Adaptive Unsupervised Learning Approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 2933–2943. https://doi.org/10.1109/TNSRE.2020.3039999 (2020).
22. Duong, T. T. et al. Accurate COP Trajectory Estimation in Healthy and Pathological Gait Using Multimodal Instrumented Insoles and Deep Learning Models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **31**, 4801–4811. https://doi.org/10.1109/TNSRE.2023.3338519 (2023).
23. Zhao, P. et al. Motor Function Assessment of Children with Cerebral Palsy using Monocular Video. In *2023 IEEE 19th International Conference on Body Sensor Networks, BSN 2023 - Proceedings*, https://doi.org/10.1109/BSN58485.2023.10331472 (publisherInstitute of Electrical and Electronics Engineers Inc., 2023).
24. Skaramagkas, V., Pentari, A., Kefalopoulou, Z. & Tsiknakis, M. Multi-Modal Deep Learning Diagnosis of Parkinson's Disease - A Systematic Review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **31**, 2399–2423. https://doi.org/10.1109/TNSRE.2023.3277749 (2023).
25. Gatica-Rojas, V. & Cartes-Velásquez, R. Telerehabilitation in Low-Resource Settings to Improve Postural Balance in Older Adults: A Non-Inferiority Randomised Controlled Clinical Trial Protocol. *International Journal of Environmental Research and Public Health* **20**, https://doi.org/10.3390/ijerph20186726 (2023).
26. Gatica-Rojas, V., Cartes-Velásquez, R., Soto-Poblete, A. & Lizama, L. E. C. Postural control telerehabilitation with a low-cost virtual reality protocol for children with cerebral palsy: Protocol for a clinical trial. *PLoS ONE* **18**, https://doi.org/10.1371/journal.pone.0268163 (2023).
27. Gatica-Rojas, V. et al. Does Nintendo Wii Balance Board improve standing balance? A randomized controlled trial in children with cerebral palsy. *European journal of physical and rehabilitation medicine* **53**, 535–544, https://doi.org/10.23736/S1973-9087.16.04447-6 (2017).
28. Gatica-Rojas, V. et al. Effectiveness of a Nintendo Wii balance board exercise programme on standing balance of children with cerebral palsy: A randomised clinical trial protocol. *Contemporary clinical trials communications* **6**, 17–21. https://doi.org/10.1016/J.CONCTC.2017.02.008 (2017).
29. Pimentel-Ponce, M. et al. Ludificación y neurorrehabilitación motora en niños y adolescentes: revisión sistemática. *Neurología* **39**, 63–83. https://doi.org/10.1016/J.NRL.2021.02.011 (2024).
30. Ullrich, M. et al. Detection of Unsupervised Standardized Gait Tests from Real-World Inertial Sensor Data in Parkinson's Disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 2103–2111. https://doi.org/10.1109/TNSRE.2021.3119390 (2021).
31. Eguchi, R. & Takahashi, M. Insole-Based Estimation of Vertical Ground Reaction Force Using One-Step Learning with Probabilistic Regression and Data Augmentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**, 1217–1225. https://doi.org/10.1109/TNSRE.2019.2916476 (2019).
32. Pham, T. D. Texture Classification and Visualization of Time Series of Gait Dynamics in Patients with Neuro-Degenerative Diseases. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26**, 188–196. https://doi.org/10.1109/TNSRE.2017.2732448 (2018).
33. Iosa, M., Verrelli, C. M., Gentile, A. E., Ruggieri, M. & Polizzi, A. *Gaming Technology for Pediatric Neurorehabilitation: A Systematic Review* https://doi.org/10.3389/fped.2022.775356 (2022).
34. Chakraborty, S. & Nandy, A. Automatic diagnosis of cerebral palsy gait using computational intelligence techniques: A low-cost multi-sensor approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**, 2488–2496. https://doi.org/10.1109/TNSRE.2020.3028203 (2020).
35. Gombolay, G.Y. et al. Review of Machine Learning and Artificial Intelligence (ML/AI) for the Pediatric Neurologist, https://doi.org/10.1016/j.pediatrneurol.2023.01.004 (2023).

36. Khanna, S., Arunmozhi, R. & Goyal, C. Neurodevelopmental Treatment in Children With Cerebral Palsy: A Review of the Literature. *Cureus* https://doi.org/10.7759/cureus.50389 (2023).
37. Diagramación, D.Y., Avendaño, P. II Estudio Nacional de la Discapacidad en Chile, Edición Ingrid Medel.
38. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).
39. Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111, https://doi.org/10.3115/v1/W14-4012 (publisherAssociation for Computational Linguistics, Stroudsburg, PA, USA, 2014).
40. Schuster, M. & Paliwal, K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* **45**, 2673–2681. https://doi.org/10.1109/78.650093 (1997).
41. Box, G.E., Jenkins, G.M. *Time Series Analysis: Forecasting and Control* (1976).
42. Chen, J. X., Jiang, D. M. & Zhang, Y. N. A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification. *IEEE Access* **7**, 118530–118540. https://doi.org/10.1109/ACCESS.2019.2936817 (2019).

Acknowledgements

Author contributions

All authors drafted and reviewed the manuscript. All authors contributed to the methodology and prepared figures and tables. C.A. and V.G. supervised all the manuscripts. J.A. conducted a research process and source code. J.A. should be considered the first author.

Declarations

Confidentiality

Ethical approval for this study was approved by the Scientific Ethics Committee of the University of Talca, Chile, under protocol number 24-2018 on 26/09/2018. Furthermore, the study is registered with the Australian New Zealand Clinical Trials Registry (ANZCTR) under the number ACTRN12621000117819 with date of trial registration 05/02/2021. All participants provided their informed consent prior to inclusion in the study and their data were anonymized to ensure confidentiality.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93166-3.

Correspondence and requests for materials should be addressed to J.T.A.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.