A dual stream graph convolutional network based framework for Parkinson's disease severity assessment by analyzing gait using skeleton and gait energy image features

I. Data augmentation

Data augmentation has been applied to expand the dataset and introduce temporal variability, thereby enhancing model robustness, sensitivity to subtle PD related gait patterns, and mitigating overfitting, as detailed below.

a) Skeleton frames: Skeleton based augmentation strategy has been applied by generating five clips per video with varying frame skips, expanding the dataset fivefold. The procedure has involved two key steps: **(i) middle frame identification** and **(ii) symmetric frame extraction**. The middle frame has been determined as the ceiling of half the total frame count, i.e., for a video with N frames, the index is given by $f_m = \lceil \frac{N}{2} \rceil$ (e.g., for $N = 13$, $f_m = 7$). Around this middle frame, symmetric frames have been selected at an interval of skip size S. The selected frame set has been expressed as $\mathcal{F}(S) = \{f_m - kS,\ f_m + kS \mid k = 1, 2, \ldots, K\}$. where, K is the largest integer such that $1 \leq f_m - kS$ and $f_m + kS \leq N$. This selection has emphasized the most informative central frames while discarding boundary frames of lesser relevance. For example, with $N = 13$, $f_m = 7$, and $S = 1$, the set has become $1, 3, 5, 7, 9, 11, 13$. The process has been repeated for $S \in \{1, 2, 3, 4, 5\}$, with progressively more frames skipped as S increases. Frames have been selected until either the start or end of the video has been reached, or a class specific target frame count N_{avg} has been met, The number of selected frames has been constrained by the class specific target length as $|\mathcal{F}(S)| \leq N_{\text{avg}}^{(c)}$, where $N_{\text{avg}}^{(c)}$ denotes the average target frame count for class $c \in \{\text{NM (healthy)}, \text{PD-1 (stage 1)}, \text{PD-2 (stage 2)}\}$. As shown in Fig. 1, the 5th, 3rd, and 1st frames have been collected to the left of the 7th frame, skipping one frame each time; similarly, the 9th, 11th, and 13th frames have been collected to the right. The strategy has emphasized central informative frames while discarding boundary frames, thereby reducing video length, expanding the dataset, and improving efficiency. Table (A) I presents class wise video count, average N_{avg} values, and average frame length for augmented videos across skip sizes (S_1–S_5) on two datasets.

b) GEI frames: Class imbalance has been a key challenge in computer vision, especially for medical and gait based PD diagnosis. It has often led to biased models favoring majority classes, resulting in high overall performance but low sensitivity, which is crucial for early PD detection. In this work, both datasets have shown imbalance, with fewer PD samples than NM, causing overfitting and reduced generalization. To mitigate this, GEI based data augmentation has been applied to the minority class using geometric transformations such as rotation, scaling, and horizontal flipping. The transformation pipeline is defined as $T_{\text{final}}(x, y) = T_{\text{flip}}(T_{\text{scale}}(T_{\text{rotate}}(x, y)))$, where $T_{\text{rotate}} \in [-10°, 10°]$, $T_{\text{scale}} \in [0.9, 1.1]$, and T_{flip} is applied with probability 0.5. These augmentations have simulated natural gait variations while preserving structural characteristics. To further enhance the motion related cues within static GEI representations, a motion aware augmentation technique has been introduced using dense optical flow computed between temporally adjacent frames. Dense optical flow between two consecutive frames I_t and I_{t+1} has been estimated using the Farnebäck method [1] as $\mathbf{F}_{t \to t+1}(x, y) = (u(x, y), v(x, y))$, where u, v denote horizontal and vertical components. The motion magnitude is $M(x, y) = \sqrt{u^2 + v^2}$, and the augmented image is generated by blending motion with the original frame as $I_{\text{aug}}(x, y) = \alpha I_t(x, y) + (1 - \alpha)M(x, y)$, where $\alpha \in [0, 1]$ is a blending factor. This GEI augmentation has preserved structural gait features while enriching motion dynamics, thereby improving the model's robustness and classification performance. Table (B) I summarizes the number of GEI frames before and after applying this augmentation strategy.

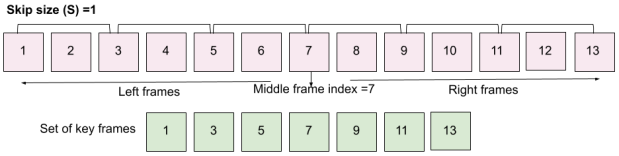

Fig. 1: Example of data augmentation using video frames with skip size $S = 1$, repeated for $S \in \{1, 2, 3, 4, 5\}$.

II. Saliency map

For skeleton and GEI based gait analysis, saliency maps have highlighted the most influential joints or regions, helping to validate key features and understand gait changes in PD. The saliency map generation, value computation, and hypothesis testing on saliency values are described below.

a) Saliency map and value computation: In this work, gradient weighted class activation mapping (Grad-CAM) [2] has been employed to interpret CNN predictions by highlighting the input regions that influence the model's decision. It has computed a class specific localization map $L_{\text{Grad-CAM}}^{c}$ as a weighted combination of feature maps A^k, where the weights α_k^c are obtained from the gradients of the target class score y^c. The weights are defined as $\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$, and the localization map is given by $L_{\text{Grad-CAM}}^{c} = \text{ReLU}\big(\sum_k \alpha_k^c A^k\big)$. In addition, saliency maps have quantified critical input regions

TABLE I: (A) Class wise video count, average N_{avg}, and mean frame length of augmented skeleton sequences for skip sizes S_1–S_5 on both datasets. (B) Total number of GEI frames before and after applying the data augmentation strategy

Dataset	(A) Skeleton sequence augmentation																					(B) GEI frame augmentation					
	Total Videos			Average of frames			S_1			S_2			S_3			S_4			S_5			Before augmentation			After augmentation		
	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM	PD-1	PD-2	NM
GAIT-IST	40	40	40	186	186	77	93	93	39	62	62	26	47	47	20	37	37	16	31	31	13	257	218	120	492	495	478
GAIT-IT	92	92	92	474	581	198	237	291	99	158	194	66	119	145	50	95	116	40	79	97	33	656	839	342	860	839	887

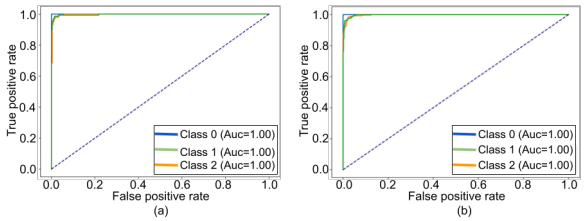

Fig. 2: ROC curve: (a) GAIT-IST and (b) GAIT-IT datasets.

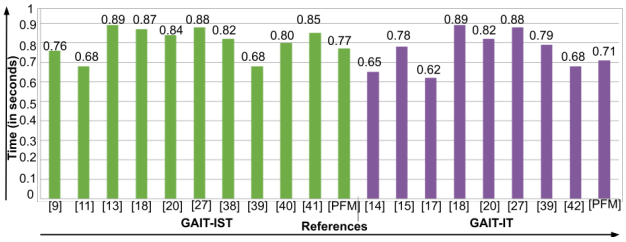

Fig. 3: Comparison of PFM and SOTA methods in terms of average test time per sample on two datasets.

as $S = \max_c \left| \frac{\partial y^c}{\partial I} \right|$. For skeleton data with the HGCN model, the saliency of the j^{th} joint has been computed as $S_j = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\partial \mathcal{L}(x_i)}{\partial x_{i,j}} \right|$, where $x_{i,j}$ denotes the input of joint j in the i^{th} frame, $\mathcal{L}(x_i)$ represents the model output, and N is the number of frames. The resulting joint wise saliency vector is expressed as $\mathbf{S} = [S_1, S_2, \ldots, S_J]$, where J is the number of joints. This formulation has quantified the model's sensitivity to perturbations at individual joints.

b) Hypothesis test on saliency values: Joint wise saliency values have been aggregated on test samples and have been grouped by class labels—NM, PD-1, and PD-2. Class wise differences have been assessed using the non-parametric Kruskal-Wallis H-test for each joint. For joints that have shown significant variation ($p < 0.05$), Dunn's post hoc test with Bonferroni correction has identified the specific class pairs that have differed significantly. Spearman's rank correlation has been used to examine associations between joint saliency and disease severity ($0 = \text{NM},\ 1 = \text{PD-1},\ 2 = \text{PD-2}$). The significance of the correlation coefficient (ρ) has been tested using a two tailed hypothesis test, with $p < 0.05$ indicating rejection of the null hypothesis and a statistically significant relationship. This framework has provided interpretable insights into the role of specific joints in PD classification.

III. PD Diagnosis Results

Fig. 2 presents the ROC curves for the proposed PFM model on the GAIT-IST and GAIT-IT datasets. The curves have shown near perfect classification performance, with the AUC reaching 1.00 for all three classes on both datasets. This indicates that the model has achieved an excellent balance between sensitivity and specificity, demonstrating its capability to accurately distinguish PD patients from healthy individuals. The consistently high AUC values on datasets have validated the robustness and generalization ability of the proposed approach for gait based PD diagnosis.

IV. Computational Complexity

Fig. 3 presents the comparison of PFM and SOTA methods in terms of average test time per sample on two datasets. The PFM has achieved competitive inference efficiency, with average test times of 0.77 s and 0.71 s on the GAIT-IST and GAIT-IT datasets, respectively. These results have demonstrated that, despite incorporating dual stream processing and Attn mechanisms, the proposed model maintains low computational overhead. Compared with other existing approaches, PFM has provided a favorable balance between accuracy and inference speed, confirming its suitability for practical and real time PD diagnosis using gait analysis.

V. Strengths of the Proposed Method

The proposed framework demonstrates several key strengths in gait based PD diagnosis:

- The dual stream integration of skeleton and GEI modalities enables complementary spatial temporal feature extraction, improving diagnostic accuracy and generalization.
- The CPAA efficiently extracts keypoints directly from binary skeletons, eliminating pretrained pose dependencies and ensuring robust joint localization.
- The HGCN, combining MultiScaleGCN, residual learning, and temporal Attn, effectively models local and global gait dynamics.
- The GEI based CNN with spatial Attn enhances interpretability by emphasizing clinically relevant gait regions.

The dual stream fusion achieves high accuracy with low inference time, validated on GAIT-IST and GAIT-IT datasets, confirming the framework's robustness, interpretability, and computational efficiency for automated PD diagnosis.

References

[1] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Scandinavian conference on Image analysis*. Springer, 2003, pp. 363–370.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.