Computer Methods and Programs in Biomedicine 282 (2026) 109412

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Video-based computational analysis of spontaneous movements in preterm infants: A longitudinal neuromotor assessment

Matteo Moro [a,b,c,*], Sofia Sigismondi [a], Laura Asia Gismondi [a], Chiara Tacchino [c], Sara Uccella [c,d], Luca Antonio Ramenghi [c,d], Paolo Moretti [c], Francesca Odone [a,b], Maura Casadio [a,c]

[a] Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genova, via Dodecaneso 35, Genova, 16146, Italy
[b] Machine Learning Genoa (MaLGa) Center, via Dodecaneso 35, Genova, 16146, Italy
[c] IRCCS Istituto Giannina Gaslini, Via Gerolamo Gaslini 5, Genova, 16147, Italy
[d] Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genova, Largo Paolo Daneo 3, Genova, 16146, Italy

ARTICLE INFO

Keywords:
Computer-aided diagnosis
Infants motion analysis
Longitudinal neuro-motor assessment
Video analysis
Prematurity

ABSTRACT

Background and Objective: Monitoring the spontaneous movements of preterm infants is crucial for the early detection of potential neuromotor deficits. Recent advances in human pose estimation have made markerless video-based methods a valid, non-intrusive, and cost-effective option for analyzing these movements. This paper aims to explore the efficacy of video-based markerless techniques in assessing infants' spontaneous movements, with a focus on identifying early signs of developmental abnormalities.
Methods: We conducted a longitudinal study with two acquisition sessions to evaluate the stability and consistency of our video-based analysis over time. Our approach builds on previous methodologies by incorporating advanced techniques for feature detection, parameter extraction, feature selection, and classification. Emphasis was placed on the interpretability and clinical relevance of the extracted motion parameters.
Results: The results highlight the effectiveness of our approach in identifying subtle changes in infants' motion patterns that may indicate neuromotor deficits. We observed differences in the detection of these deficits across the acquisition sessions, with our method achieving a maximum test accuracy of 90%.
Conclusion: Our findings support the potential of markerless video-based analysis as a valuable tool in the support of the early detection of neuromotor deficits in preterm infants. The high accuracy and clinical relevance of our approach suggest it could play a critical role in early intervention strategies.

1. Introduction

Understanding and analyzing infants' spontaneous movements play a crucial role in the early detection of neuromotor deficits, especially in the context of preterm birth [1,2]. Approximately 5%–15% of premature infants born with a birth weight of less than 1500 g exhibit motor alterations, with 25%–50% of them developing cognitive, behavioral, and/or learning deficits [3]. Among the prevalent neurological disorders affecting infants, particularly those born prematurely, are conditions related to lesions in brain areas controlling movement and posture, commonly called Cerebral Palsy (CP) [4].

Despite the challenges, advancements in intensive care techniques have increased preterm survival rates in high-income countries [5]. Traditional neurological examinations, including observations of spontaneous motor behavior, are typically employed to understand infants' neurological status [6,7]. However, visual analysis methods, such as Prechtl's General Movement Assessment (GMA) [8], though highly sensitive (98% sensitivity), are dependent on highly specialized personnel and may not be feasible for widespread early diagnosis [9,10]. Moreover, other diagnostic tools like magnetic resonance imaging (86%–89% sensitivity) and the Hammersmith Infant Neurological Examination (90% sensitivity), while effective, also have limitations in accessibility and ease of use [7]. Given these constraints, there is a pressing need for non-intrusive, accessible, and automated methods that can assist in the early diagnosis of neuromotor deficits in preterm infants.

* Correspondence to: DIBRIS, University of Genova, via Dodecaneso 35, Genova, 16146, Italy.
E-mail addresses: matteo.moro@unige.it (M. Moro), chiaratacchino@gaslini.org (C. Tacchino), sara.uccella@unige.it (S. Uccella), lucaramenghi@gaslini.org (L.A. Ramenghi), paolomoretti@gaslini.org (P. Moretti), francesca.odone@unige.it (F. Odone), maura.casadio@unige.it (M. Casadio).

https://doi.org/10.1016/j.cmpb.2026.109412
Received 28 December 2024; Received in revised form 22 December 2025; Accepted 26 April 2026
Available online 29 April 2026

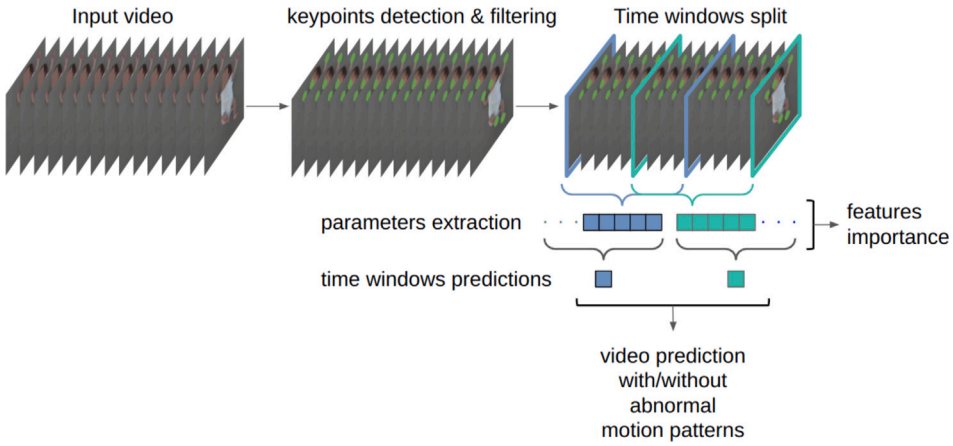

Fig. 1. Summary of the pipeline. Starting from the input video (left), the first steps were the detection of the 5 keypoints (*i.e.*, nose, hands and feet) in each frame and the temporal filtering. Then, each video was split in time windows and a set of quantitative motion parameters were obtained from each time window. These parameters were used for two tasks. (i) Classification: each time window was classified as positive or negative (*i.e.*, with or without neuromotor deficits); then, time windows prediction were combined to reach a final video classification. (ii) Feature selection: we highlighted the most influential features in the classification process to increase interpretability.

Research into human motion analysis has traditionally relied on sophisticated motion capture systems and wearable sensors [11]. However, recent advancements in computer vision and deep learning have facilitated the adoption of less intrusive and cost-effective techniques [12,13]. Markerless video-based methods have emerged as an attractive alternative, offering increased flexibility and natural motion without the constraints of physical markers [14,15]. This becomes particularly advantageous in the study of infants' spontaneous movements, where the use of marker-based techniques can be cumbersome [10,16].

Our research addresses this need by proposing a comprehensive computer-aided methodology aimed at predicting neuromotor disorders directly from video recordings of infants' spontaneous movements. In this domain, non-intrusive video-based methods for infants' motion characterization can adopt classical computer vision algorithms and/or recent deep learning approaches. Early attempts, such as those by Adde et al. [17,18], Cattani et al. [16], Tacchino et al. [19], and Tsuji et al. [20], utilized change detection techniques. Optical flow methods, as demonstrated by Stahl et al. [21] and Rahmati et al. [22], were used to track body parts to focus the analysis on specific motor patterns. All these works reached promising results in the early detection of neuromotor disorders (all of them close to 90% accuracy), but highlighting limits dependent on background changes and lighting, reducing generalizability.

Progress in the definition and extraction of human pose estimation from images leveraging deep learning models [23] allowed for a more fine-grained analysis [24], less influenced by the limits of classical computer vision algorithms. Pose estimators have been adapted in three different works — Reich et al. [25], Chambers et al. [26] and Gao et al. [27] - to study and classify GMs directly from the infants' pose and in Ledwoń et al. [28] for determining the infant's positional asymmetry. However, also these algorithms present limitations, mainly due to the fact that they lack interpretability and the pose estimators adopted are usually tailored for adult poses (*e.g.*, Openpose [29]). To overcome this problem, recent efforts have explored RGB-D data, with Hesse et al. [30] developing a 3D infant body model, and Moccia et al. [31] introducing spatio-temporal features for limb detection using depth videos. However, extracting comprehensive infants' pose information from 2D images remains a challenge, also due to a noticeable lack of open-source datasets (babyPose [32] and MINI-RGBD [33]).

Our study introduces several novel aspects that distinguish it from existing research in the field. Firstly, building on previous consideration presented in [34], we leverage the simplicity and accessibility of videos. Secondly, we employ state-of-the-art machine learning and computer vision techniques specifically fine-tuned with examples from our dataset, underscoring our commitment to leverage modern methodologies for robust and accurate analyses. Furthermore, we prioritize interpretability as a key aspect of our early diagnosis analysis, by incorporating the extraction of quantitative parameters derived from clinical practice [10], offering a more comprehensive understanding of infants' motion patterns. These parameters have proven to be crucial in refining the automatic early diagnosis process. Fig. 1 summarizes the pipeline we propose.

A key aspect of our study is the acquisition of motion patterns at distinct time points: 40 weeks of corrected age (CA), or term equivalent age (*TEA*), and 3 months of corrected age (*3mCA*). This temporal dimension enhances our understanding of the development of infants' motion patterns, enabling us to discern changes over time. The longitudinal aspect not only contributes to the reliability of early detection, but also allows a deeper exploration of the dynamic nature of neuromotor development and of the characteristics (*i.e.*, early detection accuracy and interpretability) of our implemented procedure.

Our analysis is based on a dataset acquired at the Giannina Gaslini Hospital in Genova (Italy). The dataset includes videos of: (i) 142 preterm infants (of which 59 have been diagnosed with neuromotor deficits) acquired at *TEA* and (ii) 118 preterm infants (53 with neuromotor deficits) acquired at *3mCA*. In this dataset, videos of 95 infants have been acquired and evaluated at both time points.

2. Dataset and preprocessing

2.1. Dataset

Data acquisition. The data acquisition occurred at two time points: (i) term of equivalent age (TEA) and (ii) 3 months of corrected age (3mCA) (see Fig. 2). To be eligible for inclusion in the study, infants had to meet specific criteria: they had to be preterm, in stable clinical conditions, without recent pharmacological sedative treatment or respiratory support in the preceding 4 weeks, and with birth occurring no later than the 33th gestational week.

First acquisition: *TEA*. We collected videos of spontaneous movement data of 142 preterm infants (59 of them had a clinical diagnosis of neuromotor deficits), including 81 females. Globally, the infants were born at gestational age of 29.4 ± 2.0 (mean ± standard deviation) weeks, with an average weight of 1221.3 ± 319.1 g (mean ± standard deviation).

Second acquisition: *3mCA*. We collected videos of spontaneous movement data of 118 preterm infants (53 of them had a clinical

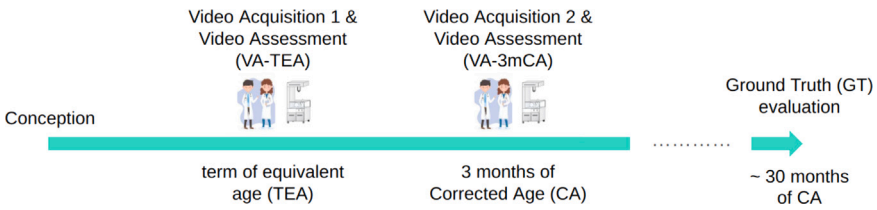

Fig. 2. Acquisition, video evaluations and ground truth timeline.

Table 1
Demographic characteristics (including gestational age (GA) at birth, birth weight, and number of females (F) and males (M)) reported globally and separately for each class (infants *with* and *without* neuromotor deficits) at both term-equivalent age (TEA) and 3 months of corrected age (3mCA). We also report the statistical tests performed and the corresponding p-values for the comparisons between infants with and without neuromotor deficits.

		Overall (142)	Without (83)	With (59)	p-value	Test
TEA	Sex (F/M)	81/61	48/35	33/26	0.957	Chi-squared
	GA (weeks)	29.4 ± 2.0 weeks	29.6 ± 2.0 weeks	29.2 ± 2.1 weeks	0.139	Mann–Whitney
	weight (g)	1221.3 ± 319.1 g	1260.2 ± 330.6 g	1166.6 ± 296.3 g	0.080	t-test
		Overall (118)	Without (65)	With (53)	p-value	Test
3mCA	Sex (F/M)	68/50	39/26	29/24	0.696	Chi-squared
	GA (weeks)	29.2 ± 2.0 weeks	29.5 ± 2.0 weeks	28.8 ± 2.0 weeks	0.048	t-test
	weight (g)	1157.6 ± 316.2 g	1209.0 ± 328.1g	1094.7 ± 291.8 g	0.051	t-test

diagnosis of neuromotor deficits), including 68 females. The infants were born at a mean gestational age of 29.2 ± 2.0 (mean ± standard deviation) weeks, with an average weight of 1157.6 ± 316.2 g (mean ± standard deviation).

The demographic characteristics for each class are reported in Table 1. We performed a statistical analysis to assess whether these characteristics differed significantly between the two groups. For sex, we applied a chi-squared test, which is appropriate for categorical variables. For birth weight and gestational age, we first used the Shapiro–Wilk test to assess normality; normally distributed variables were compared using a t-test, while non-normal variables were analyzed using the Mann–Whitney test. All analyses were conducted in Python with the scipy library, and statistical significance was set at $p < 0.05$. The corresponding results are presented in Table 1. Sex was not associated with neuromotor outcome at either acquisition time point, with high p-values observed at both TEA ($p = 0.975$) and 3mCA ($p = 0.696$). At TEA, neither gestational age ($p = 0.139$) nor birth weight ($p = 0.080$) reached statistical significance. At 3mCA, gestational age ($p = 0.048$) and birth weight ($p = 0.051$) were close to the significance, with gestational age slightly below and birth weight slightly above the $p < 0.05$ threshold. This trend is reasonable from a clinical perspective, as lower gestational age and birth weight are known risk factors for neuromotor impairments, and the reduced 3mCA size may influence the slightly stronger effect observed at this time point (118 infants compared to 142 at TEA). Overall, these results suggest that while routine demographic variables show expected trends, they do not strongly differentiate the two groups. 95 infants (44 with neuromotor deficits) have been acquired at both time points.

The data acquisition setup consisted of a single camera (Canon Legria HF R37), recording at 25 frames per second (fps) with a resolution of 1080 × 1920 pixels. The camera was securely mounted on a stable support positioned above a cradle or a physiotherapy treatment table, at a height approximately between 1 and 1.5 m from the infant [35]. This setup allowed infants to move freely while facing the camera during the recording session. The whole dataset included videos with a mean ± standard deviation duration of 9 ± 3 min. To ensure the quality and consistency of the dataset, we excluded video sequences in which unintentional interventions by operators obscured parts of the scene and in which the infants were crying or using the pacifier because this can influence their spontaneous movements.

The study and the consent form signed by parents were approved by the Giannina Gaslini Hospital Institutional Review Board (protocol number: IGGPM01 20/06/2013).

Ground truth (GT). The neuromotor assessment encompassed various evaluations and tests [19]. All preterm infants underwent a brain Magnetic Resonance Imaging (MRI) assessment at TEA. The standard clinical protocol included T1 and T2 anatomical sequences, in order to detect acquired or prenatal brain lesions that are associated with risk of CP [7,36,37]. In addition, infants participated in a longitudinal clinical follow-up, including routine neurological examinations at 0, 3, 9, 12, and 24 months of corrected age, followed by a final comprehensive assessment at 30 months using the Bayley Scales [38] together with a standardized neuromotor examination [6,39]. The ground truth (GT) label was defined through an integrated longitudinal clinical evaluation ($GT = 1$ and $GT = 0$ denote infants with and without neuromotor deficits, respectively). MRI findings at TEA, repeated neurological examinations, and the final developmental outcome were jointly considered by the expert clinical team to determine a single, unified GT label for each infant. As a result, no discordant cases between MRI findings and clinical follow-up were present. Neuromotor impairments were subsequently characterized by severity: major impairments corresponded to a diagnosis of Cerebral Palsy (CP), established through comprehensive clinical evaluation in conjunction with consistent MRI findings, whereas minor impairments referred to infants without CP who nonetheless exhibited neuromotor impairment, primarily developmental delay (DD), assessed at three years of age using the Griffiths developmental scale [40]. In our cohort, 12 infants were diagnosed with CP, while the remaining infants classified as impaired presented developmental delay.

Video assessment (VA). Two trained physicians assessed the movement of each infant enrolled in the study, at the same time of video acquisition: $VA - TEA$ and $VA - 3mCA$ for the video assessment at TEA and 3mCA respectively. These assessments were administered utilizing only the video recordings and adhering to the General Movements Assessment protocol [8]. Even for expert physicians, identifying abnormal motion patterns at these stages presents challenges and it is more complex at TEA with respect to 3mCA. The two experts collaboratively determined whether the infant exhibited neuromotor deficits, assigning a positive or negative verdict at both acquisition time points independently. The two evaluators showed a high level of agreement in their GMA assessments. In the few cases (27 videos across both timepoints) where their evaluations differed, a third senior clinician reviewed the recordings, and the final decision was reached by majority consensus. Overall, the observed agreement was 89.6% (233/260 assessments), and Cohen's k was 0.8, reflecting substantial/almost perfect agreement.

Table 2
Mean error (in pixels) ± standard deviation (SD) for each keypoint computed considering 350 manually labeled ground truth images do not used during model training. For each keypoint, we also report the error expressed as a mean percentage of the infant's height in pixels (shown in round brackets).

Keypoints	Mean error ± SD in pixels (% of pixels with respect infants height)
Nose	7 ± 3 (1.1%)
Right hand	8 ± 4 (1.2%)
Left hand	8 ± 5 (1.2%)
Right foot	9 ± 7 (1.4%)
Left foot	11 ± 6 (1.7%)

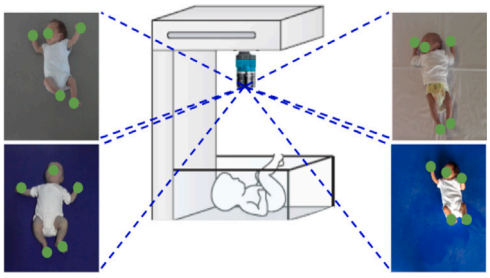

Fig. 3. Examples of camera viewpoint and keypoints detection. The keypoints actual position is identified with the center of the green dots.

2.2. Infants pose estimation

Our approach to video-based analysis primarily focused on extracting the (x, y) position on the image plane of significant keypoints from each video frame. To achieve this, we leveraged a deep architecture specifically designed for semantic features detection (ResNet-50 [41] pre-trained on Imagenet [42]), fine-tuning it to identify key landmark points on infants' bodies (*i.e.*, nose, hands and feet). This decision was motivated by several factors: firstly, by focusing solely on keypoints providing meaningful information regarding infants' motion, we aimed to ensure higher per-point accuracy and greater interpretability of results. Secondly, semantic feature detectors typically necessitate a limited number of annotated examples during the fine-tuning phase [43]. Lastly, we were interested in focusing precisely on the movement of those 5 body parts because they can reflect the majority . To manually label the data (10 randomly selected frames for 200 videos) and to train the deep architecture we relied on DeepLabCut (DLC) [43].

Once trained, we employed the model to extract the five landmarks from each frame of every video in the dataset. The output of this step is the position of each keypoint in the image plane (see Fig. 3) and the confidence the keypoint has been detected correctly $(x_k^t, y_k^t, c_k^t)_{t=0}^T$, where k represents the specific keypoint (nose, left hand, right hand, left foot and right foot), (x_k^t, y_k^t) denotes the position of the kth point in the tth frame, and c_k^t, ranging from 0 to 1, represents the corresponding confidence. Upon obtaining image coordinates for the keypoints in each frame, we proceeded with a filtering step to enhance stability across time and reduce localization errors, as described in [34].

Pose-estimation error analysis. Pose-estimation error analysis. The accuracy of keypoint detection was initially evaluated in our previous study on infant motion analysis [15], where the same DeepLabCut architecture was trained on a subset of the present data and demonstrated reliable performance. Since the current model was retrained using an expanded set of annotated images, we performed an additional ad hoc assessment on an independent test set of 350 manually labeled frames not used during the training of the pose estimation architecture. For each anatomical keypoint, we computed the mean pixel error between predicted and ground-truth locations (see Table 2). This analysis confirmed that the extended model maintains comparable accuracy to the results reported in [15], with no substantial degradation in detection quality despite increased dataset variability.

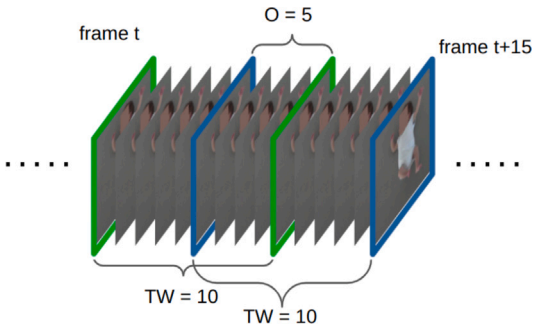

Fig. 4. Examples of time windows TW and overlap O. In this case, just as an example to better understand the meaning of TW and O, we set $TW = 10$ and $O = 5$.

3. Data analysis

3.1. Motion parameters

Starting from the filtered signals describing the temporal evolution of each keypoint position, we derived quantitative parameters characterizing infants' motion patterns in the image plane. Our decision to operate within the image plane, rather than the real-world domain, aimed to underscore the video analysis potential without introducing additional complexities, such as multiple view acquisitions and camera calibration. As the duration of the videos varied and since we aimed to characterize spontaneous movements localized in time, we segmented the video clips into temporal windows (TW) and computed motion features for each time window within every video. To minimize information loss and augment the number of time windows per video, we introduced an overlap O between consecutive time windows (as illustrated in Fig. 4). We set $TW = 500$ (correspondent to time windows of 20 s) and $O = 400$ (16 s) [34].

For each time window, starting from the filtered 2D keypoints coordinates *i.e.*, the (x, y) positions of the five landmark points in the image plane, we computed 156 quantitative parameters describing spontaneous movements, categorized into general and specific motor parameters.

General motor parameters (Pg). We derived general motor parameters following the methodology implemented in [44]. Within each window TW, we extracted 125 features (here referred to as Pg) - 25 parameters per detected keypoint (*i.e.*, 5 keypoints {nose, right hand, left hand, right foot, left foot} $*$ 25 = 125) - divided into spatio-temporal kinematic parameters and parameters in the frequency domain. Kinematic parameters encompassed total path length of the keypoint motion, keypoints' speed, acceleration and jerk, along with their respective statistical descriptors (*i.e.*, mean, maximum, minimum, median and standard deviation). In the frequency domain, we computed the power spectral density of the position, the speed and the acceleration of each keypoint, and derived entropy, peak magnitude, sum of the spectrum and spectral half-point.

Specific motor parameters (Ps). Specific motor parameters were computed to capture symmetric behaviors, variability, smoothness and complexity of motion [10,45]. These 32 parameters were derived from qualitative assessments commonly employed by physicians in evaluating infants' spontaneous motor activity, translating into quantitative descriptors. A subset of these specific motor parameters were already adopted in other works [10,34]:

- Cross-Correlation. Pearson correlation coefficient between the speed and the acceleration of left and right hand and of left and right foot. It provides an indication of motion variability and coordination symmetry between left and right parts of the body [10].
- Skewness. Fisher–Pearson coefficient of skewness for the speed profile of each keypoint [46]. It provides an indication of smoothness and symmetry [10].
- Area out of moving average (AMA). It is the area of the region between the trajectory over time of a keypoint coordinate (x and y) and its moving average. For each keypoint, the values for x and y coordinates are summed [10]. Additionally, we considered the area out of the standard deviation of the moving average (ASTDMA), *i.e.*, the area between the trajectory of a point's coordinate and its moving average ± its standard deviation. These parameters reflect the smoothness [10].
- Periodicity. It quantifies the number of intersections between a landmark point coordinate and its temporal average. For a given keypoint, the periodicity of the x and y coordinates are summed together [10]. It is an indirect measure of motion complexity.

In this work, we added new parameters that, although they are utilized in clinical practice [7,47], for the best of our knowledge, they have never been all extracted using computer-aided techniques:

- Hands and feet mid-line crossing. It is the number of times that a keypoint (right hand, left hand, right foot and left foot) cross the mid-line of the body. This parameter evaluates the symmetry and complexity of the motion [47,48]. This parameter is also developmentally meaningful, as midline crossing movements become part of the infant motor repertoire by 3mCA, which is reflected in the higher feature relevance observed at this later age. Although the coordinated use of both hands at the midline typically emerges around 4 months of age, infants at 3mCA already show movements increasingly oriented towards the midline and the upper limbs acting in front of the body.
- Hands and feet crossing. These 2 parameters represent the number of times the hands and feet cross each other. They are computed considering the sign of the difference between the x coordinates of the left and right limbs (*e.g.*, $x_{left\,hand} - x_{right\,hand}$). It provides insights into the symmetry and complexity of motor activities [47].
- High frequency head movements. For each time window TW, this parameter indicates the number of times the following condition is verified:

  $$\sqrt{(x_t - x_{t+5})^2 + (y_t - y_{t+5})^2} > \tau$$

  where (x_t, y_t) and (x_{t+5}, y_{t+5}) are the nose coordinates at frame t and frame $t+5$ and $\tau = 20$ pixels. We set τ equal to the 75th quartile of the empirical distribution of the computed distances in the image plane. This parameter gives insights into variability of infants' spontaneous motion patterns. In fact, higher complexity and variability may indicate a broader range of motor activities and exploratory behaviors exhibited by infants [49,50].

In summary, we derived a 156-dimensional feature vector:

$$Ptot_i^n = concat(Pg_i^n, Ps_i^n)$$

with i indicating the ith infant and n spanning the total number of time windows for each ith video.

To ensure consistency across infants' body sizes and camera-to-acquisition plane distances, for each infant, we normalized every spatial parameter with respect to the infant height in the image plane (expressed in pixels). Furthermore, we normalize each parameter in the interval $[0, 1]$ following a minimum–maximum approach.

3.2. Classification

We utilized the computed parameters $Ptot_i^n$ for training and evaluating two binary classifiers aimed at distinguishing between video clips containing the presence or the absence of neuromotor deficits. To select the most appropriate classifier, we started from the considerations highlighted in [34], and we considered: (1) Random forest (RF) and (2) an architecture based on Long Short-Term Memory (LSTM) implemented partially following the consideration done by [44] (two hidden layers with 64 units each and a last dense layer). Although we focused on these two models, we also experimented with other classifiers (*e.g.*, Support Vector Machines and fully connected Neural Networks), which yielded comparable results not reported here.

We computed the accuracy of our analysis on a test set that we did not consider for training and validation. We initially divided our dataset into training and test sets. We considered 20 videos as test set (10 for each recording time) and, notably, we ensured that the same 10 infants (comprising 5 with and 5 without neuromotor deficits) were present in both acquisition time instances. This approach aimed to maintain consistency and integrity in evaluating classifier performance across different acquisition times. The dataset partitioning was performed at the infant level. All time windows belonging to the same infant were assigned entirely to a single set (training, validation, or test) and were never shared across sets. The infants included in the test set were never used during training or validation, ensuring full independence between data subsets.

For model validation, we employed a 5-fold cross-validation technique grouped by infant on the training set. During each fold, the dataset was divided into five subsets, with each subset serving as a validation set once while the remaining four subsets were used for training. This process was repeated five times, ensuring that windows from the same infant never appeared in different folds. Following cross-validation, we selected the hyperparameters that maximized validation accuracy and trained the model on the entire training set. This approach enabled us to obtain a more robust and generalized model configuration. The finalized model was subsequently evaluated on the held-out test set to assess performance in predicting neuromotor deficits across different acquisition times while ensuring robustness and generalization.

During training and validation, a label L was assigned for each nth time window TW corresponding to the label of the ith video indicating the absence ($L_i^n = 0$) or the presence ($L_i^n = 1$) of abnormal motion patterns. In the same way during test, a prediction indicating either the absence or the presence of abnormal motion patterns was generated for each time window ($PredTW_i^n$). Then, to evaluate the accuracy of the pipeline in correctly recognizing infants with neuromotor deficits, for each ith video (infant), we computed the percentage of time windows where abnormal motion patterns were detected (TWAP). The ith infant is classified as positive (with neuromotor deficit) if

$$TWAP_i = (1/N_i) * \sum_{n=1}^{N_i} PredTW_i^n > thr$$

with N_i representing the total number of time windows for the ith video. The threshold thr is a relevant parameter of the method, in Session 4 we discuss the results we obtain for different thr values.

To assess the classifiers' predictive efficacy, we calculated the mean accuracy, sensitivity (*i.e.*, the proportion of correctly identified infants with neuromotor deficits relative to the GT), and specificity (*i.e.*, the proportion of correctly identified infants without neuromotor deficits relative to the GT). In addition, we reported the corresponding 95% Wilson confidence intervals (CI), the full confusion matrix including true positives, true negatives, false positives, and false negatives, as well as the area under the receiver operating characteristic curve (AUC). The classification process was performed separately for the two time points (TEA and 3mCA).

3.3. Assessment of classifier independence from demographic covariates

To assess whether the classifier provides predictive value independently of demographic covariates, we fitted multivariable logistic regression models for each age-specific dataset (TEA and 3mCA). Logistic regression was used to model the binary outcome. For each dataset, two models were estimated:

- Base Model with covariate only

  $$\text{Outcome} = \beta_0 + \beta_1 \cdot \text{GA} + \beta_2 \cdot \text{Weight} + \beta_3 \cdot \text{Sex}$$

- Full Model containing the covariate and the classifier score

  $$\text{Outcome} = \beta_0 + \beta_1 \cdot \text{GA} + \beta_2 \cdot \text{Weight} + \beta_3 \cdot \text{Sex} + \beta_4 \cdot \text{Classifier Score}$$

where the classifier score corresponds to the percentage of windows classified as abnormal. Because two classification models were evaluated (LSTM and RF), the Full Model was fitted separately for each classifier. The classifier's independent contribution was quantified by examining the statistical significance of its coefficient in the Full Model after adjusting for gestational age, birth weight, and sex. Discriminative performance was additionally evaluated by comparing the AUC of the Base and Full Models. Associations between the classifier score and demographic variables were assessed to evaluate potential confounding. Spearman's rank correlation was used for continuous variables (GA, weight), while differences for sex were evaluated using the Mann–Whitney U test. To quantify the strength of this association, the effect size r was computed as $r = \frac{Z}{\sqrt{N}}$ where Z is the standardized test statistic, and N is the sample size. Given the limited sample size of the independent test set, statistical inference and multivariable modeling were not appropriate. Instead, the added predictive value of the classifier on the test set was assessed by fitting logistic regression models on the test data and comparing the AUC with and without the classifier's scores, providing a direct evaluation of predictive improvement.

3.4. Features importance

To discern the most influential features in the classification process and to increase the interpretability of our pipeline, we employed two distinct algorithms: Minimum Redundancy Maximum Relevance (MRMR) [51] and Relief [52]. We opted for these algorithms, neglecting, for instance, the inherent parameter selection process of random forests, because they offer complementary approaches to feature selection, aiding in the identification of discriminative features while mitigating redundancy. The MRMR algorithm [51] highlights features that exhibit high relevance to the target variable (ground truth) while minimizing redundancy among selected features. Features with high relevance and low redundancy are prioritized. In contrast, the Relief algorithm [52] assesses feature importance by estimating the ability of individual features to distinguish between different classes. It accomplishes this by iteratively sampling instances and updating feature weights based on their ability to correctly classify neighboring instances of the same and different classes.

4. Results

4.1. Infants classification

Firstly, the accuracy of various classifiers in distinguishing between infants with and without neuromotor deficits was assessed. As highlighted in Section 3.2, we adopted Random Forests (RF) and a Long Short-Term Memory (LSTM) architecture.

To evaluate the classifiers, our dataset was divided into training and test sets. We treated the two acquisition sessions separately, performing the classification and evaluation steps for TEA and 3mCA independently. The infants used to test the classifiers were randomly selected among those that had the recording at both time points, considering the same 10 infants in both acquisitions. This strategy ensured a robust evaluation across different acquisition session and allowed to highlight changes in the classification accuracy and sensitivity. A 5-fold cross-validation methodology was employed on the training set to evaluate model performance. This iterative process facilitated the identification of optimal hyperparameters, enhancing model robustness and generalization.

Table 3
Validation and test accuracy on the time windows ($PredTW$) of Random Forest (RF) and Long Short-Term Memory (LSTM) for infants at Term of Equivalent Age (TEA) and at 3 months of Corrected Age (3mCA). For the validation accuracy, we report the mean and the standard deviation (in round brackets) obtained from the 5 folds.

	$PredTW$ TEA		$PredTW$ 3mCA	
	Validation	Test	Validation	Test
RF	62.9% (5.5%)	64.5%	65.9% (2.8%)	66.3%
LSTM	64.2% (9.2%)	66.4%	65.2% (6.1%)	67.1%

Table 3 reports the validation and test accuracy obtained when performing classification at the time-window level ($PredTW$) using RF and LSTM models at both TEA and 3mCA. Overall, both classifiers achieved comparable performance across acquisition times, with slightly higher accuracies observed at 3mCA. Validation accuracies were consistent with test results, indicating limited overfitting and good generalization. At TEA, LSTM achieved marginally higher test accuracy than RF (66.4% vs. 64.5%), while at 3mCA both models showed similar performance, with LSTM reaching the highest test accuracy (67.1%). The relatively low standard deviations across the five validation folds suggest stable performance across different training subsets, particularly at 3mCA.

Then, we examined the role of the threshold thr, a critical parameter influencing classification accuracy and clinical relevance. Because the dataset includes infants with minor impairments and since few sign of abnormal movement could be present, we change the threshold thr between 0.2 and 0.5. In fact, abnormal motion patterns may only be visible for a short time less than half of the video. The results obtained are reported in Table 4. As we can notice, at TEA, both models achieved moderate classification performance, with accuracies ranging from 70% to 90% depending on the decision threshold, reflecting the intrinsic difficulty of neuromotor assessment at this early developmental stage. At 3mCA, performance was slightly higher overall, with both classifiers reaching maximum accuracies of 90% and perfect sensitivity (100%) at lower thresholds, suggesting improved discriminative capability as motor patterns mature. As expected, given the limited test sample size, Wilson confidence intervals were relatively wide; nevertheless, the observed trends indicate a modest improvement in performance at 3mCA, further supported by higher AUC values (up to 0.93) compared to TEA.

To further analyze model performance with respect to impairment severity, we examined the clinical profiles of all infants in the test set, which included two infants diagnosed with Cerebral Palsy (CP) and the remaining impaired infants presenting developmental delay (DD). Across all experiments and decision thresholds, both CP infants were consistently and correctly classified at both acquisition time points. All misclassified cases belonged exclusively to the DD subgroup, indicating that classification errors occurred only among infants with milder impairments. This pattern suggests that the proposed models are particularly sensitive to the more pronounced motor abnormalities characteristic of CP, while the subtler and more heterogeneous manifestations associated with developmental delay represent the primary source of residual misclassification.

Table 4
We highlight the test accuracy, sensitivity and specificity and the correspondent 95% Wilson Confidence Interval (CI)on the whole video classification task ($TWAP$) considering different threshold thr values for 10 test videos. We report also the complete confusion matrix with True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) and the value of the area under the receiver operating characteristic curve (AUC). RF: Random Forest; LSTM: Long Short-Term Memory; TEA: Term of Equivalent Age; 3mCA: 3 months of Corrected Age.

		TEA				3mCA			
		$thr = 0.5$	$thr = 0.4$	$thr = 0.3$	$thr = 0.2$	$thr = 0.5$	$thr = 0.4$	$thr = 0.3$	$thr = 0.2$
LSTM	Accuracy	70%	80%	80%	70%	70%	80%	90%	80%
	Wilson CI	[39–89]%	[49–94]%	[49–94]%	[39–89]%	[39–89]%	[49–94]%	[60–98]%	[49–94]%
	Sensitivity	60%	80%	80%	80%	60%	80%	100%	100%
	Wilson CI	[23–88]%	[38–96]%	[38–96]%	[38–96]%	[23–88]%	[38–96]%	[57–100]%	[57–100]%
	Specificity	80%	80%	80%	60%	80%	80%	80%	60%
	Wilson CI	[38–96]%	[38–96]%	[38–96]%	[23–88]%	[38–96]%	[38–96]%	[38–96]%	[23–88]%
	TP	3	4	4	4	3	4	5	5
	FP	1	1	1	2	1	1	1	2
	TN	4	4	4	3	4	4	4	3
	FN	2	1	1	1	2	1	0	0
	AUC	72%	78%	78%	74%	74%	80%	90%	86%
RF	Accuracy	70%	80%	90%	80%	90%	80%	90%	80%
	Wilson CI	[39–89]%	[49–94]%	[60–98]%	[49–94]%	[60–98]%	[49–94]%	[60–98]%	[49–94]%
	Sensitivity	60%	80%	100%	100%	80%	80%	100%	100%
	Wilson CI	[23–88]%	[38–96]%	[57–100]%	[57–100]%	[38–96]%	[38–96]%	[57–100]%	[57–100]%
	Specificity	80%	80%	80%	60%	100%	80%	80%	60%
	Wilson CI	[38–96]%	[38–96]%	[38–96]%	[23–88]%	[57–100]%	[38–96]%	[38–96]%	[23–88]%
	TP	3	4	5	5	4	4	5	5
	FP	1	1	1	2	0	1	1	2
	TN	4	4	4	3	5	4	4	3
	FN	2	1	0	0	1	1	0	0
	AUC	75%	82%	90%	85%	92%	86%	93%	88%

4.2. Assessment of classifier independence from demographic covariates

No significant differences in gestational age, birth weight, or sex were observed between infants with and without neuromotor deficits at TEA. At 3mCA, gestational age differed significantly between groups ($p = 0.048$), while birth weight showed a borderline, non-significant difference ($p = 0.051$); sex remained non-significant at both time points. To more rigorously evaluate the potential influence of demographic characteristics, we conducted multivariable logistic regression analyses on the whole dataset. These models of covariate data, which included gestational age, birth weight, and sex did not identify any demographic variable as a significant predictor of the outcome (all $p > 0.25$). The discriminative performance of the covariate-only model was limited in both age-specific cohorts, with AUCs of approximately 0.6 (0.58 at TEA and 0.6 at 3mCA), indicating that routine demographic variables alone were insufficient to reliably distinguish infants with and without deficits. Adding the classifier score substantially improved model performance in both cohorts. For the LSTM classifier, the score was a significant independent predictor ($p < 0.001$ in both age groups) and the AUC increased to 0.80 at TEA and 0.83 at 3mCA. A similar pattern was observed for the RF classifier; the classifier score remained significant ($p = 0.001$ at 3mCA and $p = 0.0006$ at TEA), and model performance increased to AUC 0.98 in both age groups. Across classifiers and age groups, the classifier scores consistently emerged as the strongest independent predictor. Across cohorts, the classifier score had weak correlations with demographic variables. At 3mCA, no significant associations were observed with birth weight ($\rho = -0.097, p = 0.297$), gestational age ($\rho = -0.017, p = 0.851$), or sex ($\rho = 0.139, p = 0.131$) when considering the LSTM model. For the RF classifier, correlations with birth weight ($\rho = -0.154, p = 0.096$) and sex ($\rho = 0.103, p = 0.265$) were again non-significant, while gestational age showed a small but significant association ($\rho = -0.225, p = 0.014$). At TEA, no associations were observed with the classifier score and any demographic covariate for either model: birth weight ($\rho = -0.046, p = 0.591$ for LSTM; $\rho = -0.074, p = 0.382$ for RF), gestational age ($\rho = -0.050, p = 0.548$ for LSTM; $\rho = -0.104, p = 0.217$ for RF) and sex ($\rho = 0.028, p = 0.74$ for LSTM; $\rho = 0.008, p = 0.925$ for RF). Overall, these findings indicate that the classifier output is not strongly driven by routine demographic characteristics. Finally, we evaluate the predictive performance on the independent test set of ten infants. The demographic-only model achieved an AUC of 0.6. Adding the classifier score improved predictive performance: the LSTM-based model reached an AUC of 0.84 at both ages, while the RF-based model achieved an AUC of 0.92 at 3mCA and 1.00 at TEA. Although statistical inference was not performed due to the small sample size, the consistent improvement in AUC indicates that the classifier contributes meaningful predictive value beyond demographic covariates. Our analysis indicates that the classifiers' output is independent from routine demographic variables, suggesting that it does not merely encode demographic risk factors traditionally associated with neuromotor outcomes. Rather, it captures aspects of spontaneous motor behavior that provide information beyond basic standard clinical variables.

4.3. Features importance

Regarding feature importance, we leveraged MRMR and Relief algorithms to highlight the role of each parameter in the classification process. We summed the weights derived from both methods for each parameter and we normalized all the weights in the interval $[0, 1]$. In Table 5 we report the first 10 most important parameters obtained considering the 95 infants acquired at both time points. As we can notice, the parameters considered in the feature selection procedure are different between *TEA* and *3mCA*. In general, those considered more important in distinguishing between normal and abnormal motion patterns at *3mCA* are related to specific motion parameters (*e.g.*, quantity of motion and mid-line crossing), while at *TEA* are more represented by general motion parameters (*e.g.*, speed and acceleration).

5. Discussion and conclusion

Our study contributes to the growing body of work on the automated analysis of infants' spontaneous movements using markerless video-based techniques. The slightly improved performance observed in the second acquisition session (3mCA) suggests that the proposed pipeline is able to extract features that are increasingly informative for classifying motion patterns associated with neuromotor development.

Table 5
Relevant parameters that guided the classification between infants with and without neuromotor deficits for both acquisition instances (*TEA* and *3mCA*) focusing only on the 95 infants acquired at both time points. Abbreviations: nose (N), right hand (RH), left hand (LH), right foot (RF), left foot (LF). The parameters in italics are those introduced with this work.

TEA	3mCA
Minimum acceleration (LH)	*Mid-line crossing (RF)*
Maximum jerk (N)	*Mid-line crossing (LH)*
Minimum jerk (LF)	Spectral half-point acceleration (RH)
Minimum jerk (LH)	Spectral half-point acceleration (RF)
Minimum jerk (RF)	*Mid-line crossing (RH)*
Minimum speed (LF)	Hands speed cross-correlation
Minimum jerk (N)	Periodicity (N)
Maximum jerk (RH)	*Hands crossing*
Minimum acceleration (LF)	*Feet crossing*
Minimum acceleration (N)	Speed spectral entropy (N)

This trend is consistent with clinical practice, as abnormal movement patterns are generally easier to identify at 3mCA than at TEA, reflecting the progressive organization of motor behavior as infants mature. Comparing our results to the state-of-the-art, we highlight advancements in the classification of neuromotor deficits and in the identification of meaningful parameters, increasing the interpretability of the whole process. Compared to [34], we highlight similar performance in the classification of *TEA* infants, while extending the analysis through a longitudinal design that explicitly evaluates consistency across developmental stages. Notably, our results highlight that parameters related to movement symmetry and complexity (such as midline-crossing features) become increasingly informative at 3mCA, in agreement with established developmental milestones. Furthermore, severity-specific analysis revealed that all Cerebral Palsy cases included in the test set were consistently identified across experiments, with misclassifications limited to infants presenting milder developmental delay, suggesting a higher sensitivity of the proposed approach to more pronounced neuromotor impairments. Together, these findings support the clinical relevance of the proposed pipeline and its potential for early screening and longitudinal monitoring of neuromotor development.

Future research involves further refinement of the classification process, particularly in delineating distinct classes based on the severity and nature of the neuromotor impairments, as well as extending data acquisition towards real at-home or mobile recording scenarios to enhance ecological validity and facilitate broader clinical applicability. Moreover, future studies will overcome the current limitation of having a single-center, single-camera dataset by performing external validation across multiple sites and using different acquisition devices, by including different mobile platforms, to evaluate the robustness and generalizability of the proposed methodology under diverse recording conditions such as varying lighting, clothing, and backgrounds. In addition, expanding the cohort of infants diagnosed with Cerebral Palsy represents an important priority for future data acquisition. Although CP cases were consistently identified in the current study, their limited number restricts the possibility of performing balanced, severity-specific analyses or developing multi-class classification models. Future work will therefore focus on collecting a larger and more clinically diverse sample of CP infants, in collaboration with multiple neonatal follow-up centers, to better capture the heterogeneity of CP phenotypes and motor manifestations. Such an expanded dataset will enable more robust modeling of impairment severity, support the development of multi-class classification frameworks distinguishing CP from developmental delay, and allow for stronger validation of the system's ability to characterize neuromotor outcomes across different levels of severity. In conclusion, our study represents a significant step forward in the automated analysis of infants' spontaneous movements, offering a promising approach for early detection and intervention of neuromotor deficits. Through continued innovation and collaboration, we can strive towards more effective and generalizable tools to help clinicians in assessing and promoting infants' neuromotor development, ultimately improving long-term outcomes for children.

CRediT authorship contribution statement

Matteo Moro: Writing – original draft, Software, Methodology, Formal analysis. **Sofia Sigismondi:** Formal analysis, Software, Writing – original draft. **Laura Asia Gismondi:** Writing – original draft, Formal analysis. **Chiara Tacchino:** Writing – review & editing, Data curation. **Sara Uccella:** Writing – review & editing, Data curation. **Luca Antonio Ramenghi:** Writing – review & editing, Conceptualization. **Paolo Moretti:** Writing – review & editing, Conceptualization. **Francesca Odone:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Maura Casadio:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Institutional review board statement

The study and the consent form signed by parents were approved by the Giannina Gaslini Hospital Institutional Review Board (protocol number: IGGPM01 20/06/2013).

Funding information

This work has been funded by the European Union - NextGenerationEU and by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.5, project "RAISE - Robotics and AI for Socio-economic Empowerment" (ECS00000035).

Declaration of competing interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

This work was funded by the European Union - NextGenerationEU. However, the views and opinions expressed are those of the authors alone and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. The authors thank Pietro Morasso, Psiche Giannoni and Marco Massimo Fato for their support to the study.

References

[1] G. Cioni, F. Ferrari, C. Einspieler, P.B. Paolicelli, T. Barbani, H.F. Prechtl, Comparison between observation of spontaneous movements and neurologic examination in preterm infants, J. Pediatr. 130 (5) (1997) 704–711.
[2] C. Einspieler, P.B. Marschik, H.F. Prechtl, Human motor behavior: Prenatal origin and early postnatal development, Z. Psychologie/J. Psychol. 216 (3) (2008) 147–153.
[3] M. Bax, M. Goldstein, P. Rosenbaum, A. Leviton, N. Paneth, B. Dan, B. Jacobsson, D. Damiano, Proposed definition and classification of cerebral palsy, april 2005, Dev. Med. Child Neurol. 47 (8) (2005) 571–576.
[4] I. Krägeloh-Mann, C. Cans, Cerebral palsy update, Brain Dev. 31 (7) (2009) 537–544.
[5] O. Lincetto, A. Banerjee, World prematurity day: improving survival and quality of life for millions of babies born preterm around the world, 2020.
[6] F.B. Palmer, Strategies for the early diagnosis of cerebral palsy, J. Pediatr. 145 (2) (2004) S8–S11.

[7] I. Novak, C. Morgan, L. Adde, J. Blackman, R.N. Boyd, J. Brunstrom-Hernandez, G. Cioni, D. Damiano, J. Darrah, A.-C. Eliasson, et al., Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment, JAMA Pediatr. 171 (9) (2017) 897–907.
[8] H.F. Prechtl, Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction., Early Hum. Dev. (1990).
[9] F. Ferrari, R. Frassoldati, A. Berardi, F. Di Palma, L. Ori, L. Lucaccioni, N. Bertoncelli, C. Einspieler, The ontogeny of fidgety movements from 4 to 20 weeks post-term age in healthy full-term infants, Early Hum. Dev. 103 (2016) 219–224.
[10] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, C. Disselhorst-Klug, Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy, Hum. Mov. Sci. 25 (2) (2006) 125–144.
[11] S.L. Colyer, M. Evans, D.P. Cosker, A.I. Salo, A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system, Sport. Medicine-Open 4 (1) (2018) 24.
[12] R. Poppe, Vision-based human motion analysis: An overview, Comput. Vis. Image Underst. 108 (1–2) (2007) 4–18.
[13] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A.A. Osman, D. Tzionas, M.J. Black, Expressive body capture: 3d hands, face, and body from a single image, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10975–10985.
[14] M. Moro, G. Marchesi, F. Hesse, F. Odone, M. Casadio, Markerless vs. marker-based gait analysis: A proof of concept study, Sensors 22 (5) (2022) 2011.
[15] L. Garello, M. Moro, C. Tacchino, F. Campone, P. Durand, I. Blanchi, P. Moretti, M. Casadio, F. Odone, A study of at-term and preterm infants' motion based on markerless video analysis, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 1196–1200.
[16] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, F. Pisani, Monitoring infants by automatic video processing: A unified approach to motion analysis, Comput. Biol. Med. 80 (2017) 158–165.
[17] L. Adde, J.L. Helbostad, A.R. Jensenius, G. Taraldsen, R. Støen, Using computer-based video analysis in the study of fidgety movements, Early Hum. Dev. (2009).
[18] L. Adde, J.L. Helbostad, A.R. Jensenius, G. Taraldsen, K.H. Grunewaldt, R. Støen, Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study, Dev. Med. Child Neurol. 52 (8) (2010) 773–778.
[19] C. Tacchino, M. Impagliazzo, E. Maggi, M. Bertamino, I. Blanchi, F. Campone, P. Durand, M. Fato, P. Giannoni, R. Iandolo, et al., Spontaneous movements in the newborns: a tool of quantitative video analysis of preterm babies, Comput. Methods Programs Biomed. 199 (2021) 105838.
[20] T. Tsuji, S. Nakashima, H. Hayashi, Z. Soh, A. Furui, T. Shibanoki, K. Shima, K. Shimatani, Markerless measurement and evaluation of general movements in infants, Sci. Rep. 10 (1) (2020) 1–13.
[21] A. Stahl, C. Schellewald, Ø. Stavdahl, O.M. Aamo, L. Adde, H. Kirkerod, An optical flow-based method to predict infantile cerebral palsy, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (4) (2012) 605–614.
[22] H. Rahmati, O.M. Aamo, Ø. Stavdahl, R. Dragon, L. Adde, Video-based early cerebral palsy prediction using motion segmentation, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 3779–3783.
[23] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, Deep learning-based human pose estimation: A survey, ACM Comput. Surv. 56 (1) (2023) 1–37.
[24] M.T. Irshad, M.A. Nisar, P. Gouverneur, M. Rapp, M. Grzegorzek, Ai approaches towards prechtl's assessment of general movements: A systematic literature review, Sensors 20 (18) (2020) 5321.
[25] S. Reich, D. Zhang, T. Kulvicius, S. Bölte, K. Nielsen-Saines, F.B. Pokorny, R. Peharz, L. Poustka, F. Wörgötter, C. Einspieler, et al., Novel AI driven approach to classify infant motor functions, Sci. Rep. 11 (1) (2021) 1–13.
[26] C. Chambers, N. Seethapathi, R. Saluja, H. Loeb, S.R. Pierce, D.K. Bogen, L. Prosser, M.J. Johnson, K.P. Kording, Computer vision to automatically assess infant neuromotor risk, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (11) (2020) 2431–2442.
[27] Q. Gao, S. Yao, Y. Tian, C. Zhang, T. Zhao, D. Wu, G. Yu, H. Lu, Automating general movements assessment with quantitative deep learning to facilitate early screening of cerebral palsy, Nat. Commun. 14 (1) (2023) 8294.
[28] D. Ledwoń, M. Danch-Wierzchowska, I. Doroniewicz, K. Kieszczyńska, A. Affanasowicz, D. Latos, M. Matyja, A.W. Mitas, A. Myśliwiec, Automated postural asymmetry assessment in infants neurodevelopmental evaluation using novel video-based features, Comput. Methods Programs Biomed. 233 (2023) 107455.
[29] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime multi-person 2d pose estimation using part affinity fields, in: Proceedings of the IEEE CVPR, 2017, pp. 7291–7299.
[30] N. Hesse, S. Pujades, J. Romero, M.J. Black, C. Bodensteiner, M. Arens, U.G. Hofmann, U. Tacke, M. Hadders-Algra, R. Weinberger, et al., Learning an infant body model from RGB-d data for accurate full body motion analysis, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I, Springer, 2018, pp. 792–800.
[31] S. Moccia, L. Migliorelli, V. Carnielli, E. Frontoni, Preterm infants' pose estimation with spatio-temporal features, IEEE Trans. Biomed. Eng. 67 (8) (2019) 2370–2380.
[32] L. Migliorelli, S. Moccia, R. Pietrini, V.P. Carnielli, E. Frontoni, The babypose dataset, Data Brief 33 (2020) 106329.
[33] N. Hesse, C. Bodensteiner, M. Arens, U.G. Hofmann, R. Weinberger, A. Sebastian Schroeder, Computer vision for medical infant motion analysis: State of the art and rgb-d data set, in: Proceedings of the ECCV, 2018.
[34] M. Moro, V.P. Pastore, C. Tacchino, P. Durand, I. Blanchi, P. Moretti, F. Odone, M. Casadio, A markerless pipeline to analyze spontaneous movements of preterm infants, Comput. Methods Programs Biomed. 226 (2022) 107119.
[35] I. Doroniewicz, D.J. Ledwoń, A. Affanasowicz, K. Kieszczyńska, D. Latos, M. Matyja, A.W. Mitas, A. Myśliwiec, Writhing movement detection in newborns on the second and third day of life using pose-based feature machine learning classification, Sensors 20 (21) (2020) 5986.
[36] S. Uccella, A. Parodi, M.G. Calevo, L. Nobili, D. Tortora, M. Severino, C. Andreato, E.-B.N. Group, D. Preiti, C. Traggiai, et al., Influence of isolated low-grade intracranial haemorrhages on the neurodevelopmental outcome of infants born very low birthweight, Dev. Med. & Child Neurol. 65 (10) (2023) 1366–1378.
[37] M. Malova, A. Parodi, M. Severino, D. Tortora, M.G. Calevo, C. Traggiai, P. Massirio, D. Minghetti, S. Uccella, D. Preiti, et al., Neurodevelopmental outcome at 3 years of age in very low birth weight infants according to brain development and lesions, Curr. Pediatr. Rev. 20 (1) (2024) 94–105.
[38] N. Bayley, Bayley Scales of Infant and Toddler Development: Administration Manual, Harcourt assessment, 2006.
[39] D.M.M. Romeo, A. Guzzetta, M. Scoto, M. Cioni, P. Patusi, D. Mazzone, M.G. Romeo, Early neurologic assessment in preterm-infants: integration of traditional neurologic examination and observation of general movements, Eur. J. Paediatr. Neurol. 12 (3) (2008) 183–189.
[40] R. Griffiths, M. Huntley, Griffiths mental development scales-revised: Birth to 2 years, 1996.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.
[43] A. Mathis, P. Mamidanna, K.M. Cury, T. Abe, V.N. Murthy, M.W. Mathis, M. Bethge, DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, Nature Neurosci. 21 (9) (2018) 1281.
[44] D. Ahmedt-Aristizabal, S. Denman, K. Nguyen, S. Sridharan, S. Dionisio, C. Fookes, Understanding patients' behavior: Vision-based analysis of seizure disorders, IEEE J. Biomed. Health Inform. 23 (6) (2019) 2583–2591.
[45] P.R. Butcher, K. Van Braeckel, A. Bouma, C. Einspieler, E.F. Stremmelaar, A.F. Bos, The quality of preterm infants' spontaneous movements: an early indicator of intelligence and behaviour at school age, J. Child Psychol. Psychiatry 50 (8) (2009) 920–930.
[46] D. Zwillinger, S. Kokoska, CRC Standard Probability and Statistics Tables and Formulae, Crc Press, 1999.
[47] C. Einspieler, A.F. Bos, M. Krieber-Tomantschger, E. Alvarado, V.M. Barbosa, N. Bertoncelli, M. Burger, O. Chorna, S. Del Secco, R.-A. DeRegnier, et al., Cerebral palsy: early markers of clinical phenotype and functional outcome, J. Clin. Med. 8 (10) (2019) 1616.
[48] C. Einspieler, P.B. Marschik, J. Pansy, A. Scheuchenegger, M. Krieber, H. Yang, M.K. Kornacka, E. Rowinska, M. Soloveichick, A.F. Bos, The general movement optimality score: a detailed assessment of general movements during preterm and term age, Dev. Med. Child Neurol. 58 (4) (2016) 361–368.
[49] J.L. Bruggink, C. Einspieler, P.R. Butcher, E.F. Stremmelaar, H.F. Prechtl, A.F. Bos, Quantitative aspects of the early motor repertoire in preterm infants: do they predict minor neurological dysfunction at school age? Early Hum. Dev. 85 (1) (2009) 25–36.
[50] J.L. Bruggink, G. Cioni, C. Einspieler, C.G. Maathuis, R. Pascale, A.F. Bos, Early motor repertoire is related to level of self-mobility in children with cerebral palsy at school age, Dev. Med. Child Neurol. 51 (11) (2009) 878–885.
[51] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[52] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of relieff and rrelieff, Mach. Learn. 53 (2003) 23–69.