Patient-Level Biomechanical Pose Encoding and Temporal Gait Modeling for Pediatric EVGS Assessment

Muhammad Ibrahim Qasmi
Government Graduate College Sahiwal
Affiliated with Bahauddin Zakariya University, Multan
Department of Information Technology
ibrahimqasmi@gpgcs.edu.pk

Abstract

Pediatric gait assessment is clinically important for children with cerebral palsy, but manual visual scoring requires expertise and can be difficult to scale. This report describes our solution for the CVPR 2026 Children Gait Challenge, where the goal is to predict bilateral Edinburgh Visual Gait Score (EVGS) labels and cerebral palsy gait subtypes from multi-view 2D pose sequences. We use a two-stage design. First, a compact temporal convolutional network models fixed-length multi-view pose tensors for left and right gait subtype prediction. Second, a patient-level clinical feature ensemble predicts the 34 bilateral EVGS labels. The feature representation summarizes joint angles, trunk and pelvis posture, foot geometry, cadence proxies, frequency-domain ankle motion, left-right symmetry, and stance/swing phase statistics. A Gradient Boosting, Extra Trees, and Random Forest ensemble is trained with repeated cross-validation and per-label threshold tuning. The final selected submission, `submission_stage3_cgc_t1_stage2t2.csv`, achieved a public leaderboard score of 0.54177. The released code, model artifacts, feature tables, and selected submission are publicly available for reproduction.

1. Introduction

Gait analysis provides a structured way to quantify walking patterns and functional mobility in children with cerebral palsy. Instrumented three-dimensional gait analysis is clinically informative, but it is expensive, specialized, and not always available in low-resource settings. Observational scoring systems such as the Edinburgh Visual Gait Score (EVGS) provide a practical alternative by allowing clinicians to score gait deviations from video [1]. Recent progress in markerless pose estimation and temporal skeleton modeling makes it possible to approximate parts of this workflow from two-dimensional pose sequences [2, 3].

The CVPR 2026 Children Gait Challenge evaluates computer vision systems for pediatric gait analysis [4, 5, 6]. The challenge contains two tasks. Track 1 requires prediction of 17 EVGS labels for each leg, producing 34 binary outputs and a total deviation score. Track 2 requires prediction of left and right gait subtype labels. These tasks are related, but they have different statistical structure: Track 1 is a multi-output clinical scoring task, while Track 2 is a small-sample multi-class classification task.

Our main design decision is to separate the two tracks. For Track 2, we use a compact temporal convolutional network (TCN) over multi-view pose tensors. For Track 1, we use explicit patient-level biomechanical features and tree-based multi-output classifiers. This design reflects the small data regime: temporal neural modeling is useful for subtype dynamics, while interpretable clinical features and threshold control are more stable for the 34 EVGS outputs.

2. Related Work

Clinical gait analysis is widely used to evaluate children with cerebral palsy and to guide treatment planning [7]. EVGS was introduced as a validated observational score for cerebral palsy gait assessment, reducing complex gait deviations into a structured set of clinically meaningful parameters [1]. Robinson et al. further showed that EVGS correlates strongly with 3D gait profile scores and GMFCS levels, supporting its use as a compact clinical description of gait pathology [8]. Automated EVGS work has shown that pose estimation and rule-based event detection can recover several EVGS parameters from video, although some pelvis and hindfoot parameters remain challenging from two-dimensional views [2, 3].

Markerless pose estimation provides the representation used by many low-cost gait systems. OpenPose demonstrated reliable multi-person 2D keypoint estimation with part affinity fields [9]. Kidzinski et al. showed that deep models using OpenPose keypoints from single-camera videos can predict clinical gait metrics such as walking speed, cadence, knee flexion, and gait deviation index in cerebral palsy patients [10]. Lin et al. reported reliable pediatric gait measurements from a video-based markerless system [11]. These results motivate using 2D pose trajectories when full instrumented gait analysis is unavailable.

Temporal modeling of skeleton sequences is a natural fit for gait. Spatial-temporal graph convolutional networks learn motion patterns directly from skeleton graphs [12], and pathological-gait recognition work has applied ST-GCN variants with attention to classify gait abnormalities [13]. Temporal convolutional networks are also strong sequence models and can be more stable than recurrent baselines across diverse sequence tasks [14]. Our approach borrows this temporal modeling idea, but keeps the Track 2 network compact because only a limited number of labeled subtype patients are available.

For Track 1, we favor clinical pose features over a single end-to-end deep model. Recent work on cerebral palsy severity assessment also argues for combining skeleton dynamics with interpretable gait features such as joint angles, symmetry, and step descriptors [15]. Tree ensembles are robust for tabular clinical features and small sample sizes. Random Forests [16], Extremely Randomized Trees [17], and Gradient Boosting [18] provide complementary bias-variance behavior, which we average at probability level before final thresholding.

3. Challenge Data and Tasks

The challenge dataset and baseline framework are introduced by Shen et al. [4]. Each patient contains multiple pose JSON files grouped by camera view. The available views are backward, forward, left, and right. Each frame contains 2D body keypoints and confidence scores. The solution uses only pose-level data and challenge labels; raw RGB frames are not used.

Component	Value
Patients in pose set	110
Pose sequences	1185
Pose frames	339236
Track 1 train patients	94
Track 2 train patients	22
Track 1 test patients	16
Track 2 test patients	9

Table 1: Dataset summary from the released pose JSON files and challenge labels.

Track 1 labels are binary EVGS outputs $L_1, \dots, L_{17}$ and $R_1, \dots, R_{17}$. The total score is the sum of the 34 binary predictions. Track 2 labels are categorical gait subtypes for each side.

4. Method

4.1. Pipeline Overview

The final system is a two-stage pipeline. First, the pose JSON files are indexed by patient and camera view, and frames are sampled uniformly to create fixed-length sequences. Stage 2 uses these sequences as multi-view pose tensors for left and right gait subtype prediction. Stage 3 extracts patient-level clinical and biomechanical descriptors from the same pose streams and predicts the 34 bilateral EVGS labels. The selected submission concatenates the Stage 3 Track 1 rows with the Stage 2 Track 2 rows.

4.2. Stage 2: Temporal Pose Model for Track 2

For each patient and view, frames are sampled uniformly to a fixed length of $T = 128$. Each frame is converted into a tensor with 23 joints and 6 channels: bounding-box-normalized coordinates, hip-centered coordinates, keypoint confidence, and a confidence mask. Missing or short sequences are padded by repeating the last valid frame.

The Track 2 model is a compact temporal convolutional network. It reshapes each view into a sequence of joint-channel vectors, applies one-dimensional convolutional blocks over time, and pools each view by temporal mean and maximum. View descriptors are concatenated and passed to separate left and right classification heads. Cross-entropy loss is used for both sides. Class weighting is applied to reduce the effect of subtype imbalance.

4.3. Stage 3: Clinical Feature Encoding for Track 1

Track 1 is modeled as a patient-level multi-output prediction problem. We aggregate all available sequences for each view, sample up to 500 frames per view, and extract view-wise clinical descriptors. The final feature vector contains view-specific features plus cross-view averages and view standard deviations.

Group	Examples
Joint angles	knee, hip, ankle, ankle-heel angles
Posture	trunk angle, pelvis tilt, shoulder tilt
Foot geometry	toe-heel offsets, foot angle, foot clearance proxies
Motion	ankle/knee/toe speed and acceleration statistics
Cadence	autocorrelation step period, dominant ankle frequency
Symmetry	left-right angle, range, and timing asymmetry
Phase summaries	stance and swing angle statistics

Table 2: Main patient-level feature groups used for Track 1 EVGS prediction.

The feature groups are selected to align with clinical EVGS concepts while remaining computable from noisy 2D pose. Knee, hip, and ankle angle summaries approximate sagittal-plane gait deviations. Trunk, pelvis, and shoulder summaries capture posture. Foot and toe-heel geometry provide proxies for foot progression, heel lift, and clearance. Cadence and frequency features summarize repeated ankle motion. Symmetry features capture bilateral differences that are often important in cerebral palsy gait assessment [7, 10, 15].

4.4. Track 1 Ensemble and Thresholding

The Track 1 feature matrix is filtered with a zero-variance filter and standardized. We train three multi-output classifiers: Gradient Boosting, Extra Trees, and Random Forest. Their positive-class probabilities are averaged. Repeated five-fold cross-validation over five random seeds is used to estimate out-of-fold probabilities. A per-label threshold is then selected from a grid to maximize label accuracy on out-of-fold predictions. The final model is retrained on all Track 1 labeled patients.

Per-label thresholding is important because Track 1 labels have different positive rates. A single threshold tends to over-predict frequent labels and under-predict rare labels. The threshold grid is intentionally coarse to reduce overfitting to the small validation set.

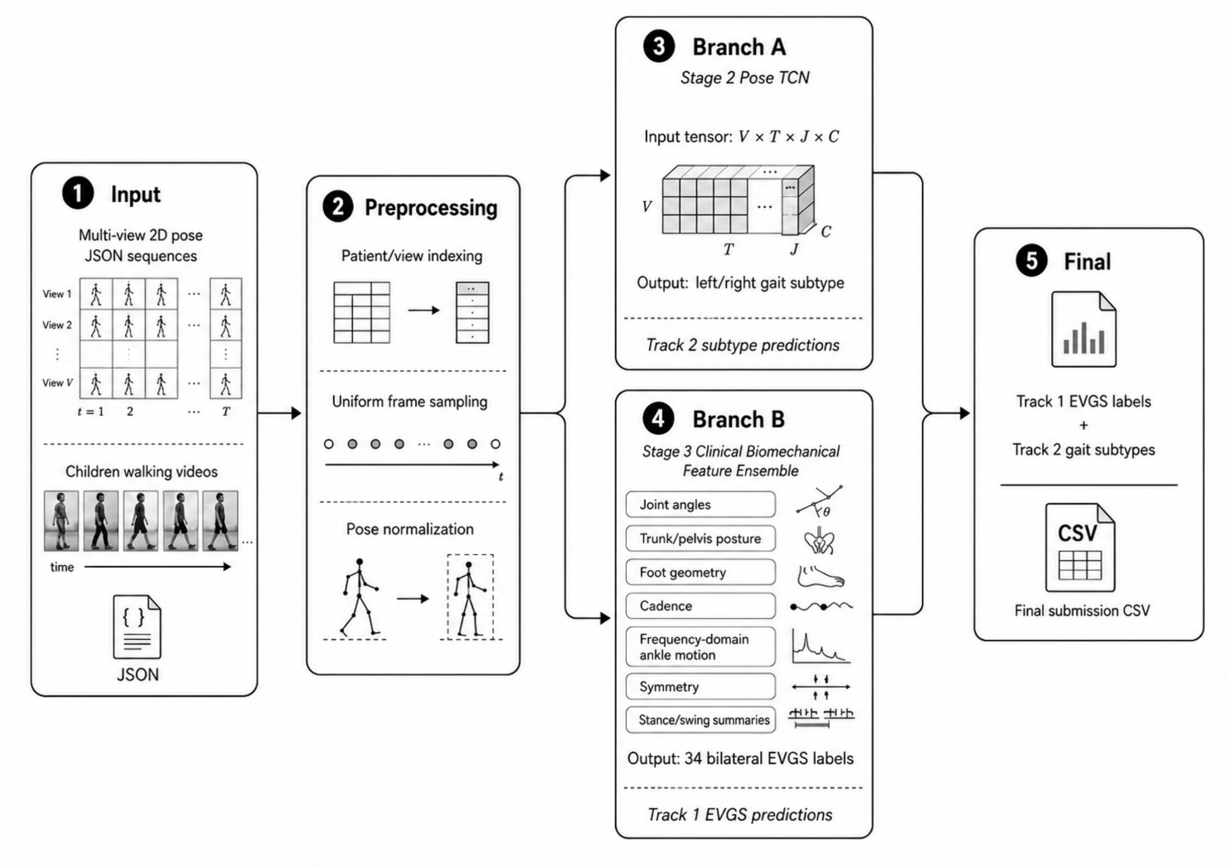

Figure 1: Overview of the proposed two-stage pose-based gait assessment pipeline.

5. Experiments and Results

5.1. Validation Protocol

For Track 1, we report the challenge-style proxy score on out-of-fold predictions:

$$S_1 = \frac{1}{2}\left(\text{Accuracy} + 1 - \frac{\text{RMSE}_{total}}{34}\right). \qquad (1)$$

Accuracy is computed over the 34 binary labels, and RMSE is computed over patient total scores. Track 2 is validated with side-level accuracy and macro-F1, averaged into an S_2 proxy. Because the public leaderboard uses hidden labels, these local scores are used only for model selection.

Model component	Score	Accuracy	RMSE/F1
Track 1 Stage 1 baseline	0.71958	0.63611	6.69630
Track 1 Stage 3 ensemble	0.82619	0.77409	4.13856
Track 2 Pose TCN OOF	0.37340	0.47727	0.26952

Table 3: Local validation summary. Track 1 reports S_1, label accuracy, and total-score RMSE. Track 2 reports S_2, side-level accuracy, and mean macro-F1.

The Stage 3 clinical feature ensemble substantially improves the Track 1 proxy score over the earlier sampled-feature baseline. The improvement comes from richer view-level biomechanics, cadence descriptors, and repeated-seed model averaging.

Selected submission	Public score
Stage 1 + Stage 2 hybrid	0.53195
Final Stage 3 + Stage 2	0.54177

Table 4: Public leaderboard scores for the strongest intermediate and final selected submissions.

5.2. Discussion

The final system benefits from treating the two tracks differently. The Track 2 labels are side-specific gait subtype classes, so temporal dynamics over pose sequences are useful despite the small number of labeled examples. Track 1 contains many correlated binary labels and a clinically meaningful total score, so explicit biomechanical summaries and threshold tuning provide a more stable solution.

The method also has limitations. First, all modeling is based on 2D pose, so depth-dependent gait properties such as pelvic rotation and some foot-plane measures are difficult to estimate. Second, the Track 2 training set is very small, making subtype validation noisy. Third, the model uses patient-level aggregation and does not explicitly segment full gait cycles. More robust gait-event detection, raw video features, or 3D pose reconstruction could improve future systems.

6. Reproducibility

The code release contains an end-to-end Kaggle notebook that trains both stages and writes the selected submission. The notebook requires two Kaggle inputs: the official competition labels and the released pose dataset. The selected model artifacts, feature tables, run summaries, and submission CSV are released separately as a Kaggle artifact dataset.

Item	Link
Code	GitHub repository
Artifacts	Kaggle artifact dataset
Competition	Kaggle competition page

Table 5: Public materials for reproducing the selected submission.

7. Conclusion

We presented a two-stage pose-based solution for pediatric gait assessment in the CVPR 2026 Children Gait Challenge. A compact temporal convolutional model predicts gait subtypes, while a patient-level biomechanical feature ensemble predicts bilateral EVGS labels. This separation matches the structure of the two challenge tracks and improves the final public leaderboard score to 0.54177. The approach emphasizes reproducibility, clinical feature design, and robustness in a small-sample setting.

References

[1] H. S. Read, M. E. Hazlewood, S. J. Hillman, R. J. Prescott, and J. E. Robb, "Edinburgh visual gait score for use in cerebral palsy," *Journal of Pediatric Orthopaedics*, vol. 23, no. 3, pp. 296–301, 2003.

[2] S. H. Ramesh, E. D. Lemaire, A. Tu, K. Cheung, and N. Baddour, "Automated implementation of the edinburgh visual gait score (evgs) using openpose and handheld smartphone video," *Sensors*, vol. 23, no. 10, p. 4839, 2023.

[3] I. Somasundaram, A. Tu, R. Olleac, N. Baddour, and E. D. Lemaire, "Automated implementation of the edinburgh visual gait score (evgs)," *Sensors*, vol. 25, no. 10, p. 3226, 2025.

[4] Y. Shen, B. Li, M. Huang, Y. Liu, X. Cao, J. Jin, Z. Li, A. Liu, J. Kim, J. Zhu, L. Fangzhou, J. Cao, J. Chen, I. Lourentzou, and J. M. Rehg, "Decoding children's gait behavior," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2026. arXiv:2608.00371.

[5] CV4CHL Organizers, "CVPR 2026 Workshop on Computer Vision for Children (CV4CHL)." `https://pediamedai.com/cv4chl/`, 2026. Accessed: 2026-05-05.

[6] PediaMedAI, "CVPR 2026 The First AI for Children Challenge." `https://www.kaggle.com/competitions/cvpr-2026-the-first-ai-children-challenge`, 2026. Accessed: 2026-05-05.

[7] S. Armand, G. Decoulon, and A. Bonnefoy-Mazure, "Gait analysis in children with cerebral palsy," *EFORT Open Reviews*, vol. 1, no. 12, pp. 448–460, 2016.

[8] L. W. Robinson, N. D. Clement, M. Fullarton, A. R. Richardson, J. Herman, G. Henderson, J. E. Robb, and M. S. Gaston, "The relationship between the edinburgh visual gait score, the gait profile score and GMFCS levels I–III," *Gait & Posture*, vol. 41, no. 2, pp. 741–743, 2015.

[9] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, 2017.

[10] L. Kidzinski, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Communications*, vol. 11, p. 4054, 2020.

[11] J. Lin, Y. Wang, J. Sha, Y. Li, Z. Fan, W. Lei, and Y. Yan, "Clinical reliability and validity of a video-based markerless gait evaluation method," *Frontiers in Pediatrics*, vol. 11, p. 1331176, 2023.

[12] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[13] J. Kim, H. Seo, M. T. Naseem, and C. S. Lee, "Pathological-gait recognition using spatiotemporal graph convolutional networks and attention model," *Sensors*, vol. 22, no. 13, p. 4863, 2022.

[14] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[15] K. Yang, X. Chen, and J. He, "Multimodal fusion of skeleton dynamics and clinical gait features for video-based cerebral palsy severity assessment," *arXiv preprint arXiv:2603.22371*, 2026.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.