Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Occlusion-Robust Pose Sequence Aware GAN based Classification using CNN for Early Detection of Cerebral Palsy in Infants

[1]Rajalekshmy K.D., [2]Dr. Thomson Fredrik,
[1]Research Scholar,Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore ,India, rajidevaraj16@gmail.com
[2]Professor, Department of Technology, Karpagam Academy of Higher Education,Coimbatore ,India, thomson500@gmail.com

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Early detection and diagnosis of Cerebral Palsy (CP) are crucial for minimizing its impact. Previous studies have used pose estimation techniques called the OpenPose for CP detection, but these methods have limitations in annotating large infant movement datasets. To address this, a new method called Pose Sequence-aware Generative Adversarial Network (PS-GAN) has been developed to create high-quality skeleton images, followed by the Convolutional Neural Network (CNN) with softmax classifier for CP detection. However, pose estimation techniques can still have recognition errors in the left upper limbs since the left upper limbs have extra movements and a few movements were rigorously obstructed with the torso segment. This results in missing values in the Feature Matrix (FM) used for CP classification. To solve this, this article proposes an Occlusion-Robust PS-GAN-CNN (OR-PSGAN-CNN) model for detecting CP in infants from video. First, it uses an OpenPose model to estimate infant skeletal joint positions, which are then augmented through the PS-GAN. Then, the coordinates of the infant's joints are extracted into FM based on the matrix encoding, along with the extraction of joint motion complexity and joint motion correlation features. These extracted FM are fed to the CNN with a softmax classifier to detect CP. Thus, this model handles occlusion by transforming the skeletons into the FM, substituting the missing coordinates with zeros, resulting in high accuracy. Finally, experiments results show that the OR-PSGAN-CNN model achieves 93.7%, 93.3%, and 93.2% accuracy on the MINI-RGBD, babyPose, and MIA datasets, respectively, outperforming existing CP detection models.

Keywords: Cerebral palsy, PS-GAN, Occlusion, Matrix encoding, Joint motion complexity, Joint motion correlation

I. INTRODUCTION

Cerebral Palsy is a brain damage condition affecting muscle control in infants, affecting movement, posture, and communication. It affects 17 million people worldwide, with higher incidences in preterm newborns, with 32.4 cases per 1000 births between 28-32 weeks [1-2]. Timely recognition of CP is crucial for the early intervention and brain adaptation [3]. Pose estimation and GMA are primary CP prevention strategies. Two types of newborn posture estimation are RGB-based and depth information-based [4]. Deep learning models like CNN can predict infant 2D poses [5]. However, using Kinect for motion capture in similar positions is challenging [6]. A Model-based Recursive Matching (MRM) approach has been developed for learning-free posture estimation but has limited computational effectiveness [7].

The General Movement Assessment (GMA) is a reliable and cost-effective method used to detect abnormal movement patterns in infants. It identifies potential risks by detecting the absence or infrequency of fidgety movements [8]. However, manual analysis can be time-consuming and inaccessible for many infants. Automated GMA methods, using machine learning models, have been explored for diagnosing cerebral palsy [9]. In normal development, body movements remain coordinated, with muscle contractions and relaxations controlling overall movements. To address the challenges of pose estimation and GMA techniques, Wu et al. [10] developed a novel approach using RGB-D videos to predict cerebral palsy in newborns. This method combines RGB images with depth information to generate 3D coordinates of an infant in a prone position, allowing for a comprehensive assessment of the infant's CP risk. However, the Part Affinity Field (PAF) technique used in this approach had limitations in estimating skeletal images, and annotating a large dataset of newborn movements was challenging. Various GAN model variations were explored [11], but they encountered limitations in capturing spatial relationships among joints and temporal characteristics across frames without an expensive pre-learning stage. To create high-quality skeleton images for CP identification, a PS-GAN was developed [12]. The PS-GAN utilized self-attention to capture long-range dependencies in continuous frames and a Graph Convolutional Network (GCN) to encode spatial joints and temporal properties. The selection of the best PS-GAN structure was optimized using Reinforcement Learning (RL). Finally, the CNN with a softmax classifier was trained to detect CP using the generated skeleton images.

1.1 Problem Description

While this pose estimation module is efficient at estimating the infant pose, it still has a few detection inaccuracies. For example, occlusion can affect the CP detection performance. Occlusion occurs when the human body is partially obscured on the screen by other people or objects. The PS-GAN-CNN model, which is skeleton-based, encounters occlusion when the infant's body skeleton information is not fully visible. Current CP detection models do not have a specific strategy for handling occlusion and endure calculation with the unoccluded segment. This results in a huge amount of absent values in the FM, which can lead to insufficient features for the subsequent classifier to accurately classify CP infants.

1.2 Major Contributions of the Paper

This paper proposes the OR-PSGAN-CNN model for detecting CP infants from video sequences. The main aim is to address the occlusion issue in extracting FM without missing values during pose estimation and to significantly increase the detection accuracy. The primary contributions of this paper include:

1. Initially, RGB-D videos of infant movements are collected, and the OpenPose technique is applied for infant pose estimation. This allows the estimation of joint positions or the skeleton of the infant from the RGB-D videos. The estimated skeleton images are augmented by the PS-GAN, which generates high-quality skeleton images.
2. Then, the skeleton images are transformed into the FM by extracting the coordinates of the infant's joints based on the matrix encoding approach. Additionally, joint motion complexity and joint motion correlation features are extracted, which can differentiate the infant's spontaneous motion patterns.
3. Moreover, those extracted FM are fed to the CNN followed by a softmax classifier for detecting infants with CP disorder. According to these pose estimation and feature extraction, the occlusion problem can be resolved, and the detection accuracy can be increased with the help of adequate features.
4. Finally, extensive experiments demonstrate that the proposed model outperforms existing CP detection models using various datasets. Accordingly, this model could improve automated monitoring of CP and other movement disorders in infants when full body visibility cannot be guaranteed.

The following article is prepared as follows: Section II covers literature survey. Section III explains the OR-PSGAN-CNN and Section IV discusses its efficacy. Section V précises the study and offers suggestions for future developments.

II. LITERATURE SURVEY

This section discusses recent deep-learning models for detecting CP using video sequences of normal and CP infants. These models analyze infant movements and poses from video data to enable early detection of CP. Sakkos et al. [13] developed a CNN-Long Short-Term Memory (LSTM) model to categorize infant body movements related to CP using pose characteristics from RGB video frames. They also used a visualization feature to identify body parts most relevant for CP recognition. But the model's performance was hindered by issues such as camera movements, illumination changes, noise, and occlusion, leading to low sensitivity in estimating infant motion and detecting CP.

A new Spatio-Temporal Attention-based Model (STAM) [14] was developed to detect CP by analyzing fidgety motions. Initially, individual postures were extracted from video clips and the spatiotemporal GCN was applied to represent newborn movements. A spatiotemporal attention mechanism was then applied to select body areas containing discriminatory features related to nerve movements for CP prediction. However, the accuracy of the pose estimation algorithm was crucial since limb occlusion can introduce noise and reduce effectiveness.

Wu et al. [15] developed an innovative technique using joint characteristic coding to predict infant pose from depth images captured with Kinect V2. The method encodes the offset vector in the confidence zone of all joints and exports joint locating and link encoding via CNN model. However, it only extracts the infant's 2D pose and did not fully utilize the spatial data in the depth picture. Zhang et al. [16] developed a Pytorch-based Attention-informed GCN called CP-AGCN to recognize infants at risk of CP using skeletal information from RGB videos. They created a frequency-binning unit to capture CP movements in the frequency domain and filter out noise. Additionally, they developed a Frequency Attention-Informed GCN (FAI-GCN) [17] to utilize the frequency information of infant's movement for CP prediction. However, the accuracy of the pose estimation task, which can be affected by occlusion or camera movements, impacts the model performance.

Li et al. [18] developed a Knowledge-Based Recurrent Neural Network (KBRNN) using the CP information graph. They used an evolution scheme to extract knowledge from the graph and then fed the data into an RNN with tensor embedding to determine the relationship between symptoms and disorders in medical records for CP prediction. However, the model's accuracy was affected by noise, missing values, and limited labeled data.

Gao et al. [19] developed a deep learning motion assessment model that uses infant videos and basic characteristics to analyze the GMA at the fidgety movements stage. The model employs 3D pose estimation to predict joint coordinates and a distance representation approach to capture motion patterns. However, the model's performance was impacted by noise and occlusion. Also, it only included high-risk infants, so the model's effectiveness on normal infants was unknown.

Wu et al. [20] introduced an innovative learning-free technique for quantifying infant impulsive movements to predict CP from videos. The technique involved extracting the joints of the infant using a pose estimation scheme, segmenting the skeleton sequence into several clips using a sliding window, and clustering them to recognize infant CP. However, its accuracy was affected by noise or occlusion errors that occurred during the pose estimation of input infants.

Recent research has focused on using deep learning models such as CNN-LSTM, attention mechanisms, and GCNs to analyze infant movements and poses from video data to detect early signs of CP. These models extract important spatiotemporal features from the input infants' videos. However, challenges such as noise, occlusion, camera movement, and limited labeled training data can hinder performance. Therefore, a more robust model is needed to accurately analyze subtle infant motions for reliable early CP prediction.

III. PROPOSED METHODOLOGY

This section delivers a detailed description of the OR-PSGAN-CNN framework. The presented study is outlined in Figure 1.

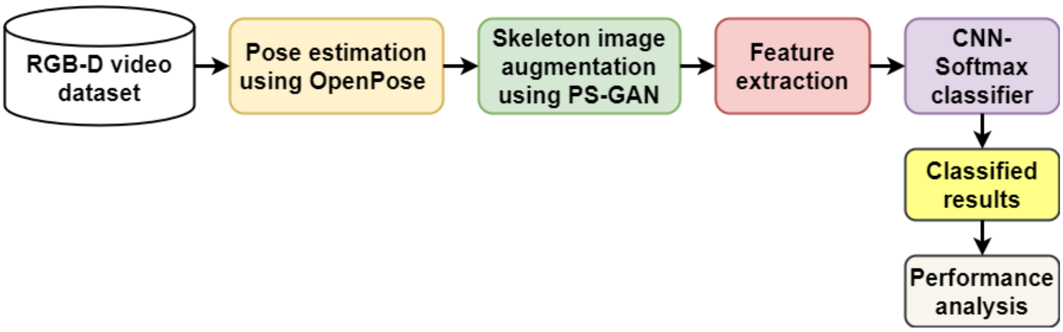


Figure 1. An Outline of the Presented Study

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

3.1 Data Collection

This study utilizes three publicly available infant motion analysis datasets:

1. MINI-RGBD dataset [21]: A collection of 12 RGB-D video sequences of 6-month-old infants, providing accurate silhouette, texture, and motion data. It anonymizes information and labels videos using the GMA scheme to determine nervous motions.
2. BabyPose dataset [22]: Contains sixteen depth videos of preterm infants' motion in NICUs, with annotated limb-joint positions for twelve joints.
3. Motion Infant Analysis (MIA) dataset [23]: Comprises state vector and timestamp data derived from depth measurements from an RGB-D sensor positioned above the child in a supine position on the crib.

After acquiring the dataset, OpenPose [24] is used to estimate the coordinates of infant joints, resulting in skeleton images, which are then given as input to the PS-GAN [12] for augmenting the number of infant skeleton images. Then, the skeleton information is encoded into the FM using the matrix encoding approach. Additionally, features such as joint angle, joint motion complexity, and joint motion correlation are extracted. Finally, these characteristics are fed to the CNN with a softmax categorizer to detect infant CP. The following sections provide the details of feature extraction and detection.

3.2 Feature Extraction

3.2.1 Skeleton to Matrix Encoding Approach

The model uses the GCN to learn features and determine skeleton coordinates. While the body is obstructed, the GCN is unable to determine coordinates owing to missing features. Infants' lower body skeleton coordinates are occluded by right/left limbs, as shown in Figure 2. To address this issue, the study adopts an encoding approach that can convert skeleton information into the FM, making it easier for the GCN to handle missing information.

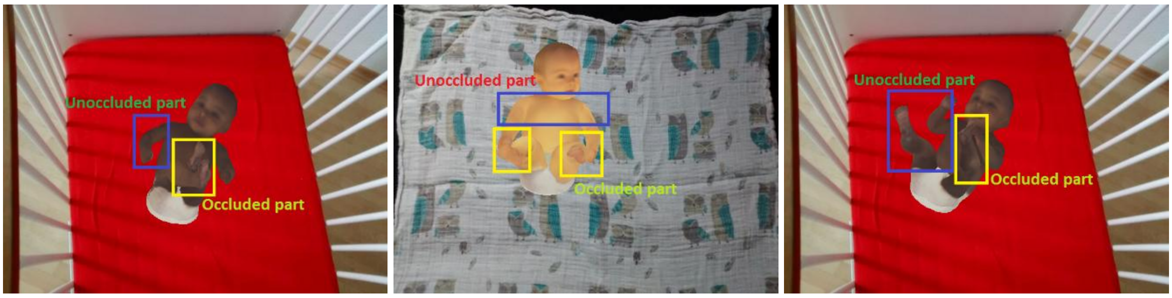


Figure 2. Infant Movement Images under Occlusion

As depicted in Figure 3, the encoding approach utilizes the following types of features: (i) skeleton length γ_i^t, signifies the length of 2 nearby joints, (ii) skeleton cosine λ_i^t, signifies the cosine angle between the skeleton and the horizontal direction, (iii) the length vector between the head and all joints ς_i^t, signifies the distance between the head and all joints, and (iv) the cosine of the vector between the head and all joints κ_i^t, signifies the cosine angle from the line of head and all joints and the horizontal direction. The certain conversions of such different types of characteristics are presented in Eqns. (1) – (11).

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

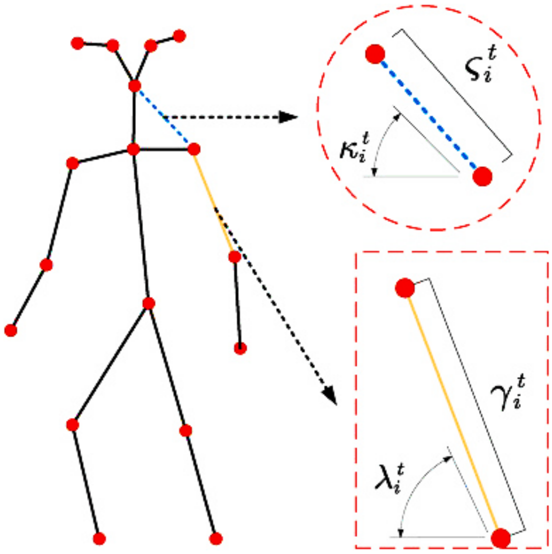


Figure 3. Schematic Representation of Skeleton Features

The skeleton length is defined by

$$\gamma_i^t = \|(x_i^t, y_i^t) - (x_j^t, y_j^t)\| \quad (1)$$

The length vector is defined by

$$F_\gamma^t = [\gamma_1^t \; \gamma_2^t \dots \gamma_n^t] \quad (2)$$

To avoid innumerable, $\lambda_i^t = 0$ when $\gamma_i^t = 0$, therefore the skeleton angle is given as:

$$\lambda_i^t = \frac{|x_i^t - x_j^t|}{\gamma_i^t} \quad (3)$$

The angle feature vector is provided by

$$F_\lambda^t = [\lambda_1^t \; \lambda_2^t \dots \lambda_n^t] \quad (4)$$

The length feature of the head to the body joints is given as:

$$\varsigma_i^t = \|(x_i^t, y_i^t) - (x_0^t, y_0^t)\| \quad (5)$$

The length feature vector is defined as:

$$F_\varsigma^t = [\varsigma_1^t \; \varsigma_2^t \dots \varsigma_n^t] \quad (6)$$

To avoid innumerable, $\kappa_i^t = 0$ when $\varsigma_i^t = 0$, therefore ς_i^t is given by

$$\kappa_i^t = \frac{|x_i^t - x_0^t|}{\varsigma_i^t + e^\beta} \quad (7)$$

The angle feature vector is defined as:

$$F_\kappa^t = [\kappa_1^t \; \kappa_2^t \dots \kappa_n^t] \quad (8)$$

Therefore, the total FM is defined by

$$F = [F^1 \; F^2 \; \vdots \; F^t] \quad (9)$$

In conclusion, four distinct FMs such as $F_\lambda, F_\gamma, F_\varsigma$, and F_κ are created. The Sobel operator is used to calculate the gradient matrix of four FMs by multiplying them together. The operator is defined in Eq. (10) and is used to determine the level of skeleton data between two frames.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

$$G_y = [1\ 2\ 1\ 0\ 0\ 0\ -1\ -2\ -1] \quad (10)$$

The absolute FM utilized for detection is $\hat{F}$, as given by

$$\hat{F} = [F_\lambda\ F_\gamma\ F_\varsigma\ F_\kappa\ G_\lambda\ G_\gamma\ G_\varsigma\ G_\kappa] \quad (11)$$

3.2.2 Joint Motion Complexity Features

When identifying infant CP, angle parameters are more accurate than position parameters. Limb length differences can cause people to move differently, but angle changes will always be the same. To deal with specific variations, joint angle is used as the features for further processing. Figure 4 illustrates the delineation of joint angles. The left subfigure shows the joints and the vector link among them (in red), while the right subfigure defines the angle characteristics $\{\lambda_i, i = 1, \ldots, 14\}$ among vectors.

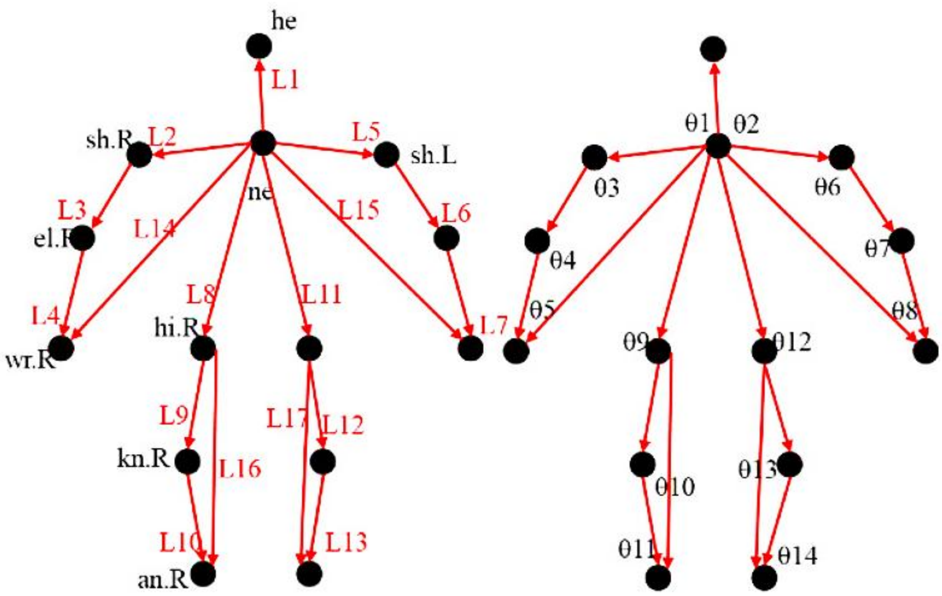


Figure 4. Delineation of Joint Angles in Infant Skeleton

Using angle characteristics can help reduce the impact of individual variations in limb length. In this study, the angular feature time series $\{[\lambda_i]_t\}$ includes the following independent variables: interval t and keypoint number i.

The performance of joint motion features in motion patterns varies, with simple repeated motions being straightforward and complex motions resulting in diverse and unpredictable performances. The system's nonlinear complexity characterizes these ordered, changeable, and simple traits. Normal spontaneous movement is complex due to its changeability and irregularity, while aberrant spontaneous movement lacks complexity and is monotonous. So, the GP time-series relationship is used to quantity the nonlinear difficulty of movement forms.

The time-series $J_t = \{[\lambda_i]_t, i = 1, \ldots, 14; t = 1, \ldots, N\}$ is reconstructed and embedded in an m-dimensional Euclidean space R^m to obtain a vector set $J(m)$. The components of $J(m)$ are represented by

$$X_n(m, \tau) = \left([\lambda_i]_n, [\lambda_i]_{n+\tau}, \ldots, [\lambda_i]_{n+(m-1)\tau}\right), n = 1, \ldots, N_m \quad (12)$$

In Eq. (12), τ denotes the delay, which is an integer multiple of 2 nearby sampling periods $N_m = N - (m-1)\tau$. The X_i is selected from X_n as a reference point, and the distance from another $N_m - 1$ point to X_i is calculated as:

$$d_{ij} = d(X_i, X_j) = \|X_i - X_j\|_2 \quad (13)$$

For each data point, this procedure is repeated. After that, the correlation integral function is calculated as follows:

$$C_m(r) = \frac{2}{N_m(N_m-1)} \sum_{i,j=1}^{N_m} H(r - d_{ij}), r > 0 \quad (14)$$

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

In Eq. (14), H is Heaviside function and represented by

$$H(x) = \{1, x > 0\ 0, x \leq 0 \tag{15}$$

After that, m-dimensional time-related dimension is defined by

$$D_2(m) = \frac{ln ln\ C_m(r)}{ln ln\ r} \tag{16}$$

If $D_2(m)$ doesn't modify with an increase in the phase space element m, then it indicates that the complexity of the full-body movement of the joint characteristics $\{[\lambda_i]_t\}$ is becoming stable. In this case, D_2 represents the dimension related to the sequence.

$$D_2 = D_2(m) \tag{17}$$

3.2.3 Joint Motion Correlation Features

For $J_i = \{[\lambda_t]_i, i = 1, \dots, 14\}$, the Spearman Correlation Coefficient Matrix (SCCM) is computed among all feature dimensions θ_t. The SCCM is utilized to define the relationship among joints in body movement. Consider that multiple keypoint feature sequences $[\lambda_t]_i$ and $[\lambda_t]_j$ have an equal rank, then the SCC $(\rho_{i,j})$ between $[\lambda_t]_i$ and $[\lambda_t]_j$ is determined as follows:

$$\rho_{i,j} = \frac{\sum_t \left([\lambda_t]_i - \underline{[\lambda_t]_i}\right)\left([\lambda_t]_j - \underline{[\lambda_t]_j}\right)}{\sqrt{\sum_t \left([\lambda_t]_i - \underline{[\lambda_t]_i}\right)^2 \sum_t \left([\lambda_t]_j - \underline{[\lambda_t]_j}\right)^2}} \tag{18}$$

In Eq. (18), $\underline{[\lambda_t]_i}$ refers to the mean of i^{th} λ_i in t sequence. By computing the cross-correlation factor among each feature, the SCCM is obtained. If the infant's movements are complex, the connections between the angle characteristics are moderately fragile, indicating that the SCCM can efficiently reveal the relationship among the movement feature sequences of each joint in the infant's body.

3.3 Infant Cerebral Palsy Detection Using CNN-Softmax Classifier

Moreover, the extracted features such as $\hat{F}$, D_2, and relative SCCM are given to the CNN-Softmax classifier. The CNN-Softmax classifier architecture is illustrated in Figure 5, containing 3 convolutional (conv) units, 3 pooling units, 2 Fully Connected (FC) units and a softmax. Table 1 presents more information about the CNN-Softmax classifier design.

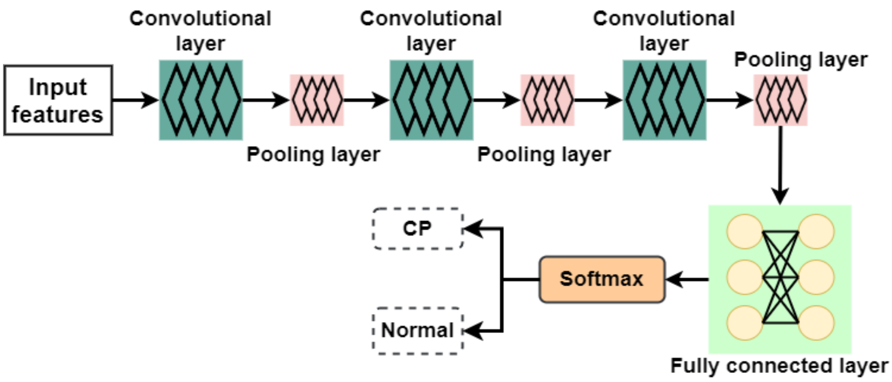


Figure 5. Structure of CNN-Softmax for Infant CP Detection

Table 1. Architecture of CNN-Softmax Classifier

Layers	Output size	Filter size/Stride
Input	128×64×1	-
Conv 1	128×64×64	3×3/1

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Pooling 1	64×32×64	2×2/2
Conv 2	64×32×128	3×3/1
Pooling 2	32×16×128	2×2/2
Conv 3	32×16×256	3×3/1
Pooling 3	16×8×256	2×2/2
FC 1	1×1×512	-
FC 2	1×1×256	-
Softmax	2	-

This CNN-Softmax classifier is trained using the parameters listed in Table 2 to detect infant CP even under occlusion conditions.

Table 2. Training Parameters for CNN-Softmax Classifier

Parameter	Value
Training rate	0.001
Batch size	64
Epochs	120
Optimizer	Adam
Momentum	0.99
Dropout rate	0.5
Weight decay	0.0005
Loss function	Mean square error

Algorithm 1: OR-PSGAN-CNN Model for Infant CP Detection

Input: RGB-D videos of infant movements

Output: CP detection

1. **Begin**
2. Collect RGB-D video dataset of infant movements;
3. $for(each\ video)$
4. Apply OpenPose to estimate infant skeletal joint positions;
5. Get skeleton joint coordinates;
6. Augment the estimated skeleton images using the PS-GAN;
7. $end\ for$
8. $for(each\ skeleton\ image)$
9. Convert skeleton to FM $\hat{F}$ using the matrix encoding approach;
10. Compute the joint motion complexity feature D_2;

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

11. Calculate the joint motion correlation features SCCM;
12. $end\ for$
13. Feed the extracted features to the CNN-Softmax classifier;
14. Train the classifier to detect normal and CP infants;
15. Evaluate the model performance;
16. **End**

The time complexity of the OR-PSGAN-CNN model is $O(n \times j^2)$, where n represents the amount of frames, j represents the amount of skeleton joints. This is dominated by the matrix encoding process. The space complexity is $O(nm)$, where m is the feature dimension,.

IV. EXPERIMENTAL RESULTS

The efficiency of the OR-PSGAN-CNN is evaluated with existing models such as PS-GAN-CNN [12], STAM [14], CP-AGCN [16], and FAI-GCN [17]. All models are implemented in Python 3.7.8 using three distinct datasets as described in Section 3.1 for evaluation.

4.1 Performance Evaluation Metrics

- Accuracy: It is the proportion of properly detected CP infants to the overall infants tested.

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \tag{19}$$

  In Eq. (19), TP represents CP infants correctly detected as CP, TN represents healthy infants correctly identified as healthy, FP represents healthy infants incorrectly identified as CP, and FN represents CP infants incorrectly recognized as healthy.

- Precision: It is calculated by

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

- Recall: It is calculated by

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

- F1-score (F1): It is determined as:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{22}$$

4.2 Analysis of Various CP Detection Models on MINI-RGBD Dataset

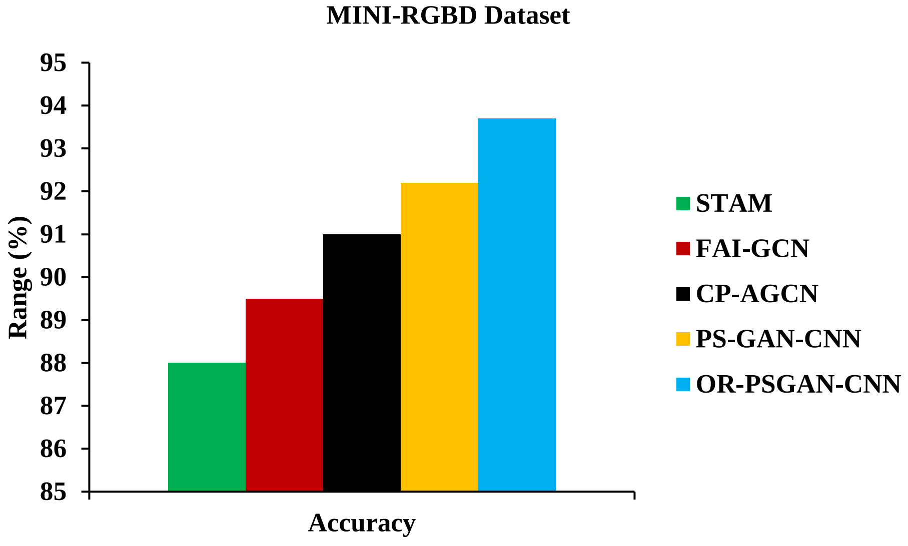
Table 3 presents the test results of various models on the MINI-RGBD dataset for CP detection.

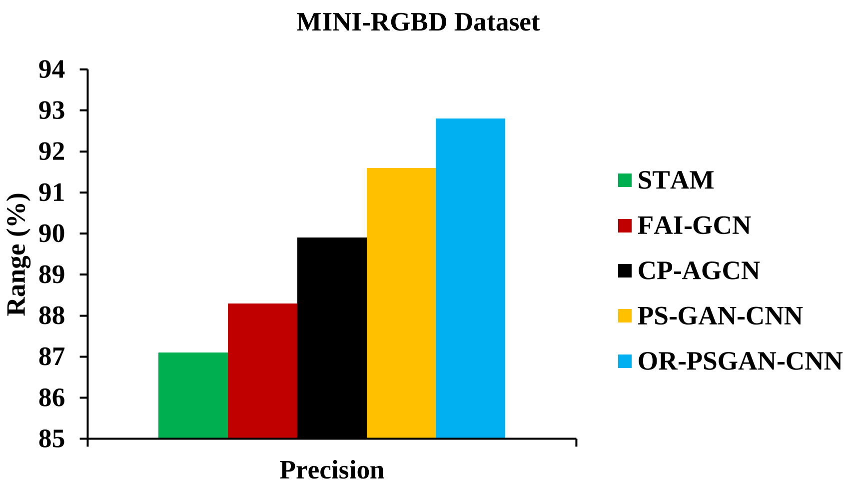
Table 3. Comparison of Different CP Detection Models on MINI-RGBD Dataset

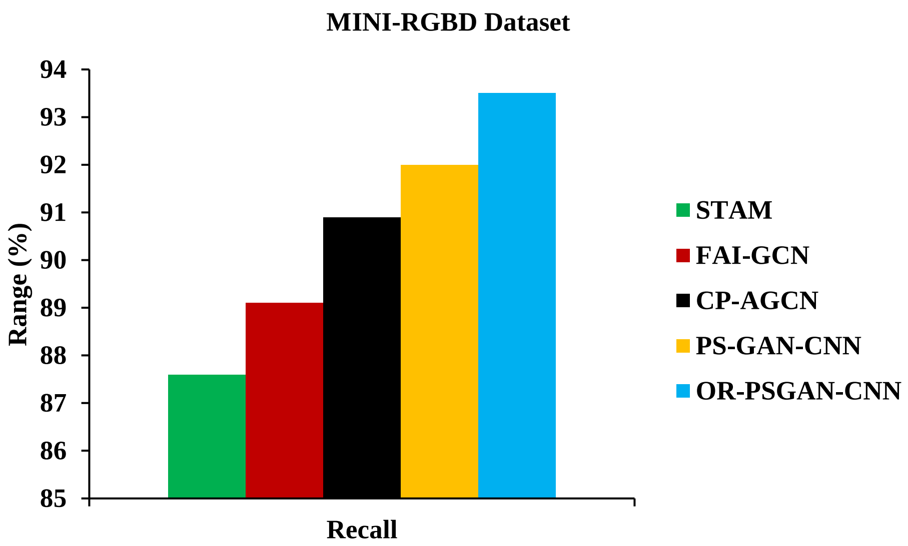
Metrics	STAM	FAI-GCN	CP-AGCN	PS-GAN-CNN	OR-PSGAN-CNN
Accuracy (%)	88.0	89.5	91.0	92.2	93.7
Precision (%)	87.1	88.3	89.9	91.6	92.8
Recall (%)	87.6	89.1	90.9	92.0	93.5
F-score (%)	87.4	88.7	90.4	91.8	93.2

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article







Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

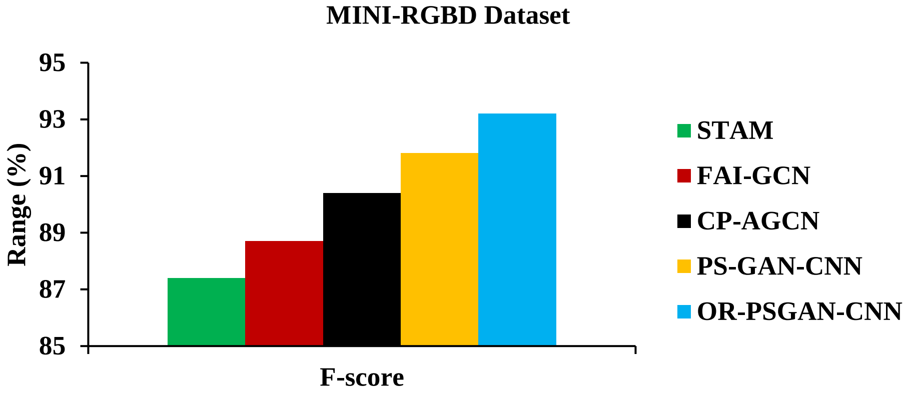


Figure 6. Comparison of OR-PSGAN-CNN Model against Existing CP Detection Models on MINI-RGBD Dataset

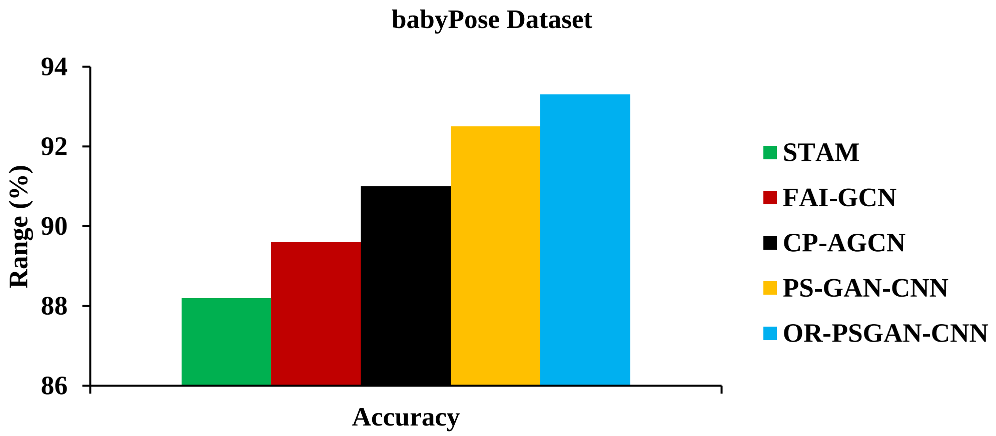
Figure 6 compares the proposed OR-PSGAN-CNN model with the conventional models for infant CP detection on the MINI-RGBD dataset. The OR-PSGAN-CNN model achieves the highest accuracy of 93.7%, outperforming STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN by margins of 6.48%, 4.69%, 2.97%, and 1.63% respectively. This demonstrates the occlusion handling capability of the proposed model through robust skeleton feature encoding. For precision, OR-PSGAN-CNN obtains 92.8% compared to 87.7%, 89.1%, 90.9%, and 92.9% for the other methods. The recall follows a similar trend with OR-PSGAN-CNN achieving 93.5%, highlighting the improved TP rate. The F1-score of OR-PSGAN-CNN is 93.2%, which is 6.64%, 5.07%, 3.1%, and 1.53% higher than the STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN models, respectively.

4.3 Analysis of Various CP Detection Models on babyPose Dataset

Table 4 displays the test results of various models on the babyPose dataset for CP detection.

Table 4. Comparison of Different CP Detection Models on babyPose Dataset

Metrics	STAM	FAI-GCN	CP-AGCN	PS-GAN-CNN	OR-PSGAN-CNN
Accuracy (%)	88.2	89.6	91.0	92.5	93.3
Precision (%)	87.5	88.8	90.2	91.8	92.6
Recall (%)	88.1	89.4	90.9	92.2	93.0
F-score (%)	87.8	89.1	90.6	92.0	92.8



Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

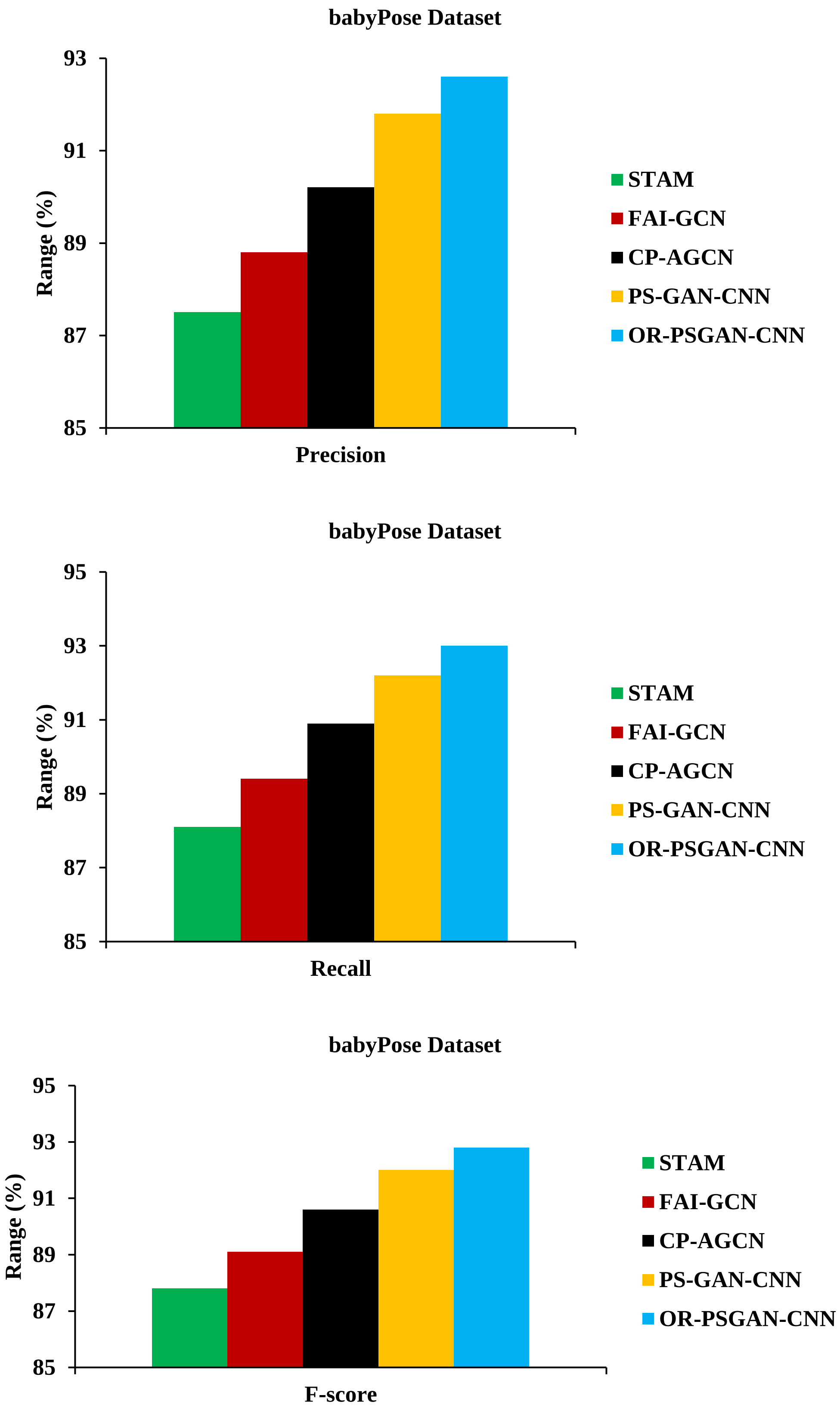


Figure 7. Comparison of OR-PSGAN-CNN Model against Existing CP Detection Models on babyPose Dataset

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

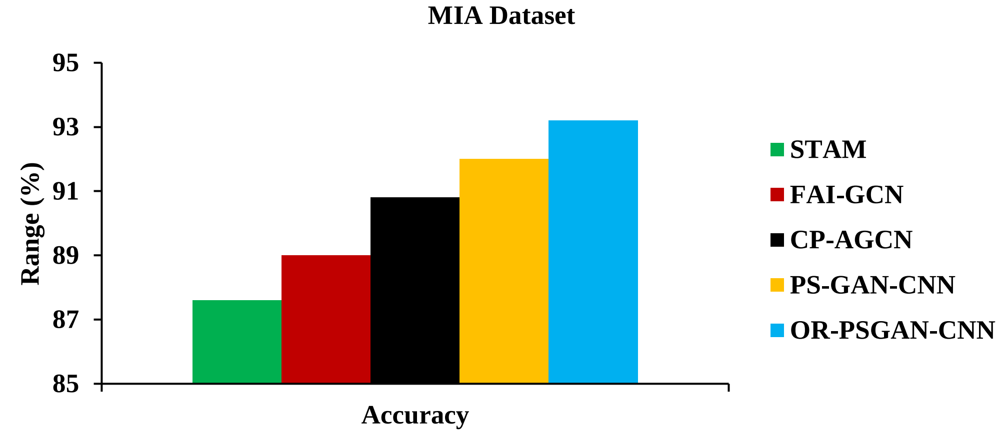
Figure 7 shows the efficiency of the OR-PSGAN-CNN compared to conventional models for CP detection in infants on the babyPose dataset. The OR-PSGAN-CNN model achieves the highest accuracy of 93.3%, outperforming STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN by margins of 5.78%, 4.13%, 2.53%, and 0.86% respectively. The precision of OR-PSGAN-CNN is 92.6%, which is 5.83%, 4.28%, 2.66%, and 0.87% higher than the other models, demonstrating its lower FP rate. The OR-PSGAN-CNN also achieves a recall of 93%, compared to 88.1%, 89.4%, 90.9%, and 92% for the other models, showing an improved TP rate. The F1-score gained by the OR-PSGAN-CNN is 92.8%, exceeding STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN models by 5.69%, 4.15%, 2.43%, and 0.87% correspondingly. Such outcomes highlight the efficiency of the suggested occlusion-handling model in extracting discriminative features from infant skeletal motions for reliable CP detection on the babyPose dataset.

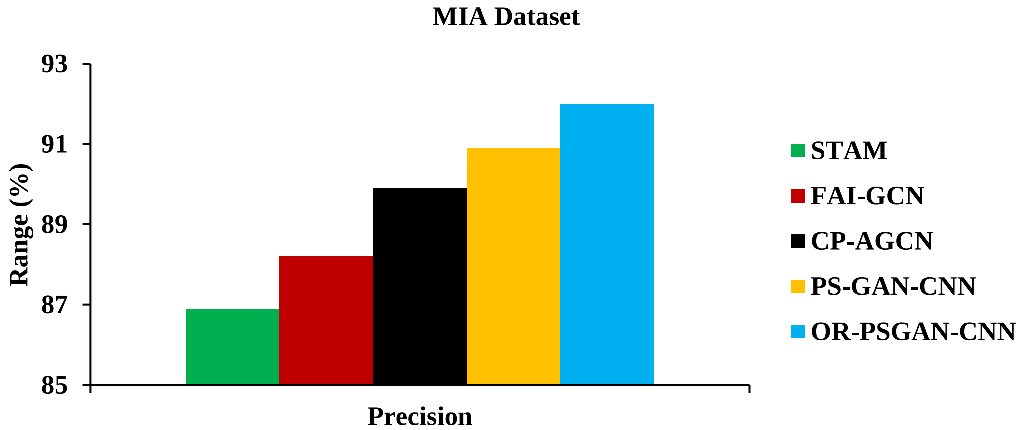
4.3 Analysis of Various CP Detection Models on MIA Dataset

Table 5 presents the performance outcomes of various models on the MIA dataset for CP detection.

Table 5. Comparison of Different CP Detection Models on MIA Dataset

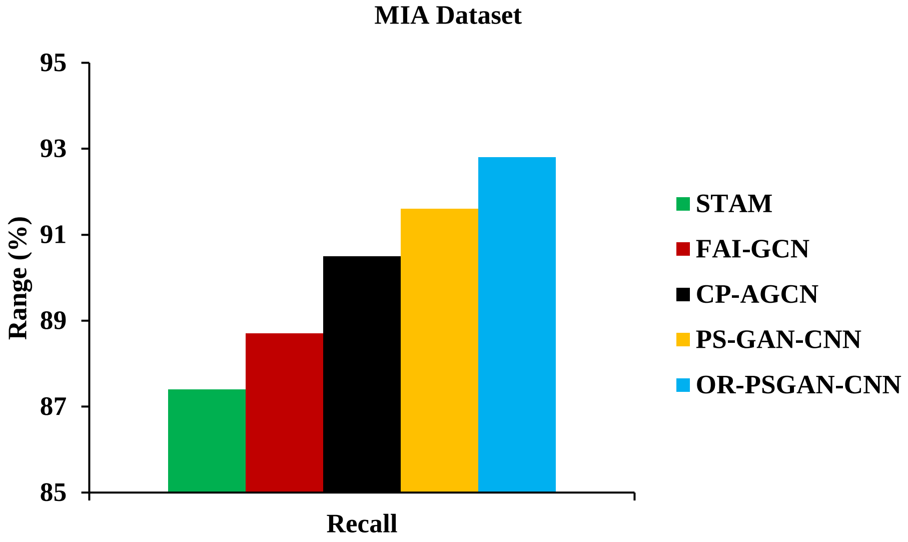
Metrics	STAM	FAI-GCN	CP-AGCN	PS-GAN-CNN	OR-PSGAN-CNN
Accuracy (%)	87.6	89.0	90.8	92.0	93.2
Precision (%)	86.9	88.2	89.9	90.9	92.0
Recall (%)	87.4	88.7	90.5	91.6	92.8
F-score (%)	87.2	88.5	90.2	91.3	92.4





Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



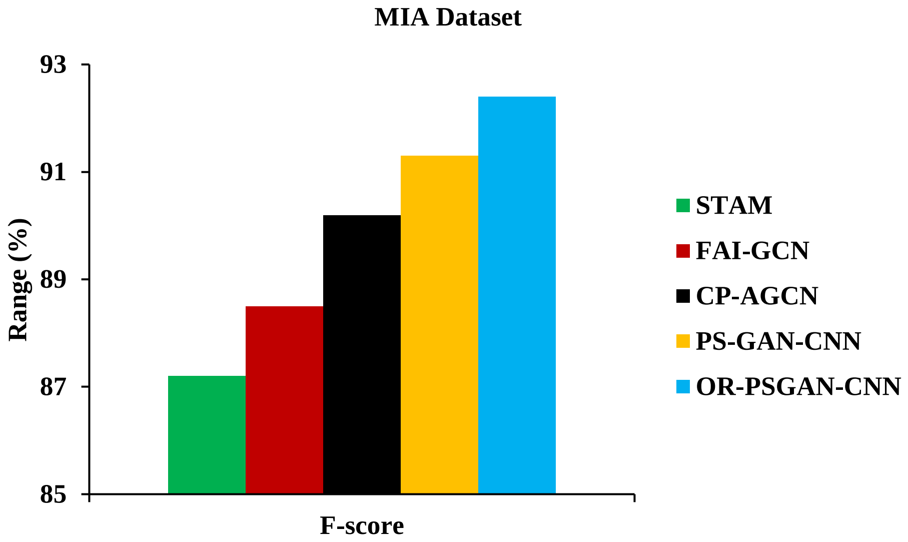


Figure 8. Comparison of OR-PSGAN-CNN Model against Existing CP Detection Models on MIA Dataset

Figure 8 illustrates the performance evaluation of the proposed OR-PSGAN-CNN and state-of-the-art models using the MIA database for CP detection. The OR-PSGAN-CNN achieves the highest accuracy of 93.2%, outperforming STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN by margins of 6.39%, 4.72%, 2.64%, and 1.3% respectively, demonstrating the robustness of the proposed model respectively. For precision, the OR-PSGAN-CNN obtains 92% compared to 86.9%, 88.2%, 89.9%, and 90.9% for other models, showing reduced FP rates. Similarly, the OR-PSGAN-CNN has the top recall of 92.8%, exceeding STAM, FAI-GCN, CP-AGCN, and PS-GAN-CNN by 6.18%, 4.62%, 2.54%, and 1.31%, highlighting detection of more TP rates. The F1-score achieved by the OR-PSGAN-CNN is 92.4%, which is 5.96%, 4.41%, 2.44% and 1.2% higher than the STAM, FAI-GCN, CP-AGCN and PS-GAN-CNN respectively.

V. CONCLUSION

This paper presents the OR-PSGAN-CNN model for improved detection of cerebral palsy in infants using video data. The model applies PS-GAN for skeleton image augmentation and uses a matrix encoding approach to transform joint coordinates into FMs. It also extracts joint motion complexity and correlation features. The CNN-Softmax classifier then learns these features to detect CP infants. Extensive experiments show that the OR-PSGAN-CNN model outperforms existing models on various datasets. The results demonstrate 93.7%, 93.3%, and 93.2% accuracy on the MINI-RGBD, babyPose, and MIA datasets, respectively, outstanding other models. The algorithmic complexity is quadratic in the number of video frames and linear in the number of CNN layers. As future work, this occlusion-robust framework can be extended to related problems of detecting developmental disorders using full-body motion analysis under imperfect visibility, enabling automated screening of high-risk infants for timely intervention through intelligent video analysis techniques.

REFERENCES

[1] Mushta, S. M., Khandaker, G., Power, R., & Badawi, N. (2021). Cerebral palsy in the Middle East: epidemiology, management, and quality of life. In *Handbook of Healthcare in the Arab World*, Cham: Springer International Publishing, pp. 2539-2572.

[2] Mushta, S. M., King, C., Goldsmith, S., Smithers-Sheedy, H., Badahdah, A. M., Rashid, H., ... & McIntyre, S. (2022). Epidemiology of cerebral palsy among children and adolescents in Arabic-speaking countries: a systematic review and meta-analysis. *Brain Sciences*, *12*(7), 859.

[3] Glass, H. C., Li, Y., Gardner, M., Barkovich, A. J., Novak, I., McCulloch, C. E., & Rogers, E. E. (2021). Early identification of cerebral palsy using neonatal MRI and general movements assessment in a cohort of high-risk term neonates. *Pediatric Neurology*, *118*, 20-25.

[4] Paul, S., Nahar, A., Bhagawati, M., & Kunwar, A. J. (2022). A review on recent advances of cerebral palsy. *Oxidative Medicine and Cellular Longevity*, *2022*, 1-20.

[5] Leo, M., Bernava, G. M., Carcagnì, P., & Distante, C. (2022). Video-based automatic baby motion analysis for early neurological disorder diagnosis: state of the art and future directions. *Sensors*, *22*(3), 866.

[6] Lin, Y. (2023). A deep learning algorithm-based visual strategy intervention study for children with autism spectrum disorders–extraction and detection of children's behavioral features. *European Review for Medical & Pharmacological Sciences*, *27*(11), 4914-4928.

[7] Adama, D. A., Lotfi, A., & Ranson, R. (2021). A survey of vision-based transfer learning in human activity recognition. *Electronics*, *10*(19), 2412.

[8] Redd, C. B., Karunanithi, M., Boyd, R. N., & Barber, L. A. (2021). Technology-assisted quantification of movement to predict infants at high risk of motor disability: A systematic review. *Research in Developmental Disabilities*, *118*, 104071.

[9] Silva, N., Zhang, D., Kulvicius, T., Gail, A., Barreiros, C., Lindstaedt, S., ... & Marschik, P. B. (2021). The future of general movement assessment: The role of computer vision and machine learning–a scoping review. *Research in Developmental Disabilities*, *110*, 103854.

[10] Wu, Q., Xu, G., Wei, F., Chen, L., & Zhang, S. (2021). RGB-D videos-based early prediction of infant cerebral palsy via general movements complexity. *IEEE Access*, *9*, 42314-42324.

[11] Farahanipad, F., Rezaei, M., Nasr, M. S., Kamangar, F., & Athitsos, V. (2022). A survey on GAN-based data augmentation for hand pose estimation problem. *Technologies*, *10*(2), 1-13.

[12] Devarajan, R. K., & Khader, S. S. (2023). Pose sequence-aware generative adversarial network for augmenting skeleton sequences to improve cerebral palsy detection by deep learner. *International Journal of Intelligent Engineering & Systems*, *16*(5), 512-522.

[13] Sakkos, D., Mccay, K. D., Marcroft, C., Embleton, N. D., Chattopadhyay, S., & Ho, E. S. (2021). Identification of abnormal movements in infants: A deep neural network for body part-based prediction of cerebral palsy. *IEEE Access*, *9*, 94281-94292.

[14] Nguyen-Thai, B., Le, V., Morgan, C., Badawi, N., Tran, T., & Venkatesh, S. (2021). A spatio-temporal attention-based model for infant movement assessment from videos. *IEEE Journal of Biomedical and Health Informatics*, *25*(10), 3911-3920.

[15] Wu, Q., Xu, G., Wei, F., Kuang, J., Qin, P., Li, Z., & Zhang, S. (2022). Supine infant pose estimation via single depth image. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11.

[16] Zhang, H., Ho, E. S., & Shum, H. P. (2022). CP-AGCN: Pytorch-based attention informed graph convolutional network for identifying infants at risk of cerebral palsy. *Software Impacts*, *14*, 100419.

[17] Zhang, H., Shum, H. P., & Ho, E. S. (2022). Cerebral palsy prediction with frequency attention informed graph convolutional networks. In *44th IEEE Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 1619-1625.

[18] Li, D., Qu, J., Tian, Z., Mou, Z., Zhang, L., & Zhang, X. (2022). Knowledge-based recurrent neural network for TCM cerebral palsy diagnosis. *Evidence-Based Complementary and Alternative Medicine*, *2022*, 1-10.

[19] Gao, Q., Yao, S., Tian, Y., Zhang, C., Zhao, T., Wu, D., ... & Lu, H. (2023). Automating General Movements Assessment with quantitative deep learning to facilitate early screening of cerebral palsy. *Nature Communications*, *14*(1), 8294.

[20] Wu, Q., Qin, P., Kuang, J., Wei, F., Li, Z., Bian, R., ... & Xu, G. (2023). A training-free infant spontaneous movement assessment method for cerebral palsy prediction based on videos. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1670-1679.

[21] Hesse, N., Pujades, S., Black, M. J., Arens, M., Hofmann, U. G., & Schroeder, A. S. (2019). Learning and tracking the 3D body shape of freely moving infants from RGB-D sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(10), 2540-2551.

[22] Migliorelli, L., Moccia, S., Pietrini, R., Carnielli, V. P., & Frontoni, E. (2020). The babyPose dataset. *Data in Brief*, *33*, 106329.

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

[24] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Real-time multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291-7299.