A Training-Free Infant Spontaneous Movement Assessment Method for Cerebral Palsy Prediction Based on Videos

Qingqiang Wu, Penglin Qin, Jiachen Kuang, Fan Wei, Zejiang Li, Ruping Bian,
Chengcheng Han, and Guanghua Xu

***Abstract*— *objective*: Early diagnosis of infant cerebral palsy (CP) is very important for infant health. In this paper, we present a novel training-free method to quantify infant spontaneous movements for predicting CP. *Methods*: Unlike other classification methods, our method turns the assessment into a clustering task. First, the joints of the infant are extracted by the current pose estimation algorithm, and the skeleton sequence is segmented into multiple clips through a sliding window. Then we cluster the clips and quantify infant CP by the number of cluster classes. *Results*: The proposed method was tested on two datasets, and achieved state-of-the-arts (SOTAs) on both datasets using the same parameters. What's more, our method is interpretable with visualized results. *Conclusion*: The proposed method can quantify abnormal brain development in infants effectively and be used in different datasets without training. *Significance*: Limited by small samples, we propose a training-free method for quantifying infant spontaneous movements. Unlike other binary classification methods, our work not only enables continuous quantification of infant brain development, but also provides interpretable conclusions by visualizing the results. The proposed spontaneous movement assessment method significantly advances SOTAs in automatically measuring infant health.**

Manuscript received 21 November 2022; revised 14 February 2023; accepted 7 March 2023. Date of publication 10 March 2023; date of current version 15 March 2023. This work was supported in part by the Shaanxi Province Natural Science Foundation under Grant 2023-JC-QN-0678, in part by the Key Projects in Shaanxi Province under Grant 2022ZDLSF04-10, and in part by the Xi'an Key Laboratory of Brain Computer Interaction and Neuro Regulation Digital Medicine. *(Corresponding author: Guanghua Xu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Xi'an Jiaotong University under Approval No. 2021-1357.

Qingqiang Wu, Penglin Qin, Jiachen Kuang, Fan Wei, Zejiang Li, and Chengcheng Han are with the School of Mechanical Engineering, Institute of Engineering and Medicine Interdisciplinary Studies, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wuqingqiang@xjtu.edu.cn; qplqplqpl@stu.xjtu.edu.cn; kjc1331@stu.xjtu.edu.cn; wf3117370016@stu.xjtu.edu.cn; 3120301193@stu.xjtu.edu.cn; hanchengcheng@xjtu.edu.cn).

Ruping Bian is with the College of Information Engineering, Shaanxi Institute of International Trade and Commerce, Xi'an 710049, China (e-mail: 20221066@csiic.edu.cn).

Guanghua Xu is with the State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: ghxu@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3255639

***Index Terms*— Cerebral palsy, general movement assessment, clustering, visualization, pattern recognition, human pose.**

I. Introduction

INFANT cerebral palsy (CP) is a permanent impairment of motor and postural development caused by non-progressive damage that occurs in the developing fetus or infant's brain [1]. CP, the most common movement disorder in children, is one of the major diseases that seriously affects children's lives [2]. At least 17 million children are currently affected by CP worldwide. Early detection and recovery during the highly plastic stage of infants mean a lot to infants. Unfortunately, most positive children are discovered after they have obvious symptoms of abnormal brain development (about 18 months or later), missing the best time for treatment [3].

Studies found that in the early stages of infant brain development (before 16 weeks) there is no self-consciousness without stimulation, so the infant's spontaneous movements can reflect the extent of brain development [4], [5]. Therefore, some scholars [6], [7], [8] have designed touch sensors (such as accelerometers) to quantify infant movements. However, using markers to touch the baby's skin may significantly increase the baby's pressure, discomfort and pain. At the same time, it can affect the baby's normal movement pattern, which may be difficult to practice in actual clinical practice. To overcome these problems, researchers seek new reliable and non-contact surveillance alternatives, which are mainly based on video analysis.

General movements assessment (GMA) [9], proposed by Prechtl, is a typical tool that observes infant movement videos to predict the risk of infants with CP. In Prechtl's theory, the spontaneous movements of normal infants give a complex impression of variable speed and acceleration, while the spontaneous movements of infants with abnormal brain development are simple and monotonous [10]. Although GMA has proven to be an effective tool for the early prediction of CP, it requires skilled specialists and their time to study the videos and provide assessments. This makes assessments impossible or delayed for many infants. Therefore, it is necessary to develop tools that can automatically assess infant movements in videos.

At first, scholars directly made infant movement videos as input and extracted motion features for classification using traditional classifiers such as Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Adde et al. [11] used the frame difference method to extract the range of motion of each limb for direct classification. In [12], [13], Adde et al. further extended their method to classify by features such as the centroid of each limb movement. Rahmati et al. [14] used the movement frequency of each limb as a feature for classification by the partial least square regression. Stahl et al. [15] applied optical flow to extract the motion features of infants. The extraction features of these methods are relatively rough and easily affected by background noise.

With the development of deep learning, image-based human pose estimation methods emerge in an endless stream, and researchers began to extract the baby's motion for analysis. At this time, the classifier were still traditional machine learning classifiers. McCay et al. [16] proposed a classifier that fuses multiple pose motion histogram features and obtained satisfactory results after training on different datasets. There are more and more methods using deep learning as classifiers. Sakkos et al. [17] fused Convolutional Neural Network (CNN) and long short-term memory network (LSTM) models to classify trajectories of 8 joints of infants. McCay et al. [18] combined the Histogram of Joint Orientation 2D (HOJO2D) and the Histogram of Joint Displacement 2D (HOJD2D) and proposed a CNN model for classification. Nguyen-Thai et al. [19] explored the interpretability of the results, using a Spatio-temporal Attention-based Model (STAM) with an attention mechanism to classify pose sequences, and the visualization of attention weights can highlight the location of normal motion. Tsuji et al. [20] built an Artificial Neural Network to classify the dataset into 4 classes. Groos et al. [21] made the joint points of the human body as nodes and the bones as edges to construct a graph network, classifying each clip separately. Since deep learning is currently in a black box state, interpretability is still in the research stage. There are also some training-free quantitative assessment methods for spontaneous movement quality. Wu et al. [22], [23] respectively proposed a complexity-based motion complexity index (MCI) for the spontaneous movement of infants in 2d and 3d postures.

Nowadays, there are so many excellent AI-assisted assessment methods for CP [24]. Most of them are based on deep learning for training to obtain binary classification results of video input. However, these methods have the following shortcomings. First, limited by a small number of training samples, it is very likely that the training will be over-fitting, which makes the reliability of the results worse. Second, it's difficult to directly use a unified model for cross-dataset prediction, that's to say, different datasets need to be retrained, which is inconsistent with reality. Third, the development of infants is a dynamic process, and the degree of abnormal brain state continuously changes. Current methods can only judge whether the input video is abnormal or not, but cannot judge the severity. It is inappropriate to use binary classification rather than continuous indicators to assess the developmental extent. Fourth, these methods only provide classification results and lack interpretability, which makes it impossible to assist physicians in targeted treatment.

According to GMA theory, normal spontaneous movements give people a complex and changeable impression. Normal infant movement videos tend to contain more types of movement patterns, while abnormal spontaneous movements are often manifested as the absence of fidgety motion or fidgety motions are simply monotonous, that is, abnormal infant motion videos contained fewer categories of motion patterns. Then we propose a method called the affinity propagation clustering model (APCM) to quantify the spontaneous movement based on the characteristics of the input video itself. Unlike other classification tasks, our method treats the quantitative evaluation of spontaneous movements as a clustering task. First, we divide the input infant pose sequences into many clips. Then we cluster these clips. The higher the abnormal movement degree of the infant, the less the number of clusters of clips.

This proposed method is a training-free method and has no potential impact on overfitting. At the same time, our method analyzes each video independently, which can adapt to the individual development of infants. Further, our results can be visualized by which clips in the video are repeated. If there are too many clips of the same type in the input video, these clips have more possibility of abnormal movement patterns. At the same time, we separate the limbs of infants for cluster analysis, which can further demonstrate the motor correlation and complexity of each limb through interpretable clustering results, which can assist physicians in targeted therapeutic interventions.

In this paper, we mainly make the following contributions:

First, we propose a training-free quantization method based on the characteristics of the infant movement video itself, which can realize the evaluation of infant self-adaptation and can be used directly between different datasets.

Second, to the best of our knowledge, this is the first training-free method with interpretable results, and the proposed method can assist doctors in conducting targeted interventions;

Third, unlike other methods, the results of our method are quantifiable indicators that can show how much the infant is at risk, and can also give hints for some infants with mild developmental abnormalities.

The rest of the sections of this paper are organized as follows: Section II shows the proposed infant spontaneous movement assessment method for CP prediction. Section III presents experiments and comparisons with SOTAs on two public infant movement datasets. Section IV discusses the limitations of our approach and future work, and the last Section draws out conclusions.

II. Methodology

In [16] and [19], the authors extracted the motion features of different subjects for training and classification, focusing on the feature differences across all samples. The proposed APCM focuses more on the intrinsic characteristics of each sample, which is training-free. Compared with other methods,

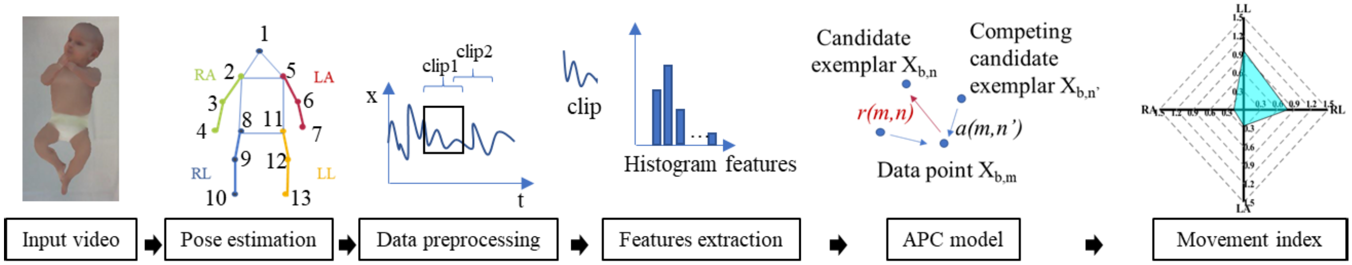

Fig. 1. An overview of the proposed method.

our method is training-free, has good portability, and can be directly applied across different data sets. At the same time, our method provides visual results, through which doctors can locate abnormal positions in infant motion videos and further confirm diagnostic results.

A. The Proposed Method Framework

Fig. 1 shows an overview of the proposed method. First, we use a current pose estimation algorithm to extract the 2D or 3D skeleton sequence of the supine infant from the input video. Next, the skeleton sequence is preprocessed, mainly including interpolation and filter smoothing. We divide the infant's limbs into four body parts and cut the four body parts sequences with a sliding time window to obtain a series of clips. Then, the histogram features of these clips are extracted, and the affinity propagation clustering model is applied for clustering. Finally, the clustering categories of each body part of the infant are counted, and our movement indicators are proposed.

B. Pose Estimation

We employ the Joint feature coding (JFC) method [25] to estimate infant poses from color videos. The JFC is used to estimate infant pose from depth images. Since the JFC method combines regression and classification tasks, and it considers different scales of the human body, it is more suitable for small-scale body (such as an infant) pose estimation. The JFC method has higher pose estimation accuracy than the commonly used OpenPose algorithm [26]. We transfer the JFC method to color images and used the MS COCO dataset [27] for training and got satisfactory results, as shown in Fig. 2. There are 18 joint points marked in the MS COCO dataset, while the proposed method APCM mainly accesses the motion quality of the limbs, so only 12 joints are considered. To clearly show the shape of the human body, the nose (J_1) is also marked to represent the head in Fig. 1. We define the Right Arm body part RA $= \{J_3, J_4\}$, Left Arm body part LA $= \{J_6, J_7\}$, Right Leg body part RL $= \{J_9, J_{10}\}$, Left Leg body part LL $= \{J_{12}, J_{13}\}$.

C. Data Preprocessing

The JFC method is based on a single image, making the estimated pose discontinuous or having slight jitter between adjacent frames. Before extracting features, the pose sequence needs to be preprocessed. We refer to the steps of [19] to preprocess the joint sequence $\{J_{i,t}\}$ (i represents the i-th joint, and t is the t-th frame), including: (1). coordinate interpolation. We use linear interpolation to compute the missing coordinate into each joint time series $\{J_{i,t}\}$. (2) outlier removal. We remove outliers by median filtering using a sliding time window with a window length of 15 frames, the same length as [16], [19]. (3). filter smoothing. We use the mean filter with a sliding window to smooth the interpolated joint sequence, and the window length is 15 frames. Fig. 3 shows the preprocessed changes in the x-coordinate sequence of an infant's joint.

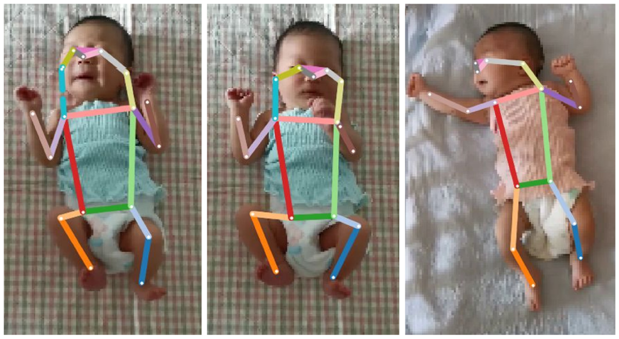

Fig. 2. Some infant pose estimation results using the JFC method.

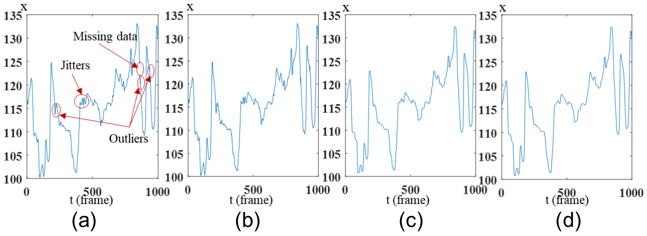

Fig. 3. An example of the data pre-processing. (a). the original time sequence of a joint x coordinate with missing data, jitters, and outliers. (b). the time sequence after interpolation. (c). the time sequence after median filtering to remove outliers. (d). the data after preprocessing.

Due to factors such as the distance or views from the camera to the infant, the scales of infants vary in different datasets. In general, normalization for different infant sequences is also required. However, since we focus on the internal regularity of the single infant video, we don't need to normalize different infant videos.

D. Feature Extraction

1) Clips Generation: For a given joint sequence $\{J_{i,t}\}$, we split the sequence into a set of clips with a certain window length. To ensure that there are enough clips for clustering,

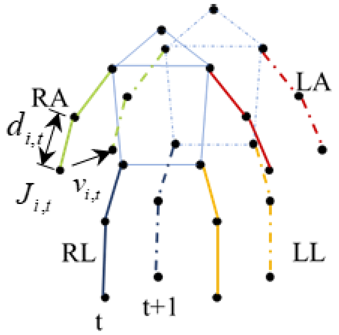

Fig. 4. The feature description of each clip, including pose coordinate $J_{i,t}(x,y)$, each joint's velocity $v_{i,t}(x,y)$, and pixel length of each bone $d_{i,t}$.

there is an overlap between each clip. Due to the duration of different movement patterns of infants are different, the window length of the clip should be long enough to ensure that each clip can fully cover the movement. Through experimental parameter optimization in Section III-H, we set the window length to 90 frames, and the step size to 40 frames. Then the overlapping part is 50 frames.

According to [16], [18], [19], and [21], the differences in infant movement patterns can be reflected in the pose and velocity of limbs, so we select joint coordinates $J_{i,t}$ and velocity $v_{i,t}$ as features. $v_{i,t}$ can be calculated by Equation (1):

$$v_{i,t} = J_{i,t+1} - J_{i,t}, \tag{1}$$

where t refers to the *t-th* frame and i is the *i-th* joint defined in Fig. 1, and $v_{i,t}$ represents the velocity of the corresponding joint.

Pose estimation methods such as JFC and OpenPose can only get the infant's 2D pose. In fact, the infant's spatial motion is more suitable for analysis. The movement of the baby's limbs in the z-axis direction is difficult to describe with $J_{i,t}(x,y)$ and $v_{i,t}(x,y)$. To make up for this deficiency, we introduce a new feature: $d_{i,t}$, the pixel length of each bone of the limbs. The $d_{i,t}$ represents the projection of the z-axis motion of each limb on the x-y plane and can be calculated by Equation (2).

$$d_{i,t} = ||J_{i+1,t} - J_{i,t}||_2, \tag{2}$$

Fig. 4 shows the feature description of each clip. For example, for RA body part, the feature at time t can be expressed as $X_{t,RA}$:

$$X_{t,RA} = [J_{3,t}(x, y), v_{3,t}(x, y), d_{3,t}, J_{4,t}(x, y), v_{4,t}(x, y), d_{4,t}], \tag{3}$$

2) Clip Normalization: Infants have a limited range of motion for each limb, and the ranges of joint coordinates vary from infant to infant. To fully reflect the difference of motions in the limited range, we normalize the feature sequence $\{X_{b,t}| b = RA,LA,RL,LL\}$ as Equation (4).

$$X_{b,t} = \frac{X_{b,t} - \min\{X_{b,t}\}}{\max\{X_{b,t}\} - \min\{X_{b,t}\}}, \tag{4}$$

The normalization of other methods [16], [19], [21] is mainly to ensure that different infants are on the same scale. Unlike other normalization purposes in preprocessing, we aim to ensure that the movement ranges are fully reflected in histogram encoding.

3) Histogram Encoding: In the description of infant motion features, the authors in [19] directly used joint coordinates and velocity parameters as the features of STAM for classification, McCay et al. [16] compared the impact of different feature combinations on the results. According to their research, histogram encoding has excellent performance. Therefore, we also use the histogram to encode our features. Histogram encoding also brings another benefit: it mitigates the effects of jitter between adjacent frames in pose estimation results. Since the feature sequence $\{X_{b,t}\}$ contains three parts ($J_{t,i}$, $v_{t,i}$ and $d_{t,i}$), the bin of the pose part is set to bin1, the bin of the velocity part is set to bin2, and the bin of the bone length is set to bin3. In Section III we will experiment with different combinations of bins.

E. Affinity Propagation Clustering Model

Affinity propagation clustering (APC) was proposed by Brendan and Delbert [28]. APC selects exemplars by continuously passing information between different points. The algorithm does not need to define the number of classes in advance but continuously searches for appropriate exemplars in the iterative process, and automatically identifies the location and number of exemplars from data points. In our method, each clip is treated as a data point. We first construct the similarity matrix ***s*** of all data points by Equation (5).

$$s(m, n) = -||X_{b,m} - X_{b,n}||_2, \quad m \neq n \tag{5}$$

where s(m,n) represents the similarity between points $X_{b,m}$ and $X_{b,n}$. $X_{b,m}$ and $X_{b,n}$ represents two clips in the clips set $\{X_{b,t}\}$. The larger s, the closer from point m to point n (for simplicity, we use [m, n] to represent [$X_{b,m}$, $X_{b,n}$] respectively).

Since the number of exemplars is unknown, each point is initially treated as an exemplar, and $p(m)$ is used to measure the possibility of the point n as an exemplar. Due to no prior information, $p(m)$ is initialized to the median value of the {s(m,n)} matrix, as shown in Equation (6).

$$p(m) = s(m, m) = median\{s(m, n)\}, \quad m \neq n \tag{6}$$

We use the responsibility matrix *r(m,n)* to describe the degree how the point n is suitable as an exemplar for the point m. The responsibility *r(m,n)* can be calculated in Equation (7).

$$r(m, n) \leftarrow s(m, n) - \max_{n', s.t. n' \neq n} \{a(m, n') + s(m, n')\} \tag{7}$$

where *a(m,n')* represents the available value of other points except the point n to the point m, and be initialized to 0. *s(m,n')* is the similarity of other points to the point m except the point n. If the value of *r(m,n)* is greater than 0, it means that the point n has a strong ability to become the exemplar of the point m.

The availability matrix *a(m,n)* is used to describe the suitability of point m to select point n as its exemplar. The availability matrix *a(m,n)* can be calculated by

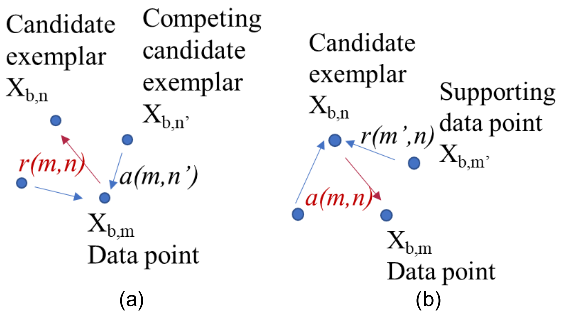
Candidate exemplar X_{b,n}
Competing candidate exemplar X_{b,n'}
r(m,n)
a(m,n')
X_{b,m} Data point
(a)
Candidate exemplar X_{b,n}
r(m',n)
Supporting data point X_{b,m'}
a(m,n)
X_{b,m} Data point
(b)

Fig. 5. The affinity propagation clustering model. (a). Responsibility r(m,n) represents the responsibility value of the point n as the exemplar of the point m. (b). Availability matrix a(m,n) describes the suitability of the point m to select the point n as its exemplar.

Equations (8) and (9).

$$a(m,n) \leftarrow \min\{0, r(n,n) + \sum_{m',s.t.m' \notin \{m,n\}} \max\{0, r(m',n)\}\}, \tag{8}$$

$$a(n,n) \leftarrow \sum_{n',s.t.n' \notin n} \max\{0, r(n',n)\}, \tag{9}$$

where *r(m',n)* represents the responsibility value of the point n as the exemplar of other points except the point m. It means point n is supported by the other points except for the point m with responsibility value greater than 0. The algorithm model is shown in Fig. 5. Fig. 5a shows the transmission from point m to the candidate exemplar point n, which reflects the degree how the point n is suitable as the exemplar of the point m after considering other potential exemplar point n'. Fig. 5b shows the transmission from the candidate exemplar point n to the data point m, reflecting the suitability of the point m to select point n as the exemplar after considering the support of other point n'.

During affinity propagation, availabilities and responsibilities of every point can be combined to identify exemplars. When $\{a(m,n) + r(m,n)\}$ takes the maximum value, the point n can be used as the exemplar of the point m. We update availability matrix a and responsibility matrix r iteratively, count the number of all exemplars, and end the iteration when the result remains unchanged for 10 iterations.

We implemented the above APC model using the Matlab toolkit [29]. And got the final numbers of clustered categories for 4 body parts $\{\mathrm{N_b} | \mathrm{b} = \mathrm{RA, LA, RL, LL}\}$.

F. Spontaneous Movement Index

As we all know that the longer the video is, the more clips there are, then the more categories tend to exist. To unify different durations, we obtained the mean ratio k by Equation (10).

$$k = mean\{K_{b,i}/N_{b,i}\}, \tag{10}$$

where $K_{b,i}$, $N_{b,i}$ refer to the number of clips, and the number of clustered categories of i-th infant's b-th body part, respectively.

We counted the number of categories of all normal samples in the RVI-38 dataset and got the mean ratio of $\mathrm{k} = 3.12$.

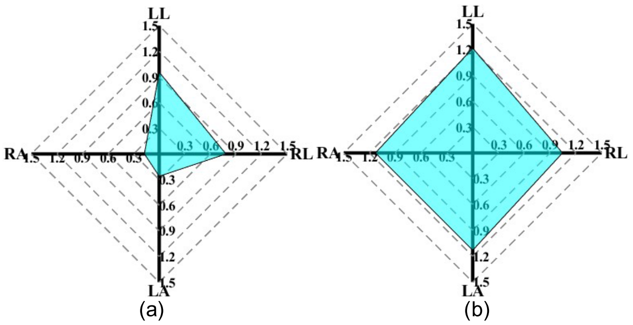

Fig. 6. The $\{\mathrm{P_b}\}$ spider chart of 2 samples. (a) abnormal infant. (b) normal infant.

Then we define the level of the infant's b-th body part motion pattern as $\mathrm{P_b}$. $\mathrm{P_b}$ can be calculated by Equation (11).

$$P_b = k\frac{N_b}{K_b}, \tag{11}$$

where $\mathrm{K_b}$ is the number of clips of input infant's b-th body part. Through Equation (11), the $\{\mathrm{P_b}\}$ of 4 body parts can be obtained respectively. We set the threshold *th* to limit the normal movement probability. When $\mathrm{P_b} < th$, it means that this body part movement is abnormal. To increase the robustness of the indicators, we set that when there is more than 1 body part movement abnormality, the global spontaneous movement of the baby is abnormal, that is, the baby has a high risk of CP. Fig. 6 shows the 4 body parts $\{\mathrm{P_b}\}$ spider chart of different infants. The significant difference between the $\{\mathrm{P_b}\}$ of normal and abnormal infants can be seen in Fig. 6.

III. Experiments and Results

To test the effect of the proposed method, we verified our method on 2 datasets. The detailed experimental process and results are as follows.

A. Dataset

1) MINI-RGBD: The MINI-RGBD dataset [30] contains 12 infants. A color video and a depth video were recorded for each baby. The videos were collected from the local children's hospital during the half year of their life. In this paper, we use color videos as our input. The resolution of the color videos is 640 × 480. Each video contains 1000 frames, and the video frame rate is 30 frames per second (FPS). To protect the infants' privacy, the dataset adopts the Skinned Multi-Infant Linear model to hide the baby's identity and facial information. The dataset has been labeled by experienced experts (including 4 abnormal babies, and 8 normal babies).

2) RVI-38 Dataset: The RVI-38 dataset [16] is a real patient video dataset collected at the Royal Victoria Infirmary (RVI), containing 38 baby samples, each with a color video captured, for a total of 38 videos. All infants were aged from 3 to 5 months post-term. The videos were recorded with a handheld Sony DSC-RX100 advanced compact premium camera. The resolution of the videos is 1920 × 1080. The duration of each video is not fixed. The shortest is 40 seconds, and the longest is 5 minutes. The average duration is 3 minutes and 36 seconds. To protect infants' privacy, the videos have been processed by OpenPose to get 38 skeleton sequences. The dataset was independently classified by two experienced evaluators using GMA (contains 32 normal samples, and 6 abnormal samples).

B. Detailed Implementation

The original data provided by the two datasets are different. For the MINI-RGBD dataset, we use the JFC method for pose estimation to reduce errors. For the RVI-38 dataset, the dataset has estimated pose using OpenPose in advance, resulting in many fatal errors, and it's a more challenging dataset. Although the frame rates of the two are different, we did not delve into the effect of the capture frame rate on the results. Both datasets take the same parameters.

The implementation parameters of the pose estimation part are the same as [25]. From preprocessing to the end, it is implemented in the Windows 10 system. The experimental environment is Matlab r2020b, 6xCPU (i5-11400f@2.6GHz). The GPU is not involved in subsequent computations.

C. Performance Metrics

Consistent with [16], [18], [19], we also use the accuracy (Acc for short), sensitivity (Sen), and specificity (Spe) evaluation metrics to access the performance of the proposed method. The metrics are calculated by Equations (14), (13), and (14).

$$Acc = \frac{TN + TP}{TN + FN + TP + FP}, \tag{12}$$

$$Sen = \frac{TP}{TP + FN}, \tag{13}$$

$$Spe = \frac{TN}{TN + FP}, \tag{14}$$

where TN is short for true negative and means healthy infants are correctly classified as healthy infants. TP is short for true positive and represents the case in which abnormal infants are correctly classified as unhealthy infants. FN is short for false negative and stands for abnormal infant incorrectly as healthy infants. FP is short for false positive, which means healthy infants are incorrectly classified as abnormal infants.

We also use the receiver operating characteristic (ROC) curve to evaluate our model, which is the most popular method to measure the model's preference.

D. Compare With SOTAs on MINI-RGBD Dataset

The classification performance of the MINI-RGBD dataset is presented in Table I. As can be seen from Table I, our method achieves an accuracy of 91.67%, a sensitivity of 100%, and a specificity of 87.5%.

In terms of accuracy, our method is lower than the Pose&Vel method [16]. But it is worth noting that the MINI-RGBD dataset only includes 12 samples. Only one sample out of the actual 12 samples was misclassified with our method. Overall, the training-based methods are generally better than the training-free methods, but the possibility of overfitting cannot be ignored because of the small dataset size.

In the sensitivity metric, our method achieves 100%, which means that all positive babies can be screened. It shows that APCM can meet the needs of screening. In terms of specificity, our method is lower than the Pose&Vel method [16] about 12.5%. Only one sample was misclassified.

TABLE I
CLASSIFICATION PERFORMANCE OF THE MINI-RGBD DATASET

	Methods	Acc(%)	Sen(%)	Spe(%)
Training-based	Qm [13]	58.33	50	62.5
	Qsd [13]	75	75	75
	FFT_m [14]	91.67	75	**100**
	HOJO2D [18]	91.67	75	**100**
	STAM [19]	91.67	100	87.5
	FFT-JO [16]	**100**	**100**	**100**
	Pose and vel. [16]	**100**	**100**	**100**
Training-free	MCI [23]	91.67	**100**	87.5
	APCM	91.67	**100**	87.5

TABLE II
CLASSIFICATION PERFORMANCE OF THE RVI-38 DATASET

	Methods	Acc(%)	Sen(%)	Spe(%)
Training-based	Qm [13]	52.63	50	53.13
	Qsd [13]	86.84	50	93.75
	FFT_m [14]	84.21	50	90.63
	STAM [19]	81.58	33.33	90.63
	FFT-JO [16]	84.21	83.33	84.38
	HOJO2D [18]	92.11	66.67	96.88
	MWC [15]	83.33	75	87.5
	Pose and vel. [16]	**97.37**	83.33	**100**
Training-free	MCI [23]	84.21	**100**	81.25
	APCM (th=0.5293)	89.47	**100**	87.5
	APCM (th=0.1040)	94.74	83.33	96.88

In the comparison of training-free methods, our results are consistent with the MCI [23], but our method can provide more visual information, which will be shown later.

E. Compare With SOTAs on the RVI-38 Dataset

Table II presents the classification results of the RVI-38 dataset. On the RVI-38 dataset, our method achieved an accuracy of 94.74%, a sensitivity of 83.33%, and a specificity of 96.88% when th $= 0.104$. In terms of accuracy, our method is 2.63% lower than the Pose&Vel method [16] with one more sample classification error. Although the size of the RVI-38 dataset has increased compared to the MINI-RGBD dataset, it is still a small sample dataset.

Compared with the MINI-RGBD dataset, the accuracy of most training methods is reduced, indicating that the size of the dataset has a great impact on training. However, the accuracy of our method has improved, because the interference of small sample accidental phenomena is excluded, and the accuracy is more accurate. Notably, all training methods are retrained on the RVI-38 dataset, while our method uses the same parameters as the MINI-RGBD dataset.

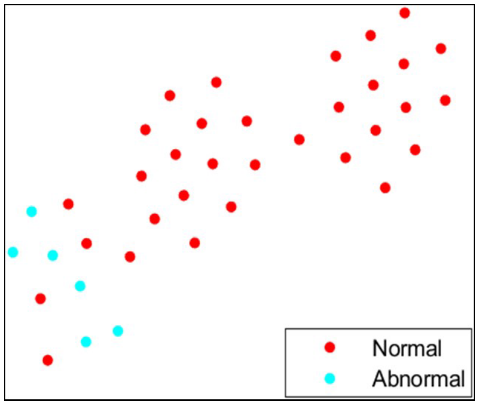

Fig. 7. T-SNE result of {P_b} of RVI-38 dataset.

In terms of sensitivity, our method achieves the SOTAs when $th = 0.104$. When we increase the threshold th to 0.5293, we can achieve 100% sensitivity, at the cost of sacrificing some accuracy. Different th leads to different performance of the proposed method, we use ROC curve to optimize th, and the specific discussion will be presented in Section III-G. It explains that all positive babies can be screened at this threshold.

Different infants have different fidgety motion trajectories, limited by the size of the dataset, and unbalanced samples of two class movements, resulting in a serious decrease in the sensitivity of the STAM method [19] compared to MINI-RGBD. It shows that the STAM has a strong dependence on the dataset and weak generalization ability. In terms of specificity, the training-free methods are still lower than training-based methods. Because the positive samples are much lower than the normal samples (the proportion is 3:16), in extreme cases, if all 38 samples are classified as normal samples, it can also reach 84.21% accuracy and 100% specificity. Therefore, for screening, sensitivity is more important than specificity.

Because we use 4 body parts and do more detailed quantification, our indicators are better than MCI in the training-free protocol. As the threshold th increases, the sensitivity increases, but the accuracy and specificity decrease.

To further demonstrate our results, we visualize the results using the t-Distributed Stochastic Neighbor Embedding method (t-SNE) [31]. Since the weights of the 4 body parts are the same, we sort the {P_b} of each sample and then use t-SNE for visualization. Fig. 7 shows the t-SNE visualization result. We can see that all the abnormal samples (blue dots) are clustered together, while the normal samples (red dots) are clustered into 2 groups. This is because, in normal samples, some infants' 4 limbs have uniform spontaneous movements, while other infants have one limb to show abnormal movement patterns. Nonetheless, there is a good distinction between normal and abnormal samples. This also proves that our training-free method can effectively quantify the different qualities of infant spontaneous movement.

When the threshold is more stringent, the results of our method in MINI-RGBD and RVI-38 are very close, which shows that our method has good generalization ability.

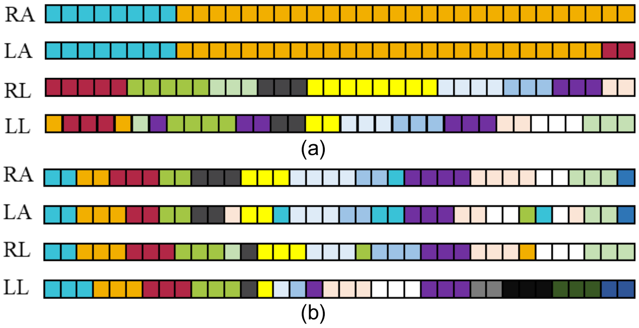

Fig. 8. Visualization of two samples' clustering results. (a) abnormal sample. (b). normal sample. Each row corresponds to a body part and each cell corresponds to a clip. The same color in the same row indicates the same category.

F. Results Visualization

In the STAM [19], the authors train the network using attention encoding, and finally visualize the attention weights, highlighting body parts and frames that contain discriminative information about normal fidgety movements. There are obvious differences between attention weights and real labels, and the association performance of different joints is not clear. Inspired by them, we visualize the clustering results, as shown in Fig. 8. Fig. 8a shows the clustering results of an abnormal infant with 4 limbs. We can see that the entire sequences of RA and LA body parts have few clustering categories, indicating that the movements are relatively simple and there are obvious abnormalities. For the RL part, the yellow part may have abnormal movement patterns. The location of the yellow clips can assist experts in locating the time sequence location for further analysis. Fig. 8b shows the clustering results of the limbs of a normal infant. It can be seen from Fig. 8b that the movements of the limbs of normal infants are relatively complex, and most of the categories contain 3 clips, which is close to the ratio k value (k = 3.17) in Section II-F. Further, there is a certain correlation in the first half of the clips of the four limbs, and the correlation of the latter half of the clips are weak in Fig. 8b, which explains that infant movements are random and complex.

G. ROC Curve on RVI-38 Dataset

The selection of the threshold th affects the accuracy, sensitivity, and specificity of the results. To choose an appropriate threshold, we investigated the relationship between sensitivity and specificity under different thresholds using the ROC curve as shown in Fig. 9. As can be seen from Fig. 9, when $th = 0.1040$, the accuracy rate is 94.74%, and the sensitivity is less than 1. After increasing the threshold, the accuracy rate is reduced to 89.47%, and the sensitivity is 100%. Since the purpose of our method is for large-scale mass screening, to ensure that there is no missed diagnosis, we choose a stricter threshold to build the model.

H. Ablation Study

Important parameters affecting the accuracy of APCM are the number of bins of histogram encoding, step length, and

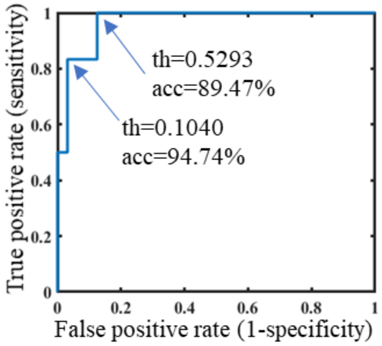

Fig. 9. ROC curve of the APCM on RVI-38 dataset.

TABLE III
BINS OPTIMIZATION RESULT

[bin1, bin2, bin3]	Acc(%)	Sen(%)	Spe(%)
[16, 16, 16]	81.58	83.33	81.25
[32, 16, 16]	**89.47**	**100**	**87.5**
[32, 32, 16]	86.84	**100**	84.38
[32, 32, 32]	81.58	83.33	81.25

window length of the sliding window in clip generation. To simplify parameter optimization, we use the experience to initialize and fix other parameters to optimize specific parameters. The first is the optimization of bins.

1) Bins Optimization: In [18], the authors optimized the parameters with bins = 8 and 16. Based on their experiments, when bins = 8, the differences in movements cannot be characterized. We choose 16 and 32 as the candidate for [bin1, bin2, bin3] to perform experiments. Table III shows the experimental results of different bins configuration.

As we adopt the similar feature description with [18], many experiments were done on the optimization of bins of [18]. Therefore, we also refer to the similar configuration of [18] on the selection of bins. The difference is that [18] chooses 8 and 16 as candidates, we use 16 and 32 as our candidates. We found that when bins = 8, different features [$J_{i,t}(x,y)$, $v_{i,t}(x,y)$, $d_{i,t}$] have great interference. In our feature descriptors, bin1 is the coordinate segmentation, bin2 is the velocity segmentation, and bin3 is the bones segmentation. Different bins can affect the weight and precision of the three features. Since both velocity and bone length are calculated by coordinates, the estimation errors of coordinates will be transferred to the two features. Taking small values in bin2 and bin3 can reduce the influence of error and improve the accuracy. It can be seen from Table III that when (bin1, bin2, bin3) = (32, 16, 16), the result can reach the optimal solution.

2) Sliding Window Step Length Optimization: The step length affects the number of clips and then affects the results of the cluster. We choose (30, 40, 50, and 60) as the candidate to perform experiments, and the accuracy is shown in Fig. 10.

From Fig. 10, we can see that when step length = 40 frames, the result is optimal. When the step length is too short, there are too many overlapping parts in adjacent clips, making the results of the APCM of abnormal infants and normal infants similar. However, when the step is too long, the number of clips is too small, resulting in inaccurate cluster results.

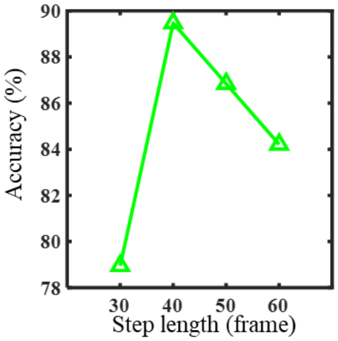

Fig. 10. The accuracy results on the RVI-38 dataset with different step lengths.

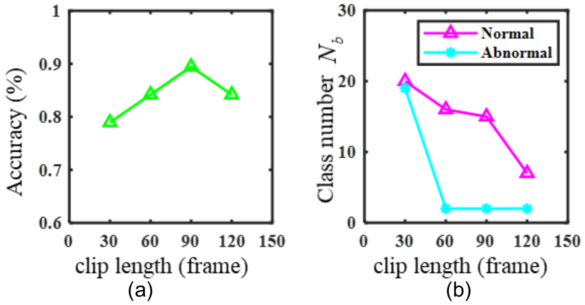

Fig. 11. The trend of accuracy and number of cluster classes N_b with clip length increasing. (a). the accuracy of the RVI-38 dataset with different clip lengths. (b). N_b of normal and abnormal sample with changes in the length of the clip.

3) Sliding Window Length Optimization: The infants' movement velocities are diverse. The too short window length isn't enough to include a complete infant movement process, while too long window lengths will cause too much overlap. Fig. 11a presents the accuracy with different clip lengths on the RVI-38 dataset. As shown in Fig. 11a, with the window length increases, the accuracy of classification is the first improved and then decreased. The best clip length is 90.

To study the influence of different window lengths on the final clustering number, we selected some samples, and normalized them to the same proportion, then obtained the results of Fig. 11b. From Fig. 11, we can see that with the increase in clip length, both N_b of the normal and abnormal sample have decreased, but the reduction of the normal infant is slow, and abnormal infant is reduced relatively fast. This is because the movements of abnormal infants are relatively monotonous and simple, and the small window length is enough to include all movements mode. While the spontaneous movements of the normal infant are complicated, the large window length can still maintain the diversity of the clips.

IV. DISCUSSION

Monitoring infants, especially premature infants, are essential to evaluate the health status of the baby and early brain development state. However, we found that there is currently a lack of quantitative assessment of infant brain development. This is mainly due to the lack of current monitoring technology and information processing. The current mainstream diagnosis relies on the subjective visual judgment of clinicians on the side of the crib. Of course, some methods use wearable sensors (such as acceleration meters) or automated GMA. The wearable sensor may have a certain stimulus on the baby, affecting the spontaneous motion mode of the baby. For automated GMA methods, most of the evaluations are completed as classification tasks, which is limited by the current small sample datasets. Most of the classification training models may have problems with over-fitting or generalization ability. Infant movement using data-driven indeed is one of the future development trends, but the current small sample dataset is not enough to support the establishment of perfect models. And just doing the binary classification of the input videos cannot meet the needs of fine quantification.

Compared with other methods [13], [14], [16], [18], [19], our method is training-free. After optimizing model parameters with RVI-38 dataset, we directly apply the optimized parameters model to MINI-RGBD dataset and obtained SOTA result, indicating that the proposed method has good portability.

In our study, we are committed to adopting the training-free method to cluster single infant movement video essential features. Unlike other classification tasks, our method has implemented a fine quantitative assessment of infant four limbs movements.

In terms of the accuracy of the two public datasets, although our method has not reached the optimal, it is only one sample that is different from the optimal result. Note that our method is training-free, and no parameters are adjusted separately in the two datasets. All model parameters are the same across the two datasets. This ensures the reliability of the model and the huge prospect of a migration to actual clinical applications.

In terms of sensitivity, by adjusting a stricter threshold, our method achieves 100% sensitivity in both datasets. In other words, our method can ensure that all babies with high risk can be screened. Further testing for high-risk infants can avoid misdiagnosis.

In terms of explanation of results, our method can directly display the different states of the four limbs through the final visualization of the clustering result. At the same time, the quantitative indicators {Pb} are given. The higher the {Pb}, the higher the probability of normal limb movement, and the higher the probability of the normal baby's normal brain development visualization results, experts can directly locate the corresponding limbs and corresponding moments of abnormal motion in the baby's movement video, and it can assist the experts for personalized intervention.

With the help of advanced machine learning theory and deep learning methods, digital information of infants can be used to assist the early diagnosis of cerebral palsy in infants. In our previous study [23], complexity can be used to describe the quality of spontaneous movement in infants, but this evaluation is not comprehensive. By APCM, the self-clustering idea can be used to clarify the regularity of the same movement pattern in the infant movement, and then visualize the quality of the infant's spontaneous movement. Because the APCM method can realize the movement evaluation of different limbs, it can be used to further quantify the developmental processes of different limbs in infant brain development.

The proposed method still has some limitations. First, the proposed method relies on the accuracy of pose estimation of input infant video. Most of the current pose estimation algorithms are based on a single image, which will introduce noise or identification errors. We can improve the accuracy of recognition through multiple image postures between continuous frames in the future. Second, the APCM method has only analyzed the number of cluster classes, and the distribution of each category has not yet been analyzed. In the future, it will consider analyzing the cluster distribution of 4 body parts and propose a more specific evaluation plan. The third limitation is the lack of data. At present, due to the privacy of infants and other factors, there are few public available data sets. In the later stage, we will build our own private dataset to further verify our method.

V. Conclusion

We have proposed a new training-free video-based spontaneous movements assessment method for infant CP screening.

This method uses JFC to estimate infant pose, then split the joint sequences into clips, use the APCM cluster class and form multi-dimensional assessment indicators. By testing on the two public datasets, the effectiveness and generalization ability of the method is verified. The main discoveries of this article are: (1). Converting traditional classification ideas into clustering can get rid of the dependence on the dataset size. (2). The number of categories of clustering can quantify the quality of infant spontaneous movement. (3). Visualization of clustering results can explain the exception in the input infant video.

References

[1] P. Rosenbaum et al., "A report: The definition and classification of cerebral palsy April 2006," *Developmental Med. Child Neurol.*, vol. 109, no. 109, pp. 8–14, 2007.

[2] M. Oskoui, F. Coutinho, J. Dykeman, N. Jetté, and T. Pringsheim, "An update on the prevalence of cerebral palsy: A systematic review and meta-analysis," *Develop. Med. Child Neurol.*, vol. 55, no. 6, pp. 509–519, Jun. 2013.

[3] C. Marcroft, A. Khan, N. D. Embleton, M. Trenell, and T. Plötz, "Movement recognition technology as a method of assessing spontaneous general movements in high risk infants," *Frontiers Neurol.*, vol. 5, p. 284, Jan. 2015.

[4] H. F. R. Prechtl and B. Hopkins, "Developmental transformations of spontaneous movements in early infancy," *Early Hum. Develop.*, vol. 14, nos. 3–4, pp. 233–238, Dec. 1986.

[5] J. P. Piek and R. Carman, "Developmental profiles of spontaneous movements in infants," *Early Hum. Develop.*, vol. 39, no. 2, pp. 109–126, Oct. 1994.

[6] M. Singh and D. J. Patterson, "Involuntary gesture recognition for predicting cerebral palsy in high-risk infants," in *Proc. Int. Symp. Wearable Comput. (ISWC)*, Oct. 2010, pp. 1–8.

[7] D. Gravem et al., "Assessment of infant movement with a compact wireless accelerometer system," *J. Med. Devices*, vol. 6, no. 2, pp. 1–7, Jun. 2012.

[8] D. Karch, K.-S. Kim, K. Wochner, J. Pietz, H. Dickhaus, and H. Philippi, "Quantification of the segmental kinematics of spontaneous infant movements," *J. Biomechanics*, vol. 41, no. 13, pp. 2860–2867, Sep. 2008.

[9] H. F. Prechtl, C. Einspieler, G. Cioni, A. F. Bos, F. Ferrari, and D. Sontheimer, "An early marker for neurological deficits after perinatal brain lesions," *Lancet*, vol. 349, no. 9062, pp. 1361–1363, May 1997.

[10] C. Einspieler and H. F. R. Prechtl, “Prechtl’s assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system,” *Mental Retardation Developmental Disabilities Res. Rev.*, vol. 11, no. 1, pp. 61–67, 2005.

[11] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, and R. Støen, “Using computer-based video analysis in the study of fidgety movements,” *Early Hum. Develop.*, vol. 85, pp. 541–547, May 2009.

[12] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, and R. Støen, “Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study,” *Develop. Med. Child Neurol.*, vol. 52, no. 8, pp. 773–778, Feb. 2010.

[13] L. Adde et al., “Characteristics of general movements in preterm infants assessed by computer-based video analysis,” *Physiotherapy Theory Pract.*, vol. 34, no. 4, pp. 286–292, Apr. 2018.

[14] H. Rahmati, H. Martens, O. M. Aamo, Ø. Stavdahl, R. Stoen, and L. Adde, “Frequency analysis and feature reduction method for prediction of cerebral palsy in young infants,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1225–1234, Nov. 2016.

[15] A. Stahl, C. Schellewald, Ø. Stavdahl, O. M. Aamo, L. Adde, and H. Kirkerød, “An optical flow-based method to predict infantile cerebral palsy,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 20, no. 4, pp. 605–614, Jul. 2012.

[16] K. D. McCay et al., “A pose-based feature fusion and classification framework for the early prediction of cerebral palsy in infants,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 8–19, 2021.

[17] D. Sakkos, K. D. Mccay, C. Marcroft, N. D. Embleton, S. Chattopadhyay, and E. S. L. Ho, “Identification of abnormal movements in infants: A deep neural network for body part-based prediction of cerebral palsy,” *IEEE Access*, vol. 9, pp. 94281–94292, 2021.

[18] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton, “Abnormal infant movements classification with deep learning on pose-based features,” *IEEE Access*, vol. 8, pp. 51582–51592, 2020.

[19] B. Nguyen-Thai, V. Le, C. Morgan, N. Badawi, T. Tran, and S. Venkatesh, “A spatio-temporal attention-based model for infant movement assessment from videos,” *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3911–3920, Oct. 2021.

[20] T. Tsuji et al., “Markerless measurement and evaluation of general movements in infants,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Jan. 2020.

[21] D. Groos et al., “Development and validation of a deep learning method to predict cerebral palsy from spontaneous movements in infants at high risk,” *JAMA Netw. Open*, vol. 5, no. 7, Jul. 2022, Art. no. e2221325.

[22] Q. Wu, G. Xu, F. Wei, L. Chen, and S. Zhang, “RGB-D videos-based early prediction of infant cerebral palsy via general movements complexity,” *IEEE Access*, vol. 9, pp. 42314–42324, 2021.

[23] Q. Wu et al., “Automatically measure the quality of infants’ spontaneous movement via videos to predict the risk of cerebral palsy,” *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[24] M. T. Irshad, M. A. Nisar, P. Gouverneur, M. Rapp, and M. Grzegorzek, “AI approaches towards Prechtl’s assessment of general movements: A systematic literature review,” *Sensors*, vol. 20, no. 18, p. 5321, Sep. 2020.

[25] Q. Wu et al., “Supine infant pose estimation via single depth image,” *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[26] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7291–7299.

[27] T.-Y. Lin et al., “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 740–755.

[28] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[29] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, “Adaptive affinity propagation clustering,” 2008, *arXiv:0805.1096*.

[30] N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. Weinberger, and A. S. Schroeder, “Computer vision for medical infant motion analysis: State of the art and RGB-D data set,” in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, p. 1–17.

[31] M. C. Cieslak, A. M. Castelfranco, V. Roncalli, P. H. Lenz, and D. K. Hartline, “T-distributed stochastic neighbor embedding (t-SNE): A tool for eco-physiological transcriptomic analysis,” *Mar. Genomics*, vol. 51, Jun. 2020, Art. no. 100723.