A pose estimation for motion tracking of infants cerebral palsy

Manal Mostafa Ali[1] · Samer I. Mohamed[2]

Abstract

The General Movements Analysis (GMA) has demonstrated noteworthy promise in the early detection of infantile Cerebral Palsy (CP). However, it is subjective and requires highly trained clinicians, making it costly and time-consuming. Automation of GMA could potentially enhance accessibility and further our comprehension of infants' full-body movements. This paper investigates the feasibility of using 2D and 3D pose estimation strategies to observe and scrutinize the infant's comprehensive body movement attributes to improve our perspective to consider joint movement and positions over time as an alternative to GMA for early CP prediction. The study includes comprehensive movement analysis from video recordings for accurate and efficient analysis of infant movement by computing various metrics such as angle orientations at different predicted joint locations, postural information, postural variability, movement velocity, movement variability, and left–right movement coordination. Along with antigravity movements are assessed and tracked as indicators of CP. We employed a variety Machine Learning (ML) algorithms for CP classification based on a series of robust features that have been developed to enhance the interpretability of the model. The proposed approach is assessed through experimentation using the MINI-RGBD and RVI-38 datasets with a classification accuracy of 92% and 97.37% respectively. These results substantiate the efficacy of employing pose estimation techniques for the precocious prediction of infantile CP, highlighting the importance of monitoring changes in joint angles over time for accurate diagnosis and treatment planning.

Keywords Cerebral Palsy · General Movements Assessment · Pose Estimation · Angle Detection · Pose-Based Features

✉ Manal Mostafa Ali
  manal.mustafa@azhar.edu.eg

  Samer I. Mohamed
  saibrahim@msa.edu.eg

1 Al Azhar University, Computer & System Engineering, Cairo, Egypt

2 October University for Modern Sciences and Arts (MSA), Computer Systems Engineering, Giza, Egypt

 ⓐ Springer

1 Introduction

Cerebral Palsy (CP) denotes a spectrum of enduring neurologic anomalies [1, 2]. These anomalies are conventionally attributed to harm inflicted on the neonatal brain either antecedent to, concurrent with, or in the immediate aftermath of birth due to factors such as infection, cerebrovascular accident, cranial trauma, or cerebral hypoxia [3, 4]. The disorder predominantly compromises an individual's motor function, postural control, and coordination [5]. It may also precipitate an array of ancillary complications encompassing challenges with swallowing, deglutition, delayed crawling, speech, vision, and cognitive impairments [6]. The intensity of these manifestations can exhibit substantial variation, with some individuals encountering only trivial symptoms while others may be profoundly incapacitated [7]. CP affects approximately 2.11 out of every 1000 live births [4, 8]. The risk of developing CP is higher for infants born prematurely [9] i.e., a delivery that occurs prior to the completion of thirty-seven weeks of gestation [10].

General movement assessment (GMA) has been widely investigated to detect motor dysfunction, especially CP [11–14]. Prechtl [15] discovered that the state of an infant's nervous system can be accurately determined by observing the quality of their spontaneous movements [2, 5], particularly their general movements (GMs), at the corrected age of 3–5 months [16]. GMs are spontaneous body movements that constitute a unique pattern of movements, termed Fidgety Movements (FMs) [6, 7]. These movements can be discerned from early intrauterine life until approximately 20 weeks post-term [17]. They provide valuable insights into the evolution and operation of the infant's nervous system [18]. As presented by [16], 95% of CP infants did not have FM. Physiotherapists or doctors observe the spontaneous movements of an infant along with their family medical history to diagnose movement disorders [17]. However, the subjective nature of these clinical tests, which rely on gestalt visual perception of movement, may affect their accuracy and reliability [9] and lacking in discernible quantitative diagnostic features [3, 11]. Additionally, they rely heavily on the infant being in an appropriate behavioral condition [13], rendering them a time-intensive process [19] and potentially leading to observer fatigue [6]. Therefore, there is a need for an automatic system to objectively analyze the movements and standardize criteria for assessing infant movement. Several researchers have favored marker-based techniques [20, 21] which, while precise, necessitate a specific arrangement that may not be feasible [9]. In the last decade, computer vision techniques have revolutionized the medical sector and achieved highly encouraging outcomes in the early identification of various diseases. Due to the progress made in deep learning, there are now many dependable neural models that can be used for automatic estimations of human poses in both 2D and 3D from 2D images [9, 19, 22]. Furthermore, the development of markerless motion capture (MoCap) has been facilitated by the widespread utilization of traditional video recordings for documenting infant movements. This technology allows for the non-invasive collection of movement data and offers a more affordable option compared to sensor-based MoCap. Its application is feasible in both clinical environments and home settings [7]. This research aims to utilize video recordings to recognize the movement patterns of infants lying on their backs by obtaining the 2D and 3D coordinates using different pose estimation algorithms. Then, analyzes the infant's whole-body characteristics for early prediction of CP. The subsequent sections of this study are organized as follows: Sect. 2 explores the related works. The proposed methodology is introduced in Sect. 3. In this section, we explain 2D and 3D pose estimation, preprocessing, movement feature extraction, and the underlying benchmark datasets. Section 4 discusses the experimental evaluation for predicting infant CP, along with the analysis and interpretation of the results. Finally, Sect. 5 delivers concluding remarks and proposes potential avenues for further studies.

2 Related work

Over the past two decades, there has been considerable advancement in the development of sensor-based metrics for infant motion. These metrics quantify the characteristics of movement, providing a more objective and accurate assessment of infant movement [12]. Such techniques include movement tracking using markers or wearable motion sensors [20, 21] situated on the toddler's body parts. Nonetheless, having a multitude of sensors or markers can be a challenge that requires calibration and re-positioning [13] and can be uncomfortable for infants which may hinder the infant's spontaneous movements and affect the accuracy of the tests [3, 10, 23]. Other techniques, such as those using the Microsoft Kinect [5], have been developed to investigate movement characteristics of the infant body. These methods suffer from many limitations such as the subject being standing and being larger than one meter for body tracking, which renders them inappropriate for infants [11, 13].

To attenuate this problem, significant work has been put into creating non-invasive monitoring solutions, such as using video information from RGB or RGB-D cameras. Wang et al. [22] propose a novel multi-task Convolutional Neural Networks (CNN) framework for simultaneously estimating multiple infant body joints based on RGB-depth images and evaluate their method on two publicly available infant pose estimation datasets namely Infant Body Pose Dataset (IBPD) and Infant Multiple View Dataset (IMV). Groos et al. [7] focus on the creation and verification of a deep learning technique to forecast CP from spontaneous movements in high-risk infants aged between 9 to 18 weeks corrected age. Khan et al. [3] propos a deformable part-based model (DPM) to identify movement abnormalities in infants between 2 to 6 months old, using video data and a process called skeletonization. They evaluate their approach on a dataset of infants with various movement disorders, including CP, spina bifida, and microcephaly.

As deep learning techniques continue to advance, numerous studies have been developed that can precisely determine human poses and postures from 2D images [9, 14, 19, 22]. These frameworks use advanced ML algorithms to analyze images and extract information about the position and movement of the human body. By analyzing posture and orientation of infants, trained pediatricians can use this information to early predict and diagnose CP. McCay et al. [8] establish histogram-oriented pose characteristics, such as Histogram of Joint Orientation 2D (HOJO2D) and Histogram of Joint Displacement 2D (HOJD2D) for early detection of infant CP. The authors also propose the challenging RVI-38 dataset which compromises a collection of videos recorded during regular clinical procedures. In the context of this research, we have employed the RVI-38 dataset which was obtained upon request from the corresponding authors. Ningrum et al. [18] describe the development of a classifier that utilizes OpenPose and Deep Neural Networks (DNN), including a Long Short Term Memory (LSTM), ID CNN integrated with an LSTM, and Gated Recurrent Unit (GRU). The classifier is designed to estimate the probability of infant CP, given the unavailability of data and imbalances in categories. ID CNN – LSTM records the best accuracy of 96%. In another work, [4] combine different deep learning techniques with 2D pose estimation methods to categorize infant movements. This is used for independent tracking of neurological development when the infant interacts with a toy. The study demonstrates that the Bi-LSTM-CNN model outperforms others in positional experiments, achieving an F_1 score of 84.5%. Chambers et al. [12] create a standard reference dataset for infantile motion by analyzing 85 online video clips. The motions of various

body parts were then quantified using the OpenPose estimator algorithm. Ni et al. [2] present SiamParseNet, a semi-supervised system designed for parsing infant bodies, and demonstrate its effectiveness in estimating infant poses and superior performance in predicting GMA and risky CP. Groos et al. [19] train four different types of CNN and evaluate on a novel dataset of infant poses sourced from a clinical international community to precisely localize segments and joints in the movements of infants aged 2 to 5 months post-term.

To accurately assess infant movements and perform a comprehensive analysis of their movements in space, it is essential to perform 3D pose estimation of infants. Soualmi et al. [9] present 3D infants pose estimation method which uses stereoscopic images, specifically tailored for premature babies in the Neonatal Unit. The researchers used a substantial dataset of over 88k images, which were gathered from 175 videos featuring 53 premature infants. In their work, Hesse and colleagues [11] devise a technique for predicting the 3D locations of infant joints and develop a classifier that identifies body parts by utilizing random ferns and achieve an average distance error of 41 mm. The authors incorporate a feature selection process and a scheme for reweighting the kinematic chain and also ensure rotation invariance by applying Principal Component Analysis (PCA) to the input image. In 2018 [24], Hesse et al. developed the MINI-RGBD dataset, which comprises 12 sequences that capture authentic infant movements. It showcases a diverse range of textures, shapes, and backgrounds, providing a realistic representation of infant movements. Li et al. [13] introduce a 3D pose estimation method that is cost-effective in terms of training and is appropriate for infants in reclining positions. This method employs an established 2D human body keypoints detection technique, which is integrated with RGB-D data gathered from a Kinect sensor. The effectiveness of this approach is assessed employing the MINI-RGBD sequences with an average error of 13.76 mm in the estimated length of body parts.

In this context, we extract new GMA-relevant features using a variety of 2D and 3D pose-estimation algorithms and calculate the movements of the body parts generated from standard 2D RGB video for early prediction of CP. The proposed approach calculates the angular orientation at the identified body part locations and encodes this data in a manner that can precisely depict their movement. During the processing stage, a graph is created that represents the detected joints and their respective angles, which change based on the viewing angle. By integrating the input with this graph, we can extract important information about the posture and alignment of an infant.

The significant contributions made by this study are outlined as follows:

- Employing 2D and 3D pose estimation techniques for tracking and analyzing the infant's whole-body movement characteristics to revolutionize our comprehension of joint movements and positions over time which serves as an alternative to GMA for early prediction of CP.
- Performing comprehensive movement analysis in videos for accurate and efficient analysis of infant movement by computing various metrics such as angle orientations at different predicted joint locations, postural information, variability, velocity, acceleration, left–right movement coordination, antigravity movements, and postural patterns.
- A series of robust features have been developed to enhance the interpretability of the model used for CP prediction. It does this by visually emphasizing the contribution of each body segment in the video. This allows for a clearer understanding of how the model makes its predictions and finally the outcomes are validated across two different datasets.

3 Methodology

In this section, we delve into the primary procedures of the proposed methodology. Primarily, pose estimation using different techniques is explained. The joint angles computation and tracking were clearly explained. Pre-processing, movement characteristics, and classification techniques were elaborated. Deceptively, the seemingly simple movements of humans, such as moving an arm or lifting a leg, require complex coordination between the brain and the musculoskeletal system. Any disruption in this system can lead to involuntary movements, challenges in executing desired movements, or a combination of both. To study this, we extract a set of features to capture a holistic view of infant movement and posture which include:

- Postural information: This includes the position and angle of various body parts, providing a snapshot of the body's overall posture at any given moment.
- Postural variability: This measures the variability of both position and angle, capturing the range and diversity of movements.
- Movement velocity: The absolute velocity of movement provides insight into the speed and fluidity of the subject's actions.
- Anti-gravity movements: These are movements of the legs and arms that work against gravity, providing important information about strength and coordination.
- Left–Right symmetry of movement: By cross-correlating the positions and angles of movements on the left and right sides of the infant body, we can assess the symmetry of movement, which is a key aspect of coordinated motor function.

This method begins by pinpointing various body parts according to the location of their joints. Following this, it computes the angles at predicted joints, such as the elbow and knee. By monitoring these angles over time, the method can effectively characterize the motion and detect any potential anomalies. Our predictions of infants with CP were based on kinematic variables. The time-series data extracted from body-landmark positions served as the source for computing these kinematic features. Subsequently, the method encodes the movements at various joints by monitoring the angle orientations over time and determining the degree of deviation of each infant from the standard. The classification performance of the proposed framework was assessed employing both the RVI-38 [8] and the publicly available MINI-RGBD [24] datasets. This adjusted data is then utilized to create features grounded in GMA for further examination. The extracted features are subsequently inputted into a classification framework for evaluation. This comprehensive approach allows for a detailed and nuanced understanding of infant motor function. A block diagram that illustrates the primary stages involved in this study is demonstrated in Fig. 1. The detailed specifics of each step are elaborated in the following subsections.

3.1 Pose estimation

Estimating the pose of infants in a supine position is a vital step in the creation of an automated system for GMA [13]. Utilizing markerless vision-based data is a more appealing approach for examining infant movement compared to the use of wearable sensors. Features reflecting posture, kinematic variables, acceleration, and symmetry (left–right cross-correlation) are extracted from videos using pose estimation. While infant movement data is recorded as RGB
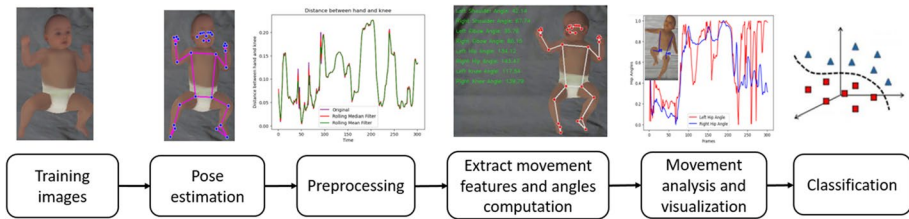
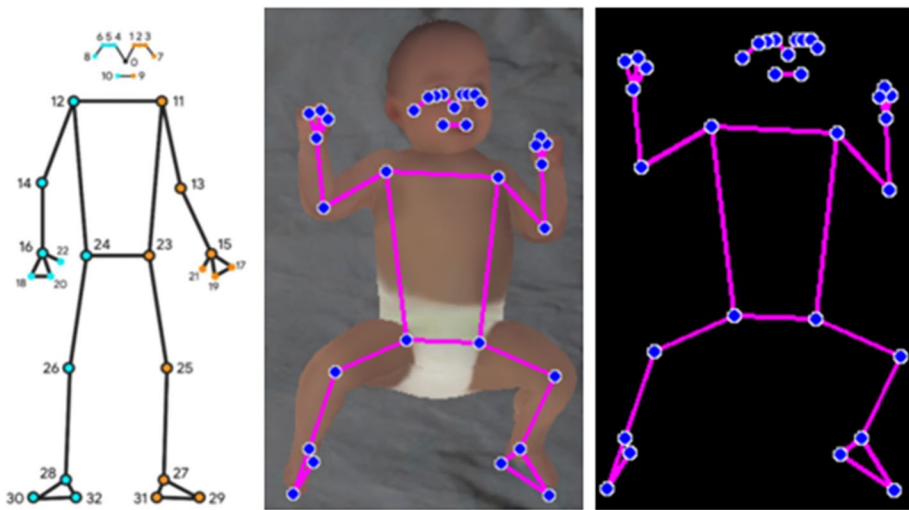**Fig. 1** Schematic of the proposed methodology

Fig. 2 The estimated poses using MediaPipe

video, direct video analysis can be difficult due to intra-class variations including changes in lighting, background, and the infant's appearance, among other factors. Consequently, each frame is treated as an individual image that undergoes processing for the estimation of posture and detection of joint angles by exploring three different techniques namely MediaPipe,[1] OpenPose [25], and MeTRAbs [26]. These approaches continue to be highly ranked in numerous human pose estimation competitions. They employ parallel networks with varying resolutions, as opposed to the conventional high-to-low networks in series.

MediaPipe Google's MediaPipe is a powerful tool that can accurately track various movements, including the dexterous and positional movements of infants. MediaPipe Pose, based on the BlazePose CNN, surpasses current leading approaches and delivers real-time performance on the majority of modern devices. It is a bottom-up multi-person pose estimation that employs a two-step detection-tracking ML approach. Initially, it identifies the region of interest (ROI) within the frame. Following this, it forecasts the poses and the segmentation mask within the identified ROI [29] as illustrated in Fig. 2.
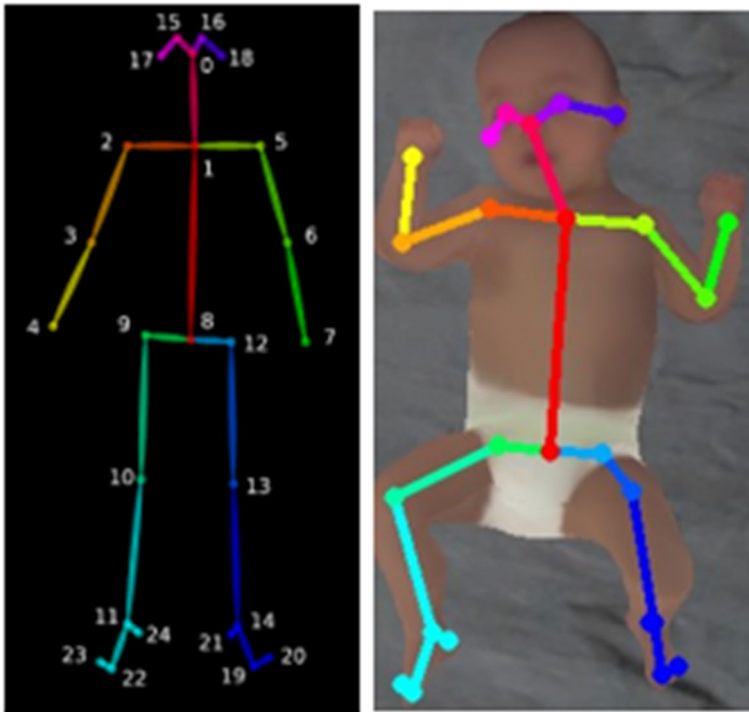
---

[1] https://pypi.org/project/mediapipe/

Fig. 3 Output skeleton extracted by OpenPose [25]

OpenPose [25] Is a leading method for identifying human body parts and their appearance in images. It operates on the principle of Part Affinity Fields (PAFs) and is known as a 'bottom-up' approach. This means it first identifies lower-level features, which are then used to reconstruct the higher-level skeletal posture. The OpenPose framework is utilized to extract the positions of joints from 2D RGB video data which results in a skeletal pose representation that comprises 25 predefined joints, along with a confidence score associated with the prediction. This approach provides a comprehensive and reliable method for human pose estimation. Figure 3 displays an example of the key joints that have been extracted using OpenPose.

MeTRAbs To accurately assess infant movements and conduct a thorough spatial analysis, it is crucial to perform 3D pose estimation for infants [9]. MeTRAbs [26] is a technique that estimates 3D human postures using volumetric heatmaps that are metric-scale truncation-robust (MeTRo). These heatmaps are defined in metric 3D space, not in alignment with the image space. This method overcomes the constraints of 2.5D volumetric heatmaps, which are unable to pinpoint body joints beyond the image borders and necessitate an additional post-processing phase to eliminate scale uncertainty. MeTRAbs can detect joints defined by the SMPL (Skinned Multi-Person Linear model), H36M (Human3.6M), or COCO (Common Objects in Context) skeleton conventions. The SMPL skeleton convention defines 24 joints, while the H36M skeleton convention defines 32 joints, both including the head, neck, torso, hips, shoulders, elbows, wrists, knees, and ankles. The COCO skeleton convention defines 17 keypoints for the human body. Figure 4 shows the results obtained after exploring MeTRAbs as a 2D and 3D pose estimator.
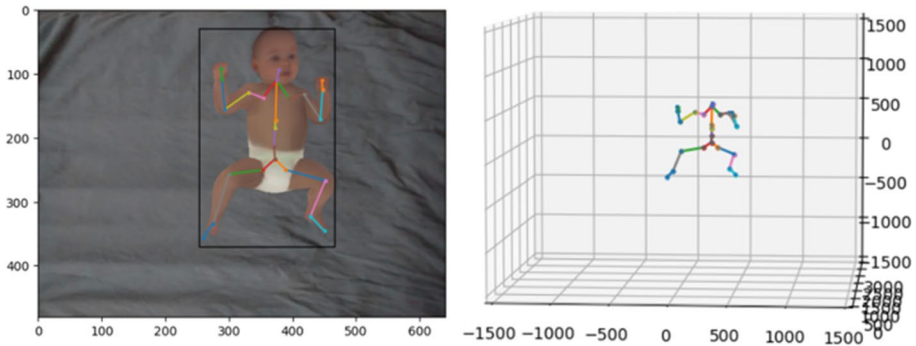
Fig. 4 Example of 2D and 3D keypoints extraction using MeTRAbs [26]

3.2 Data preprocessing and correction

Pose estimation data can often contain missing values, false positive detections, break-points in the time series, and irregular serrations. To obtain clean signals, it is necessary to preprocess the time series data extracted by different pose estimators. One approach is to initially apply linear interpolation [27] to the raw time series to remove missing data. Subsequently, outliers can be discarded using a rolling-median filter with a 1-s smoothing window [28]. In the final act of refining the data, a rolling-mean filter was investigated across the data points with a 1-s smoothing window, smoothing out the rough edges and bringing harmony to the signals. Post interpolation and filtering, the high-frequency noise diminishes, resulting in smoother data. The disparities in the keypoint coordinate pre and post-the automatic interpolation and smoothing correction procedures are illustrated in Fig. 5.

Finally, the confidence threshold t_i is derived from the predicted confidence scores by computing the average of these scores and subtracting 5% per joint as inspired by [8]. All joints that did not surpass the confidence threshold were ignored. As such, t_i value for joint i can be mathematically given by:

$$t_i = (\frac{1}{n}\sum_{j=1}^{n} c_{i,j}) * 95\% \tag{1}$$

where n represents the total count of frames, $c_{i,j}$ is the confidence score of joint i at frame j predicted by the estimator. The preprocessed signals are then used for determining kinematic variables, joint angles calculation, and motion tracking.

3.3 Movement analysis and angles calculation

To provide a clear and concise explanation of the infant movement, the measurements of crucial angles were performed. Some of the more informative angles can be computed using the coordinates of the landmarks detected by different pose estimation techniques. These angles can provide valuable information about the movement and posture of different body parts of infants. By analyzing these angles over time, it may be possible to identify patterns or abnormalities in movement that could be indicative of underlying health conditions or injuries. A basic collection of characteristics, derived from the movements of the arms, hips, and legs, have been calculated using the positions of the shoulder, elbow, hip, knee, and ankle.
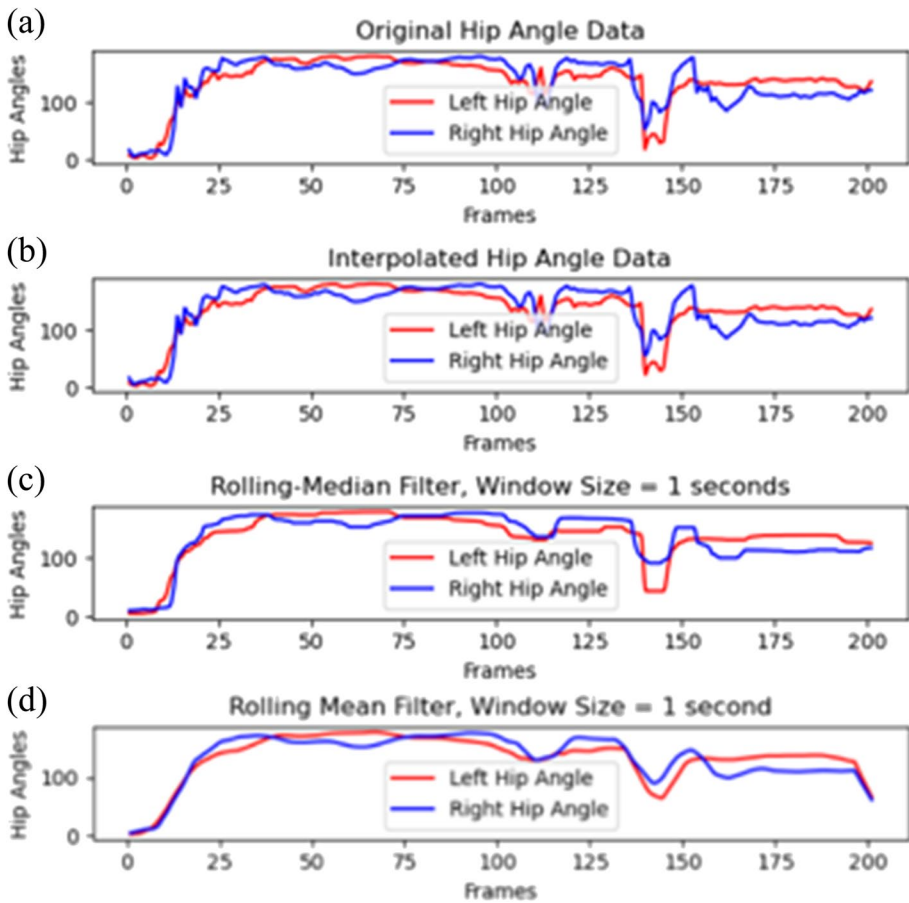
(a)

(b)

(c)

(d)

Fig. 5 Keypoints coordinates of **a** Original data **b** Linear interpolation, **c** Rolling median filter **d** and Rolling mean filter

3.3.1 Angles calculation

Each angle's computation is based on three previously identified keypoints. The central keypoint is employed to estimate the angle under observation, while the other two keypoints, which are the nearest to the observed angle, aid in this calculation.

Shoulder angle The angle formed by the upper arm and the torso, with the shoulder joint acting as the pivot point is commonly known as the shoulder angle, also named the torso-upper arm angle. This angle is formed by three points: the shoulder, hip, and elbow. So, θ_{Shld} can be computed as the dot product of the upper arm vector $(\overrightarrow{V_1})$ and the torso vector $(\overrightarrow{V_2})$, where $\overrightarrow{V_1}$ begins at the shoulder and ends at the elbow, and the $\overrightarrow{V_2}$ begins at the shoulder and ends at the hip as demonstrated by the subsequent equations:

$$\theta_{Shld} = arccos\left(\frac{\overrightarrow{V_1}.\overrightarrow{V_2}}{|\overrightarrow{V_1}||\overrightarrow{V_2}|}\right) \quad (2)$$

Where

$$\vec{V_1} = Shoulder\ coordinate\ (x, y) - Hip\ coordinate\ (x, y)$$

$$\vec{V_2} = Elbow\ coordinate\ (x, y) - Shoulder\ coordinate\ (x, y)$$

Radians to degrees conversion can be accomplished using the following formula:

$$\theta_{Shld\ (deg)} = \left| \frac{\theta_{Shld(rad)} * 180}{\pi} \right| \tag{3}$$

Similarly, the angles for other joints including arms, hips, and knees can be calculated and monitored over time in a video sequence to characterize the corresponding movements of infants.

Elbow angle The elbow angle was calculated as the angle between the upper arm and forearm vectors with the elbow joint acting as the pivot point. This angle is formed by three points: the shoulder, elbow, and wrist [29]. To calculate the angle, we first calculated the vectors representing the upper arm and forearm by subtracting the coordinates of the elbow from the coordinates of the shoulder and wrist, respectively. Then, the cosine of the angle between these two vectors was computed using the dot product.

Hip angles The hip angle is computed as the angle formed between the vectors of the thigh and the torso. This angle is shaped by three points: the knee, hip, and shoulder. The thigh vector is outlined by the line joining the knee and hip, whereas the vector of the torso is outlined by the line joining the hip and shoulder.

Knee angles The knee angle is quantified as the angle intersecting the vectors of the thigh and the lower leg. This angle is constituted by a triad of anatomical landmarks of the hip, knee, and ankle. Definitions of these angles are summarized and visualized in Table 1.
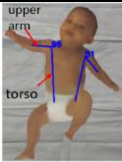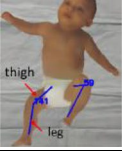
Figure 6 illustrates the estimated angle orientations, derived using the proposed algorithm, on a representative image from the test dataset as well as tracing of these angles at (a) shoulder, (b) elbow, (c) hip, and (d) knee in the subsequent frames. For visualization purposes, only a selection of 200 subsequent frames is displayed to prevent a cluttered view. However, the entirety of the videos is utilized during experimentation and evaluation.

These angles can offer valuable insights into the range of motion for the shoulder, elbow, hip, and knee landmarks. The proposed model detects and tracks the joints and angles which are subsequently utilized to create a visual depiction of the infant's movement as depicted in Fig. 7. It illustrates how the proposed system captures movement information like joint angles accurately where the orientations of the angles at predetermined joints are marked in the top-left corner of each frame for example from the MINI-RGBD dataset [24]. Results are plotted for only the left and right knee angles for the first subsequent 200 frames for simplicity.

3.3.2 Velocity and acceleration

To estimate quantities such as velocities and accelerations, it is necessary to look at multiple frames rather than just a single frame. The velocity and acceleration can be estimated as the first- and second-order differences between successive frames.

Table 1 Some important angle definitions

| Angle	Joints	Definition	Overlaid frame
Shoulder angle	θ_{RShld}: right elbow, right shoulder, and right hip landmarks	represents the angle between the upper arm and torso vectors.	
	θ_{LShld}: left elbow, left shoulder, and left hip landmarks		
Elbow angle	θ_{RElbw}: right wrist, right elbow, and right shoulder landmarks	represents the angle between the upper arm and forearm vectors.	
	θ_{LElbw}: left wrist, left elbow, and left shoulder landmarks		
Hip angle	θ_{RHip}: right shoulder, right hip, and right knee landmarks	represents the angle between the torso and thigh vectors.	
	θ_{LHip}: left shoulder, left hip, and left knee landmarks		
Knee angle	θ_{RKnee}: right hip, right knee, and right ankle landmarks	represents the angle between the thigh and lower leg vectors.	
	θ_{LKnee}: left hip, left knee, and left ankle landmarks		
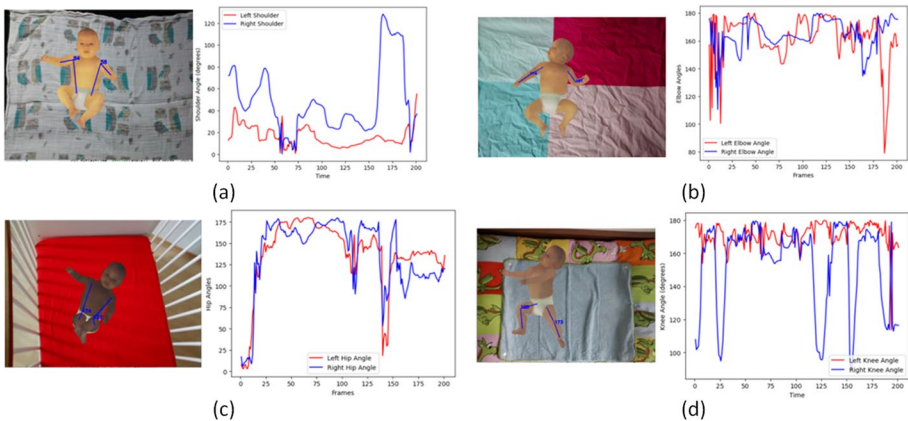
Fig. 6 Visualization and tracking of the angles at **a** Shoulder, **b** Elbow, **c** Hip, and **d** Knee

Velocity To describe the motion of the infants we calculate the velocity of their left and right hands. The velocity is calculated by evaluating the variation in the x and y coordinates of the wrist landmarks between two consecutive frames, squaring them, adding them, and then taking the square root. This results in the Euclidean distance between the
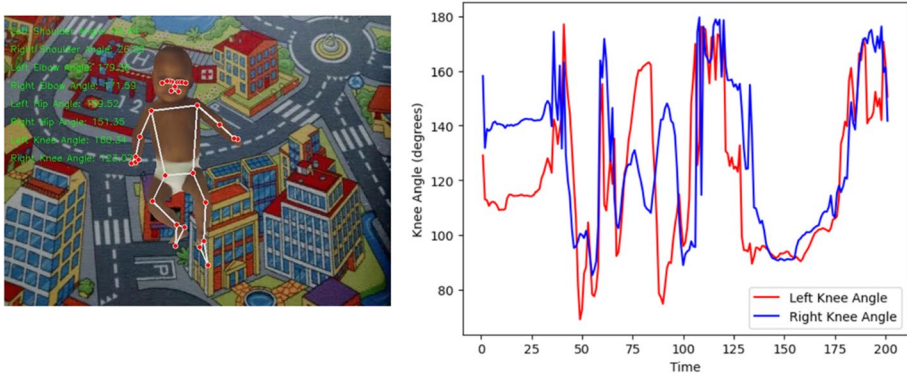
Fig. 7 Annotation of the infant movement and tracking of left and right knee angles
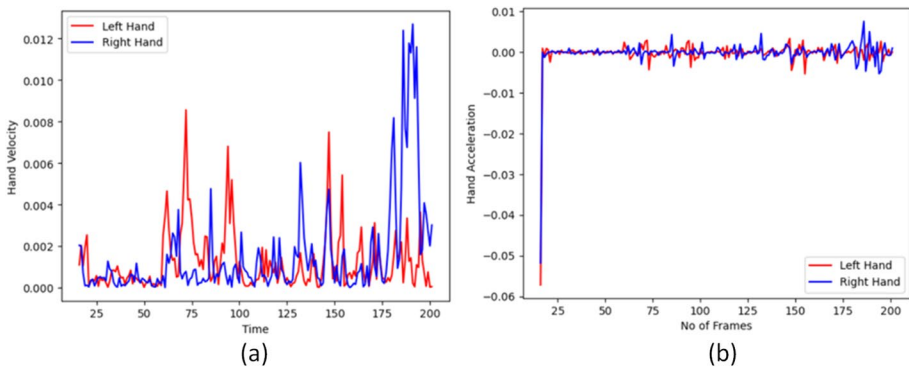

Fig. 8 **a** Velocities and **b** accelerations of the left and right hands

wrist positions in two successive frames, serving as an approximation of the velocity of the hands. The velocity of the left hand at time t can be calculated as follows:

$$v_{left}(t) = \sqrt{(x_{left}(t) - x_{left}(t-1))^2 + (y_{left}(t) - y_{left}(t-1))^2} \tag{4}$$

where $x_{left}(t)$ and $y_{left}(t)$ are the x and y coordinates of the left wrist at time t, respectively. Similarly, the velocity of the right hand at time t is calculated as follows:

$$v_{right}(t) = \sqrt{(x_{right}(t) - x_{right}(t-1))^2 + (y_{right}(t) - y_{right}(t-1))^2} \tag{5}$$

where $x_{right}(t)$ and $y_{right}(t)$ are the x and y coordinates of the right wrist at time t, respectively. This method allows us to capture the rate of change in position over time. Figure 8 (a) draws velocities of both hands against time.

Acceleration The left and right-hand acceleration can be computed by taking the difference between their velocities in two consecutive frames. The left-hand acceleration at time t can be calculated as follows:
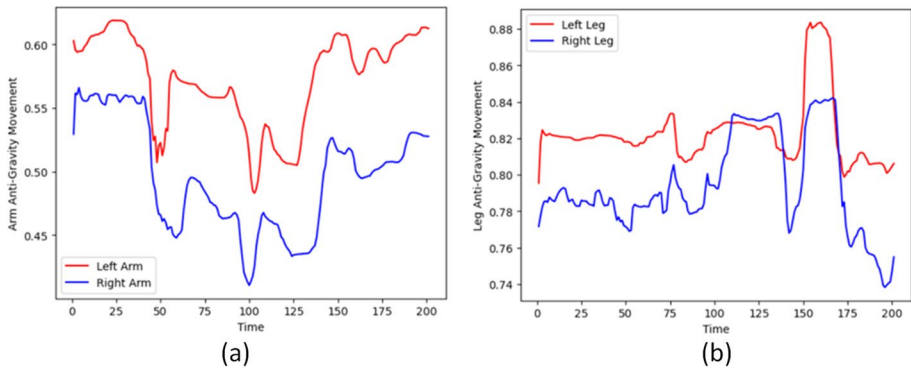
Fig. 9 Anti-gravity movements for both **a** arms and **b** legs

$$a_{left}(t) = v_{left}(t) - v_{left}(t-1) \tag{6}$$

where $v_{left}(t)$ is the velocity of the left hand at time t. Similarly, the right-hand acceleration at time t is calculated as follows:

$$a_{right}(t) = v_{right}(t) - v_{right}(t-1) \tag{7}$$

where $v_{right}(t)$ is the velocity of the right hand at time t. Examples of predicted hands accelerations were elaborated in Fig. 8 (b).

3.3.3 Anti-gravity movements

It represents the movements that work against gravity, such as lifting a limb or sitting up. These movements require the use of muscles to generate force and overcome the pull of gravity. For example, when you lift your leg or arm, you are performing an anti-gravity movement. To examine if the infant was performing these movements, anti-gravity with arms and legs were computed. Figures 9 (a) and (b) track the y-coordinate of the wrists and ankles over time and use it as measures of the arm's and leg's anti-gravity movement. The y-coordinate represents the vertical position of the wrists and ankles, so a decrease in their values indicates an anti-gravity movement (i.e., lifting the arm or leg).

3.3.4 Additional features

The proposed method assesses motor milestones by tracking the spatial position of different body landmarks. Monitoring the movements of arms and legs is especially important when assessing the cognitive and motor development of infants [10]. Other features describing symmetry between the left and right sides are identified and computed. To further investigate the development CP, we measured several parameters that are commonly used as indicators for CP markers, as presented by [16]. For instance, the method can determine if an infant is bringing its hands together, reaching for its feet or knees, and more. Figure 10 (a) measures the distance between hand and knee.
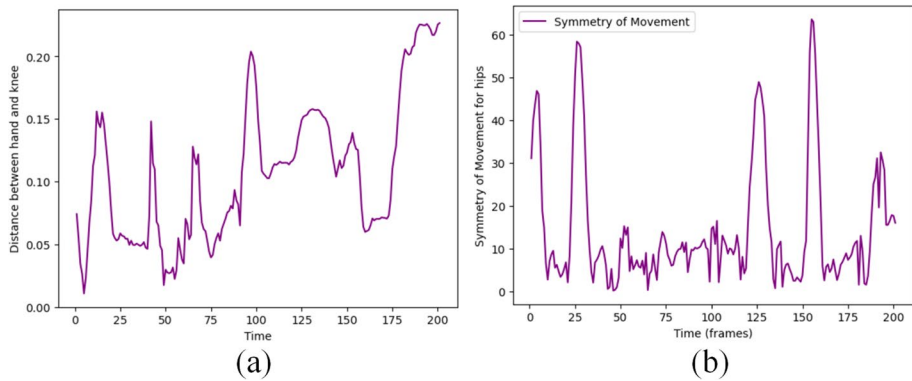
Fig. 10 **a** Distance between hand and knee **b** Symmetry of movement for hips

Symmetry of movement The correlation between the movements of opposite limbs can be used to measure the synchronization of motion on both sides for infants. This parameter takes into account the magnitude and direction of limb movements and calculates the degree of correlation between the two sides. This approach can be applied to various limbs to analyze the symmetry of movement throughout the body. Figure 10 (b) assesses the symmetry of movement for both sides using left and right hips and can be expressed mathematically as follows:

$$Symmetry(t) = LeftHipAngle(t) - RightHipAngle(t) \tag{8}$$

In this formula, Symmetry(t) represents the symmetry of movement at time t, *leftHipAngle*(t) represents the angle of the left hip joint at time t, and *RightHipAngle*(t) represents the angle of the right hip joint at time t. The absolute value function $|x|$ calculates the absolute variance between the hip angles, giving us a measure of how symmetric or asymmetric the movement is at that time point. By applying this formula at each time point, we can generate a time series of symmetry values that can be plotted and analyzed to gain insights into the infant's movement patterns.

The quantities mentioned are just a subset of the possible quantities that can be quantitatively computed from the extracted poses. The model has the ability to calculate several metrics, such as the distances from the neck to each wrist, and the distances from the left or right hip joint to the corresponding ankle, as well as other movements such as joining the hands, foot to foot contact, reaching for the feet, and rolling to the side. Moreover, a flowchart that illustrates CP prediction using the extracted feature set and pose estimation techniques is illustrated in Fig. 11. The process begins with reading frames, and then estimating the pose from these frames. Any noise in the pose estimation is smoothed out and corrections are made in the next step. A confidence score is then computed and checked against a threshold. If the score is greater than or equal to the threshold, the process moves to computing angles and extracting the underlying features. If the score is less than the threshold, the joint is simply ignored. The extracted features are then classified using various ML models and the results are evaluated and compared.
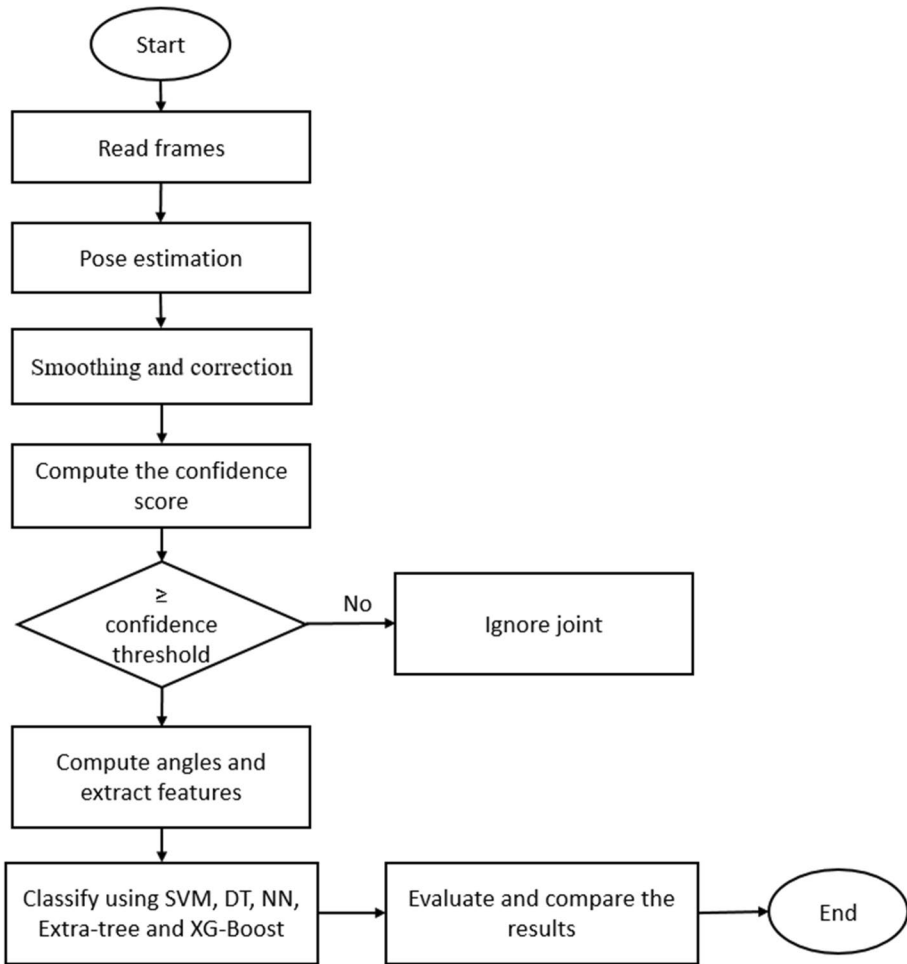
Fig. 11 Flowchart of infant CP prediction

3.4 Dataset

The MINI-RGBD[2] dataset [24] Is a synthetic dataset that was designed to map real-world 3D movements of infant, which were captured in a clinical environment, into virtual 3D infant models. This dataset was created and made publicly available for research purposes. Its primary goal is to facilitate the study and understanding of infant movements in a controlled and replicable virtual environment. This dataset includes images of infants up to 7 months old lying on their backs and facing the camera. The images were produced using the Skinned Multi-Infant Linear (SMIL) body model. This system is designed to generate lifelike sequences of infant bodies using both RGB and depth images, and it can accurately provide 2D and 3D keypoints positions [9]. The MINI-RGBD dataset contains 12

---

[2] http://s.fhg.de/mini-rgbd

sequences of continuous motion, each with a resolution of 640 X 480 at 25 FPS and a length of 1000 frames. Out of the total videos, 8 are labeled as FM + and 4 as FM − [8]. These sequences are classified into various difficulty levels depending on the infants' movements, ranging from lying flat with minimal limb movement, slight turning with limb interaction, and more complex movements such as turning to the side and grabbing their legs. To maintain the confidentiality of infants in recorded data, their actual forms and textures are not used. Instead, by averaging the data from several infants, new shapes and textures are created [5].

RVI-38 dataset [8] The assortment consists of 38 video fragments, each presenting a distinct infant aged from 2 to 5 months post-term. These footages were obtained during regular clinical check-ups using a handheld Sony DSC-RX100 Advanced Compact Premium Camera, which provides a 1920 X 1080 resolution at a speed of 25 FPS. The length of each video was varied, with the shortest being 40 s and the longest being 5 min. On average, the videos were approximately 3 min and 36 s long. Two skilled evaluators sorted the videos into two groups using the GMA; FM + and FM-. Out of the 38 videos, 32 were marked as FM + and 6 were identified as FM-. The dataset, which posed a significant challenge, was instrumental in determining the resilience of various motion features for classification, thereby contributing to the conceptualization of a feature fusion framework based on the GMA. The footage in these datasets was captured while the infants were at rest and not engaged with any toys or objects, mirroring the environmental conditions of the GMA test. It is worth mentioning that RVI-38 dataset was available only in JSON (Java Script Object Notation) format and the original video files were not accessible for this study.

3.5 Binary classification for CP prediction

Once the features are extracted, they are inputted into conventional ML classifiers. This process allows for a comprehensive analysis of movement patterns and the subsequent prediction of CP, based on the annotations given by the GMA evaluators. The classification framework employs a variety of ML algorithms, such as Support Vector Machine (SVM), Neural Networks (NN), Decision Tree (DT), Extra-Tree and eXtreme Gradient Boosting (XGBoost). A radial basis function (RBF) was utilized as the kernel for SVM, which is a common choice for its effectiveness in handling non-linearly separable data. The regularization parameter (C) was tuned within the range of 2^{-5} to 2^5. For the NN classifier, we employed a multi-layer perceptron (MLP) architecture with a hidden layer size of 50 neurons. The activation function used in the hidden layer was the Rectified Linear Unit (ReLU), known for its efficiency in handling non-linear relationships. The number of epochs was set to 100, and the learning rate was optimized using a grid search within the range of 10^{-5} to 10^{-2}. The DT classifier did not require any specific parameter tuning, as it uses a simple decision-making process based on the input features. Extra-Tree Classifier, a variant of Random Forests (RF) was also used, is ensemble learning method that combine multiple DTs to improve prediction accuracy and reduce overfitting. It bears a close resemblance to the RF, with the primary distinction being the way the DT in the forest are constructed. Each tree is given a random subset of k features from the total feature set. Each DT must then choose the optimal feature to partition the data, based on certain mathematical criteria. The performance of Extra-Tree classifiers was significantly influenced by several hyperparameters, including the number of trees (n_estimators), maximum tree depth (max_depth), minimum samples required to split a node (min_samples_split), and

minimum samples per leaf (min_samples_leaf). Finally, XGBoost is a powerful gradient-boosting algorithm that combines multiple weak prediction models (DTs) to create a strong predictive model. Similarly, XGBoost's performance depends on various hyperparameters, such as the number of boosting rounds (n_estimators), maximum depth of trees (max_depth), learning rate (eta), regularization parameters (alpha and lambda), and subsample rate (subsample and colsample_bytree). Overall, the model's ultimate goal is to determine the presence FM + or absence FM- of FMs, which in turn helps classify the infant's motion as either normal or abnormal. In simpler terms, the lack of FM serves as an indirect predictor for CP. To assess the classifiers' universality, a leave-one-subject-out cross-validation method is utilized.

4 Experimental setting, results, and evaluation

The software for estimating poses and detecting joint angles was created using Python and Jupyter Notebook. The software has been tested locally but can be used on the web with further development. The hardware configuration used includes a laptop computer with the Intel Core i7- 7660U CPU 7th Generation Processor 2.5 GHz and 16 GB of memory. The MediaPipe pose estimator was selected for 2D pose estimation because it's designed to operate on devices with restricted computational capabilities, like smartphones, and it can produce time series data [29]. Then, OpenPose [25] is considered as another 2D pose estimation algorithm due to its robustness and efficiency. It excels in applications that require multi-person settings, demonstrating its ability to perform tasks accurately. This contributes to its widespread popularity in the field of ML and computer vision. Finally, MeTRAbs [26] is customized for estimating absolute 3D human poses from RGB images. It allows for the estimation of comprehensive, metric-scale poses without requiring knowledge of distance at the time of testing or reliance on assumptions about human body proportions.

4.1 Classification results on MINI-RGBD dataset

The results obtained from the application of three different models (MediaPipe, OpenPose, MeTRAbs) using five different classification algorithms (SVM, NN, DT, Extra-Tree and XGBoost) on the MINI RGBD dataset [24] provide interesting insights as presented in Fig. 12.

Accuracy (Acc) measures the proportion of total predictions that are correct [6]. Sensitivity (Se) measures the ability of a binary classifier to correctly identify positive cases while Specificity (Sp) is the number of true negative instances divided by the total number of actual negative ones. In the analysis of the classification results, it is observed that the performance of the pose estimation methods varies significantly depending on the classifier used. The MediaPipe model, for instance, achieves its highest accuracy of 91.67% when paired with the NN classifier, but it shows the lowest sensitivity,75%, and specificity,62.5% when paired with the DT and XGBoost classifiers. On the other hand, the OpenPose model demonstrates consistent performance across all classifiers, with its highest accuracy of 91.67% achieved when combined with SVM, DT, and XGBoost. It also achieves perfect specificity,100%, when paired with SVM and DT classifiers, although its sensitivity is slightly lower when paired with SVM and NN classifiers. The MeTRAbs model also exhibits consistent performance across all classifiers. It achieves the highest sensitivity of 100% when paired with SVM, DT, Extra-Tree, and XGBoost classifiers.
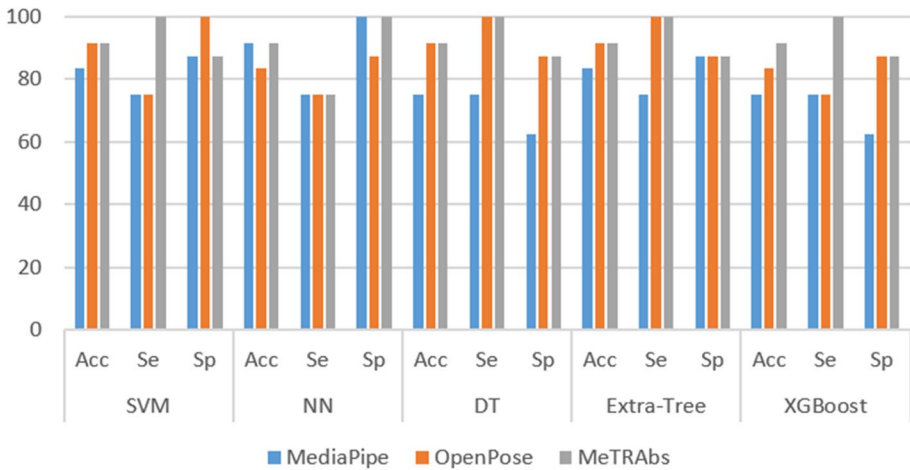
Fig. 12 Classification results on MINI RGBD dataset

These results highlight the significant impact of the choice of pose estimation method and classifier on the performance metrics. While some combinations achieve high accuracy, they may not perform as well on other metrics. This underscores the importance of considering all relevant performance metrics when evaluating different approaches. Therefore, the choice of classifier should be carefully considered based on the specific requirements of the task. For instance, if high sensitivity is required, MeTRAbs paired with SVM, DT, or XGBoost would be a good choice. Conversely, if high specificity is required, Open-Pose paired with SVM or DT would be preferable.

4.2 Classification results on RVI-38 dataset

In addition to the MINI RGBD dataset, we also applied the OpenPose model to the RVI-38 dataset using the same classification algorithms. Despite the constraints of the dataset, a comprehensive analysis demonstrated the efficacy of OpenPose in this context. The analysis of the OpenPose model across the underlying ML algorithms reveals a range of performances as depicted in Fig. 13. The results presented herein were derived from the application of the OpenPose algorithm to the RVI 38 dataset, which was available in JSON format. The original video files were not accessible for this study, thereby precluding the use of alternative algorithms such as Mediapipe and MeTRAbs. Despite these constraints, a comprehensive analysis was conducted using the available resources, demonstrating the efficacy of OpenPose in this context. For instance, SVM exhibits high overall accuracy at 97.37% and perfect specificity at 100%, albeit with a slightly lower sensitivity at 83.33%. NN sees a drop in all three metrics, with accuracy at 84.21%, sensitivity at 66.67%, and specificity at 87.5%, indicating a lower overall performance. However, DT shows an improvement, with accuracy and specificity rising to 94.74% and 96.88% respectively, and sensitivity returning to 83.33%. Extra-Tree maintains the same accuracy but sees a drop in sensitivity to 66.67%, even as specificity reaches 100%. XGBoost mirrors the DT performance, with identical values for all three metrics. These results underscore the variability in the model's performance across different ML algorithms, particularly in its ability to correctly identify positive cases, as reflected in the sensitivity metric.
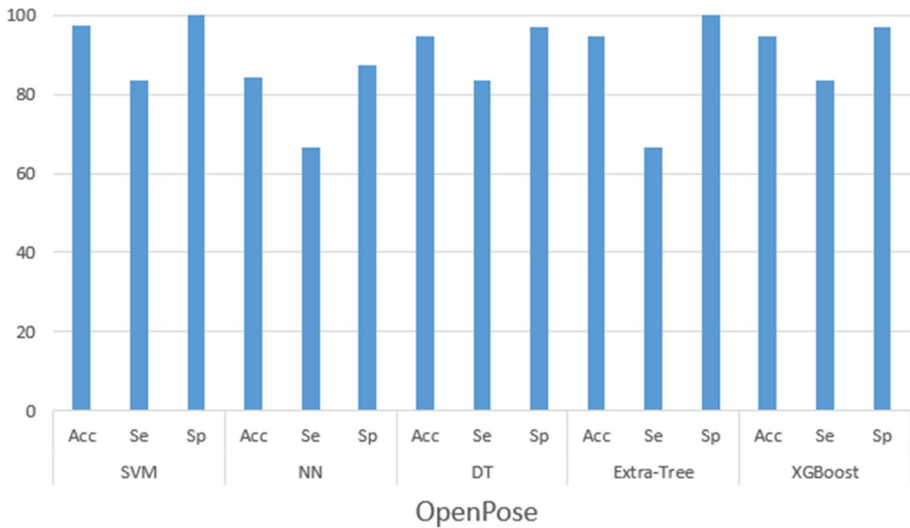
Fig. 13 Classification results on RVI-38 dataset

In general, the 2D and 3D joint coordinates returned by (MediaPipe, OpenPose [25], and MeTRAbs [26] are not the same, as they represent the joint positions in different coordinate systems. The 2D coordinates represent the keypoint positions in the image plane, while the 3D coordinates represent the keypoint positions in the real world. Typically, 3D pose estimators are expected to provide more accurate results than 2D pose estimation methods. 3D pose estimation methods consider the depth information of the scene, which allows for a more precise representation of the spatial relationships between different body parts. In contrast, 2D pose estimation techniques only consider the image plane and do not have access to depth information, which can result in ambiguities and inaccuracies in the predicted poses. Overall, the results suggest that the choice of pose detection techniques, exploring the more informative features, and selection of suitable classification methods can greatly impact the classification results. Thus, it is essential to carefully evaluate and compare different approaches to determine the most effective and efficient solution for pose estimation and classification.

4.3 Comparison with previous work

This study demonstrates a significant improvement in the early detection of infantile CP over existing methods. Table 2 presents a comparison of different methods applied to various datasets and algorithms, evaluated based on accuracy, sensitivity, and specificity. The use of 2D and 3D pose estimation strategies to observe and scrutinize the infant's comprehensive body movement attributes has shown to be a promising alternative to the GMA for early CP prediction. In comparison to the method presented in [1], which achieved an accuracy of 83.33% using a fully connected neural network (FCNet) and 90.28% with Conv1D-2 on the RGB-D dataset, the proposed method achieved a higher accuracy of 97.37% on the same dataset using ML algorithms. This indicates the effectiveness of the proposed method in classifying infant body movements into normal and abnormal categories. Furthermore, the proposed method outperforms the detection

Table 2 Comparison with the state-of-the-art methods

Method	Dataset	Algorithm	Accuracy %	Sensitivity %	Specificity %
McCay et al. [1]	RGB-D	Deep Learning (FCNet)	83.33	N/A	N/A
		Deep Learning (Conv1D-2)	90.28	N/A	N/A
		ML	83.33	N/A	N/A
Wu et al. [5]	RGB-D	GMA-ML	91.67	87.5	100
Groos et al. [7]	557 infant	Deep Learning	90.6	71.4	94.1
		GMA	85.9	70.0	88.7
		ML	72.7	71.4	72.9
Proposed	**RGB-D, RVI-38**	**ML**	**91.67**	**100**	**87.5**
			97.37	**83.33**	**100**

technique introduced by [5] which achieved an accuracy of 91.67% on the RGB-D dataset. Despite the high sensitivity (87.5%) and specificity (100%) of the GMA-ML method, the proposed method still shows superior performance. When compared to the deep learning-based method for predicting CP on 557 infants presented in [7], which demonstrated a sensitivity of 71.4% and a specificity of 94.1%, the proposed method exhibits a higher accuracy. Although the deep learning method showed higher sensitivity and specificity values than the GMA tool, the accuracy of the deep learning method was not significantly improved (90.6% vs 85.9%). The ML method also applied to the 557-infant dataset, had the lowest accuracy (72.7%), but its sensitivity (71.4%) was similar to the other methods applied to this dataset, and its specificity (72.9%) was the lowest among all methods. In contrast, the proposed method achieved a higher accuracy, further substantiating the efficacy of employing pose estimation techniques for the precocious prediction of infantile CP.

In conclusion, the proposed method provides a more accurate and efficient analysis of infant movement by computing various metrics such as angle orientations at different predicted joint locations, postural information, postural variability, movement velocity, movement variability, anti-gravity movements, and left–right movement coordination. These results highlight the importance of monitoring changes in joint angles over time for accurate diagnosis and treatment planning in the early detection of CP.

Furthermore, the proposed approach for early CP prediction offers several enhancements over existing methods, particularly in the areas of interpretability, computational efficiency, and versatility.

Interpretability While deep learning techniques can often achieve high accuracy, they are typically seen as "black boxes" due to their complex internal workings. Our approach, on the other hand, relies on well-understood ML algorithms like SVM, NN, DT, Extra-tree, and XG-Boost which are more transparent and interpretable. This makes it easier for clinicians and researchers to understand the model's predictions and trust its results.

Computational efficiency Deep learning models, especially those that work with video data, can be computationally intensive and require significant resources to train and run. In contrast, our pose estimation approach is less resource-intensive, making it more accessible for use in a wider range of settings, including those with limited computational resources.
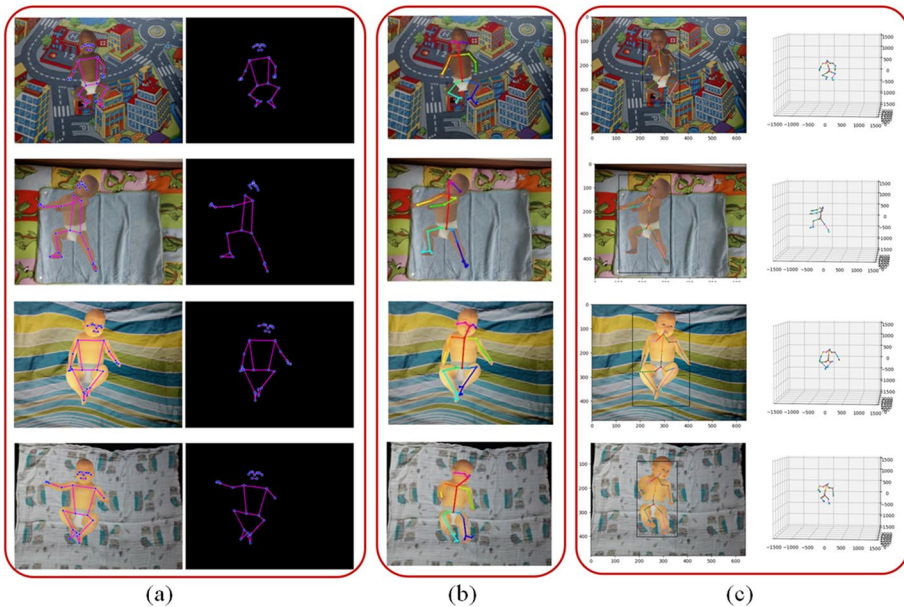
Fig. 14 Random frames from MINI-RGBD [24] dataset overlaid with the estimated poses using **a** Media-Pipe, **b** OpenPose, and **c** MeTRAbs

Versatility Our approach is versatile in that it can work with both 2D and 3D pose estimation strategies. This allows it to analyze a wide range of infant body movement attributes and adapt to different data sources and conditions.

5 Discussion

GMA is a method for evaluating the neurological development of infants, but it requires a trained expert and suffers from human variability. In order to tackle these challenges, it's necessary to implement automated analysis through non-intrusive, affordable techniques that can be utilized in a patient's home. This method would enhance the accessibility of testing and offer the possibility of analyzing data remotely without the requirement for specialized equipment. We have evaluated the performance computational efficiency and interpretability to highlight the strengths and limitations of each approach. Our approach offers a detailed analysis through various metrics providing comprehensive insights but might require extensive feature engineering compared to deep learning methods which can automatically extract features. However, our ML classifiers might be faster and require fewer computational resources compared to deep learning models which are typically resource-intensive. The proposed approach significantly enhances the interpretability of the model compared to deep learning models, which are often referred to as "black box" due to their complex internal workings. The ML algorithms we use, such as SVM, NN, DT, Extra-tree, and XG-Boost, are more transparent and their decision-making processes can be more easily understood. On the other hand, deep learning models involve multiple layers of computation, which makes

it difficult to trace how the input data is transformed into output predictions. While they can often achieve high accuracy, their lack of interpretability can be a drawback, especially in medical applications where understanding the reasoning behind a prediction can be as important as the prediction itself. The main goal of our research is to show the viability of monitoring infant CP at home using 2D video recordings and 2D and 3D pose estimation frameworks. Figure 14 illustrates the predicted locations of key joints in randomly chosen frames from the MINI-RGBD sequences using three distinct techniques.

According to the GMA theory, the patterns of movements for infants with typical development are more intricate, whereas those of infants with abnormalities are usually simple and repetitive [5]. In infants who are developing normally, spontaneous whole-body movements are coordinated, and the movements of each limb are relatively autonomous. However, there's a coordination of whole-body movements resulting from the concurrent contraction and muscle relaxation.

As elaborated in Fig. 15, the knee angles for the normal infant appear to be more consistent and smoother compared to the CP infant which indicates better coordination in the normal infant. The CP infant's graph shows more variability, sharper peaks, and troughs compared to the normal infant. This could be indicative of sudden, uncontrolled movements, which are common in CP patients. Looking at the timing of the peaks in the knee angles, we can make some observations about the coordination between the limbs. If the peaks for the left and right knee angles occur at similar times, this would suggest that there is a simultaneous or mirrored movement happening in both limbs. This is often indicative of a well-coordinated and balanced motion. On the other hand, if the peaks occur at different times, this might indicate a lag or delay in one limb compared to the other. This could be a sign of asymmetrical motion, where one limb is not moving in synchronous with the other. Sudden, rapid changes in joint angles can be used to characterize spasmodic, abrupt, and sporadic movements. By using pose estimation techniques to analyze the changes in joint angles over time, it may be possible to detect these types of movements. For instance, a rapid and abrupt change in the angle of a joint such as the elbow or knee could indicate a spasmodic movement. Similarly, if multiple joints change their angles rapidly and sporadically, this could indicate sporadic movements. By examining the patterns of joint angle changes over time, these types of movements can be detected and characterized.
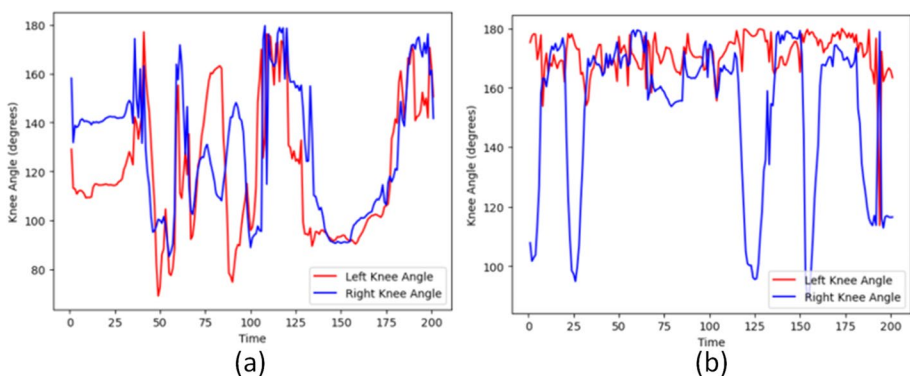

Fig. 15 Left and right knee angles for **a** Normal and **b** CP infants

In the context of infant movement analysis, antigravity movements refer to the ability of an infant to move a body part against the force of gravity. For example, lifting an arm or leg while lying on the back would be considered an antigravity movement. In a normal infant, antigravity movements of the ankles and wrists would typically be smooth and coordinated as elaborated in Fig. 16 (a). The infant would be able to lift their limbs against gravity and move them in a controlled manner. This is part of the normal development of motor skills and muscle strength. In contrast, CP infants might show different patterns as shown in Fig. 16 (b). Their antigravity movements could be less controlled and more restricted. They might struggle to lift their limbs against gravity, or their movements might appear stiff or jerky. This is due to the motor impairment associated with CP, which can affect muscle control and coordination.

This graph provides a visual representation of the infant's anti-gravity movements for both sides of ankles and wrists over time for one of the MINI RGBD dataset. From the graph, it appears that the velocity of left and right ankles and wrists in normal infants (left panel) shows a smooth curve, indicating a steady and consistent movement. However, in the CP infant (right panel), the curve is more jagged, suggesting variability and inconsistency in the movement. The findings of this research indicate that the pose estimation of spontaneous infant movements holds significant promise for aiding research efforts in the early identification of CP in infants. The proposed pose and velocity-based method is easier to comprehend, maintains intelligible information, and requires tuning of fewer parameters compared to similar methods, thereby making it more suitable for use in a clinical environment. By simply one glance at
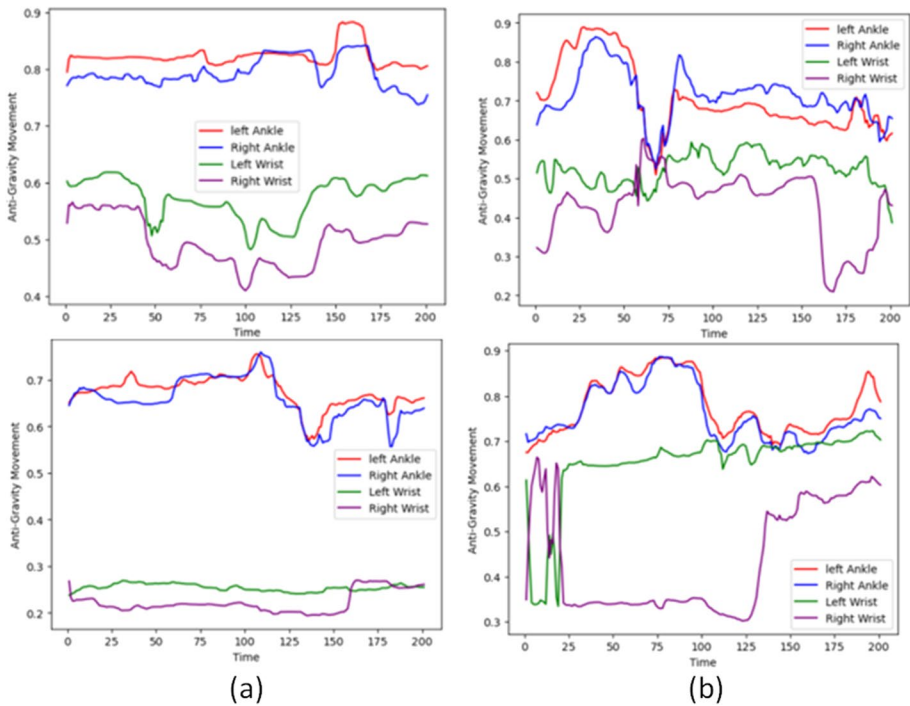

Fig. 16 Examples of antigravity movement **a** normal infants **b** CP infants

the plotted angles, doctors can get an impression of the quality of movement. For instance, it would be quite noticeable if there's a lack of movement on one side or limb of the body. Motion characteristics derived from angles, paths, speeds, accelerations, and movements against gravity have proven effective in detecting and forecasting CP. Moreover, the system is cost-effective, simple to install, can be operated by non-specialists, and is non-disruptive for infants.

6 Conclusion

This study introduces an innovative approach for the early prediction of CP by utilizing pose estimation strategies to observe and analyze comprehensive infant movement attributes. This method offers a valuable alternative to the GMA by monitoring a variety of features and specific markers indicative of CP, including postural information, movement variability, left–right movement coordination, and antigravity movements. The effectiveness of this approach has been demonstrated through an experimental evaluation using the MINI-RGBD and RVI-38 sequences. One of the key contributions of this study is the development of a comprehensive method for automated analysis of infant motor development. This method has the potential to aid in the early prediction of CP and offers a non-invasive, cost-effective, and accessible testing solution. Furthermore, this study presents a promising approach that can contribute to expanding our understanding of joint movements and positions over time. For future work, there is potential to expand the dataset and incorporate additional features to improve the accuracy and generalizability of the model. Moreover, extending this approach to other pathologies and developmental disorders such as Torticollis, Ataxia, and Developmental Dysplasia of the Hip (DDH) could broaden our understanding of pediatric motor development beyond CP. This could lead to the development of more comprehensive and effective diagnostic tools for a range of conditions, ultimately improving outcomes for affected children.

Declarations

References

1. McCay KD, Ho ESL, Shum HPH, Fehringer G, Marcroft C, Embleton ND (2020) Abnormal infant movements classification with deep learning on pose-based features. IEEE Access 8:51582–51592. https://doi.org/10.1109/ACCESS.2020.2980269

2. Ni H, Xue Y, Ma L, Zhang Q, Li X, Huang SX (2023) Semi-supervised body parsing and pose estimation for enhancing infant general movement assessment. Med Image Anal 83:102654. https://doi.org/10.1016/j.media.2022.102654

3. Khan MH, Schneider M, Farid MS, Grzegorzek M (2018) Detection of infantile movement disorders in video data using deformable part-based model. Sensors 18(10):E3202. https://doi.org/10.3390/s18103202

4. Turner A, Hayes S, Sharkey D (2023) The classification of movement in infants for the autonomous monitoring of neurological development. Sensors 23:4800. https://doi.org/10.3390/s23104800

5. Wu Q, Xu G, Wei F, Chen L, Zhang S (2021) RGB-D videos-based early prediction of infant cerebral palsy via general movements complexity. IEEE Access 9:42314–42324. https://doi.org/10.1109/ACCESS.2021.3066148

6. Sakkos D, Mccay KD, Marcroft C, Embleton ND, Chattopadhyay S, Ho ESL (2021) Identification of abnormal movements in infants: a deep neural network for body part-based prediction of cerebral palsy. IEEE Access 9:94281–94292. https://doi.org/10.1109/ACCESS.2021.3093469

7. Groos D, Adde L, Aubert S, Boswell L, de Regnier RA, Fjørtoft T, Gaebler-Spira D, Haukeland A, Loennecken M, Msall M, Möinichen UI, Pascal A, Peyton C, Ramampiaro H, Schreiber MD, Silberg IE, Songstad NT, Thomas N, Van den Broeck C, Øberg GK, Ihlen EAF, Støen R (2022) Development and validation of a deep learning method to predict cerebral palsy from spontaneous movements in infants at high risk. JAMA Netw Open 5(7):e2221325. https://doi.org/10.1001/jamanetworkopen.2022.21325

8. McCay KD et al (2022) A Pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants. IEEE Trans Neural Syst Rehabil Eng 30:8–19. https://doi.org/10.1109/TNSRE.2021.3138185

9. Soualmi A, Ducottet C, Patural H, Giraud A, Alata O (2023) A 3D pose estimation framework for preterm infants hospitalized in the Neonatal Unit. Multimed Tools Appl Springer. https://doi.org/10.1007/s11042-023-16333-6

10. Moccia S, Migliorelli L, Carnielli V, Frontoni E (2020) preterm infants' pose estimation with spatio-temporal features. IEEE Trans Biomed Eng 67(8):2370–2380. https://doi.org/10.1109/TBME.2019.2961448

11. Hesse N, Schröder AS, Müller-Felber W, Bodensteiner C, Arens M, Hofmann UG (2017) Body pose estimation in depth images for infant motion analysis. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), pp 1909–1912. https://doi.org/10.1109/EMBC.2017.8037221

12. Chambers C et al (2020) Computer vision to automatically assess infant neuromotor risk. IEEE Trans Neural Syst Rehabil Eng 28(11):2431–2442. https://doi.org/10.1109/TNSRE.2020.3029121

13. Li M, Wei F, Li Y, Zhang S, Xu G (2021) Three-dimensional pose estimation of infants lying supine using data from a kinect sensor with low training cost. IEEE Sens J 21(5):6904–6913. https://doi.org/10.1109/JSEN.2020.3037121

14. Cao X, Li X, Ma L, Huang Y, Feng X, Chen Z, Zeng H, Cao J (2022) AggPose: deep aggregation vision transformer for infant pose estimation. In: Proceedings of the thirty-first international joint conference on artificial intelligence (IJCAI-22) Special Track on AI for Good. https://api.semanticscholar.org/CorpusID:248693530

15. Prechtl HFR (1990) Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. Early Hum Dev 23(3):151–158. https://doi.org/10.1016/0378-3782(90)90011-7

16. Einspieler C, Bos AF, Krieber-Tomantschger M, Alvarado E, Barbosa VM, Bertoncelli N, Burger M, Chorna O, Del Secco S, DeRegnier R-A et al (2019) Cerebral palsy: early markers of clinical phenotype and functional outcome. J Clin Med 8:1616. https://doi.org/10.3390/jcm8101616

17. Adde L, Brown A, van den Broeck C, DeCoen K, Eriksen BH, Fjørtoft T, Groos D, Ihlen EAF, Osland S, Pascal A, Paulsen H, Skog OM, Sivertsen W, Støen R (2021) In-Motion-App for remote General Movement Assessment: a multi-site observational study. BMJ Open 11(3):e042147. https://doi.org/10.1136/bmjopen-2020-042147

18. Ningrum ES, Yuniarno EM, Purnomo MH (2023) Early detection of infant cerebral palsy risk based on pose estimation using openpose and advanced algorithms from limited and imbalance dataset. 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Jeju, Korea, Republic of 2023, pp 1–6. https://doi.org/10.1109/MeMeA57477.2023.10171951

19. Groos D, Adde L, Støen R, Ramampiaro H, Ihlen EA (2022) Towards human-level performance on automatic pose estimation of infant spontaneous movements. Comput Med Imaging Graph 95:102012. https://doi.org/10.1016/j.compmedimag.2021.102012

20 Abrishami MS et al (2019) Identification of developmental delay in infants using wearable sensors: full-day leg movement statistical feature analysis. IEEE J Trans Eng Health Med 7:1–7. https://doi.org/10.1109/JTEHM.2019.2893223

21. Redd CB, Barber LA, Boyd RN, Varnfield M, Karunanithi MK (2019) Development of a wearable sensor network for quantification of infant general movements for the diagnosis of cerebral palsy. Annu Int Conf IEEE Eng Med Biol Soc. Berlin, Germany, pp 7134–7139. https://doi.org/10.1109/EMBC.2019.8857377

22. Wang Y, Jin S, Ou Y (2019) A multi-task learning convolutional neural network for object pose estimation★. 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, pp 284–289. https://doi.org/10.1109/ROBIO49542.2019.8961594.

23. Airaksinen M, Räsänen O, Ilén E et al (2020) Automatic posture and movement tracking of infants with wearable movement sensors. Sci Rep 10:169. https://doi.org/10.1038/s41598-019-56862-5

24. Hesse N, Bodensteiner C, Arens M, Hofmann UG, Weinberger R, Sebastian Schroeder A (2019) Computer vision for medical infant motion analysis: state of the art and RGB-D data set. In: Leal-Taixé L, Roth S (eds) Computer Vision – ECCV 2018 Workshops. ECCV 2018. Lecture Notes in Computer Science(), vol 11134. Springer, Cham. https://doi.org/10.1007/978-3-030-11024-6_3

25 Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y (2021) OpenPose: realtime multi-person 2d pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 43(1):172–186. https://doi.org/10.1109/TPAMI.2019.2929257

26. Sárándi I, Linder T, Arras KO, Leibe B (2021) MeTRAbs: metric-scale truncation-robust heatmaps for absolute 3d human pose estimation. IEEE Trans Biom Behavior Identity Sci 3(1):16–30. https://doi.org/10.1109/TBIOM.2020.3037257

27. SciPy community (2021) Interpolation (scipy.interpolate). Available from: https://docs.scipy.org/doc/scipy/reference/interpolate.html. Accessed Sept 2023

28. pandas 1.3.1 Documentation (2021) pandas.DataFrame.rolling. Retrieved from pandas.DataFrame.rolling — pandas 2.1.1 documentation (pydata.org). Accessed Sept 2023

29. Mitrović K, Milošević D (2023) Pose estimation and joint angle detection using mediapipe machine learning solution. In: Filipovic N (ed) Applied Artificial Intelligence: Medicine, Biology, Chemistry, Financial, Games, Engineering. Springer Nature Switzerland AG, pp 109–120