Weakly-Supervised Temporal Decomposition with Graph Neural Networks for Parkinsonian Turning Assessment

Jieming Zhang[1], Tai-Myoung Chung[2], and Hogun Park[1](✉)

[1] Department of Computer Science and Engineering, Sungkyunkwan University, Suwon, Republic of Korea
jieming2021@skku.edu; hogunpark@skku.edu
[2] Hippo T&C Inc., Seoul, Republic of Korea
tmchung@hippotnc.com

Abstract. Turning difficulty is an early and clinically significant indicator of Parkinson's disease (PD). Skeleton-based approaches are already widely used in clinical movement analysis, as they offer a markerless and privacy-preserving representation of human motion from ordinary video. However, automated turning assessment from skeleton sequences remains challenging due to the temporal complexity of turning movements. Existing skeleton-based methods either treat the entire sequence globally, ignoring its inherent temporal structure, or require expensive frame-level phase annotations. We propose a weakly-supervised temporal decomposition framework that discovers discriminative temporal patterns from video-level diagnostic labels alone. The framework integrates a transition detector operating on motion gradients with spatial-temporal graph convolutional networks via learned soft masks, decomposing the turning sequence into multiple temporal instances with diagnostic relevance learned through attention aggregation. A learnable adjacency mechanism further adapts the skeleton graph to emphasize clinically relevant body connections. Decoupling boundary detection from the classification objective allows the learned phase structure to serve as an independent clinical interpretation, unbiased by diagnostic labels. Evaluated on the clinical PD dataset, our method achieves an AUC-ROC of 91.05% under 5-fold subject-level cross-validation, while providing interpretable outputs including instance-level importance weights and anatomical relevance maps consistent with clinical knowledge of Parkinsonian turning symptoms. The code is available at `https://github.com/jieming1113/TempoDecompGCN-PD`

Keywords: Parkinson's disease · Graph neural networks · Temporal decomposition · Weakly-supervised learning

1 Introduction

Parkinson's disease (PD) affects millions of people worldwide and is characterized by progressive motor deterioration including bradykinesia, rigidity, and postural

instability [4]. Among functional motor tasks, turning is a particularly sensitive indicator of PD severity, demanding complex multi-segmental coordination involving anticipatory postural adjustments, axial rotation, and dynamic balance control [25,2]. Clinically, turning difficulty manifests as a characteristic rigid-body turning pattern, and correlates with fall risk [5,6,14], making automated turning assessment valuable for both diagnosis and longitudinal monitoring.

Markerless skeleton extraction from standard video has enabled scalable automated clinical assessment [1,26], and skeleton-based deep learning methods have been applied to general action recognition and automated PD evaluation [28,30,29]. However, these methods rely on global temporal pooling with fixed skeleton topology, without mechanisms to exploit the task-specific temporal structure and the spatially heterogeneous motor patterns inherent to turning assessment. Specifically, a turning movement progresses through biomechanically distinct stages, namely Early-Turn, Mid-Turn, and Late-Turn [31,8,20], that differ in diagnostic relevance for PD. Notably, Parkinsonian deficits are particularly pronounced in the Late-Turn phase, manifesting as prolonged postural recovery [22,18]. Existing approaches to temporal decomposition of clinical movement sequences either require frame-level phase annotations impractical to obtain in clinical settings [10], or classify the entire sequence without uncovering its internal temporal structure [30,29,27]. We address this through weakly-supervised temporal decomposition, learning phase boundaries, i.e., transition points between motion stages, directly from video-level diagnostic labels alone.

We propose a weakly-supervised temporal decomposition framework that partitions turning sequences into temporal instances via learned boundary scores. Inspired by multiple instance learning (MIL) [7], an attention-based aggregation module assigns diagnostic importance weights to each instance. An adaptive adjacency mechanism further modulates the skeleton graph to emphasize clinically relevant connections [24]. Under subject-level cross-validation on clinical turning sequences, our method achieves an AUC-ROC of 91.05% with clinically interpretable outputs.

Our main contributions are: (1) a weakly-supervised temporal decomposition framework that discovers discriminative temporal patterns from video-level diagnostic labels alone; (2) a dual-branch architecture integrating soft temporal decomposition with instance-level attention aggregation, jointly optimized with decoupled gradient flows; and (3) evaluation on a clinical dataset demonstrating strong diagnostic performance with clinically interpretable outputs.

2 Related Work

Recent advances in pose estimation enable markerless skeleton extraction from standard video [1,26], facilitating scalable clinical deployment. Building on this, ST-GCN [27] and extensions [24,12] have achieved strong performance on action recognition benchmarks [23] and have been applied to clinical movement assessment [13] and automated PD evaluation [30,29,19]. However, all these methods

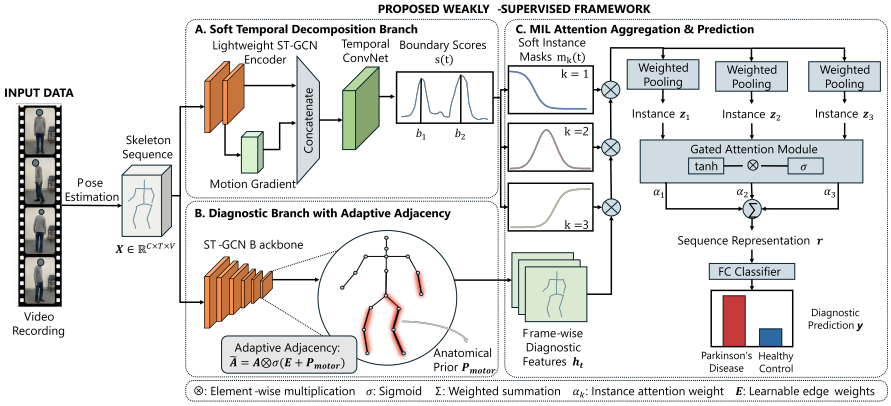

Fig. 1. Overview of the proposed framework. (A) The soft temporal decomposition branch computes boundary scores $s(t)$ and generates soft instance masks $m_k(t)$. (B) The diagnostic branch extracts frame-wise features $\boldsymbol{h}(t)$ with adaptive adjacency. (C) Instance representations $\boldsymbol{z}_k$ are obtained via mask-weighted pooling and aggregated through gated attention weights α_k into a sequence representation $\boldsymbol{r}$ for classification.

rely on global temporal pooling, discarding the phase-specific temporal structure that is diagnostically relevant in turning assessment.

Temporal decomposition of clinical movement sequences has been approached through supervised phase segmentation requiring frame-level annotations [10], which is impractical in clinical settings. Weakly-supervised approaches reduce annotation cost but still require partial temporal labels [11,21]. Multiple instance learning with attention-based aggregation [7] enables bag-level classification over unordered instances, yet does not model the structured temporal progression of clinical movement tasks. Our framework adapts attention-based aggregation to ordered temporal instances discovered from video-level labels alone.

3 Method

As shown in Fig. 1, our framework consists of two parallel branches: a soft temporal decomposition branch that partitions the sequence into temporal instances via learned soft masks, and a diagnostic branch that extracts spatial-temporal features with an adaptive adjacency mechanism.

3.1 Problem Formulation and Input Representation

Given a skeleton sequence $\boldsymbol{X} \in \mathbb{R}^{C\times T\times V}$ with C input channels, T frames, and V joints, along with a video-level diagnostic label $y \in \{0, 1\}$, we aim to learn a classifier $F(\boldsymbol{X}) \to y$ that simultaneously discovers discriminative temporal structure without frame-level annotations. The skeleton is modeled as a

spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where vertices represent body joints and edges represent skeletal connections. Each sequence is normalized by centering on the hip midpoint and scaling by torso length, then interpolated to T frames. The input features consist of joint position, velocity $\boldsymbol{v}(t) = \boldsymbol{p}(t) - \boldsymbol{p}(t-1)$, and acceleration $\boldsymbol{a}(t) = \boldsymbol{v}(t) - \boldsymbol{v}(t-1)$, where $\boldsymbol{p}(t) \in \mathbb{R}^2$ denotes the joint position vector at frame t, yielding C input channels per joint.

3.2 Soft Temporal Decomposition Branch

Turning movements progress through biomechanically distinct stages, yet obtaining frame-level phase annotations is prohibitively expensive in clinical settings. To address this, this branch learns to partition the turning sequence into K temporal instances in a data-driven manner, detecting motion transitions via auxiliary regularization losses, decoupled from the classification objective. A lightweight ST-GCN encoder processes the input to obtain frame-wise features $\boldsymbol{f}(t) \in \mathbb{R}^d$ after spatial pooling. We compute motion gradient features as $\Delta\boldsymbol{f}(t) = \boldsymbol{f}(t) - \boldsymbol{f}(t-1)$, with $\Delta\boldsymbol{f}(0) = \mathbf{0}$. The concatenation $[\boldsymbol{f}(t); \Delta\boldsymbol{f}(t)] \in \mathbb{R}^{2d}$ encodes both the instantaneous motor state and its rate of change. A temporal convolutional network maps these to boundary likelihood scores:

$$s(t) = \sigma\left(\text{Conv1D}([\boldsymbol{f}(t); \Delta\boldsymbol{f}(t)]) + \pi(t)\right), \tag{1}$$

where $\sigma(\cdot)$ denotes the sigmoid function, and $\pi(t)$ is a learnable position prior initialized with Gaussian peaks.

Boundary positions $b_1, b_2 \in [0, 1]$ are the top-2 local maxima of $s(t)$, sorted temporally. Specifically, local maxima are frames satisfying $s(t) > s(t-1)$ and $s(t) > s(t+1)$; the two local maxima with the highest $s(t)$ values are selected and normalized as $b_k = t_k/T$, yielding $b_1 \leq b_2$. These define soft instance masks:

$$\begin{aligned} m_1(t) &= \sigma(\beta(b_1 - t/T)), \\ m_2(t) &= \sigma(\beta(t/T - b_1)) \cdot \sigma(\beta(b_2 - t/T)), \\ m_3(t) &= \sigma(\beta(t/T - b_2)), \end{aligned} \tag{2}$$

where β controls boundary sharpness. The masks are normalized as $\hat{m}_k(t) = m_k(t) / \sum_j m_j(t)$, allowing each frame to contribute probabilistically to multiple instances.

3.3 Diagnostic Feature Extraction with Adaptive Adjacency

Parkinsonian motor impairment exhibits characteristic spatial patterns across body regions, yet the standard fixed skeleton topology treats all joint connections equally. To capture clinically relevant spatial patterns, the diagnostic branch consists of stacked ST-GCN blocks with progressive channel expansion and residual connections, equipped with an adaptive adjacency mechanism motivated by the diagnostic significance of lower extremity kinematics [16,15] and trunk coordination [6]. The mechanism computes a soft importance mask:

$$\boldsymbol{M} = \sigma\left(\boldsymbol{E} + \boldsymbol{P}_{\text{motor}}\right), \tag{3}$$

where $\boldsymbol{E} \in \mathbb{R}^{V \times V}$ is a learnable matrix initialized to zero, and $\boldsymbol{P}_{\text{motor}}$ is an anatomical prior with higher values for lower limb and trunk joint pairs. $\boldsymbol{P}_{\text{motor}}$ provides a soft initialization bias encoding clinical knowledge, while $\boldsymbol{E}$ is free to override it through training. The effective adjacency matrix is $\tilde{\boldsymbol{A}} = \boldsymbol{A} \odot \boldsymbol{M}$, where $\boldsymbol{A}$ is the physical skeleton adjacency matrix and $\odot$ denotes element-wise multiplication.

3.4 Aggregation and Training

With the sequence decomposed into temporal instances, we require a mechanism to aggregate instance-level information into a video-level diagnostic decision while identifying which instances are most diagnostically relevant. We formulate classification as a MIL problem [7]. Given diagnostic features and soft instance masks $\hat{m}_k(t)$, instance representations are computed via weighted pooling:

$$\boldsymbol{z}_k = \frac{\sum_t \hat{m}_k(t) \cdot \bar{\boldsymbol{h}}(t)}{\sum_t \hat{m}_k(t) + \epsilon}, \tag{4}$$

where $\bar{\boldsymbol{h}}(t)$ is the spatially-averaged feature at frame t and ϵ is a small constant for numerical stability. Each $\boldsymbol{z}_k$ undergoes an instance-specific transformation to obtain $\tilde{\boldsymbol{z}}_k$, and importance weights are computed via gated attention: $\alpha_k = \text{softmax}_k(\boldsymbol{w}^\top(\tanh(\boldsymbol{W}_1\tilde{\boldsymbol{z}}_k) \odot \sigma(\boldsymbol{W}_2\tilde{\boldsymbol{z}}_k)))$. The video-level representation $\boldsymbol{r} = \sum_k \alpha_k \tilde{\boldsymbol{z}}_k$ is passed to a classifier.

The training objective combines focal loss with class-dependent weights to address class imbalance:$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_s\mathcal{L}_{\text{sparse}} + \lambda_m\mathcal{L}_{\text{smooth}}$, where $\mathcal{L}_{\text{sparse}} = \frac{1}{T}\sum_{t=1}^{T} s(t)$ and $\mathcal{L}_{\text{smooth}} = \frac{1}{T}\sum_{t=2}^{T}(s(t) - s(t-1))^2$ encourage sparse and temporally smooth boundary scores, respectively. Together, these losses concentrate $s(t)$ at a small number of temporally localized peaks corresponding to frames of rapid motor state change, thereby enabling motion-salient boundary detection without frame-level supervision.

Since b_1, b_2 are obtained via a non-differentiable top-2 local maxima operation, $\mathcal{L}_{\text{cls}}$ does not back-propagate through boundary positions. The decomposition branch is trained exclusively by $\mathcal{L}_{\text{sparse}}$ and $\mathcal{L}_{\text{smooth}}$, while $\mathcal{L}_{\text{cls}}$ trains the diagnostic branch and aggregation module. This decoupling prevents the classification objective from biasing boundary positions, lending interpretability to the discovered soft temporal transition candidates.

4 Experiments

4.1 Dataset and Setup

We evaluate on a clinical dataset from the REMAP study [17], comprising 1,294 turning sequences (521 healthy controls (HC) and 773 PD patients) from 12 HC and 12 PD participants recorded during daily activities. PD diagnosis followed the UK Brain Bank Criteria; participants had H&Y 2.3 ± 0.8 (entry criterion ≤ 3) and MDS-UPDRS III score 36.8 ± 23.0. Videos are processed with a pretrained

Table 1. Results and ablation study of the proposed framework. All metrics are reported in percentage (%). The best results are highlighted in bold.

Configuration	Acc.	Sens.	Spec.	F1	AUC-ROC
ST-GCN baseline†	59.07 ± 3.96	60.39 ± 5.02	55.74 ± 5.74	67.88 ± 3.90	$61.71 \pm 2.53^{***}$
+ Global temporal attention	62.80 ± 1.88	66.20 ± 1.84	54.10 ± 2.32	71.80 ± 1.59	$64.00 \pm 2.04^{***}$
+ Fixed decomposition (independent encoder)	49.45 ± 15.27	51.29 ± 31.41	44.39 ± 29.89	54.62 ± 21.09	$46.14 \pm 5.19^{***}$
+ Fixed decomposition (soft mask)	72.88 ± 2.76	70.38 ± 6.26	76.58 ± 4.98	75.61 ± 3.38	$82.43 \pm 2.40^{***}$
+ Weakly-supervised (w/o adaptive adjacency)	80.15 ± 2.72	84.52 ± 5.27	75.08 ± 4.46	83.27 ± 2.71	$88.36 \pm 0.60^{**}$
+ Weakly-supervised decomposition (ours)	$\mathbf{81.54 \pm 1.92}$	$\mathbf{85.86 \pm 3.25}$	$\mathbf{75.19 \pm 5.96}$	$\mathbf{84.68 \pm 0.66}$	$\mathbf{91.05 \pm 1.13}$

$^{**}p < 0.01$, $^{***}p < 0.001$ vs. our model (paired t-test across folds).
† Lightweight 2-block encoder used as ablation baseline.

HRNet [26] model for 2D pose estimation, yielding skeleton sequences with 17 keypoints per frame. Only video-level diagnostic labels are available, with no frame-level temporal annotations. All experiments use subject-level 5-fold cross-validation. Sequences are temporally resampled to a uniform length of $T = 100$ frames with 6-channel joint features. The decomposition branch uses $K = 3$ instances ($\beta = 5$) and a lightweight encoder with 2 ST-GCN blocks ($d = 64$), while the diagnostic branch uses 6 ST-GCN blocks with channels up to 128. The two branches adopt separate encoders to avoid biasing boundary detection toward diagnostic cues. We employ an anatomical prior $\boldsymbol{P}_{\text{motor}}$ that assigns 1.0 to lower-limb joint pairs, 0.5 to lower-limb–trunk connections, and 0 elsewhere. We use AdamW with learning rate 3×10^{-4}, weight decay 0.01, focal loss with $\gamma = 2.0$, boundary regularization loss weights $\lambda_s = 0.05$ and $\lambda_m = 0.1$, and train for a maximum of 200 epochs.

4.2 Results and Ablation Study

Table 1 reports the full model performance and ablation study of progressively adding modules to a baseline ST-GCN. The baseline achieves only 59.07% accuracy with global pooling, and global temporal attention yields marginal improvement to 62.80%. Fixed decomposition(soft mask) improves accuracy by 13.81% and AUC-ROC by 20.72%, confirming temporal structure as the key factor, while fixed decomposition(independent encoder) yields substantially degraded performance as separate encoders produce incomparable instance representations that cannot be effectively aggregated. Replacing fixed decomposition with weakly-supervised decomposition further improves accuracy by 7.27% and AUC-ROC by 5.93%, confirming that motion transition-driven boundaries provide more discriminative temporal partitioning than fixed segmentation. Adding adaptive adjacency yields a further 2.69% gain in AUC-ROC, bringing the full model to 91.05%, confirming that learnable skeleton topology better captures the spatially heterogeneous nature of Parkinsonian motor impairment.

Table 2. Comparison of instance aggregation strategies. All metrics are reported in percentage (%). The best results are highlighted in bold.

Aggregation Strategy	Acc.	Sens.	Spec.	F1	AUC-ROC
Mean pooling	80.31 ± 2.74	83.28 ± 3.02	$\mathbf{75.95 \pm 10.06}$	83.48 ± 1.67	$87.89 \pm 1.34^{**}$
Max pooling	80.82 ± 1.47	85.00 ± 3.21	74.68 ± 1.79	84.03 ± 1.54	$88.12 \pm 0.82^{**}$
Attention pooling	79.08 ± 0.88	84.31 ± 2.64	71.39 ± 4.05	82.73 ± 0.83	$87.19 \pm 0.97^{***}$
Gated attention	$\mathbf{81.54 \pm 1.92}$	$\mathbf{85.86 \pm 3.25}$	75.19 ± 5.96	$\mathbf{84.68 \pm 0.66}$	$\mathbf{91.05 \pm 1.13}$

$^{**}p < 0.01$, $^{***}p < 0.001$ vs. our model (paired t-test across folds).

Table 3. Comparison with existing skeleton-based methods. All metrics are reported in percentage (%). The best results are highlighted in bold.

Method	Acc.	Sens.	Spec.	F1	AUC-ROC
ST-GCN [27]	61.28 ± 2.57	93.81 ± 12.07	13.08 ± 24.23	74.16 ± 1.44	$69.94 \pm 2.03^{***}$
AGCN [24]	64.30 ± 1.71	82.49 ± 16.77	37.15 ± 27.04	72.74 ± 4.68	$74.62 \pm 2.61^{***}$
2s-AGCN [24]	63.06 ± 6.22	79.59 ± 27.48	38.58 ± 31.13	68.97 ± 15.05	$74.87 \pm 2.74^{***}$
WM-STGCN [30]	62.27 ± 5.12	81.03 ± 10.68	42.55 ± 22.16	75.62 ± 2.87	$75.15 \pm 2.26^{***}$
SkateFormer [3]	71.49 ± 1.20	74.48 ± 4.52	67.09 ± 6.75	75.61 ± 1.45	$78.97 \pm 1.53^{***}$
MS-TCN++ [9]	77.74 ± 3.58	83.28 ± 1.50	62.28 ± 7.94	82.56 ± 2.45	$85.65 \pm 2.26^{***}$
Ours	$\mathbf{81.54 \pm 1.92}$	$\mathbf{85.86 \pm 3.25}$	$\mathbf{75.19 \pm 5.96}$	$\mathbf{84.68 \pm 0.66}$	$\mathbf{91.05 \pm 1.13}$

$^{***}p < 0.001$ vs. our model (paired t-test across folds).

We further compare instance aggregation strategies while keeping the temporal decomposition branch constant, as shown in Table 2. Gated attention achieves the best AUC-ROC of 91.05%, confirming that the learned gating mechanism effectively identifies the most diagnostically relevant temporal instances.

4.3 Comparison with Existing Methods

We compare with established skeleton-based methods under an identical training protocol, as reported in Table 3. Our method outperforms all compared methods across all metrics, with 81.54% accuracy and 91.05% AUC-ROC. MS-TCN++, which explicitly models temporal structure, achieves 85.65% AUC-ROC yet still falls 5.40% short of ours, underscoring the added value of weakly-supervised phase discovery and learnable skeleton topology. The compared methods exhibit severe sensitivity-specificity imbalance due to global temporal pooling, while our framework achieves balanced classification with 85.86% sensitivity and 75.19% specificity by focusing on specific temporal phases where Parkinsonian impairment manifests most distinctively, demonstrating its clinical value for PD screening.

4.4 Clinical Interpretability Analysis

Fig. 2 shows the interpretable outputs of the proposed framework. The adaptive adjacency mechanism converges to edge importance patterns consistent with

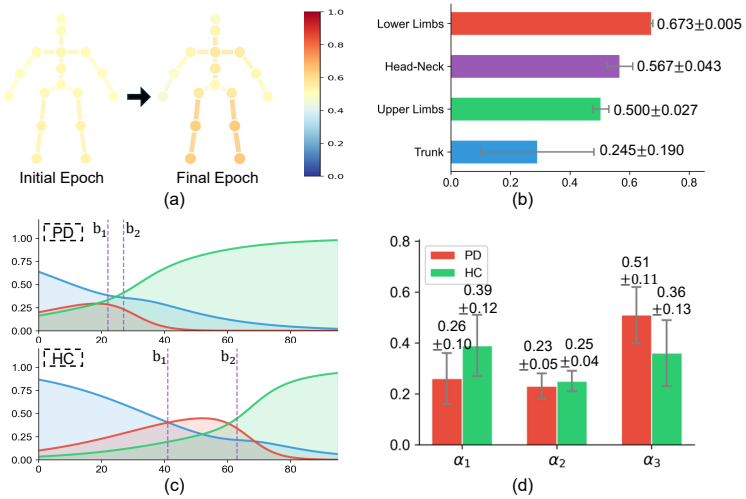

Fig. 2. Interpretability analysis of the proposed framework. (a) Skeleton edge importance at initial and final training epochs. (b) Learned edge group importance at the final epoch. (c) Soft phase masks for a representative PD patient (top) and healthy control (bottom); blue, red, and green curves denote Early-Turn ($k = 1$), Mid-Turn ($k = 2$), and Late-Turn ($k = 3$), respectively; dashed lines mark detected boundaries b_1, b_2. (d) Mean phase attention weights for PD and HC over the test set.

clinical knowledge (Fig. 2(a)–(b)), assigning the highest importance to lower limb connections while giving comparatively lower importance to trunk connections. Although axial rigidity and reduced trunk rotation are key symptoms of PD during turning, axial rigidity severely constrains trunk dynamics in PD patients, yielding limited discriminative contrast relative to healthy controls. In contrast, lower extremity kinematics exhibit more pronounced inter group differences, consistent with established evidence that these features are primary biomarkers of Parkinsonian gait impairment [16,15].

The temporal decomposition branch identifies soft temporal transition candidates that differ systematically between groups (Fig. 2(c)). In the PD example, the two boundaries are detected early and in close proximity, allowing the Late-Turn phase to dominate the sequence. In the HC example, boundaries are later and more widely spaced, yielding a balanced three-phase structure. The instance-level attention weights (Fig. 2(d)) reveal that PD patients exhibit higher attention on the late temporal instance with $\alpha_3 = 0.51$ compared to HC with $\alpha_3 = 0.36$, indicating that the Late-Turn phase is particularly discriminative for PD, consistent with observations of prolonged postural adjustment after turning [22,18]. The labels "Early-Turn," "Mid-Turn," and "Late-Turn" are temporal-ordering names for the discovered soft instances, not validated frame-level biomechanical phase annotations; direct phase-level validation remains future work.

5 Conclusion

We propose a weakly-supervised temporal decomposition framework for skeleton-based Parkinsonian turning assessment. The dual-branch architecture integrates soft temporal decomposition with instance-level attention aggregation, enabling the discovery of discriminative temporal patterns from video-level diagnostic labels alone. Evaluated on clinical turning sequences, the framework achieves an AUC-ROC of 91.05% while providing clinically interpretable outputs consistent with established understanding of PD motor symptoms. While the current evaluation is limited to a single clinical dataset due to patient privacy constraints on data sharing, multi-site validation is identified as a priority for future work, alongside severity grading and 3D pose estimation.

Acknowledgments. This work was supported by grants from the Institute of Information & Communications Technology Planning & Evaluation (IITP), the National Research Foundation of Korea (NRF), and the National IT Industry Promotion Agency (NIPA), funded by the Ministry of Science and ICT (MSIT), under Grant Nos. H0601-24-1023, IITP-2025-RS-2020-II201821, RS-2019-II190421, RS-2025-25442569, and RS-2025-24803185.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: 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)
2. Crenna, P., Carpinella, I., Rabuffetti, M., Rizzone, M., Lopiano, L., Lanotte, M., Ferrarin, M.: Impact of subthalamic nucleus stimulation on the initiation of gait in parkinson's disease. Experimental Brain Research **172**(4), 519–532 (2006)
3. Do, J., Kim, M.: Skateformer: skeletal-temporal transformer for human action recognition. In: European Conference on Computer Vision. pp. 401–420. Springer (2024)
4. Dorsey, E.R., Sherer, T., Okun, M.S., Bloem, B.R.: The emerging evidence of the parkinson pandemic. Journal of Parkinson's Disease **8**(s1), S3–S8 (2018)
5. Hong, M., Earhart, G.M.: Rotating treadmill training reduces freezing in parkinson disease: preliminary observations. Parkinsonism & Related Disorders **14**(4), 359–363 (2008)
6. Hong, M., Perlmutter, J.S., Earhart, G.M.: A kinematic and electromyographic analysis of turning in people with parkinson disease. Neurorehabilitation and Neural Repair **23**(2), 166–176 (2009)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2127–2136. PMLR (2018)

8. Kawabata, R., Yokoyama, M., Matsumoto, Y., Kubota, K., Kosuge, S., Sunaga, Y., Kanemura, N.: Lower limb biomechanics and control of center of mass during turning phases in daily gait. Journal of Electromyography and Kinesiology **80**, 102959 (2025)
9. Li, S., Farha, Y.A., Liu, Y., Cheng, M.M., Gall, J.: Ms-tcn++: Multi-stage temporal convolutional network for action segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(6), 6647–6658 (2023)
10. Li, T., Chen, J., Hu, C., Ma, Y., Wu, Z., Wan, W., Huang, Y., Jia, F., Gong, C., Wan, S., et al.: Automatic timed up-and-go sub-task segmentation for parkinson's disease patients using video-based activity classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(11), 2189–2199 (2018)
11. Li, Z., Abu Farha, Y., Gall, J.: Temporal action segmentation from timestamp supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8365–8374 (2021)
12. Liu, Z., Zhang, H., Chen, Z., Wang, Z., Ouyang, W.: Disentangling and unifying graph convolutions for skeleton-based action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 143–152 (2020)
13. Mehrizi, R., Peng, X., Zhang, S., Liao, R., Li, K.: Automatic health problem detection from gait videos using deep neural networks. arXiv preprint arXiv:1906.01480 (2019)
14. Mellone, S., Mancini, M., King, L.A., Horak, F.B., Chiari, L.: The quality of turning in parkinson's disease: a compensatory strategy to prevent postural instability? Journal of NeuroEngineering and Rehabilitation **13**(1), 39 (2016)
15. Miri, A.L., Laskovski, L., Bueno, M.E.B., Rodrigues, D.C., Moura, F.A., Smaili, S.M.: A biomechanical analysis of turning during gait in individuals with different subtypes of parkinson's disease. Clinical Biomechanics **112**, 106166 (2024)
16. Mitchell, T., Conradsson, D., Paquette, C.: Gait and trunk kinematics during prolonged turning in parkinson's disease with freezing of gait. Parkinsonism & Related Disorders **64**, 188–193 (2019)
17. Morgan, C., Tonkin, E.L., Masullo, A., Jovan, F., Sikdar, A., Khaire, P., Mirmehdi, M., McConville, R., Tourte, G.J., Whone, A., et al.: A multimodal dataset of real world mobility activities in parkinson's disease. Scientific Data **10**(1), 918 (2023)
18. Nguyen, A., Roth, N., Ghassemi, N.H., Hannink, J., Seel, T., Klucken, J., Gassner, H., Eskofier, B.M.: Development and clinical validation of inertial sensor-based gait-clustering methods in parkinson's disease. Journal of NeuroEngineering and Rehabilitation **16**(1), 77 (2019)
19. Nouredanesh, M., Kukreja, S.L., Tung, J.: Detection of compensatory balance responses using wearable electromyography sensors for fall-risk assessment. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 1680–1683. IEEE (2016)
20. Rehman, R.Z.U., Klocke, P., Hryniv, S., Galna, B., Rochester, L., Del Din, S., Alcock, L.: Turning detection during gait: algorithm validation and influence of sensor location and turning characteristics in the classification of parkinson's disease. Sensors **20**(18), 5377 (2020)
21. Richard, A., Kuehne, H., Gall, J.: Weakly supervised action learning with rnn based fine-to-coarse modeling. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 754–763 (2017)
22. Shah, V.V., McNames, J., Mancini, M., Carlson-Kuhta, P., Nutt, J.G., El-Gohary, M., Lapidus, J.A., Horak, F.B., Curtze, C.: Digital biomarkers of mobility in

parkinson's disease during daily living. Journal of Parkinson's Disease **10**(3), 1099–1111 (2020)
23. 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. pp. 1010–1019 (2016)
24. Shi, L., Zhang, Y., Cheng, J., Lu, H.: Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12026–12035 (2019)
25. Stack, E., Ashburn, A.: Dysfunctional turning in parkinson's disease. Disability and Rehabilitation **30**(16), 1222–1229 (2008)
26. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5693–5703 (2019)
27. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
28. Zhang, J., Chung, T.M., Park, H.: A kinematic-guided dual-branch framework for parkinson's disease assessment in sit-to-stand tasks. Engineering Applications of Artificial Intelligence **176**, 114679 (2026)
29. Zhang, J., Hur, S., Lee, Y., Chung, T.M.: Investigation on ai-powered digital health in parkinson's disease diagnosis. In: 2025 IEEE 15th International Conference on System Engineering and Technology (ICSET). pp. 312–317. IEEE (2025)
30. Zhang, J., Lim, J., Kim, M.H., Hur, S., Chung, T.M.: Wm–stgcn: A novel spatiotemporal modeling method for parkinsonian gait recognition. Sensors **23**(10), 4980 (2023)
31. Zhang, Q., Xie, H., Zhao, B., Zhuang, Y., Wang, H., Liu, Y., Gao, D., Zhang, H., Meng, F., Yang, A., et al.: Neural signatures of turn-induced freezing of gait in parkinson's disease: insights from phase-specific cortico-subthalamic dynamics. npj Parkinson's Disease **11**(1), 305 (2025)