https://doi.org/10.2352/EI.2026.38.7.IMAGE-276

Multimodal Fusion of Skeleton Dynamics and Clinical Gait Features for Video-Based Cerebral Palsy Severity Assessment

Kaiyuan Yang[a], Xupeng Chen[b], Jiangpeng He[c]

[a]Department of Intelligent Systems Engineering, Luddy School, Indiana University Bloomington, Bloomington, USA
[b]Department of Electrical and Computer Engineering, Tandon School, New York University, Brooklyn, USA
[c]Department of Computer Science, Luddy School, Indiana University Bloomington, Bloomington, USA
Corresponding author: jhe2@iu.edu

Abstract

Video-based gait analysis has become a promising approach for assessing motor impairment in children with cerebral palsy (CP). However, existing methods usually rely on either pose sequences or handcrafted gait features alone, making it difficult to simultaneously capture spatiotemporal motion patterns and clinically meaningful biomechanical information. To address this gap, we propose a multimodal fusion framework that integrates skeleton dynamics with contribution-guided clinically meaningful gait features. First, Grad-CAM analysis on a pre-trained ST-GCN backbone identified the most discriminative body keypoints, providing an interpretable basis for subsequent gait feature extraction. We then build a dual-branch architecture, with one branch modeling skeleton dynamics using ST-GCN and the other encoding gait features derived from the identified keypoints. Fusing the two branches through feature cross-attention improved four-level CP motor severity classification to 70.86%, outperforming the baseline by 5.6 percentage points. Overall, we demonstrate that integrating skeleton dynamics with clinically meaningful gait descriptors can improve both prediction performance and biomechanical interpretability for video-based CP severity assessment.

Keywords: Cerebral Palsy, Gait Analysis, Spatiotemporal Graph Convolutional Networks, Multimodal Fusion, Clinical Interpretability

Introduction

Cerebral palsy (CP) is one of the most common childhood motor disability, affecting approximately 2–3 per 1,000 live births worldwide [1], with substantial regional variation in prevalence [2] and significant economic burden on healthcare systems [3]. Because gait dysfunction is a core manifestation of CP, scalable and objective video-based gait assessment has become an important direction for screening, follow-up, and treatment evaluation. Traditional clinical assessment remains widely used [4], but it requires in-person clinical visit, and the evaluation depends on expert judgment, which may introduce inter-rater variability and limit sensitivity to subtle changes over time. Instrumented gait analysis can provide detailed biomechanical measurements [5], but it usually requires specialized laboratories and expensive motion capture sensor systems, which limits its accessibility. Caregiver-reported questionnaires can reflect daily functional impact [6], but they are also indirect and may not provide sufficiently objective measurements.

Recent advances in computer vision have made video-based gait assessment a promising option for motor function evaluation, with the potential to improve accessibility and reduce the cost of traditional assessment [7]. A typical pipeline extracts 2D body keypoints from walking videos using pose estimation methods such as OpenPose [8], and then models motion patterns to predict clinically meaningful outcomes. Existing studies have shown that monocular videos can provide useful gait information [7], while recent skeleton-based deep learning methods have improved the prediction of motor severity from pose sequences [9, 10]. Work in other movement disorders has also shown that markerless video analysis can support interpretable motor assessment [11]. These advances suggest that video-based assessment is promising, but they also highlight the need to better connect learned motion representations with clinically meaningful knowledges.

Despite these advances, an important gap still remains in existing video-based methods. Skeleton-based models are effective at learning spatiotemporal motion patterns directly from pose sequences, but they mainly rely on raw keypoint trajectories that makes their predictions harder to interpret in clinically meaningful terms [12]. In contrast, clinical gait features such as joint angles, step length, and symmetry indices are easier to interpret and are more closely related to established gait assessment practice. Existing studies have also pointed out the value of interpretable and clinically grounded gait features in movement analysis [13, 14]. However, these handcrafted features provide only low-dimensional summaries of motion and may miss useful temporal information. As a result, skeleton dynamics and clinical gait features are often modeled separately, even though they offer naturally complementary information.

In this work, to address the gap mentioned between these two modalities, we propose a two-step multimodal fusion framework for video-based CP severity assessment, as shown in Fig. 1. In Step 1, we apply Grad-CAM [15] to a pre-trained ST-GCN backbone to identify the body keypoints that contribute most strongly to the severity level prediction. These keypoints are then used to guide the selection and extraction of clinically informed gait features. In Step 2, we build a dual-branch architecture consisting of a skeleton branch for modeling spatiotemporal dynamics and a clinical feature branch for encoding gait features derived from the selected keypoints. The two branches are fused for final classification. In this study, CP severity is formulated as a four-class GMFCS label [16], which is a common standard for motor function assessment in children with CP. Overall, the framework is

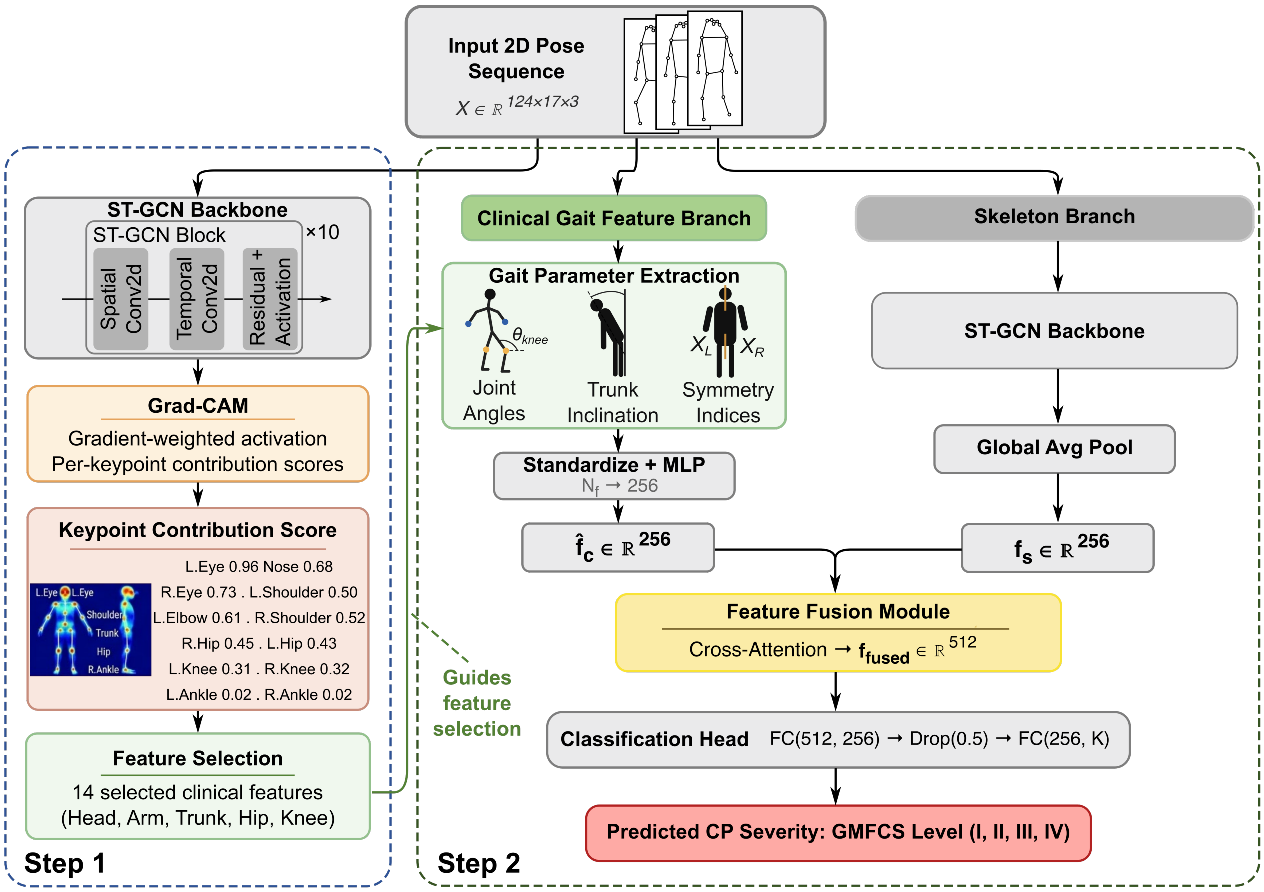

Figure 1: Overview of the proposed two-step framework. **Step 1**: An ST-GCN backbone is pre-trained and analyzed via Grad-CAM to identify discriminative keypoints, guiding gait feature selection. **Step 2**: A dual-branch architecture combines skeleton dynamics (ST-GCN backbone $\rightarrow$ global average pooling $\rightarrow \mathbf{f}_s \in \mathbb{R}^{256}$) with extracted clinical gait features (MLP projection $\rightarrow \hat{\mathbf{f}}_c \in \mathbb{R}^{256}$) via feature fusion for GMFCS classification.

designed to improve both predictive performance and clinical interpretability by combining learned skeleton dynamics with clinically meaningful gait features.

Methods

Our framework operates in two steps, as shown in Fig. 1, to identify which body keypoints are most important for prediction, and form the unified model to combine the information derived from those keypoints with learned skeleton dynamics. In the first step, we analyze a pre-trained ST-GCN backbone using Grad-CAM to estimate the contribution of individual body joints to the classification decision. These keypoint importance scores are used to identify body regions that are most relevant to CP severity prediction and to guide the extraction of clinically informed gait features. In the second step, we construct a dual-branch fusion model. One branch processes the original pose sequence using ST-GCN to learn spatiotemporal skeleton dynamics, while the other branch encodes gait features computed from the selected keypoints. The two branches are then fused for final GMFCS classification. The input to both steps is a 2D pose sequence $X \in \mathbb{R}^{T \times V \times C}$, where $T = 124$ is the number of frames, $V = 17$ is the number of COCO keypoints [17], and $C = 3$ corresponds to $(x, y, \text{confidence})$.

Step 1: Keypoint Importance Analysis

ST-GCN Architecture

We adopt the ST-GCN architecture [9, 10] as our skeleton backbone. The 17-keypoint COCO skeleton is represented as a spatial graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where each node corresponds to a body joint and each edge represents a natural anatomical connection. The network follows the standard ST-GCN design, which combines spatial graph convolution with temporal convolution to capture both body structure and motion over time.

Each block includes a spatial graph convolution layer, batch normalization, ReLU activation, temporal convolution, and residual connection. The input pose sequence first passes through data batch normalization and then through 10 stacked ST-GCN blocks with channel dimensions progressing from 64 to 128 and 256. Global average pooling is applied at the end to obtain a skeleton feature embedding $\mathbf{f}_s \in \mathbb{R}^{256}$. To provide a stronger motion representation, the backbone is initialized with weights pre-trained on NTU RGB+D dataset [18].

Grad-CAM-Based Keypoint Attribution

To understand which joints contribute most strongly to CP severity prediction, we apply Grad-CAM [15] to the final convolutional layer of the pre-trained ST-GCN model. For a target class c, Grad-CAM first computes the importance weight of the k-th feature map by globally averaging the gradients of the class score y^c with respect to the activation map A^k:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \tag{1}$$

where Z is the number of spatial-temporal locations in the feature map. The class activation map is then obtained as

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{2}$$

This produces a gradient-weighted activation map that provides an intuitive estimate of which body regions the model relies on most during classification.

To obtain joint-level importance scores, we aggregate the Grad-CAM responses over the temporal dimension and map them back to the corresponding body joints. The resulting keypoint scores are visualized as a body-joint importance map (Fig. 2). Based on these scores, we group informative joints into clinically meaningful body regions and use them to guide downstream gait feature extraction. Rather than extracting features from all joints equally, we focus on regions that receive stronger model attention and are also meaningful in clinical gait analysis. This step connects neural network interpretability with feature engineering and serves as the basis for the second stage of the framework.

The ST-GCN backbone used in this step is trained for four-class GMFCS classification with standard cross-entropy loss,

$$\mathscr{L}_{\text{CE}} = -\sum_{c=1}^{C} y_c \log \hat{y}_c, \tag{3}$$

where $C = 4$, y_c is the ground-truth label indicator, and $\hat{y}_c$ is the predicted probability for class c. Grad-CAM is then applied to the trained model to analyze which joints contribute most strongly to the classification decision.

Clinical Gait Feature Extraction

We first group the 17 skeletal keypoints into anatomically meaningful regions to improve interpretability: *head* (nose, left/right eye, left/right ear), *upper limb* (left/right shoulder, elbow, and wrist), *trunk/pelvis* (shoulders and hips), and *lower limb* (hips, knees, and ankles). These regions help summarize which body parts the ST-GCN model attends to most strongly.

For downstream feature engineering, we focus on clinically established gait descriptors rather than extracting features uniformly from every attended joint. Specifically, guided by the Grad-CAM analysis, we found that the whole head region and trunk region are very important, so we use region-level features to describe the movement. While for other regions since not all of the keypoints are very strong, we use keypoint-level features to summarize them. The selected features are therefore organized by anatomical source (knee, trunk) and global gait descriptor category (spatiotemporal, and symmetry), as summarized in Table 1.

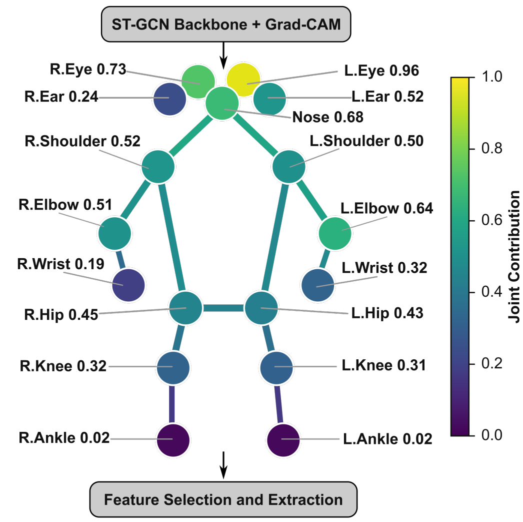

Figure 2: Grad-CAM keypoint contribution analysis on the ST-GCN backbone. Joint circle sizes and colors indicate contribution scores to GMFCS classification. Head and upper-body keypoints show the highest contributions, while ankles contribute minimally. These scores guide the selection of body regions for clinical gait feature extraction.

Among them, joint-angle features are computed from the angle between two limb vectors at each frame, trunk inclination describes upper-body posture relative to the vertical direction, and symmetry-related features are derived using a standard symmetry index.

Frame-wise joint angles are computed as

$$\theta_{\text{joint}}(t) = \arccos\left(\frac{\mathbf{a}(t)\cdot\mathbf{b}(t)}{|\mathbf{a}(t)|\,|\mathbf{b}(t)|}\right) \tag{4}$$

where $\mathbf{a}(t)$ and $\mathbf{b}(t)$ denote the two limb vectors that form the angle at frame t.

Bilateral symmetry indicess [19] are computed as

$$\text{SI} = \frac{|X_L - X_R|}{0.5(X_L + X_R)} \times 100\%. \tag{5}$$

where SI is the symmetry index, and X_L and X_R denote the values of the same gait feature measured on the left and right sides, respectively.

In total, we extract 24 candidate gait features and compare them with a clinically curated subset of 14 features (Table 1). The selected subset includes features with high Grad-CAM relevance and clear clinical meaning, while excluding features with high redundancy or limited reliability in 2D video-based analysis. All features are standardized and projected through a two-layer MLP to obtain a clinical feature embedding $\hat{\mathbf{f}}_c \in \mathbb{R}^{256}$.

Step 2: Dual-Branch Fusion Architecture

In this step, we leverage Cross-Attention Fusion to integrate skeleton and clinical feature embeddings. Specifically, a sigmoid-gated attention mechanism computes a dynamic weighting:

$$\alpha = \sigma\left(\frac{\mathbf{f}_s^\top \hat{\mathbf{f}}_c}{\sqrt{d}}\right), \quad d = 256 \tag{6}$$

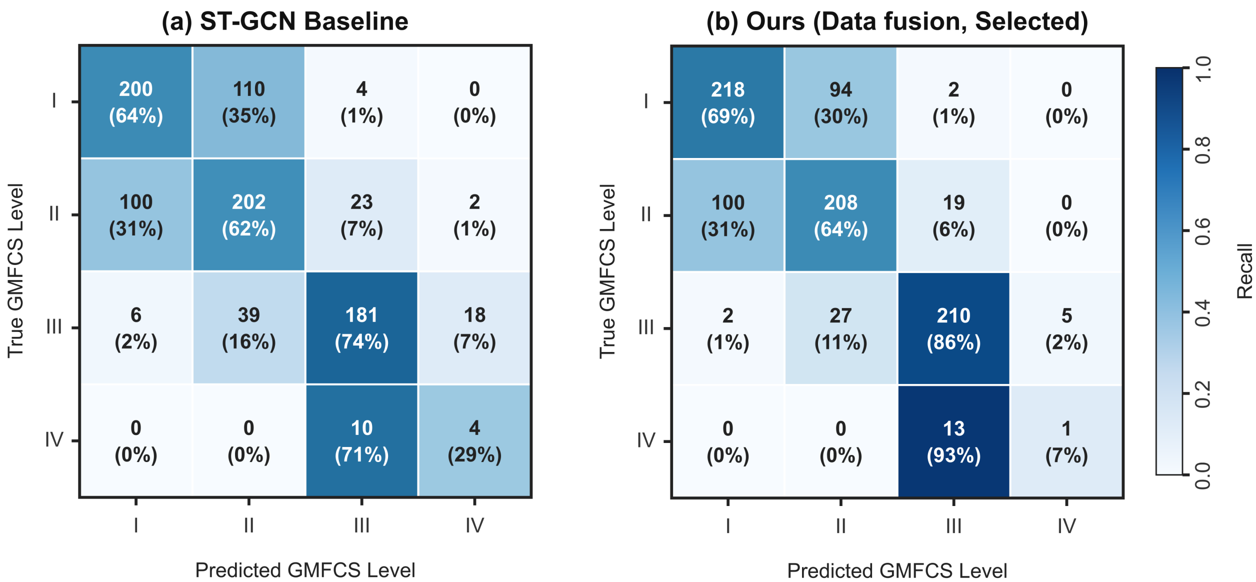

Figure 3: Confusion matrices for (a) ST-GCN baseline and (b) our framework (selected features, concatenation). Cell values show raw counts and row-normalized recall percentages.

Table 1: Selected clinical gait features (14 features) organized by anatomical source and gait feature category. Features were selected based on Grad-CAM keypoint importance and clinical relevance.

Body Region	Features	Count
Head	Head tilt, head stability	2
Elbow	L/R elbow ROM, elbow ROM sym.	2
Knee	L/R knee ROM, knee ROM sym.	2
Trunk	Trunk inclination	1
Spatial	Step len., stride len., speed	3
Temporal	Cadence, gait cycle dur.	2
Symmetry	Step len. sym., timing sym.	2

where α denotes the attention weight, $\sigma(\cdot)$ is the sigmoid function, and d is the embedding dimension.

$$\mathbf{f}_{\text{fused}} = \text{LN}\big(\mathbf{f}_s + \alpha\hat{\mathbf{f}}_c\big) \oplus \text{LN}\big(\hat{\mathbf{f}}_c + \alpha\mathbf{f}_s\big) \tag{7}$$

where $\mathbf{f}_{\text{fused}} \in \mathbb{R}^{512}$ is the fused feature representation, $\text{LN}(\cdot)$ denotes layer normalization, and $\oplus$ denotes vector concatenation. The fused representation is passed to a classification head: $\text{FC}(512 \to 256) \to \text{Dropout} \to \text{FC}(256 \to K)$ for GMFCS prediction. For optimization, training is performed in two stages. We first freeze the ST-GCN backbone and train only the fusion module and classifier. We then unfreeze the last two ST-GCN blocks for joint fine-tuning with a smaller learning rate. This strategy helps stabilize training while allowing the fused model to adapt to the CP severity classification task.

Experiments

Dataset and training

We evaluate on the public CP gait dataset from Kidziński et al. [7]. The dataset contains gait videos labeled with CP

Table 2: Dataset distribution across GMFCS levels and data splits.

	I	II	III	IV	Total
Train	556	597	390	31	1,574
Val	167	179	118	9	473
Test	314	327	244	14	899
Frac.	35.3%	37.9%	24.8%	1.6%	2,946

across GMFCS Levels I–IV, recorded using consumer-grade cameras during clinical visits. OpenPose [8] extracts 25 keypoints per frame, which are converted to COCO 17-keypoint format.After pre-processing with quality filtering ($\geq$80% keypoints with confidence$\geq$0.2) and removing the invalid videos that contain other individuals, finally we yield 2,946 clips from GMFCS Levels I–IV. We use a patient-level stratified split (Training: 1,574, Validation: 473, Test: 899) to prevent data leakage. Table 2 summarizes the dataset distribution. Note that GMFCS Level IV is severely under-represented (1.6%), reflecting the lower prevalence and fewer clinical visits at this severity level in the source cohort, in line with real-world conditions. The model is optimized using cross-entropy loss and Adam with a weight decay of 5×10^{-5}. A cosine annealing scheduler is used for learning rate adjustment. Data augmentation includes horizontal flipping with left-right joint remapping and Gaussian noise ($\sigma = 2$ px). Model selection is based on the best validation accuracy, and early stopping is applied with a patience of 5 epochs. All experiments are implemented in PyTorch and conducted on a single NVIDIA *A100* GPU.

Main Results

Table 3 presents the classification performance across all evaluated configurations (ST-GCN Baseline: only use raw keypoint sequences; +All Gait: combine the raw keypoint sequences

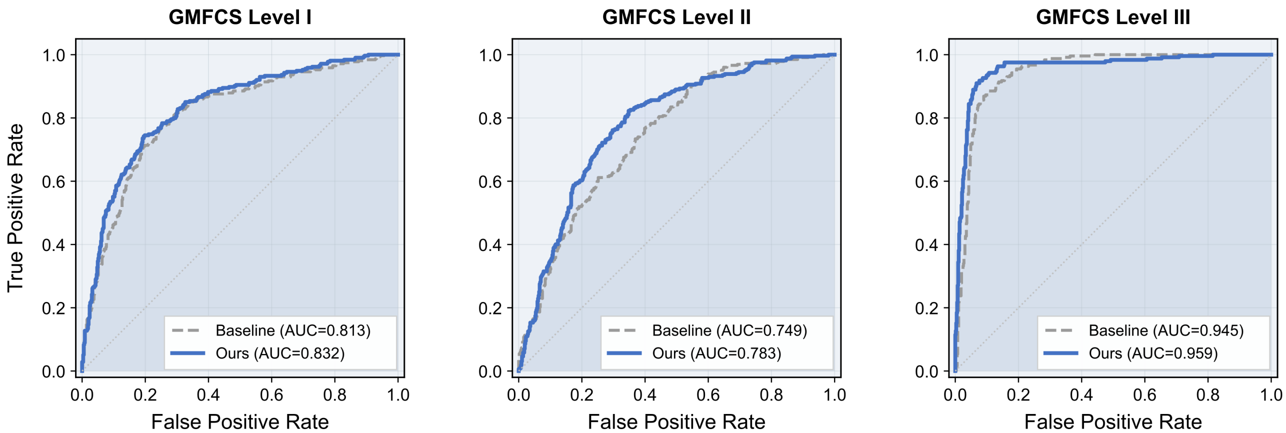

Figure 4: Per-class ROC curves for GMFCS Levels I–III, comparing ST-GCN baseline (dashed gray) and our framework (solid blue). AUC values are shown in the legend.

Table 3: GMFCS classification performance on the test set (N=899). Acc denotes overall accuracy (in %), $F1_w$ denotes weighted F1 score, and κ_l denotes linearly weighted Cohen's kappa. Best in **bold**. "Sel." denotes the clinically curated 14-feature subset and the "Sel. Gait" corresponds to our proposed framework.

Method	Acc	$F1_w$	κ_l
ST-GCN Baseline	65.29	0.658	0.597
+ All Gait	67.41	0.670	0.619
+ Sel. Gait	**70.86**	**0.706**	**0.665**

Table 4: Per-class recall (%) by GMFCS level. N denotes the number of test samples.

Method	I	II	III	IV
	N=314	N=327	N=244	N=14
ST-GCN	63.69	61.77	74.18	28.57
Ours	**69.43**	**63.61**	**86.07**	7.14†
Δ	+5.74	+1.84	+11.89	—

†Level IV contains only 14 test samples; per-class metrics at this level are not statistically reliable.

with all gait features; +Sel. Gait: combine the raw keypoint sequences with selected gait features guided by the Grad-CAM) on the 899-sample test set. We report accuracy, weighted F1 score (accounting for class imbalance), and linearly weighted Cohen's kappa (κ_l), which is appropriate for ordinal classification.

Key observations. (1) All fusion variants outperform the skeleton-only baseline, confirming the value of clinical gait features. (2) Feature selection matters: the Grad-CAM-guided 14-feature subset outperforms all 24 features by 1.7% under cross-attention, indicating that using a clinically guided feature subset reduces noise from redundant features on moderately sized clinical datasets. (3) The best configuration—selected features combined with raw gait keypoint sequences—achieves 70.86% accuracy (+5.6% over baseline), with gains in weighted F1 (+0.048) and κ_l (+0.068). We also compared with some existing studies, we tried the method from Zhao et al. [9], and reported 65.29% on our dataset; and in the paper that originally proposed the public CP dataset, Kidziński et al. [7] focus mainly on gait features prediction, and achieve 66.00%, these works contributed a lot from different perspectives.

Per-Class Analysis

Table 4 shows per-class recall (sensitivity) for the baseline and best configuration. The largest improvement is observed for GMFCS Level III, where recall increases by 11.89% , from 74.18% to 86.07%, while Level I also shows a notable gain (+5.74%). These results suggest that clinically guided gait features may be particularly helpful for separating classes in the middle of the severity spectrum. Prior studies have shown that gait deviations, asymmetry, and variability are clinically meaningful in children with CP, which supports the value of structured gait descriptors in this setting [20]. By contrast, performance on Level IV remains less stable. Although recall improves numerically, this class contains only 14 test samples, making the estimate highly sensitive to sampling variation. This limitation is consistent with the well-known difficulty of learning minority classes in small and imbalanced clinical datasets.

Confusion Matrix Analysis

Fig. 3 shows the confusion matrices. The baseline exhibits substantial adjacent-level confusion, particularly between Levels I and II (210 misclassifications). Our framework reduces this to 194 and improves Level III recall from 74.2% to 86.1% by reducing III→II errors (39→27).

Fig. 4 presents per-class ROC curves for Levels I–III (Level IV omitted due to N=14). Our framework improves AUC across all levels, with the largest gain on Level II (0.749→0.783). Level III achieves the highest AUC (0.959), consistent with its distinctive gait patterns.

Conclusion

We proposed a multimodal fusion framework that combines skeleton dynamics with clinically identified gait features for video-based CP severity assessment. By using Grad-CAM to select informative keypoints and guide gait feature extraction, the framework combines learned skeleton dynamics with clinically meaningful gait features. Experimental results show that the proposed approach consistently outperforms the skeleton-only baseline, with the best model achieving 70.86% accuracy, 5.6% improvement over ST-GCN. We also found that a clinically guided feature subset performed better than a larger unfiltered feature set, suggesting that careful feature selection is important for moderately sized clinical datasets [21].

Beyond performance, an important strength of this framework is its interpretability. [14, 13]. Grad-CAM-guided feature extraction provides a more transparent connection between model attention and clinically relevant gait descriptors. Several limitations should also be noted. GMFCS Level IV was severely underrepresented, making this class difficult to learn reliably and its results less stable. In addition, the current framework relies on 2D pose sequences and was evaluated on a single backbone and dataset. Future work should include larger and more balanced cohorts, 3D pose representations [22], and validation with additional backbone models [23] and external datasets.

References

[1] Rosenbaum P., Paneth N., Leviton A., et al., "A report: The definition and classification of cerebral palsy," Developmental Medicine & Child Neurology, 49(S109), 8–14 (2007).

[2] McIntyre S., Goldsmith S., Webb A., Ehlinger V., Hollung S. J., McConnell K., Arnaud C., Smithers-Sheedy H., Oskoui M., Khandaker G., et al., "Global prevalence of cerebral palsy: A systematic analysis," Developmental Medicine & Child Neurology, 64(12), 1494–1506 (2022).

[3] Tonmukayakul U., Shih S. T. F., Bourke-Taylor H., Imms C., Reddihough D., Cox L., Carter R., "Systematic review of the economic impact of cerebral palsy," Research in Developmental Disabilities, 80, 93–101 (2018).

[4] K. Vitrikas, H. Dalton, and D. Breish, "Cerebral palsy: An overview," *American Family Physician*, vol. 101, no. 4, pp. 213–220, 2020.

[5] H. Haberfehlner, M. Goudriaan, L. A. Bonouvrié, E. P. Jansma, J. Harlaar, R. J. Vermeulen, M. M. van der Krogt, and A. I. Buizer, "Instrumented assessment of motor function in dyskinetic cerebral palsy: A systematic review," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, p. 39, 2020.

[6] Bautista M., Whittingham K., Edwards P., Boyd R. N., "Psychometric properties of parent and child reported sleep assessment tools in children with cerebral palsy: A systematic review," Developmental Medicine & Child Neurology, 60(2), 162–172 (2018).

[7] Kidziński Ł., Yang B., Hicks J. L., et al., "Deep neural networks enable quantitative movement analysis using single-camera videos," Nature Communications, 11, 4054 (2020).

[8] Cao Z., Simon T., Wei S., Sheikh Y., "Realtime multi-person 2D pose estimation using part affinity fields," Proc. CVPR, 7291–7299 (2017).

[9] Zhao P., Alencastre-Miranda M., Shen Z., O'Neill C., Whiteman D., Gervas-Arruga J., Krebs H. I., "Computer vision for gait assessment in cerebral palsy: Metric learning and confidence estimation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 2336–2345 (2024).

[10] 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*, 2018.

[11] D. Deng, J. L. Ostrem, V. Nguyen, D. D. Cummins, J. Sun, A. Pathak, S. Little, and R. Abbasi-Asl, "Interpretable video-based tracking and quantification of parkinsonism clinical motor states," *npj Parkinson's Disease*, vol. 10, no. 1, p. 122, 2024.

[12] Halilaj E., Rajagopal A., Fiterau M., et al., "Machine learning in human movement biomechanics: Best practices, common pitfalls," Journal of Biomechanics, 81, 1–11 (2018).

[13] Kohnehshahri F. S., Merlo A., Mazzoli D., Bò M. C., Stagni R., "Machine learning applied to gait analysis data in cerebral palsy and stroke: A systematic review," Gait & Posture, 111, 105–121 (2024).

[14] Xiang L., Gao Z., Yu P., Fernandez J., Gu Y., Wang R., Gutierrez-Farewik E. M., "Explainable artificial intelligence for gait analysis: Advances, pitfalls, and challenges-a systematic review," Frontiers in Bioengineering and Biotechnology, 13, 1671344 (2025).

[15] Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE International Conference on Computer Vision, 618–626 (2017).

[16] Palisano R., Rosenbaum P., Walter S., Russell D., Wood E., Galuppi B., "Development and reliability of a system to classify gross motor function in children with cerebral palsy," Developmental Medicine & Child Neurology, 39(4), 214–223 (1997).

[17] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common objects in context," in *Computer Vision – ECCV 2014*, pp. 740–755, 2014.

[18] Shahroudy A., Liu J., Ng T.-T., Wang G., "NTU RGB+D: A large scale dataset for 3D human activity analysis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1010–1019 (2016).

[19] Robinson R. O., Herzog W., Nigg B. M., "Use of force platform variables to quantify the effects of chiropractic manipulation on gait symmetry," Journal of Manipulative and Physiological Therapeutics, 10(4), 172–176 (1987).

[20] Armand S., Decoulon G., Bonnefoy-Mazure A., "Gait analysis in children with cerebral palsy," EFORT Open Reviews, 1(12), 448–460 (2016).

[21] Jun K., Lee K., Lee S., Lee H., Kim M. S., "Hybrid deep neural network framework combining skeleton and gait features for pathological gait recognition," Bioengineering, 10(10), 1133 (2023).

[22] Jing Y., Qin P., Fan X., Qiang W., Zhu W., Sun W., Tian F., Wang D., "Deep learning-assisted gait parameter assessment for neurodegenerative diseases: Model development and validation," Journal of Medical Internet Research, 25, e46427 (2023).

[23] Chen Y., Zhang Z., Yuan C., Li B., Deng Y., Hu W., "Channel-wise topology refinement graph convolution for skeleton-based action recognition," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 13359–13368 (2021).