

Clustering-Constrained Attention Learning for Wearable-Based Rest Tremor Assessment

Theocharis Chatzis

Information Technologies Institute
Centre for Research and Technology Hellas
Thessaloniki, Greece
hatzis@iti.gr

Nikos Grammalidis

Information Technologies Institute
Centre for Research and Technology Hellas
Thessaloniki, Greece
ngramm@iti.gr

Kosmas Dimitropoulos

Information Technologies Institute
Centre for Research and Technology Hellas
Thessaloniki, Greece
dimitrop@iti.gr

Abstract—Rest tremor is a cardinal motor symptom of Parkinson’s disease, yet its assessment relies on brief observations and coarse ordinal ratings. Wrist-worn accelerometers facilitate scalable monitoring, but clinically relevant rest tremor is often sparse and intermittent. In this work, we treat wearable-based rest tremor assessment as a weakly supervised multiple-instance learning problem, where each recording is represented as a bag of short accelerometer windows supervised only at the assessment level. We propose a clustering-constrained attention framework that combines temporal modeling, class-specific gated attention and instance-level clustering to selectively aggregate tremor-relevant windows. We evaluate the proposed method on the MJFF Levodopa Response Study dataset for rest tremor detection and severity estimation. The proposed model improves held-out test performance for both detection and severity estimation, while the learned attention weights expose temporal windows contributing to the model decision and provide candidate temporal evidence for clinical interpretation.

Index Terms—Parkinson’s disease, rest tremor assessment, wearable sensors, accelerometry, multiple-instance learning, digital biomarkers.

I. INTRODUCTION

Parkinson’s disease (PD) is a progressive multisystem neurodegenerative disorder, mainly affecting people later in life [1], with rising incidence and prevalence in line with demographic change [2]. Clinically, PD involves both motor and non-motor symptoms, with tremor, bradykinesia and rigidity occurring most frequently. Tremor is a rhythmical, involuntary, oscillatory movement of a body part [3] and rest tremor is a key PD feature that can support the early diagnosis of the disease, although postural and kinetic tremors are also common [4]. Tremor severity is typically assessed using ordinal scales such as the Movement Disorder Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) [5]; however these ratings are subjective, based on short clinical observations and may not fully capture symptom fluctuations across time and daily contexts [6].

To this end, wearable inertial sensors allow for scalable and objective rest tremor monitoring [6]. Prior work has explored traditional machine learning [7], hybrid feature approaches [8], and deep learning models that learn directly from inertial signals [9]. Although recent convolutional and Transformer-based architectures have improved tremor detection and severity estimation [10], many sequence-based approaches still summa-

rize each input with a single global representation, implicitly treating all temporal segments as equally informative.

This global aggregation assumption is problematic for rest tremor, which is often sparse and intermittent. Clinically informative oscillatory activity may occur only in short segments, while other windows from the same assessment may correspond to tremor-free rest periods or movement artifacts. This challenge is amplified in clinical wearable datasets, where labels are usually assigned to an entire task or assessment rather than to individual temporal segments. As a result, a recording labeled with a nonzero tremor score may contain both tremor-relevant and tremor-free windows, while the model receives only a single assessment-level target.

To address this mismatch, we design wearable-based rest tremor assessment as a weakly supervised multiple-instance learning problem, where each assessment is treated as a bag of temporal windows supervised only by a coarse clinical label [11]. The proposed attention-based framework aims to identify informative temporal evidence without requiring manual window-level annotation and is evaluated on the Michael J. Fox Foundation Levodopa Response Study dataset (MJFF-LRS) [12].

The main contributions of this work are as follows:

- We frame rest tremor assessment from wrist-worn accelerometry as a weakly supervised multiple-instance learning task, explicitly targeting its sparse and intermittent nature.
- We propose an attention-based framework with class-specific gated attention and instance-level clustering constraints to selectively aggregate tremor-relevant temporal windows and support clinical interpretation.
- We validate the proposed approach on the MJFF-LRS dataset across both tremor detection and severity estimation tasks.

II. METHOD

A. Problem Formulation

Fig. 1 summarizes the proposed weakly supervised multiple-instance learning framework. Each recording is divided into short overlapping windows and represented as a bag $B_i = \{x_{i1}, x_{i2}, \dots, x_{iN_i}\}$ with bag-level label y_i , where x_{ij} is the

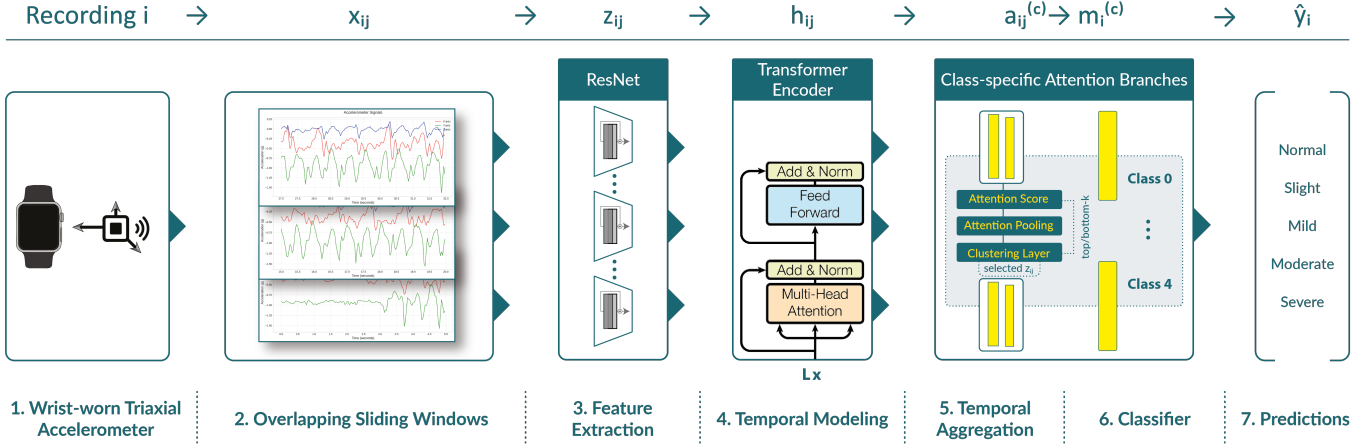


Fig. 1: Overview of the proposed clustering-constrained attention framework. Each window x_{ij} is encoded by the ResNet into z_{ij} and temporally contextualized by the Transformer into h_{ij} . Class-specific attention weights $a_{ij}^{(c)}$ form pooled representations $m_i^{(c)}$ for severity prediction $\hat{y}_i$. The attention scores also rank windows for top-/bottom- k selection, while instance clustering is applied to the corresponding ResNet embeddings z_{ij} .

j -th temporal window. Clinical supervision is available only at the assessment level through the corresponding MDS-UPDRS rest tremor score. The objective is to predict tremor severity, from which tremor presence is derived, while learning which temporal windows provide the most relevant evidence.

B. Signal Preprocessing and Window Encoding

We resample all signals to 30 Hz and segment them into overlapping windows of 5 s with a 1 s stride. To emphasize rest tremor-related oscillatory components, we apply a first-order Butterworth band-pass filter between 3 and 7.5 Hz, consistent with the characteristic frequency range of Parkinsonian rest tremor [10], before window encoding. Subsequently, we pass each window through a one-dimensional ResNet encoder, pre-trained by self-supervised learning [13] on the UK-Biobank dataset [14], producing a fixed-dimensional representation for each temporal segment. The ResNet encoder maps each window x_{ij} to an embedding z_{ij} , forming a sequence $Z_i = \{z_{i1}, z_{i2}, \dots, z_{iN_i}\}$ for each assessment.

C. Temporal Modeling

We feed the ResNet embeddings Z_i to a Transformer encoder to capture dependencies across windows, producing temporally encoded representations $H_i = \{h_{i1}, h_{i2}, \dots, h_{iN_i}\}$. The Transformer captures relationships among temporal segments through self-attention, enabling each window representation to integrate information from both nearby and distant windows. Unlike conventional sequence-level aggregation, the resulting temporal representations are not directly collapsed using global pooling. Instead, we pass them to an attention-based multiple-instance aggregation module designed to selectively emphasize tremor-relevant segments.

D. Attention-Based Multiple-Instance Aggregation

Given the temporally encoded window representations $H_i = \{h_{i1}, h_{i2}, \dots, h_{iN_i}\}$, and inspired by Clustering-constrained Attention Multiple Instance Learning (CLAM) [11], we use gated attention to compute class-specific attention weights over the temporal windows. For each tremor severity class c , a separate attention head estimates the relevance $a_{ij}^{(c)}$ of each window. The corresponding class-specific bag representation is computed as:

$$m_i^{(c)} = \sum_{j=1}^{N_i} a_{ij}^{(c)} h_{ij},$$

where h_{ij} is the temporally encoded representation of window j and $a_{ij}^{(c)}$ is its attention weight for class c . This allows the model to emphasize short intervals that are informative for severity prediction while reducing the contribution of tremor-free or noisy segments.

To encourage discriminative instance-level representations, for the ground-truth class we rank windows according to their attention scores and select the top- k and bottom- k instances. The corresponding ResNet embeddings z_{ij} are passed to a class-specific binary clustering head, with high-attention instances treated as class-positive and low-attention instances as less informative. We optimize

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{bag} + \lambda\mathcal{L}_{inst},$$

where $\mathcal{L}_{bag}$ is the multiclass cross-entropy loss between the bag-level severity prediction and y_i , while $\mathcal{L}_{inst}$ is a smooth top-1 SVM loss applied to the selected instances. The parameter λ controls the contribution of the instance-level constraint. The attention weights indicate which temporal windows contribute most strongly to the bag-level decision; they are model-derived evidence rather than window-level tremor annotations.

III. EXPERIMENTS

A. Datasets

We evaluate the proposed approach using the MJFF Levodopa Response Study dataset (MJFF-LRS) [12], which contains accelerometer recordings from people with PD, collected over four days, both in clinic and at home. The dataset includes triaxial accelerometer signals and clinical annotations for motor tasks, enabling evaluation of wearable-based tremor assessment under assessment-level supervision.

We use data from the GENEActiv wrist-worn sensor and retain only Day 1 and Day 4 recordings, where participants performed structured motor tasks with clinical annotations. Participants were assessed ON medication on Day 1, and on Day 4 after approximately 12 hours OFF medication, with motor tasks repeated before and after medication intake. We focus on tasks that do not involve upper-limb movement, including sitting, standing, sit-to-stand transitions, stairs up/down and walking tasks. These tasks are more likely to include periods in which the hand is not actively engaged in voluntary movement, making them suitable for rest tremor assessment.

The final subset contains 2,279 task instances from 27 participants, corresponding to 20.52 hours of accelerometer data. The subset is strongly class-imbalanced: 1,715 of 2,279 instances correspond to score 0, while scores 3 and 4 contain 88 and 2 instances, respectively; score 4 is represented by only one participant.

B. Experimental Protocol

We split the MJFF-LRS dataset subject-independently into 70% training, 10% validation, and 20% test sets to avoid leakage across participants and to assess generalization to unseen individuals. The validation set is used for model and hyperparameter selection, while the test set is reserved for final evaluation on unseen participants. The held-out test set contains severity scores 0–3. During training, we apply stochastic data augmentation, including Gaussian noise injection, random scaling, sequence reversal, and window permutation, each with augmentation probability 0.5. Augmentations are disabled during validation and testing.

We compare the proposed CLAM-based model against a ResNet–Transformer baseline [10] that uses the same ResNet window encoder and Transformer-based temporal modeling backbone but aggregates temporal information through global sequence-level pooling for two tasks: binary rest tremor detection and rest tremor severity classification. For detection, tremor scores greater than zero are treated as tremor-positive. For severity classification, the model predicts the MDS-UPDRS rest tremor severity score. Scores range from 0 to 4, corresponding to normal, slight, mild, moderate, and severe tremor. We report weighted and macro F1-scores for both tasks, with severity macro-F1 averaged over classes with nonzero support in the evaluated split. For severity estimation, we additionally report per-class precision and recall. Spearman’s rank correlation coefficient ρ is computed between the ground-truth MDS-UPDRS scores y_i and the corresponding

TABLE I: Held-out MJFF-LRS test performance on unseen participants.

Metric	ResNet–Transformer	Proposed
Detection weighted F1	0.81	0.87
Detection macro F1	0.74	0.81
Severity weighted F1	0.69	0.74
Severity macro F1	0.32	0.40
Spearman’s ρ	0.55	0.68
Score 0 P/R	0.87/0.90	0.90/0.94
Score 1 P/R	0.03/0.01	0.19/0.07
Score 2 P/R	0.12/0.16	0.28/0.28
Score 3 P/R	0.21/0.29	0.21/ 0.42

hard predicted severity labels $\hat{y}_i$. In qualitative analysis, we examine attention maps from correctly classified instances by mapping window-level attention weights back to the original recording timeline; the tremor-band root-mean-square (RMS) amplitude serves solely as a signal-level reference. The ResNet is fine-tuned end-to-end and produces 512-dimensional embeddings; the Transformer uses two layers and four heads, the attention hidden dimension is 128, and instance clustering uses $k = 8$ top- and bottom-ranked windows. Training uses 35 epochs of SGD (learning rate 0.005, momentum 0.9, batch size 64). Validation evaluated filter orders 1–4, overlaps 0–4 s, $\lambda \in \{0.1, 0.2, 0.3\}$, and augmentation probabilities $\{0, 0.3, 0.5\}$; the selected values were 1, 4 s, 0.1, and 0.5, respectively.

IV. RESULTS AND DISCUSSION

Table I summarizes held-out test performance. The proposed model improves detection weighted/macro F1 from 0.81/0.74 to 0.87/0.81, severity weighted/macro F1 from 0.69/0.32 to 0.74/0.40, and Spearman’s ρ from 0.55 to 0.68. Tremor-positive F1 also increases from 0.60 to 0.71. Per-class results show improvements across minority severities, although performance remains limited for slight tremor (score 1) and higher severity levels remain difficult because of their limited representation. The gain in ρ indicates improved ordinal agreement, although clinical significance cannot be established from this cohort.

Fig. 2 shows two correctly classified held-out test assessments with different severity levels (scores 2 and 3). In both cases, the class-specific attention varies substantially across the recording and assigns greater relevance to selected temporal windows rather than weighting all segments uniformly. Attention need not follow tremor-band RMS monotonically, since it reflects features contributing to the model decision rather than signal amplitude alone. Without window-level annotations, these maps constitute model-derived evidence rather than validated tremor localization. Overall, these results support the hypothesis that rest tremor assessment benefits from selective temporal aggregation. By assigning attention weights to individual temporal windows, the CLAM-based model can emphasize decision-relevant temporal information that may otherwise be diluted by global sequence pooling. This is particularly important for rest tremor, which can appear intermittently within longer assessments.

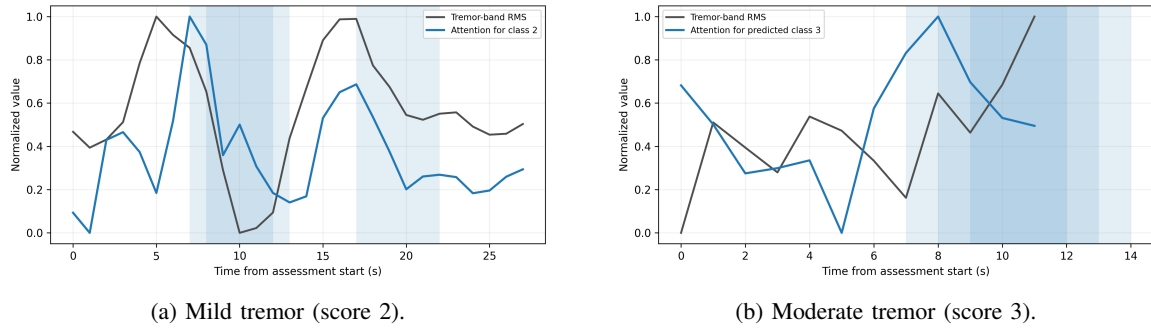


Fig. 2: Correctly classified held-out MJFF-LRS test assessments with (a) mild (score 2) and (b) moderate (score 3) tremor. Blue: normalized predicted-class attention; gray: normalized 3–7.5 Hz RMS. Overlapping shaded regions indicate the three highest-attention 5 s windows. RMS is a signal reference only; attention is not validated tremor localization.

However, several limitations remain. The cohort is small and strongly imbalanced, with very limited higher-severity data, particularly score 4. Although subject-independent evaluation, data augmentation, and validation-based model selection were used, evaluation on larger cohorts is needed to establish robustness. Furthermore, attention is model-derived rather than validated against expert temporal annotations, and evaluation is limited to MJFF-LRS and one wearable setup. Future work will investigate larger cohorts, cross-device validation, and expert temporal validation.

V. CONCLUSION

In this work, we formulate wearable-based rest tremor assessment as a weakly supervised multiple-instance learning problem and propose a CLAM-based model for selectively aggregating tremor-relevant temporal windows from wrist-worn accelerometer data. By combining temporal modeling, class-specific gated attention and instance-level clustering constraints, the model is designed to address the sparse and intermittent nature of rest tremor under assessment-level supervision.

The proposed approach improves rest tremor detection and severity estimation on MJFF-LRS while providing temporal attention maps that highlight windows assigned the greatest relevance by the model. These findings suggest that clustering-constrained attention learning is a promising direction for developing selective, interpretable wearable digital biomarkers in PD monitoring.

ACKNOWLEDGMENT

This work is part of the AI-PROGNOSIS project, which has received funding from the European Union under Grant Agreement No. 101080581.

These data were part of the L-dopa study funded by the Michael J. Fox Foundation.

REFERENCES

- [1] S. Sveinbjornsdottir, “The clinical symptoms of parkinson’s disease,” *Journal of neurochemistry*, vol. 139, pp. 318–324, 2016.
- [2] T. Pringsheim, N. Jette, A. Frolkis, and T. D. Steeves, “The prevalence of parkinson’s disease: a systematic review and meta-analysis,” *Movement disorders*, vol. 29, no. 13, pp. 1583–1590, 2014.
- [3] K. P. Bhatia, P. Bain, N. Bajaj, R. J. Elble, M. Hallett, E. D. Louis, J. Raethjen, M. Stamelou, C. M. Testa, G. Deuschl *et al.*, “Consensus statement on the classification of tremors. from the task force on tremor of the international parkinson and movement disorder society,” *Movement disorders*, vol. 33, no. 1, pp. 75–87, 2018.
- [4] J. Pasquini, R. Ceravolo, Z. Qamhawi, J.-Y. Lee, G. Deuschl, D. J. Brooks, U. Bonuccelli, and N. Pavese, “Progression of tremor in early stages of parkinson’s disease: a clinical and neuroimaging study,” *Brain*, vol. 141, no. 3, pp. 811–821, 2018.
- [5] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, “Movement disorder society-sponsored revision of the unified parkinson’s disease rating scale (mds-updrs): scale presentation and clinimetric testing results,” *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.
- [6] L. J. Evers, Y. P. Raykov, T. M. Heskes, J. H. Krijthe, B. R. Bloem, and M. A. Little, “Passive monitoring of parkinson tremor in daily life: a prototypical network approach,” *Sensors*, vol. 25, no. 2, p. 366, 2025.
- [7] F. Rodriguez, P. Krauss, J. Kluckert, F. Ryser, L. Stieglitz, C. Baumann, R. Gassert, L. Imbach, and O. Bichsel, “Continuous and unconstrained tremor monitoring in parkinson’s disease using supervised machine learning and wearable sensors,” *Parkinson’s Disease*, vol. 2024, no. 1, p. 5787563, 2024.
- [8] C. Zhou, “Developing a hybrid machine learning system for parkinson’s disease detection using a combination of freezing of gait video and inertial sensor data,” in *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*. IEEE, 2024, pp. 1–6.
- [9] C. Donié, N. Das, S. Endo, and S. Hirche, “Estimating motor symptom presence and severity in parkinson’s disease from wrist accelerometer time series using rocket and inceptiontime,” *Scientific Reports*, vol. 15, no. 1, p. 19140, 2025.
- [10] T. Chatzis, N. Grammalidis, and K. Dimitropoulos, “A deep learning approach for parkinsonian tremor assessment using wearables,” in *2025 IEEE International Conference on E-health Networking, Application & Services (Healthcom)*. IEEE, 2025, pp. 1–6.
- [11] M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, “Data-efficient and weakly supervised computational pathology on whole-slide images,” *Nature biomedical engineering*, vol. 5, no. 6, pp. 555–570, 2021.
- [12] J.-F. Daneault, G. Vergara-Diaz, F. Parisi, C. Admati, C. Alfonso, M. Bertoli, E. Bonizzoni, G. F. Carvalho, G. Costante, E. E. Fabara *et al.*, “Accelerometer data collected with a minimum set of wearable sensors from subjects with parkinson’s disease,” *Scientific Data*, vol. 8, no. 1, p. 48, 2021.
- [13] H. Yuan, S. Chan, A. P. Creagh, C. Tong, A. Acquah, D. A. Clifton, and A. Doherty, “Self-supervised learning for human activity recognition using 700,000 person-days of wearable data,” *NPJ digital medicine*, vol. 7, no. 1, p. 91, 2024.
- [14] A. Doherty, D. Jackson, N. Hammerla, T. Plötz, P. Olivier, M. H. Granat, T. White, V. T. Van Hees, M. I. Trenell, C. G. Owen *et al.*, “Large scale population assessment of physical activity using wrist worn accelerometers: the uk biobank study,” *PloS one*, vol. 12, no. 2, p. e0169649, 2017.