Enhancing Gait Video Analysis in Neurodegenerative Diseases by Knowledge Augmentation in Vision Language Model

Diwei Wang[1], Kun Yuan[1], Candice Muller[2], Frédéric Blanc[1,2], Nicolas Padoy[1], and Hyewon Seo[1]

[1]ICube laboratory, University of Strasbourg, CNRS, France

{d.wang, kyuan, f.blanc, npadoy, seo}@unistra.fr

[2] Hôpital de la Robertsau, France

{candice.muller, frederic.blan@}@chru-strasbourg.fr

Introduction

- Neurodegenerative diseases are a common cause of morbidity and cognitive impairment in older adults, with gait impairment being one of the motor symptoms.
- Our study concentrates on video-based pathological gait analysis using a limited set of clinical recordings, facilitating cost-effective monitoring and remote surveillance.

Knowledge-based Textual Prompts

- *Desc_i* is generated using ChatGPT-4, then refined by a neurologist.
- Based on **K**nowledge-**A**ware **P**rompt **T**uning [2], learnable prompt $\{C_i^k\}_{i=1,\ldots,N_{cls}} = Proj_\phi^k(RoBERTa(\{Desc_i\})) + \{X_i^k\}$, $k = 1,\ldots,8$. *Desc_i* is distilled using RoBERTa pre-trained with unified training strategy KEPLER [5], $\{X_i^k\}$ are learnable parameters.
- Keywords extracted from *Desc_i* is utilized as $\{D_i\}$, which is not learnable.

Utilize the Visual Prompts of Vita-CLIP [6]

Experiments and Results

Our approach is validated through experiments on two gait classification tasks:

- *Gait scoring:* Assess gait impairments based on MDS-UPDRS III gait score.
- *Dementia subtyping:* Differentiate the diagnostic groups (healthy / Dementia with Lewy Bodies (DLB) / Alzheimer's Disease (AD)).

Classification Results of Ablation Studies:

Configurations	Gait scoring		Dem. subtyp.	
	Acc.	Fscore	Acc.	Fscore
Baseline	64.78	60.75	86.27	79.24
Baseline+**KAPT**	65.98	61.97	87.29	78.48
Baseline+**NTE**	64.44	57.64	88.26	81.34
Ours	**67.76**	**62.59**	**90.08**	**83.86**

Classification Results Compared with SOTA:

Models	Gait scoring		Dem. subtyp.	
	Acc.	Fscore	Acc.	Fscore
OF-DDNet[3]	54.73	48.59	68.92	65.38
ST-GCN [4]	49.08	43.87	61.46	56.99
KShapeNet[1]	53.69	44.85	65.27	54.86
Ours	**67.76**	**62.59**	**90.08**	**83.86**

References

[1] R. Friji et al. Geometric deep neural network using rigid and non-rigid transformations for human action recognition. *ICCV*, 2021.

[2] B. Kan et al. Knowledge-aware prompt tuning for generalizable vision-language models. *ICCV*, 2023.

[3] M. Lu et al. Vision-based estimation of mds-updrs gait scores for assessing parkinson's disease motor severity. *MICCAI*, 2020.

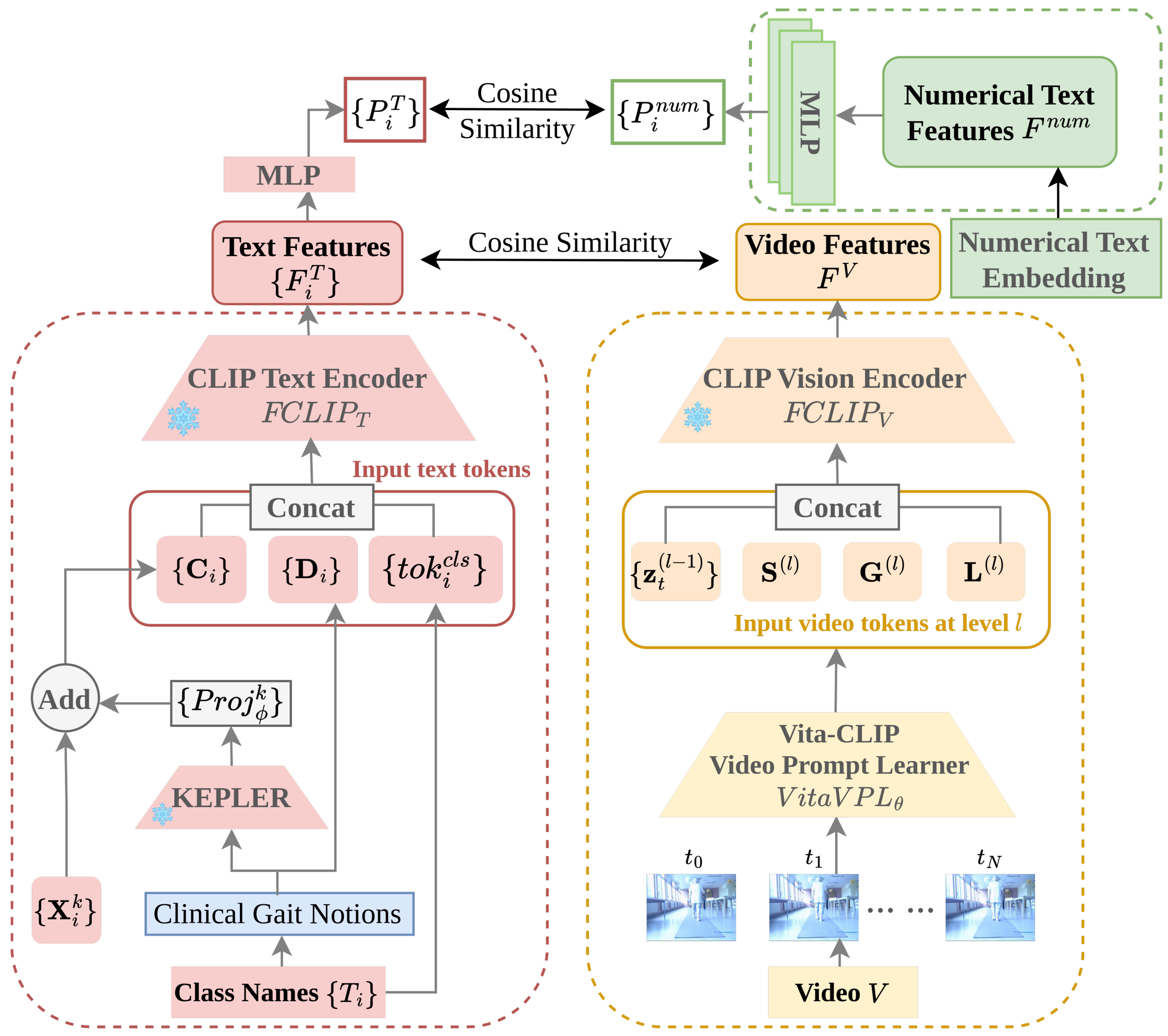
[4] A. Sabo et al. Estimating parkinsonism severity in natural gait videos of older adults with dementia. *IEEE J-BHI*, 26, 2022.

[5] X. Wang et al. Kepler: A unified model for knowledge embedding and pre-trained language representation. *TACL*, 9, 2021.

[6] S. T. Wasim et al. Vita-clip: Video and text adaptive clip via multimodal prompting. *CVPR*, 2023.

Method: Cross-modality Learning for Gait Classification Based on VLM

- We propose a knowledge augmentation strategy for diagnosing gait impairments in videos using large-scale pre-trained Vision Language Model (VLM). CLIP is utilized as backbone VLM.
- We improve visual, textual, and numerical representations learning through contrastive learning across 3 modalities: gait videos, class-specific descriptions, and numerical gait parameters.



Clinical Gait Notions $\{Desc\_i\}$

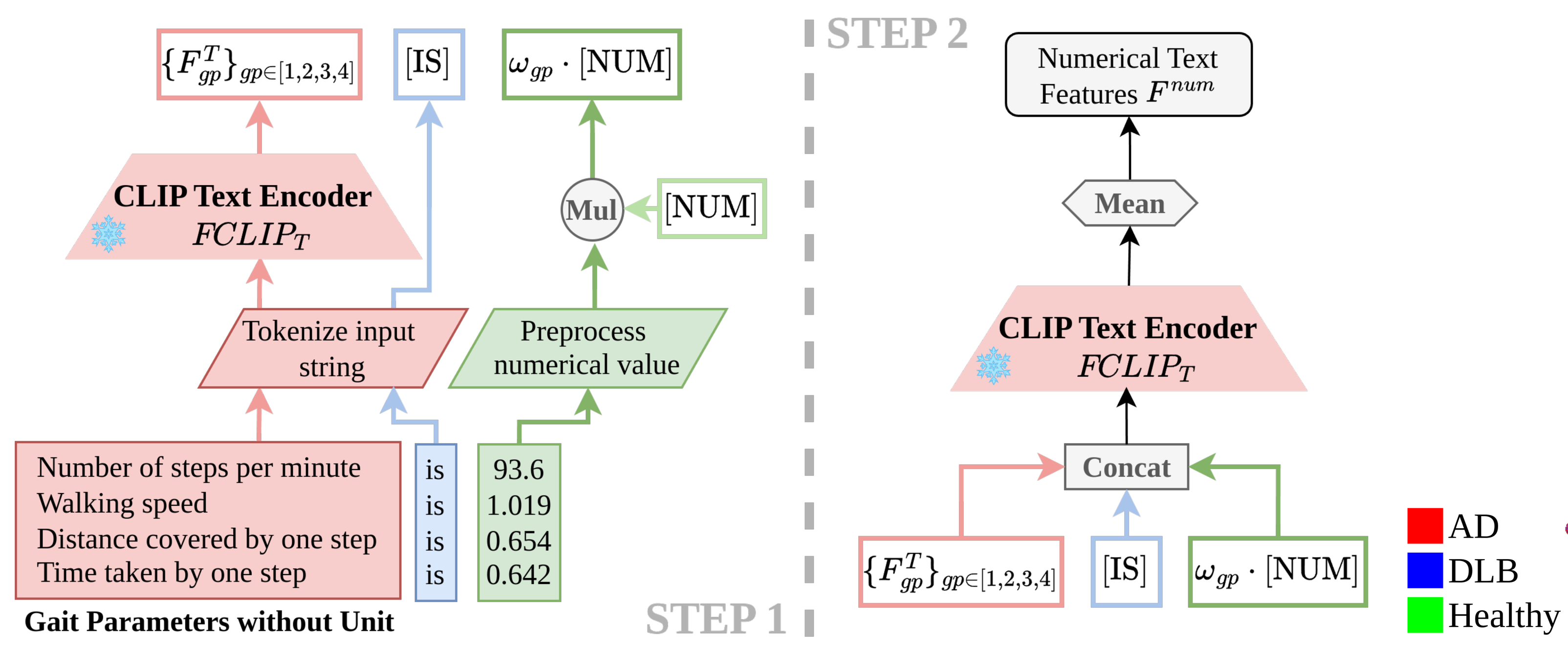
- **Normal gait** Normal gait indicates normal gait without any signs of impairment. As follows are the key features and their explanations. Normal Gait Pattern: The individual walks with (......)
- **Slight gait impairment** Slight gait impairment indicates slight or minimal impairment in gait. (......)
- **Mild gait impairment** (......)
- **Moderate gait impairment** (......)

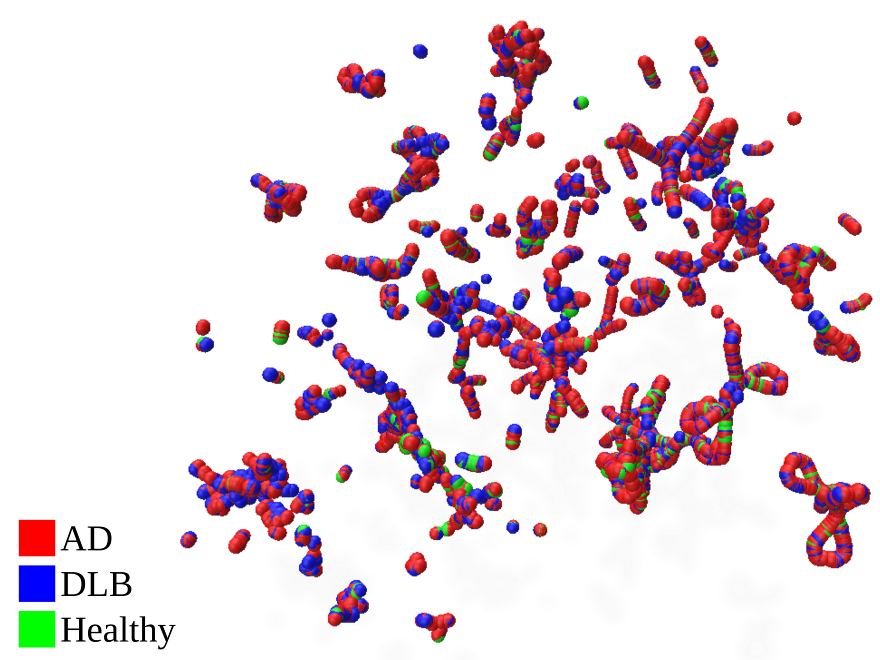
Discrete per-class prompt $\{D_i\}$

- **Normal gait** Normal Gait Pattern, No Shuffling or Dragging of Feet, Normal Arm Swing, No Balance Issues, No Assistive Device or assistant Required, Normal Speed and Rhythm, No Falling, No Freezing of Gait, No Difficulty when performing Dual Tasking.
- **Slight gait impairment** Any gait abnormalities are very subtle (......)
- **Mild gait impairment** (......)
- **Moderate gait impairment** (......)

Integrate Gait Parameters via Numerical Text Embedding (NTE)

- We employ a two-step process to embed sentences each containing 4 gait parameters.
- For numerical text embedding: $F^{num} = FCLIP_T(\{[F_{gp}^T, [\mathbf{IS}], \omega_{gp} \cdot [\mathbf{NUM}]]\})$, $gp \in \{1,2,3,4\}$.

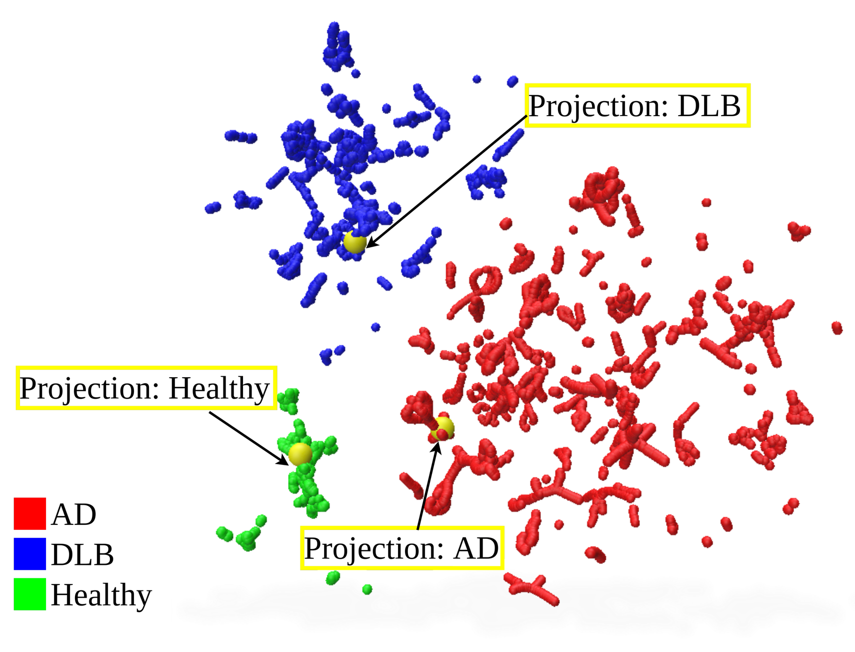




Interpretability: Per-class Text Feature Decoding

Idea: Decode $\{F_i^T\}$ to investigate whether the cross-modal alignment is formed through training.

- We train a 4-layer transformer decoder $\mathbf{D}^T$ to reverse the numercial text embedding.
- Ground-truth token IDs for numerical values [**num**]: $tok = [\mathbf{EOS}] + scale([\mathbf{num}])$.
- $F_i^T \sim \hat{F}_i^T = \sum_j softmax(\frac{p_i^T \cdot p_j^{num}}{\|p_i^T\| \cdot \|p_j^{num}\|} \cdot \frac{1}{\tau}) \cdot \frac{f_j^{num}}{\|f_j^{num}\|}$, where $\tau = 0.01$ and $f_j^{num} \in \{F^{num}\}$.
- To decode $\{F_i^T\}$ into natural language descriptions: $\{\hat{Desc_i}\} = \mathbf{D}^T(\{\frac{\hat{F}_i^T}{\|\hat{F}_i^T\|}\})$.



Healthy
Difference in distance covered between a left step and a right step is 0.002 leg,
angle between the progression line of left foot and the line from left heel to forefoot pressure center is 4.28 degree,
duration when both feet contact the ground within left walk cycle is 0.38 sec,
time when the right foot is off the ground within one walk cycle is 0.39 sec.

Dementia with Lewy Bodies (DLB)
Difference in distance covered between a left step and a right step is 0.08 leg,
angle between the progression line of right foot and the line from right heel to forefoot pressure center is -1.35 degree,
percentage of the duration when only the left foot contacts the ground within one walk cycle is 32.40 %,
time when the left foot is off the ground within one walk cycle is 0.43 sec.

Alzheimer's Disease (AD)
Difference in distance covered between a left step and a right step is 0.002 leg,
angle between the progression line of left foot and the line from left heel to forefoot pressure center is 4.28 degree,
percentage of the duration when only the left foot contacts the ground within one walk cycle is 32.40 %,
time when the right foot is off the ground within one walk cycle is 0.39 sec.