Received: June 6, 2023. Revised: July 25, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Pose Sequence-Aware Generative Adversarial Network for Augmenting Skeleton Sequences to Improve Cerebral Palsy Detection by Deep Learner

Rajalekshmy Kayiparambil Devarajan[1]* **Sheeja Shaik Khader[1]**

[1]*Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India*
*Corresponding author's Email: rajidevaraj16phd@gmail.com

Abstract: In medical analysis, early detection and diagnosis of cerebral palsy (CP) are highly helpful in stimulating brain cells and alleviating the adverse impacts of the disease. In the detection of CP, diagnostic tools such as magnetic resonance imaging (MRI) and general movements assessment (GMA) have emerged. In the past years, most studies have focused on GMA-based CP detection based on the different pose estimation methods. Amongst, the part affinity field (PAF) or OpenPose was one of the well-known methods for estimating the skeletal images from the RGB-D videos of infant general movements, which was used for CP detection. But, the estimated skeletal images were very few and it was difficult to annotate a large infant movement dataset. Therefore, in this article, a pose sequence-aware generative adversarial network (PS-GAN)-based data augmentation method is proposed that creates high-quality skeleton images for CP detection. First, long-range dependencies in continuous frames are acquired by self-attention and the dense graph is pruned to achieve efficient training. Then, spatial joints and temporal characteristics are encoded into the PS-GAN using the graph convolutional network (GCN) to map noises to high-quality skeleton images. Besides, the PS-GAN structure selection problem is defined as a Markov decision process (MDP) and solved using the new reinforcement learning (RL) to choose the optimal PS-GAN structure, which achieves effective generation. Further, the created skeleton images are used to train the convolutional neural network (CNN) with a softmax classifier and detect CP. At last, an extensive experiment is conducted on the MINI-RGBD, babyPose and motion infant analysis (MIA) databases. The results show that the PS-GAN-CNN achieves 92.2%, 92.5% and 92% accuracy for MINI-RGBD, babyPose and MIA databases in contrast with the PredictMed, fully connected network (FCNet), CNN-long short-term memory (LSTM) and knowledge-based recurrent neural network (KBRNN) methods.

Keywords: Cerebral palsy, General movement assessment, RGB-D videos, GAN, GCN, Markov decision process, Softmax.

1. Introduction

Cerebral Palsy (CP) is a group of disorders affecting 17 million individuals globally, primarily in preterm infants. It is caused by brain damage that impairs muscle control, affecting motion, pose, communication, and lifestyles. Preterm infants are more prevalent, with 32.4 cases per 1000 born prematurely and 70.6 cases per 1000 born exceedingly prematurely [1-3]. One-third of pregnancies with severe preterm births are impaired, increasing the risk of severe impairment [4]. Early detection and diagnosis of kids with CP using diagnostic tools like GMA and MRI is crucial. Early treatments aim to maximize infant brain adaptability, minimizing symptoms and damage [5-7].

Early interventions are challenging and time-consuming. Primary prevention techniques include two primary processes pose estimation and GMA.

1.1 Pose estimation

Infant pose estimation is divided into two types: RGB-based and depth information-based. Deep learning models like CNN and depth information are used for predicting 2D poses [8]. However, applying Kinect for motions is complex if the foreground and background are quite analogous, like a new-born in a prostrate position. A learning-free model-based

recursive matching (MRM) scheme is developed for estimating infant poses in prostrate positions [9], but its computational effectiveness is low.

1.2 General movement assessment

The GMA is a cost-efficient and reliable method used to detect irregular movement patterns in infants during growth phases [10-11]. It involves skilled professionals examining videos of infants, detecting fidgety movements between 9 and 20 weeks postpartum. But this method requires experienced professionals' effort, making it unaffordable for many infants. Automated detection schemes are needed to analyze infant movements in videos [12]. Academics are analyzing automated GMA schemes for CP diagnosis using machine learning models. These schemes predict new-borns based on GMA motion forms. Healthy newborns have more complications, while atypical ones are tedious [13]. The natural motion of the entire body remains synchronized in healthy brain development, but management is affected by muscle retraction and leisure [14].

To tackle the challenges in both pose estimation and GMA schemes, Wu et al. [15] developed a new scheme to predict CP in new-borns earlier from RGB-D videos. They combined RGB photos and depth details to obtain 3D coordinates of the newborn in the prostrate position, instead of 2D coordinates. The part affinity field (PAF) scheme was used to estimate infant 2D movement. The 3D coordinates of the newborn's joints were obtained after 2D coordinate conversion and orientation, along with related depth photos. The intricacy of joint motion and correlations among distinct joints were studied. A full assessment index was calculated to quantify and diagnose the infant's CP risk based on GMA. However, the estimated skeleton images using the PAF were very few and it was difficult to annotate a large infant movement dataset. So, many variants of GAN models are applied [16], but they have a few shortcomings, such as that (i) spatial correlation among joints and temporal characteristics along constant frames are not discovered and (ii) an expensive pre-learning stage is needed to obtain intra-frame constraints.

To tackle these problems, this study proposes the PS-GAN-based data augmentation method to create a large dataset and solve the class imbalance problem for CP detection.

1.3 Major contributions

The major aim of this study is to synthesize high-quality skeleton images with labels. In this PS-GAN, self-attention is used to obtain long-range dependencies in continuous frames and train to prune the dense graph for effective learning. Also, the GCN is used to encode the spatial joints and temporal characteristics into the prediction. According to this, the PS-GAN converts noises into high-quality skeleton images. Additionally, the best PS-GAN structure is explored by the new RL, in which the PS-GAN structure search problem is formulated as an MDP for structural sampling. This allows the efficient utilization of images created by past experiences. Moreover, the new large dataset is acquired from the created skeleton images, which are split into learning and test datasets. The learning dataset is fed to the CNN with a softmax classifier to learn the skeleton features for CP and the trained classifier is later used to detect CP from the test images. Thus, the accuracy of recognizing CP is increased by augmenting the number of skeleton images of high quality.

The rest of the article is structured as follows: section 2 reviews the studies focused on CP detection and diagnosis. Section 3 explains the PS-GAN-CNN method and Section 4 clarifies its performance. Section 5 concludes the study and gives future work.

2. Literature survey

A logistic regression-based framework called PredictMed [17] was developed for predicting health conditions in children with CP. It uses a brute-force algorithm for feature reduction and logistic regression for health condition prediction. However, overfitting is addressed due to limited samples and lengthy computation time.

An FCNet [18] was designed for classifying atypical infant body movements depending on the pose characteristics such as the histograms of joint angle and histograms of joint dislocation from the video sequences. However, the number of videos in the MINI-RGBD database was not sufficient, which differs from the real-time dataset for achieving efficient performance. So, a set of enhanced characteristics and a concatenation model [19] were developed for CP classification using a novel video database. But, the class imbalance problem was addressed and this model relied on the efficiency of the posture prediction.

A novel spatio-temporal attention-based model (STAM) was developed [20] to analyze fidgety movements for CP detection. It retrieves individual postures and models infant movements using the Spatio-temporal GCN (ST-GCN). The model then chooses body parts with discriminatory data for predicting CP. However, it relies on precision and requires annotated data, which can be difficult to obtain.

A new CNN-LSTM [21] model was presented for automatically classifying infant body movements associated with CP by extracting pose characteristics from RGB video frames. But, the manual annotations take a long time and it needs more video frames to solve overfitting during training. A deep CNN-based image segmentation framework [22] was integrated with 3D cranial magnetic resonance imaging for diagnosing children with CP. However, the number of samples was insufficient to achieve the highest efficiency.

A new Shallow multilayer neural network (SMNN) [23] was presented to classify an age-related movement pattern and the fidgety movements for CP detection. But, it needs a large database to enhance accuracy. A modified AlexNet [24] was presented for classifying CP from functional magnetic resonance images. But the dataset was extremely inadequate. A KBRNN was developed [25] using a CP information graph for domain understanding. An evolution scheme extracted knowledge, embedded it into tensors, and fed it into an RNN to learn correlation between signs and disorders from electronic medical records during CP diagnosis. However, lower annotated samples impacted learning performance.

3. Proposed methodology

In this section, the PS-GAN-CNN is explained briefly for CP detection and diagnosis. Fig. 1 portrays a pipeline of the presented study.

Table 1 lists the notations used in this study.

3.1 Dataset details

In this study, three most well-known publicly available infant motion analysis datasets are used:

1. MINI-RGBD dataset [26-27]: The corpus of 12 RBG-D video sequences from 6 months old is an open source dataset for infant CP, providing accurate silhouette, texture, and motion. It anonymizes information by substituting unprocessed frames with machine-processed frames, and labels videos using the GMA scheme, determining the incidence of nervous motions.
2. BabyPose dataset [28]: The data includes 16 depth videos of preterm children's motion in

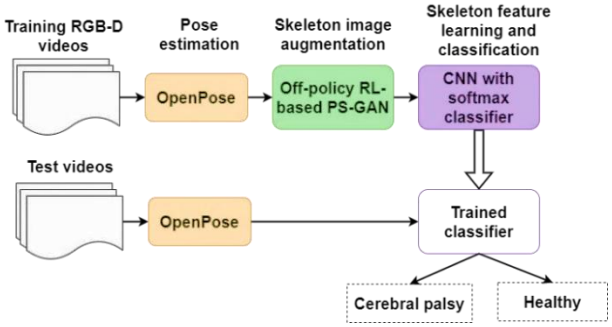
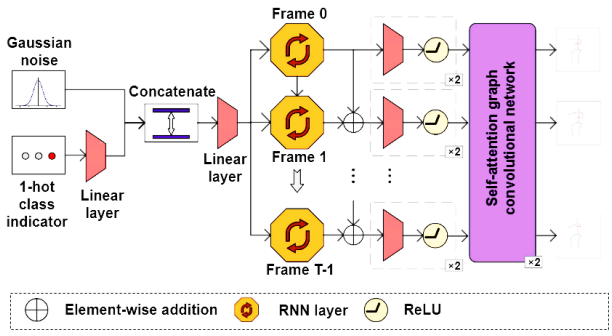


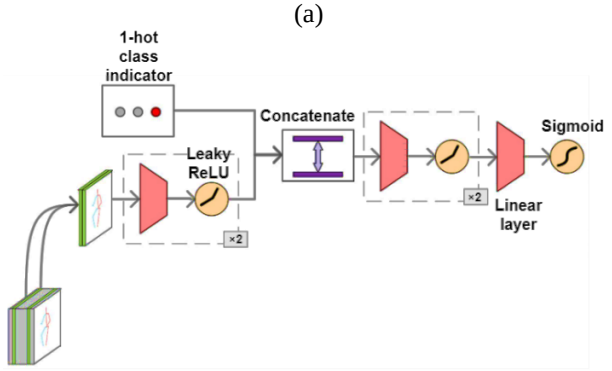
Figure. 1 Pipeline of the presented study

Table 1. Lists of notations

Notations	Explanation
x	Skeleton image
$\mathcal{G}$	Skeleton generator
$\mathcal{D}_f$	Frame-based discriminator
T	Frame (pose sequence) length
N	Number of joints
z	Gaussian random noise
y	One-hot tag (i.e., label)
o_{T-1}	Results of RNN model
H	Input vector for self-attention unit
h_t	Hidden state vector at interval t
H_{in}	Output of attention layer
S_{mask}	Attention score matrix
$\tilde{A}_s$	Adjacency matrix of an entire image
$H^{(1)}, \dots, H^{(5)}$	Hidden state output vectors
k_{frame}	Number of selected frames
$\mathcal{L}$	Objective function
$p(x)$	Ground truth distribution
$p(z)$	Normal Gaussian distribution
G	Undirected graph
V	Node set
A	Adjacent matrix of the intra-frame created based on the skeleton structure
I	Self-association
$\bar{A}$	Intra-frame adjacency matrix after adding self-association
W_q, W_k, W_v	Prediction weights
S	Attention score
$s_{m,n}$	n^{th} frame's guidance on the m^{th} frame
D^{ii}	Diagonal node degree matrix
$r(s,a)$	Remuneration value
$P(s'\|s,a)$	Switch probability
$\rho(s)$	Primary structure
a	Agent's action
$R_t(s,a)$	New reward function
α	Tradeoff variable
$J(\pi)$	Objective function of off-policy RL system
IS_{final}, FID_{final}	Final scores of the whole structure
$J(Q)$	Fitness function of critic network
Q_{target}	Approximation target for Q
f_θ	Input vector containing Gaussian noise
B	Memory buffer
β	Positive Lagrange multiplier



(a)



(b)

Figure. 2: (a) Generator network and (b) Discriminator network

NICUs, with 1000 frames per video and annotated with limb-joint positions for 12 joints: shoulder, elbow, wrist, hip, knee, and ankle.

3. Motion infant analysis (MIA) dataset [29]: It comprises the state vector and timestamp, derived from depth measurements from an RGB-D sensor situated perpendicularly above the child lying in a supine position on the crib, typically 70 cm away.

Once dataset is acquired, OpenPose or PAF is used to predict the newborn 2D motion, which is infant pose skeleton images. Those skeleton images are given as input to the PS-GAN for augmenting the number of infant skeleton images.

3.2 Pose sequence-aware GAN for high-quality skeleton generation

In the PS-GAN-based data augmentation method, the GCN is integrated to encode the rich structural data to minimize computational complexity by utilizing self-attention to train a skeleton graph. The structure of PS-GAN comprises a skeleton generator $\mathcal{G}$ (see Fig. 2 (a)) and a frame-based discriminator $\mathcal{D}_f$ (see Fig. 2 (b)).

Generator:

Consider the frame (pose sequence) length is T. The skeleton generator is initiated by an RNN with an input at every interval as the fusion of Gaussian random noise z and an entrenched category interpretation of a tag y. The results of the RNN are represented as $[o_0, o_1, \ldots, o_{T-1}]$. Assume outputting residuals rather than the accurate coordinates of various joints, i.e. $c_0 = o_0, c_1 = o_1 + c_0, \ldots, c_{T-1} = o_{T-1} + c_{T-2}$. The RNN result is passed via 3 linear transformations before providing the input to the GCN layer.

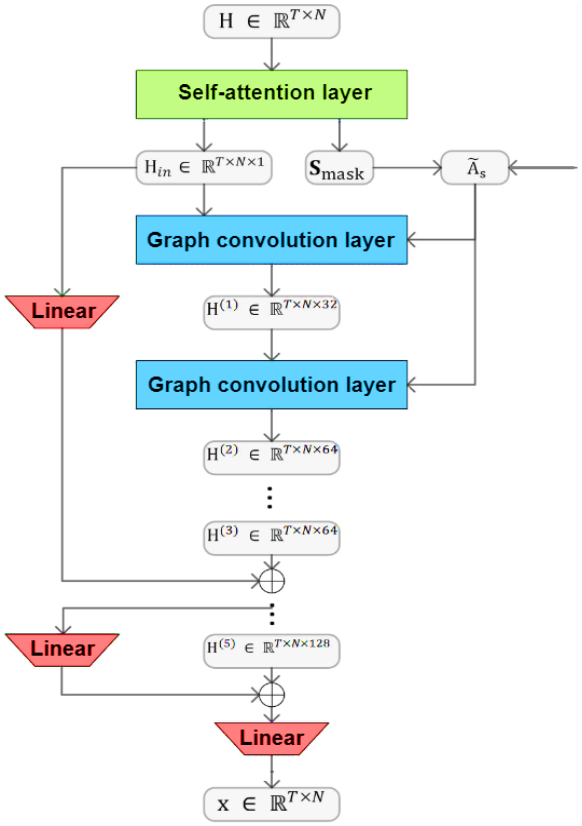


Figure. 3 Overview of self-attention graph convolution layer

Graph convolutional layer:

The major layer of this PS-GAN is the self-attention-based graph convolution as shown in Fig. 3. The input of this layer is denoted as a feature vector $H \in \mathbb{R}^{T\times N}$. Using a self-attention unit, the outcome is a novel interpretation $H_{in} \in \mathbb{R}^{T\times N\times 1}$ and a trained covered attention score matrix $S_{mask} \in \mathbb{R}^{T\times T}$. After this unit, 5 graph convolutional units are performed, wherein all layers take the final layer's hidden state vector and covered adjacency matrix $\tilde{A}_s$ as the input.

The hidden states that are outcomes of the 5 graph convolutional units, are represented as $H^{(1)} \in \mathbb{R}^{T\times N\times 32}$, $H^{(2)} \in \mathbb{R}^{T\times N\times 64}$, $H^{(3)} \in \mathbb{R}^{T\times N\times 64}$, $H^{(4)} \in \mathbb{R}^{T\times N\times 128}$ and $H^{(5)} \in \mathbb{R}^{T\times N\times 128}$, correspondingly. Also, the ResNet approach is employed on all 2 graph convolutional units, i.e. the outcome of the initial graph convolutional layer is added to the 3rd graph convolutional layer and the

outcome of the 3rd graph convolutional layer to the last outcome.

Discriminator:

The frame-based discriminator $\mathcal{D}_f$ arbitrarily chooses frames (k_{frame}) of a given image and the related tags as the input. The discriminator outcome is either real or fake. The conditional GAN objective function is formulated as:

$$\mathcal{L} = \min_{\mathcal{G}} \max_{\mathcal{D}_f} \mathbb{E}_{x\sim p(x)}\left[\log \mathcal{D}_f(x|y)\right] + \mathbb{E}_{z\sim p(z)}\left[log\left(1-\mathcal{D}_f\left(\mathcal{G}(z|y)\right)\right)\right] \quad (1)$$

In Eq. (1), $p(x)$ is the ground truth distribution, $p(z)$ defines the normal Gaussian distribution and y denotes the 1-hot tag.

3.2.1. Skeleton graph generation

Typically, the skeleton image is defined by 2D or 3D coordinates of the infant's joints in all frames. The inter-frame activity is the fixed skeleton in spatial region, whereas the inter-frame activity is the motion in temporal region. To obtain the temporal relationship, a connected graph is built for a complete skeleton image and an inter-frame association is learned by implementing self-attention training. An undirected graph $G = (V, E)$ is built on a complete skeleton image of T frames, all comprise N joints. The node set $V = \{v_{ti} | t = 0, \dots, T-1, i = 1, \dots, N\}$ comprises each joint of a skeleton image.

3.2.2. Executing self-attention-based GCN

Assume each joint in a skeleton image with a 2D adjacent matrix having row and column dimension $N * T$. Initially, $A \in \mathbb{R}^{N\times N}$ is used to represent the adjacent matrix of the intra-frame that is created based on the structure of a skeleton. Once self-associations I are added, the intra-frame adjacency matrix can be $\bar{A} = A + I$. After that, an initial adjacency matrix of an entire image is represented as:

$$\tilde{A} = \begin{pmatrix} \bar{A} & I & \cdots & I \\ I & \bar{A} & \cdots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \cdots & \bar{A} \end{pmatrix}_{(N*T)\times(N*T)} \quad (2)$$

In Eq. (2), I is utilized to define linking all nodes with each related node in the temporal region, $(N * T) \times (N * T)$ defines $\tilde{A}$ is a 2D matrix with both row and column dimension $N * T$, where N and T are integers, $*$ is a multiplication process. The adjacency matrix $\tilde{A}$ defines all nodes in a single frame are linked to the respected node in the temporal region. Simultaneously, it links to the nearby nodes in the spatial region coded by $\bar{A}$.

After, self-attention is used to prune the skeleton graph for learning a group of attention scores coding the importance of all frames with respect to the present frame and merely selecting the top-K frames in the temporal region. The self-attention unit input is denoted by $H \triangleq \{h_0, h_1, \dots, h_{T-1}\}$, where $h_t \in \mathbb{R}^N$ is the hidden state vector at interval t with N nodes. Based on this, Q, K and V are defined in Eq. (3),

$$Q = W_q H, K = W_k H, V = W_v H \quad (3)$$

In Eq. (3), W_q, W_k and W_v are prediction weights. The attention score $S \in \mathbb{R}^{T\times T}$ and the attention layer's output H_{in} are computed in Eq. (4):

$$S = softmax(QK^{\top});\ H_{in} = SV \quad (4)$$

In the task of skeleton generation, S is modified as a covered attention S_{mask} that avoids the present frame from joining to successive frames.

$$S_{mask} = \begin{pmatrix} s_{0,0} & 0 & \cdots & 0 \\ s_{1,0} & s_{1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_{T-1,0} & s_{T-1,1} & \cdots & s_{T-1,T-1} \end{pmatrix}_{T\times T} \quad (5)$$

In Eq. (5), the element $s_{m,n}$ is the n^{th} frame's guidance on the m^{th} frame and values in the upper triangle are equivalent to zero. To implement the pruning, the top-K scores are chosen in all rows of the S_{mask} and another element is assigned as zero. Observe that, when the total non-zero elements in a few rows is smaller than K, each non-zero element can be maintained. Eventually, the adjacent matrix is created as Eq. (6):

$$\tilde{A}_s = S_{mask} \odot \tilde{A} \triangleq \begin{pmatrix} s_{0,0} * \bar{A} & 0 & \cdots & 0 \\ s_{1,0} * I & s_{1,1} * \bar{A} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_{T-1,0} * I & s_{T-1,1} * I & \cdots & s_{T-1,T-1} * \bar{A} \end{pmatrix}_{(N*T)\times(N*T)} \quad (6)$$

The self-attention-based graph convolutional unit (afore activation) result is,

$$H^{(1)} = D^{-1}\tilde{A}_s H_{in} W \quad (7)$$

In Eq. (7), $D^{ii} = \sum_j \tilde{A}_s^{ij}$ is the diagonal node degree matrix for regularizing $\tilde{A}_s$, $H^{(1)}$ indicates the

hidden state next to the initial graph convolutional layer in Fig. 3. The graph convolution process is performed in Eq. (7) utilizing alike $\tilde{A}_s$ for 5 times. The outcome of the 5th graph convolutional unit is $H^{(5)}$. Following a linear unit, the outcome of the generator is obtained, which is a created skeleton image $x \in \mathbb{R}^{T\times N}$.

For training the PS-GAN, the skeleton sequence length is set at 25. The $\mathcal{D}_f$ arbitrarily chooses 20 frames from each created skeleton image and learning images as the input. The self-attention graph convolution layer chooses the first 5 earlier frames to build $\tilde{A}_s$. The learning and testing batch sizes are set to 100 and 500, respectively and the training rate is set to 0.001. Conversely, the design of good GCN structures for GANs takes more time, effort and domain knowledge, i.e., designing complex generators or discriminator backbones to well create high-resolution images.

To solve this problem, an effective automated structure search model for PS-GAN is highly essential. From this perspective, the RL-based GCN structure search model is proposed to find the best network structure of the PS-GAN.

3.3 Reinforcement learning for PS-GAN structure search

The goal is to improve training task speed by using step-wise sampling instead of whole trajectory-based sampling and utilizing historical knowledge from previous strategies. To accomplish this, the PS-GAN structure search problem is formulated as an MDP and solved using off-policy RL.

3.3.1. Problem formation

The PS-GAN structure search problem is an MDP using state-based sampling to improve training and reduce stemming from whole structures, making it easier to learn strategies using off-policy data.

An MDP is defined as a tuple (S, A, r, P, ρ), where S denotes the group of states that can accurately define the current structure (like the configuration, the efficiency of the structures, etc.), A denotes the group of actions that signifies the structure of the successive cell, $r(s,a)$ denotes the remuneration value utilized to describe how good the structure is, $P(s'|s,a)$ denotes the switch probability specifying the learning task and $\rho(s)$ denotes the primary structure. A cell is defined as a structural block used for exploration in a single epoch.

This problem formulation allows us to potentially create a new global optimizer through collective remuneration deprived of the complexity of determining remuneration across the complete trajectory simultaneously.

3.3.2. Off-policy RL solver

The off-policy RL is combined in the PS-GAN structure search using the MDP formulation. To construct the off-policy solver, the state, reward and action are designed for satisfying the requirements of the PS-GAN structure model and the MDP.

1. State: MDP needs a state interpretation that will accurately define the present system up to the present epoch. It wishes to be robust in the learning process to avoid totaling high differences in the policy system's learning. The robustness constraint is significant because the policy system depends on it to configure the successive cells. An advanced state interpretation is proposed for the PS-GAN structure search to enhance the generation efficiency of intermediate RGB results for all structure cells. For a constant batch of input noise, the mean outcome of all cells is adopted as the advanced state interpretation. This interpretation is downsampled to carry out a fixed dimension across multiple cells. Additionally, network performance and the number of units are used to obtain additional data regarding the state. To abridge, the configured state s comprises the depth, performance of the present structure and advanced state interpretation.

2. Action: The policy system determines the subsequent action for the current state that codes the data for earlier layers. It defines the structure of a single cell. For instance, when the search space is followed based on the AutoGAN, the action can include skip connections, upsampling processes, shortcut connections, various kinds of convolution units and the regularization unit. This is described by $a = [conv, norm, upsample, shortcut, skip]$. The agent's action outcome can be performed by a softmax categorizer decoding into a process. For an effective off-policy scheme, an equivalent search space called AutoGAN is used, which refers to the search for generator and the discriminator structure is pre-configured and emerges as the generator.

3. Remuneration: It is designed for efficiency enhancement after inserting additional cells. The Inception Score (IS) and Frchet Inception Distance (FID) are used as measures of network efficiency. Because the IS value is tolerant and the FID value is degressive, the new reward function is formulated as:

$$R_t(s,a) = IS(t) - IS(t-1) + \alpha\big(FID(t-1) - FID(t)\big) \quad (8)$$

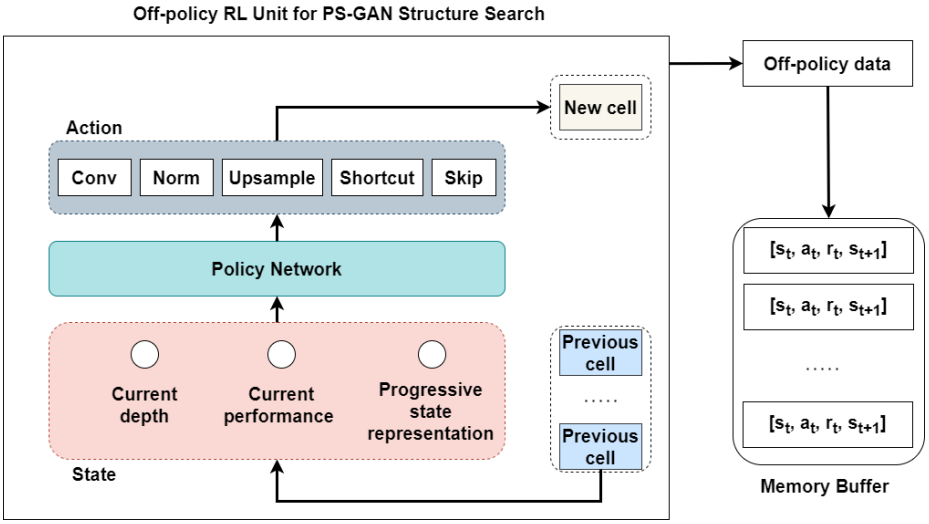


Figure. 4 Overview of off-policy RL model for PS-GAN structure search

In Eq. (8), α is a variable to balance the tradeoff between the 2 measures. By applying the performance enhancement in all steps, the RL will increase the expected sum of rewards over the whole trajectory. This allows us to target the promising global best structure with the maximum reward:

$$J(\pi) = \sum_{t=0} \mathbb{E}_{(s_t,a_t)\sim p(\pi)} R(s_t, a_t) = \mathbb{E}_{structure\sim p(\pi)} IS_{final} - \alpha FID_{final} \quad (9)$$

In Eq. (9), IS_{final} and FID_{final} are the final scores of the whole structure. After designing the state, reward and action, the off-the-shelf soft actor-critic RL scheme is applied depending on the maximum entropy RL model as the training algorithm. The actor intends at increasing expected reward and entropy. This maximizes learning stability and search ability during learning.

For the training of the critic, the fitness function is described by

$$J(Q) = \mathbb{E}_{(s,a)\sim B}\left[\frac{1}{2}\left(Q(s,a) - Q_{target}(s,a)\right)^2\right] \quad (10)$$

In Eq. (10), Q_{target} is the approximation target for Q:

$$Q_{target}(s,a) = Q(s,a) + \gamma Q_{target}\left(s' f(\epsilon, s')\right) \quad (11)$$

The fitness function of the policy system is provided as:

$$J(\pi) = \mathbb{E}_B\left[\beta\left[\log\left(\pi_\theta(f_\theta(\epsilon,s)|s)\right) - Q\left(s, f_\theta(\epsilon,s)\right)\right]\right] \quad (12)$$

In Eq. (12), π_θ is parameterized by the GCN f_θ, ϵ defines an input vector containing Gaussian noise, $B = \{(s,a,s',r)\}$ defines the restate buffer to store the MDP tuples and β defines the positive Lagrange multiplier to control the relative significance of the policy entropy.

Fig. 4 illustrates an overview of the off-policy RL model for the PS-GAN structure search. The whole procedure involves 5 tasks: (i) the agent recognizes the present state (s_t) denoted by $s = [Depth, Performance, Progressive\ state]$, (ii) the agent creates a result a_t on how to configure the cell extra to earlier cells based on the state data, whereas a_t comprises the skip connection, upsampling processes, shortcut connections, various kinds of convolution units and a regularization unit, (iii) gradually learn the novel structure, get remuneration r_t and novel state s_{t+1} data and then repeat it, (iv) store the off-policy memory tuple $[s_t, a_t, r_t, s_{t+1}]$ into the memory buffer B and (v) test a batch of information from B to update the policy system.

3.3.3. Execution of RL-based PS-GAN

1. Agent learning: As the GCN structure search problem is reformulated as a multi-phase MDP, the agent can create many results in a trajectory $\tau = [(s_1, a_1), \ldots, (s_n, a_n)]$. In all steps, the agent can gather this experience $[s_t, a_t, r_t, s_{t+1}]$ in the B. After obtaining the minimum memory size threshold, the agent is modified by the Adam optimization, which uses the fitness function given in Eq. (12), via sampling a batch of information from B in an off-policy manner. The whole search contains 2 intervals: the search interval and the exploitation interval. During the search interval, the agent can sample a

promising structure. During the exploitation interval, the agent can select the optimal structure to rapidly improve the strategy. The search interval continues for 70% of epochs and the exploitation considers 30% of epochs. After attaining the memory threshold, for all search steps, the strategy can be modified. For all exploitation steps, the strategy can be modified 10 times to converge rapidly.

2. Proxy procedure: An advanced proxy procedure efficiently gathers remunerations by training a cell trajectory for a single iteration and calculating the remuneration for the new cell. The weight of the GCN is reassigned after completing the structure trajectory design to estimate the Q-value of all state-action pairs. The training procedure for RL-based PS-GAN is presented below.

Algorithm for RL-based PS-GAN structure search

Input: hyperparameters, learning rates $\alpha_{\phi_Q}, \alpha_\theta$

Output: Optimal PS-GAN structure

1. Initialize a Q-system $Q(s,a)$ and policy system $\pi(a|s)$ with variables ϕ_Q, θ and the Lagrange multipliers β in a random manner;
2. Initialize the variables of target systems with $\bar{\phi}_Q \leftarrow \phi_Q, \bar{\theta} \leftarrow \theta$;
3. $\boldsymbol{for}(all\ iterations)$
4. Reassign the weight and cells of PS-GAN;
5. $\boldsymbol{for}(all\ time\ steps)$
6. $\boldsymbol{if}(exploration)$
7. Sample a_t from $\pi(s)$, insert the respected cell to the PS-GAN;
8. $\boldsymbol{else\ if}(exploitation)$
9. Select the optimal a_t from $\pi(s)$ and the respected cell to the PS-GAN;
10. $\boldsymbol{end\ if}$
11. Gradually train the PS-GAN;
12. Note s_{t+1}, r_t and accumulate (s_t, a_t, r_t, s_{t+1}) in B;
13. $\boldsymbol{end\ for}$
14. $\boldsymbol{for}(all\ update\ steps)$
15. Sample mini-batches of alterations from B and modify Q and π with gradients;
16. Adjust the target systems with soft substitution:
17. $\bar{\phi}_Q \leftarrow \tau\phi_Q + (1-\tau)\bar{\phi}_Q$;
18. $\bar{\theta} \leftarrow \tau\theta + (1-\tau)\bar{\theta}$;
19. $\boldsymbol{end\ for}$
20. $\boldsymbol{end\ for}$

Thus, the optimal PS-GAN structure is trained to create high-quality skeleton images, which are fed to a CNN with a softmax classifier. The trained model classifies test video sequences into CP and healthy infants for earlier diagnosis.

4. Experimental result

The efficiency of the PS-GAN-CNN is assessed by implementing it in Python 3.7.8. A comparative analysis is conducted between the proposed and methods such as PredictMed [17], FCNet [18], CNN-LSTM [21] and KBRNN [25]. For this purpose, all these methods and proposed PS-GAN-CNN are implemented and tested for the given MINI-RGBD, babyPose and MIA datasets, which helps to comprehend the successfulness of the PS-GAN-CNN method. Table 2 lists the parameter settings for existing and proposed methods.

The performance measures are described below.

- Accuracy: It is the proportion of appropriately detected CP skeleton images over the total images tested.

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (13)$$

In Eq. (13), TP is the number of healthy images detected as healthy, TN is the number of CP images detected as CP, FP is the number of CP images

Table 2. Parameter settings for proposed and existing methods

Algorithms	Parameters	Range
PredictMed [17]	Maximum number of iterations	200
	Probability threshold	0.5
FCNet [18]	Learning rate	0.0005
	Number of epochs	4000
	Batch size	3
	Dropout rate	0.5
CNN-LSTM [21]	Training rate	0.001
	Dropout rate	0.5
	Number of epochs	100
	Batch size	20
	Optimizer	Adam
KBRNN [25]	Learning rate	0.001
	Number of hidden states	100
	Number of epochs	150
	Batch size	25
	Dropout rate	0.2
	Optimizer	Adam
Proposed PS-GAN-CNN	Learning rate	0.001
	Batch size	64
	Number of epochs	120
	Momentum	0.99
	Dropout rate	0.5
	Weight decay	0.0005

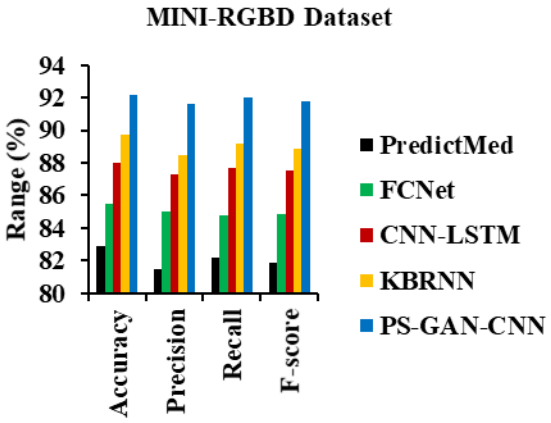


Figure. 5 Performance analysis for proposed and existing CP detection methods using MINI-RGBD dataset

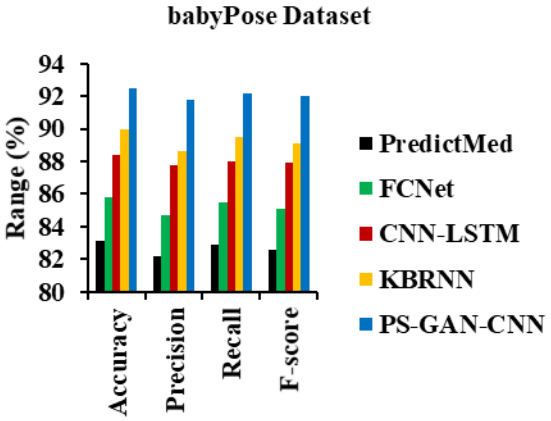


Figure. 6 Performance analysis for proposed and existing CP detection methods using babyPose dataset

detected as healthy and FN is the number of healthy images detected as CP.

- Precision: In Eq. (14), it measures the properly detected skeleton images at TP and FP rates.

$$Precision = \frac{TP}{TP+FP} \tag{14}$$

- Recall: In Eq. (15), it measures the fraction of skeleton images that are properly detected at TP and FN rates.

$$Recall = \frac{TP}{TP+FN} \tag{15}$$

- F-score (F): It is computed in Eq. (16),

$$F = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{16}$$

Fig. 5 depicts the effectiveness of various methods on the MINI-RGBD database for detecting and diagnosing CP. It observes that the accuracy of

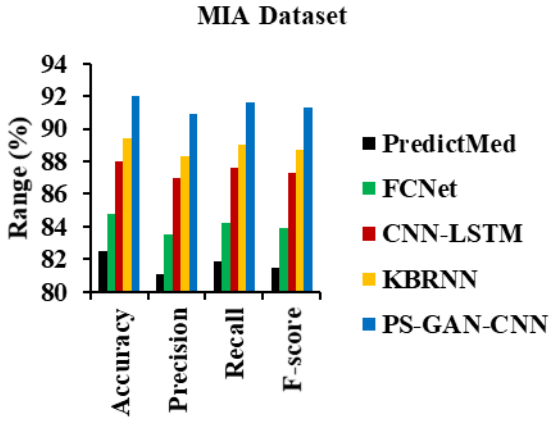


Figure. 7 Performance analysis for proposed and existing CP detection methods using MIA dataset

the PS-GAN-CNN method is higher than the other existing methods due to the use of self-attention GCN with off-policy RL for choosing the optimal structure of PS-GAN, which can generate more high-quality skeleton images for the classification process. From this scrutiny, it is addressed that the accuracy of the PS-GAN-CNN method is 11.22% superior to the PredictMed, 7.84% superior to the FCNet, 4.77% superior to the CNN-LSTM and 2.79% superior to the KBRNN methods. The precision of the PS-GAN-CNN method is 12.4%, 7.8%, 4.9% and 3.5% greater than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. The recall of the PS-GAN-CNN method is 11.9%, 8.5%, 4.9% and 3.1% better than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. As well, the f-score of the PS-GAN-CNN is 12.2%, 8.1%, 4.9% and 3.3% superior to the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively.

Fig. 6 portrays the effectiveness of various methods on the babyPose database for detecting and diagnosing CP. It is observed that the accuracy of the PS-GAN-CNN method is improved by 11.31%, 7.81%, 4.64% and 2.78% compared to the PredictMed, FCNet, CNN-LSTM and KBRNN methods. The precision of the PS-GAN-CNN method is 11.68%, 8.38%, 4.56% and 3.61% greater than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. The recall of the PS-GAN-CNN method is 11.22%, 7.84%, 4.77% and 3.02% better than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. As well, the f-score of the PS-GAN-CNN is 11.38%, 8.11%, 4.66% and 3.25% superior to the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively.

Fig. 7 illustrates the effectiveness of various methods on the MIA database for detecting and diagnosing CP. It is observed that the accuracy of the

PS-GAN-CNN method is improved by 11.52%, 8.49%, 4.55% and 2.91% compared to the PredictMed, FCNet, CNN-LSTM and KBRNN methods. The precision of the PS-GAN-CNN method is 12.08%, 8.86%, 4.48% and 2.94% greater than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. The recall of the PS-GAN-CNN method is 11.84%, 8.79%, 4.57% and 2.92% better than the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively. As well, the f-score of the PS-GAN-CNN is 12.02%, 8.82%, 4.58% and 2.93% superior to the PredictMed, FCNet, CNN-LSTM and KBRNN methods, respectively.

Thus, it is realized that the PS-GAN-CNN can augment the number of infant skeleton images for training and maximize the accuracy of detecting CP significantly.

5. Conclusion

This paper presents a PS-GAN-based data augmentation method for generating high-quality skeleton images and recognizing CP infants. The method uses self-attention and pruned dense graphs to train the PS-GAN, which encodes spatial joints and temporal properties. The off-policy RL-based model is used to choose the optimal structure for skeleton image generation. The generated images are then passed to a CNN with a softmax classifier for detecting CP. Finally, the test results proved that the PS-GAN-CNN on the MINI-RGBD, babyPose and MIA databases has an accuracy of 92.2%, 92.5% and 92%, respectively in contrast with the PredictMed, FCNet, CNN-LSTM and KBRNN methods to identify CP from the infant's general movements. But the occlusions in video frames may impact the detection performance. Hence, the future work focus on handling occlusion problem in CP detection and increasing the accuracy efficiently.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, methodology, software, validation, Rajalekshmy; formal analysis, investigation, Sheeja ; resources, data curation, writing—original draft preparation, Rajalekshmy; writing—review and editing, Rajalekshmy; visualization, supervision, Sheeja.

References

[1] D. R. Patel, M. Neelakantan, K. Pandher, and J. Merrick, “Cerebral Palsy in Children: A Clinical Overview”, *Translational Pediatrics*, Vol. 9 No. Suppl 1, pp. S125-S135, 2020.

[2] A. Jöud, A. Sehlstedt, K. Källén, L. Westbom, and L. Rylander, “Associations between Antenatal and Perinatal Risk Factors and Cerebral Palsy: A Swedish Cohort Study”, *BMJ Open*, Vol. 10, No. 8, pp. 1-8, 2020.

[3] H. S. Sheedy, E. Waight, S. Goldsmith, S. Reid, C. Gibson, L. Watson, and G. Hinwood, “Declining Trends in Birth Prevalence and Severity of Singletons with Cerebral Palsy of Prenatal or Perinatal Origin in Australia: A Population-Based Observational Study”, *Developmental Medicine & Child Neurology*, Vol. 64, pp. 1114-1122, 2022.

[4] H. C. Glass, A. T. Costarino, S. A. Stayer, C. Brett, F. Cladis, and P. J. Davis, “Outcomes for Extremely Premature Infants”, *Anesthesia and Analgesia*, Vol. 120, No. 6, pp. 1337-1351, 2015.

[5] A. T. Velde, C. Morgan, I. Novak, E. Tantsis, and N. Badawi, “Early Diagnosis and Classification of Cerebral Palsy: An Historical Perspective and Barriers to an Early Diagnosis”, *Journal of Clinical Medicine*, Vol. 8, No. 10, pp. 1-13, 2019.

[6] A. J. Spittle, C. Morgan, J. E. Olsen, I. Novak, and J. L. Cheong, “Early Diagnosis and Treatment of Cerebral Palsy in Children with a History of Preterm Birth”, *Clinics in Perinatology*, Vol. 45, No. 3, pp. 409-420, 2018.

[7] A. M. Asalu, G. Taylor, H. Campbell, L. L. Lelea, and R. S. Kirby, “Cerebral Palsy: Diagnosis, Epidemiology, Genetics and Clinical Update”, *Advances in Pediatrics*, Vol. 66, pp. 189-208, 2019.

[8] T. C. A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, “A Comprehensive Study on Deep Learning-Based 3D Hand Pose Estimation Methods”, *Applied Sciences*, Vol. 10, No. 19, pp. 1-27, 2020.

[9] Q. Wu, G. Xu, M. Li, L. Chen, X. Zhang, and J. Xie, “Human Pose Estimation Method Based on Single Depth Image”, *IET Computer Vision*, Vol. 12, No. 6, pp. 919-924, 2018.

[10] J. Seesahai, M. Luther, P. T. Church, P. Maddalena, E. Asztalos, T. Rotter, and R. Banihani, “The Assessment of General Movements in Term and Late-Preterm Infants Diagnosed with Neonatal Encephalopathy, as a Predictive Tool of Cerebral Palsy by 2 Years of Age—A Scoping Review”, *Systematic Reviews*, Vol. 10, No. 1, pp. 1-13, 2021.

[11] A. K. Kwong, J. E. Olsen, A. L. Eeles, C. Einspieler, K. J. Doyle, and A. J. Spittle, “Occurrence of and Temporal Trends in Fidgety

General Movements in Infants Born Extremely Preterm/Extremely Low Birthweight and Term-Born Controls", *Early Human Development*, Vol. 135, pp. 11-15, 2019.

[12] 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", *Sensor*, Vol. 20, No. 18, p. 5321, 2020.

[13] R. Krynauw, J. C. F. D. Preez, J. I. V. Zyl, and M. Burger, "The Trajectory of General Movements from Birth until 12-14 Weeks Corrected Age in Very Low-Birthweight and Extremely Low-Birthweight Infants Born Preterm", *South African Journal of Child Health*, Vol. 16, No. 2, pp. 99-104, 2022.

[14] M. Airaksinen, O. Räsänen, E. Ilén, T. Häyrinen, A. Kivi, V. Marchi, and S. Vanhatalo, "Automatic Posture and Movement Tracking of Infants with Wearable Movement Sensors", *Scientific Reports*, Vol. 10, No. 1, pp. 1-13, 2020.

[15] 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.

[16] F. Farahanipad, M. Rezaei, M. S. Nasr, F. Kamangar, and V. Athitsos, "A Survey on GAN-Based Data Augmentation for Hand Pose Estimation Problem", *Technologies*, Vol. 10, No. 2, pp. 1-13, 2022.

[17] C. M. Bertoncelli, P. Altamura, E. R. Vieira, S. S. Iyengar, F. Solla, and D. Bertoncelli, "PredictMed: A Logistic Regression-Based Model to Predict Health Conditions in Cerebral Palsy", *Health Informatics Journal*, Vol. 26, No. 3, pp. 2105-2118, 2020.

[18] K. D. McCay, E. S. Ho, H. P. Shum, G. Fehringer, C. Marcroft, and N. Embleton, "Abnormal Infant Movements Classification with Deep Learning on Pose-Based Features", *IEE Access*, Vol. 8, pp. 51582-51592, 2020.

[19] K. D. McCay, P. Hu, H. P. Shum, W. L. Woo, C. Marcroft, N. D. Embleton, and E. S. Ho, "A Pose-Based Feature Fusion and Classification Framework for the Early Prediction of Cerebral Palsy in Infants", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol, 30, pp. 8-19, 2021.

[20] B. N. 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 Journal of Biomedical and Health Informatics*, Vol. 25, No. 10, pp. 3911-3920, 2021.

[21] D. Sakkos, K. D. Mccay, C. Marcroft, N. D. Embleton, S. Chattopadhyay, and E. S. Ho, "Identification of Abnormal Movements in Infants: A Deep Neural Network for Body Parts-Based Prediction of Cerebral Palsy", *IEEE Access*, Vol. 9, pp. 4281-94292, 2021.

[22] R. Yang, H. Zuo, S. Han, X. Zhang and Q. Zhang, "Computer-Aided Diagnosis of Children with Cerebral Palsy Under Deep Learning Convolutional Neural Network Image Segmentation Model Combined with Three-Dimensional Cranial Magnetic Resonance Imaging", *Journal of Healthcare Engineering*, pp. 1-11, 2021.

[23] S. Reich, D. Zhang, T. Kulvicius, S. Bölte, K. N. Saines, F. B. Pokorny, and P. B. Marschik, "Novel AI Driven Approach to Classify Infant Motor Functions", *Scientific Reports*, Vol. 11, No. 1, pp. 1-13, 2021.

[24] P. Palraj and G. Siddan, "Deep Learning Algorithm for Classification of Cerebral Palsy from Functional Magnetic Resonance Imaging (fMRI)", *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 3, pp. 718-724, 2021.

[25] D. Li, J. Qu, Z. Tian, Z. Mou, L. Zhang, and X. Zhang, "Knowledge-Based Recurrent Neural Network for TCM Cerebral Palsy Diagnosis", *Evidence-Based Complementary and Alternative Medicine*, pp. 1-10, 2022.

[26] 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. of the European Conf. on Computer Vision Workshops*, pp. 32-49, 2019.

[27] N. Hesse, S. Pujades, M. J. Black, M. Arens, U. G. Hofmann, and A. S. Schroeder, "Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D Sequences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 10, pp. 2540-2551, 2019.

[28] L. Migliorelli, S. Moccia, R. Pietrini, V. P. Carnielli, and E. Frontoni, "The BabyPose Dataset", *Data in Brief*, Vol. 33, p. 106329, 2020.

[29] https://vrai.dii.univpm. it/mia-dataset