A Deep Learning Approach for Parkinsonian Tremor Assessment Using Wearables

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*—Parkinsonian tremor is one of the most common and functionally disruptive motor symptoms of Parkinson's disease (PD), particularly rest tremor due to its strong association with early stages of the disease. However, its clinical evaluation is often subjective and limited to in-clinic assessments. Wearable accelerometers allow for objective tremor tracking beyond clinical environments; however, current approaches often suffer from limited generalization, weak temporal modeling, and poor robustness to real-world variability. In this work, we present a deep learning framework for automatic tremor detection and amplitude classification using wrist-worn accelerometer data. Our method employs a ResNet encoder for spatial representation learning with a Transformer-based temporal model to capture the complex dynamics of tremor episodes. A conditional dual-head output mechanism focuses amplitude learning only when tremor is present. We evaluate the proposed method across multiple settings using data from two well-known datasets, namely the Michael J. Fox Foundation Levodopa Response Study dataset and a subset of the Verily Study Watch dataset. The results demonstrate that our framework generalizes well across wearable devices and recording protocols, highlighting its potential for continuous, real-world tremor tracking in PD.**

***Index Terms*—Parkinson's disease, Tremor Detection, Tremor Amplitude Classification, Accelerometry, Deep Learning, Wearable Sensors**

I. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects millions around the world, characterized by a variety of motor and non-motor symptoms. Motor symptoms, which include bradykinesia, rigidity, postural instability, and tremor can considerably limit daily activities and overall quality of life [1]. Tremor is defined as an involuntary, rhythmic, oscillatory movement of a body part [12] and comprises the most disruptive motor symptom [5], [6]. It constitutes one of the most common symptoms affecting up to 75% of PD patients throughout the course of the disease [3]. Rest tremor is a cardinal feature of PD, though kinetic and postural tremors may also be present [2]. Due to its strong correlation with early-stage PD, rest tremor plays a crucial role in both the detection and monitoring of the disease [2], [4].

However, tracking symptom fluctuations over time poses certain challenges [7]. In clinical settings, tremor is assessed using rating scales, such as MDS-UPDRS [8], which are based on expert observation. Thus, it is subjective to expert bias and limited to short in-clinic visits [10]. Outside of clinics, patient-reported diaries offer an alternative but are often inaccurate and prone to subjective reporting errors in real-world settings [11].

To overcome these limitations, recent studies have explored alternative modalities for Parkinson's disease motor assessment, including depth cameras and smartphone-based video analysis [28], [31], [32]. Among these, wearable inertial sensors, especially accelerometers, have gained traction as a non-intrusive means of providing continuous tremor monitoring in unconstrained settings. Accelerometers are widely employed to detect motor symptoms of PD and have been proven effective in capturing the characteristics of tremor [13], [14]. This can be attributed to their ability to monitor the movements patterns of PD patients over time, making them an essential tool for continuous monitoring in real-life applications, such as smart home healthcare systems [15]. Consequently, they can facilitate early disease detection and provide valuable insights into disease progression, enabling more informed clinical decision-making and personalized intervention strategies.

Methods for detecting and assessing Parkinsonian tremors based on accelerometers can be classified into traditional machine learning [9], [24]–[26], hybrid feature approaches [21], [27], and deep learning [26], [29], [30], [33]. Works that fall under the first category utilize techniques, such as SVM, K-means, Random Forest, etc., paired with hand-crafted features. On the other hand, some methods [27] have explored feature fusion strategies techniques or multi-modal inputs [21], e.g., video and inertial data, in order to further improve robustness, though they require the integration of different data streams and may be challenging to apply outside the clinical environment.

Recent works have explored deep learning approaches such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) networks, demonstrating improvements in automating tremor assessment [26], [29], [30]. However, many of these methods are trained on limited datasets, often involving a small number of subjects or relying on proprietary data. More recently, the authors in [33] employed RandOm Convolutional KErnel Transform (ROCKET) and InceptionTime, achieving a mean average precision of 0.544, on the Michael J. Fox Foundation Levodopa Response Study

dataset (MJFF-LRS), highlighting the potential of time-series models for this task.

In contrast to prior work, we explicitly model the complex spatio-temporal nature of tremor using a multi-task deep learning framework. Our architecture combines a pre-trained ResNet to extract rich embeddings from raw accelerometer data with a Transformer encoder to model temporal dependencies through self-attention. This design enables joint tremor detection and amplitude classification, enhancing generalization in dynamic, real-world settings. Importantly, we evaluate our framework on two well-known and publicly accessible datasets- the MJFF-LRS and a curated subset of the Verily Study Watch dataset. Although our method is designed for general tremor detection, we emphasize modeling rest tremor, given its high specificity to early-stage PD. Since rest tremor exhibits distinct frequency characteristics (3-7.5 Hz) [22], [23]. We incorporate targeted signal processing (e.g., band-pass filtering) and task-specific experimental settings (Section III-A) to improve modeling fidelity for rest tremor. In summary, our key contributions are as follows:

- We propose a deep learning architecture that combines a pre-trained ResNet encoder and Transformer-based temporal modeling for tremor detection and amplitude classification from raw wrist-worn accelerometry.
- We introduce a conditional dual-head prediction mechanism, where the amplitude loss is selectively backpropagated only when tremor is detected.
- We evaluate our method on both validation and test splits of the MJFF-LRS dataset as well as on an external clinical subset from the Verily Study Watch, demonstrating robust generalization across devices and task settings.
- We analyze the model's performance in two scenarios: (i) all-task setting and (ii) rest-related tasks, highlighting its utility for both broad and rest-specific tremor tracking.

II. Method

We propose a deep learning framework designed to detect tremor and classify its amplitude, inspired by the need for more accurate modeling of temporal dependencies among raw triaxial accelerometer signals. Our framework consists of five key stages: window segmentation, data augmentation, feature extraction using a pre-trained CNN, temporal modeling leveraging a Transformer encoder, and classification with a conditional dual-head design. The overall architecture of the proposed method is designed to handle the spatio-temporal complexity of tremor episodes and is presented in Fig. 1.

Data preprocessing and Windowing. Let the triaxial accelerometer signal for the i-th sample be denoted as $s_i^{1:T} = \{x_t, y_t, z_t\}_{t=1}^{T}$, where $s_i^{1:T} \in \mathbb{R}^{T\times 3}$ and T is the total number of timesteps. We first resample all signals to a uniform frequency f_r and segment them into overlapping windows of length w with stride o. This results in a sequence of

$$N = \left\lfloor \frac{T-w}{o} + 1 \right\rfloor$$

windows per sequence:

$$X_i = \{w_j\}_{j=1}^{N}, X_i \in \mathbb{R}^{N\times(f_r\times w)\times 3}.$$

For rest tremor analysis, we further preprocess input signals using a first-order Butterworth band-pass filter, to isolate the rest tremor-dominant frequency, 3-7.5 Hz.

Data Augmentation. To improve generalization to real-world variability and inter-subject differences, we perform various augmentation techniques with probability p_{aug}, during training:

- Gaussian noise injection:

$$w' = w + \mathcal{N}(0, \sigma^2)$$

- Random scaling:

$$w' = \alpha \cdot w, \quad \alpha \sim \mathcal{U}(\alpha_{\min}, \alpha_{\max})$$

- Sequence reversal:

$$w' = \{w_{L-t+1}\}_{t=1}^{L}, L = f_r \times w$$

- Window permutation: Given the full sequence of windows X_i, we apply random permutation along the window axis:

$$X_i' = \{w_{\pi_j}\}_{j=1}^{N}, \pi \sim UniformPerm(N)$$

These augmentations are applied only during training and are disabled during validation and testing.

Feature Extraction and Temporal Modeling. Subsequently, we pass each window w_j through a 1D Resnet-V2 encoder $f_{resnet}(\cdot)$ with 18 layers, which has been pre-trained by self-supervised learning [16] on the UK-Biobank dataset [17], producing embeddings $H_i = \{f_{resnet}(w_j)\}_{j=1}^{N}, H_i \in \mathbb{R}^{N\times D}$, where D is the dimensionality of the embeddings of the Resnet encoder. This large-scale dataset contains terabytes of real-world wearable sensor data, enabling robust feature extraction. Afterwards, we feed these embeddings into a Transformer encoder $f_{trans}(\cdot)$, consisting of L layers, to capture temporal dependencies across the sequence of window-level embeddings $\{h_1, ..., h_N\}$, leading to features $U_i = \{f_{trans}(H_i)\}, U_i \in \mathbb{R}^{N\times D}$. To obtain a unified representation for the entire signal, we apply max-pooling on the Transformer outputs, obtaining a fixed-length aggregated embedding $z_i \in \mathbb{R}^{1\times D}$, for the entire input signal $s_i^{1:T}$.

Dual-Task Prediction Head. As our final step, we pass the aggregated embeddings z_i through a shared linear projection followed by two separate task-specific classification heads. The first head, $f_{tremor}(f_{shared}(z_i))$, predicts the presence or absence of tremor, while the second, $f_{amp}(f_{shared}(z_i))$, classifies the tremor amplitude according to the MDS-UPDRS item 3.17, ranging from 1-4. To supervise these tasks, we use two standard loss functions, binary cross-entropy for tremor detection, L_{tremor}, and categorical cross-entropy for amplitude classification, L_{amp}. However, the amplitude loss is computed only when tremor is present, as determined by the ground-truth. Let λ be a scalar hyperparameter controlling

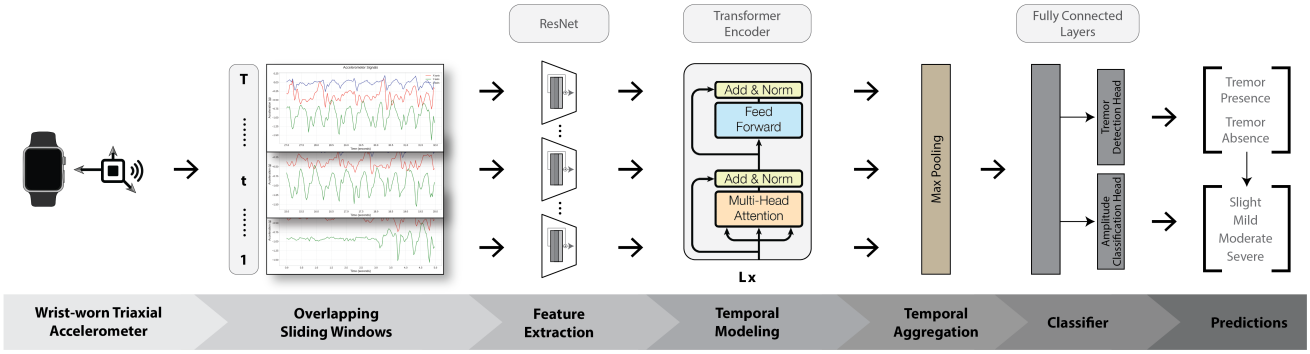

Fig. 1. **An overview of the proposed tremor tracking method.** Raw triaxial accelerometer signals are segmented into overlapping windows and passed through a pre-trained ResNet encoder for spatial feature extraction. The resulting window-level embeddings are modeled by a Transformer encoder with L layers to capture temporal dependencies. A max pooling layer aggregates temporal features into a fixed-length representation, which is fed into a dual-head classifier for tremor presence and amplitude prediction.

the weight of the amplitude loss. The total loss function is defined as follows:

$$L_{\text{total}} = L_{\text{tremor}} + \lambda \cdot \mathbb{1}[y_i^{\text{tremor}} = 1] \cdot L_{\text{amp}}$$

, where $\cdot\mathbb{1}$ is the indicator function returning 1 if tremor is present and 0 otherwise.

III. Experiments

A. Datasets & Experimental Setup

We conducted experiments using two datasets, the Michael J. Fox Foundation Levodopa Response Study dataset (MJFF-LRS) [19] and a curated subset of the Verily Study Watch dataset [20] from the Parkinson's Progression Markers Initiative (PPMI) study. MJFF-LRS was chosen for its detailed, labeled recording of wrist activity in PD patients, including accelerometer data and clinical annotations of motor tasks. Thus, it can be considered an ideal choice for accurately training our tremor tracking method. The Verily Study Watch dataset, provides longitudinal motion data from a diverse cohort, including PD patients, at-risk individuals, and healthy controls. Together, these datasets provide a robust foundation for both the development and validation of our tremor tracking framework.

MJFF-LRS dataset. This dataset includes wearable sensor data from individuals with PD, collected over 4 days both in-clinic and at home. For our analysis, we keep only the triaxial accelerometer recordings from the GENEActiv sensor, which is worn on the wrist of the most affected limb. To ensure label fidelity, we opt for data from Day 1 and Day 4, as these include motor tasks with expert clinical annotations, yielding 27 subjects and 18 tasks of mean duration of about 30 seconds. Participants were tested ON medication on Day 1, and on Day 4 after approximately 12 hours OFF medication, with motor tasks repeated before and after intake. While our model is designed to detect tremor in general, our objective is also to evaluate its performance specifically on rest tremor, given its clinical importance in early PD and its distinct characteristics from kinetic or postural tremor. We define two experimental settings:

- **All Tasks Setting** includes the entire set of task instances from the dataset (e.g., walking, standing, folding towel, etc.)
- **Rest-Related Tasks Setting** is a subset of tasks that do not involve upper limb movement. In particular, we removed classes such as repeated arm movements, drawing and writing on paper, and folding a towel. The remaining tasks—standing, sitting, sit-to-stand, stairs up, stairs down, and three walking variations—are more likely to contain periods where the hand is at rest and not actively engaged in movement, making them suitable for assessing rest tremor characteristics.

The distribution of tremor amplitude scores in both task settings is summarized in Table I (all tasks) and Table II (rest-related tasks), and visualized across all subjects in Fig. 2 and Fig. 3, respectively.

To ensure a fair evaluation, we split the MJFF-LRS dataset subject-independently into 70% training, 10% validation, and 20% test sets. The stratification preserves the tremor score distribution across sets to reflect the original class balance. This strategy is applied separately for both experimental settings, (i) all tasks and (ii) rest-related tasks.

Verily Study Watch Subset. For external validation, we use a subset of the Verily Study Watch dataset. First, we isolate clinical visit segments corresponding to MDS-UPDRS item 3.17-a or -b (rest tremor amplitude of right and left upper limb respectively), using timestamped annotation to align sensor data with clinician-observed tremor episodes. These assessments are performed while the participant is seated quietly in a chair with the hands placed on the arms of the chair, as per the MDS-UPDRS protocol. Since watch placement, left or right wrist, is unknown, we use the mean of scores from both limbs (items 3.17a and 3.17b) as the ground-truth label. After excluding subjects with incomplete data, the final subset includes 158 participants with 194 valid instances of rest tremor assessments, of mean duration approximately 10 seconds. The medication status during assessments varied among participants, with some being evaluated in the ON state

and others in the OFF state. This subset allows us to validate both task settings on a dataset that exclusively captures rest tremor in a controlled, annotated clinical context. Importantly, data in this subset are collected using a different wearable device, the Verily Study Watch, than the one used during training (GENEActiv), offering an additional layer of external robustness assessment. The score distribution is shown in Table III.

As shown in the distribution tables for both datasets, non-tremor instances constitute the vast majority of the data. In contrast, higher-severity tremor classes (scores 3 and 4) are sparsely represented.

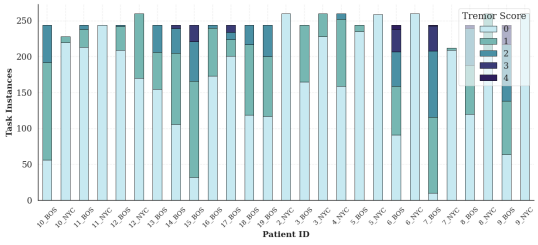

Fig. 2. All task instances. Distribution of tremor scores (severity levels 0–4) across all patients and tasks. Scores are represented as stacked color segments indicating severity.

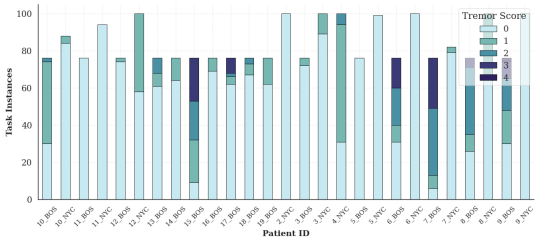

Fig. 3. Rest related tasks. Distribution of tremor scores (severity levels 0–4) across all patients during rest-related tasks. Stacked colors represent the proportion of each severity level.

TABLE I
DISTRIBUTION OF TREMOR SCORES ACROSS TASK INSTANCES AND SUBJECTS ON THE MJFF-LRS DATASET

Tremor Score (Amplitude)	Hours (% of Total)	Task Instances	Unique Subjects
0	36.40 (67.39%)	4541	27
1	12.29 (22.75%)	1433	22
2	4.09 (7.57%)	550	15
3	1.17 (2.16%)	132	7
4	0.07 (0.13%)	11	3
Total	54.02 (100%)	6667	27

TABLE II
DISTRIBUTION OF TREMOR SCORES ACROSS REST RELATED TASK INSTANCES AND SUBJECTS ON THE MJFF-LRS DATASET

Tremor Score (Amplitude)	Hours (% of Total)	Task Instances	Unique Subjects
0	15.22 (74.15%)	1715	27
1	2.96 (14.43%)	323	20
2	1.44 (7.01%)	151	10
3	0.88 (4.31%)	88	6
4	0.02 (0.09%)	2	1
Total	20.52 (100%)	2279	27

TABLE III
DISTRIBUTION OF TREMOR SCORES ACROSS CLINICAL VISIT INSTANCES ON THE VERILY STUDY WATCH SUBSET

Tremor Score (Amplitude)	Clinical Visit Instances	Unique Subjects
0	151	125
1	28	26
2	14	11
3	1	1
4	0	0
Total	194	158

B. Implementation Details

To align with the input requirements of the pre-trained feature extractor, we first resample all accelerometer triaxial signals to a fixed frequency $f_r = 30Hz$ and segment them into overlapping windows of length $w = 5$ seconds. To determine the optimal stride length, we evaluate stride values $o \in 0, 1, 2, 3, 4$ seconds on the validation set and conclude that a stride of $o = 1$ second yields the best results. Thus, this results in windows of shape 150×3 that correspond to 150 timesteps per axis of the triaxial acceleration data.

During training, we apply data augmentations with a probability of $p_{aug} = 0.5$, while we experiment with values $\{0, 0.4, 0.5\}$. Specifically, we add Gaussian noise sampled from $\mathcal{N}(0, 0.05^2)$, apply random scaling with factors drawn from the interval $[0.9, 1.1]$, reverse the temporal sequence of each window and permute the order of windows within a signal.

The pre-trained ResNet encoder outputs a $D = 512$-dimensional embeddings per window. This encoder is initialized using weights from a model trained via self-supervised learning on the UK-Biobank dataset and fine-tuned during training. Regarding temporal modeling, we use a Transformer encoder with $L = 2$ layers. To determine the most effective method for aggregating temporal features, we experiment with both max-pooling and average-pooling techniques, finding that the former achieved superior performance. Subsequently, we use a shared fully connected layer to reduce the dimensionality in the order of 4, while each dedicated head outputs a single prediction for the corresponding task. We set the weight of the amplitude loss $\lambda = 1$ for all tasks and $\lambda = 2$ for rest-related tasks. These choices are also made after experimenting with the following values: $\{1, 1.5, 2, 2.5, 3\}$.

In terms of optimization, we employ SGD optimizer with a learning rate set at 0.01 for all experiments. We train our framework in an end-to-end manner for 35 epochs using a single NVIDIA RTX 3090 GPU with 24GB VRAM, and we use the Pytorch [18] deep learning framework for implementing our method.

We evaluate our framework on two well-known datasets under two experimental conditions: the all-tasks setting and a subset restricted to rest-related tasks. In order to measure the performance on both tremor detection and tremor amplitude classification we use weighted precision, recall, and F1-score. To assess ordinal agreement with the clinician-provided tremor severity scores, we also report the Spearman's rank correlation

TABLE IV
PERFORMANCE METRICS FOR TREMOR DETECTION ON THE MJFF-LRS VALIDATION AND TEST SETS

Setting	Metric	No Tremor (Val/Test)	Tremor (Val/Test)	Weigted Avg (Val/Test)
All Tasks	**Precision**	0.80/0.90	0.74/0.71	0.78/0.84
	Recall	0.93/0.86	0.48/0.77	0.79/0.83
	F1-score	0.86/0.88	0.63/0.74	0.78/0.83
Rest-Related Tasks	**Precision**	0.87/0.87	1.00/0.65	0.91/0.81
	Recall	1.00/0.90	0.58/0.56	0.89/0.82
	F1-score	0.93/0.88	0.73/0.60	0.90/0.81

TABLE V
PERFORMANCE METRICS FOR TREMOR AMPLITUDE CLASSIFICATION ON THE MJFF-LRS VALIDATION AND TEST SETS

Setting	Metric (Weighted)	Validation	Test
All Tasks	**Precision**	0.67	0.77
	Recall	0.73	0.75
	F1-score	0.70	0.76
Rest-Related Tasks	**Precision**	0.81	0.68
	Recall	0.83	0.71
	F1-score	0.82	0.69

coefficient (ρ).

C. Results

The model achieved strong performance across both validation and test sets under the all-tasks setting on the MJFF-LRS dataset. For tremor detection, the weighted F1-score reached 0.78 on validation and 0.83 on the test set, with consistent precision and recall values, as reported in Table IV. In the amplitude classification task, the model performed F1-scores of 0.70 and 0.76 on the validation and test sets, respectively, as shown in Table V. Importantly, the Spearman correlation between predicted and ground-truth tremor scores was ρ = 0.52 ($p<0.05$) on the validation set and ρ = 0.76 ($p<0.05$) on the test set, indicating strong ordinal alignment, particularly when generalizing to unseen subjects.

Regarding the rest-related setting, we achieved a higher validation performance, especially in amplitude classification, where the F1-score reached 0.82 and Spearman correlation ρ = 0.78 ($p<0.05$), as detailed in Table V. Tremor detection on the same validation set achieved a high F1 of 0.90, reported in Table IV, indicating better discriminative performance when focusing on tasks more likely to cause rest tremor. On the test set, performance was slightly lower but still competitive, with an F1-score of 0.81 for tremor detection, 0.69 for amplitude classification, and a Spearman correlation of ρ = 0.55 ($p<0.05$). The slightly lower performance on the test set likely reflects greater variability and new patterns in unseen data, demonstrating the model's ability to generalize under realistic conditions.

On the Verily Study Watch dataset, used for external validation and recorded using a different wearable device, the model maintained robust generalization. In the all-tasks setting, the weighted F1-score for tremor detection reached 0.87, as presented in Table VI, while amplitude classification achieved 0.77, as shown in Table VII. The Spearman correlation to expert-assigned tremor scores was ρ = 0.49 ($p<0.05$). In the rest-related tasks setting on the Verily Study Watch dataset, performance remained stable. Tremor detection achieved an F1-score of 0.79 (Table VI), and amplitude classification again reached an F1 of 0.77 (Table VII). The corresponding Spearman correlation improved slightly to ρ = 0.57 ($p<0.05$). This suggests that the rest-focused setup improves the model's ability to assess rest tremor and yields even better correlation with expert labels in the Verily dataset compared to MJFF-LRS, likely due to differences in cohort characteristics and annotation protocols. Notably, these results are achieved despite uncertainty in the ground-truth labels due to unknown watch placement, which introduces label noise; even under such conditions, our method demonstrates robust ordinal alignment across both settings, as evidenced by consistent Spearman correlations on this external dataset.

Overall, these results demonstrate that the proposed framework performs reliably across both diverse activity contexts and device domains, with higher correlation scores and classification performance in rest-related setting.

TABLE VI
PERFORMANCE METRICS FOR TREMOR DETECTION ON THE VERILY STUDY WATCH SUBSET

Setting	Metric	No Tremor	Tremor	Weigted Avg
All Tasks	**Precision**	0.87	0.87	0.87
	Recall	0.98	0.47	0.87
	F1-score	0.92	0.61	0.87
Rest-Related Tasks	**Precision**	0.82	0.67	0.78
	Recall	0.97	0.23	0.80
	F1-score	0.89	0.45	0.79

TABLE VII
PERFORMANCE METRICS FOR TREMOR AMPLITUDE CLASSIFICATION ON THE VERILY STUDY WATCH SUBSET

Setting	Metric (Weighted)	Test
All Tasks	**Precision**	0.74
	Recall	0.80
	F1-score	0.77
Rest-Related Tasks	**Precision**	0.75
	Recall	0.80
	F1-score	0.77

IV. CONCLUSION

This work presents a multi-task deep learning framework for Parkinsonian tremor detection and amplitude classification from wrist-worn accelerometers. The proposed architecture combines the spatial feature learning of a pre-trained ResNet with the temporal modeling of a Transformer encoder to capture complex tremor patterns across tasks and time scales.

Our model performs well on both internal (MJFF-LRS) and external (Verily Study Watch) datasets, achieving strong classification metrics and satisfactory correlations with expert labels in both all-task and rest-related settings. These results highlight the method's potential for accurate real-world tremor tracking, including rest tremor, which is critical for early PD identification.

Despite these strengths, challenges remain, particularly class imbalance in severe tremor cases and variability across devices and subjects. Future work will focus on multiple-instance learning for weakly labeled data, pretraining on larger PD-specific datasets (e.g., the full Verily Study Watch cohort), and incorporating contextual cues (e.g., activity recognition, posture) and multi-sensor fusion to enhance robustness in real-world use.

ACKNOWLEDGMENT

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

Data used in the preparation of this article was obtained on 2024-07-08 from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-dataspecimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit www.ppmi-info.org. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research, and funding partners that can be found at https://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors. The PPMI investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit www.ppmi-info.org.

REFERENCES

[1] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[2] J. Pasquini et al., "Progression of tremor in early stages of Parkinson's disease: a clinical and neuroimaging study," *Brain*, vol. 141:3, pp. 811–821, 2018.

[3] A. H. Abusrair, W. Elsekaily, and S. Bohlega, "Tremor in Parkinson's disease: from pathophysiology to advanced therapies," *Tremor Other Hyperkinetic Movements*, vol. 12, p. 29, 2022.

[4] M.-H. Chang, T.-W. Chang, P.-H. Lai, and C.-G. Sy, "Resting tremor only: a variant of Parkinson's disease or of essential tremor," *Journal of the Neurological Sciences*, vol. 130:2, pp. 215–219, 1995.

[5] H. Zach, M. Dirkx, B. Bloem, and R. Helmich, "The clinical evaluation of Parkinson's tremor," *Journal of Parkinson's Disease*, vol. 5:3, pp. 471–474, 2015.

[6] M. Politis, K. W. Wu, S. Molloy, P. G. Bain, K. R. Chaudhuri, and P. Piccini, "Parkinson's disease symptoms: the patient's perspective," *Movement Disorders*, vol. 25:11, pp. 1646–1651, 2010.

[7] I. Ziogas, C. Lamprou, and L. J. Hadjileontiadis, "Bispectral analysis of Parkinsonian rest tremor: new characterization and classification insights pre-/post-DBS and medication treatment," *IEEE Access*, vol. 11, pp. 114459–114474, 2023.

[8] C. G. Goetz et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): process, format, and clinimetric testing plan," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23:15, pp. 2129–2170, 2008.

[9] F. Rodriguez et al., "Continuous and unconstrained tremor monitoring in Parkinson's disease using supervised machine learning and wearable sensors," *Parkinson's Disease*, vol. 2024, Art. no. 5787563, 2024.

[10] L. J. W. Evers, Y. P. Raykov, T. M. Heskes, J. H. Krijthe, B. R. Bloem, M.A. Little, "Passive monitoring of Parkinson tremor in daily life: a prototypical network approach," *Sensors*, vol. 25:2, p. 366, 2025.

[11] S. Papapetropoulos, "Patient diaries as a clinical endpoint in Parkinson's disease clinical trials," *CNS neuroscience & therapeutics*, vol. 18, no. 5, pp. 380–387, 2012.

[12] K. P. Bhatia, 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. 33:1, pp. 75–87, 2018.

[13] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D, Standaert , "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE transactions on information technology in biomedicine*, vol. 13:6, pp. 864–873, 2009.

[14] S. H. Ali, A. Akhtar, M. A. Khan, M. Bilal, "Detection of Parkinson's tremor in real time using accelerometers," in *IEEE 7th international conference on smart instrumentation, measurement and applications*, pp. 5-9, 2021.

[15] E. Rastegari, H. Ali, and V. Marmelat, "Detection of Parkinson's disease using wrist accelerometer data and passive monitoring," *Sensors*, vol. 22:23, p. 9122, 2022.

[16] H. Yuan, S. Chan, A. P. Creagh, C. Tong, A. Acquah, D. A. Clifton, A. Doherty, "Self-supervised learning for human activity recognition using 700,000 person-days of wearable data," *NPJ digital medicine*, vol. 7:1, p. 91, 2024.

[17] A. Doherty et al., "Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study," *PLoS One*, vol. 12:2, e0169649, 2017.

[18] A. Paszke et al., "Automatic differentiation in PyTorch," unpublished.

[19] J.-F. Daneault et al., "Accelerometer data collected with a minimum set of wearable sensors from subjects with Parkinson's disease," *Scientific Data*, vol. 8:1, p. 48, 2021.

[20] R. Atri et al., "Deep learning for daily monitoring of Parkinson's disease outside the clinic using wearable sensors," *Sensors*, vol. 22:18, p. 6831, 2022.

[21] 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 *IEEE 3rd International Conference on Computing and Machine Intelligence*, pp. 1–6, 2024.

[22] A. Gironell, B. Pascual-Sedano, I. Aracil, J. Marín-Lahoz, J. Pagonabarraga, J. Kulisevsky, "Tremor types in Parkinson disease: a descriptive study using a new classification," *Parkinson's Disease*, vol. 2018, Art. no. 4895243, 2018.

[23] A. S. Gala, et al., "The digital signature of emergent tremor in Parkinson's disease," *npj Parkinson's Disease*, vol. 10:1, p. 147, 2024.

[24] S. Dattola, A. Ielo, A. Quartarone, M. C. De Cola, "Integrating wearable sensor signal processing with unsupervised learning methods for tremor classification in Parkinson's disease," *Bioengineering*, vol. 12:1, p. 37, 2025.

[25] O. Bazgir, S. A. H. Habibi, A. Hassan, L. Palma, P. Pierleoni, S. Nafees, "A classification system for assessment and home monitoring of tremor in patients with Parkinson's disease," *Journal of Medical Signals & Sensors*, vol. 8:2, pp. 65–72, 2018.

[26] M. D. Hssayeni, J. Jimenez-Shahed ,M. A. Burac, B. Ghoraani , "Wearable sensors for estimation of parkinsonian tremor severity during free body movements," *Sensors*, vol. 19:19, p. 4215, 2019.

[27] X. H. Jin et al., "Incorporating prior-driven and data-driven feature extraction approaches for Parkinson's automatic tremor detection," in *Proceedings of the 2024 9th International Conference on Biomedical Signal and Image Processing.*, 2024.

[28] S. B. Dias, A. Grammatikopoulou, N. Grammalidis, J. A. Diniz, T. Savvidis, E. Konstantinidis, "Motion analysis on depth camera data to quantify Parkinson's disease patients' motor status within the framework of I-Prognosis personalized game suite," in *IEEE International Conference on Image Processing*, 2020.

[29] H. B. Kim, et al., "Wrist sensor-based tremor severity quantification in Parkinson's disease using convolutional neural network," *Computers in biology and medicine*, vol. 95, pp. 140–146, 2018.

[30] R. San-Segundo et al., "Parkinson's disease tremor detection in the wild using wearable accelerometers," *Sensors*, vol. 20:20, p. 5817, 2020.

[31] A. Stergioulas et al., "Assessing motor skills in Parkinson's disease using smartphone-based video analysis and machine learning," in *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*, 2024.

[32] I. Ioakeimidis et al., "Can bite detection algorithms substitute manual video annotations in elderly people with and without Parkinson's disease? An experimental study," in *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*, 2024.

[33] C. Donié, N. Das, S. Endo, S. Hirche, "Estimating motor symptom presence and severity in Parkinson's disease from wrist accelerometer time series using ROCKET and InceptionTime." *Scientific Reports*, vol. 15:1, pp. 1–15, 2025.