

**Wearable-Based Digital Biomarkers: An LSTM-Powered Progression  
Index for Parkinson's Disease Monitoring**



A thesis presented to the Department of Statistics and Data Science at Yale University in partial fulfillment of the requirements for the degree of Bachelor of Science

Felice Dong

Advised by Dr. Hemant Tagare

April 30, 2025

## Abstract

Parkinson’s disease (PD) is the most common age-related motor disorder, affecting more than 10 million people worldwide [1]. This paper proposes a novel progression index for PD monitoring using wearable sensor data and Long Short-Term Memory (LSTM) neural networks. Using ambulatory data obtained from the Parkinson’s Progressive Markers Initiative (PPMI), we developed a model that transforms daily physical activity patterns into a clinically meaningful disease progression metric. Our approach leverages LSTM networks trained on a custom loss function to capture temporal dependencies in time series data, enabling the detection of subtle changes in activity patterns that differentiate PD patients from healthy controls (HC). Two feature representations were evaluated: conventional weekday features, Monday through Sunday, and weekly activity levels sorted by intensity. Results demonstrate that the progression index consistently shows visual separation between PD and HC subjects across cross-validation splits, with sorted activity features providing more robust discrimination than features represented by conventional weekday labels. Weekly peak activity capability present a particularly salient indicator of disease status. While statistical significance was inconsistent across the 26-week observation period due to sample size limitations and disease heterogeneity, this approach represents a promising step toward objective, continuous, and remote PD monitoring that could enhance both clinical research and patient care.

# **Introduction**

Parkinson's disease (PD) is a progressive neurodegenerative disease characterized by declining dopamine production. Decreased dopamine levels manifest in movement-related symptoms, including bradykinesia, muscle tremors, rigidity, stiffness, and unstable gait, making it the most common age-related motor disorder.

Clinical assessments of the disease most commonly utilize the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) to track disease progression and evaluate the effectiveness of clinical trials [2]. It is a scale comprised of a series of questionnaires and a motor examination. Although widely employed, the UPDRS has limitations as it is subjective, can pose difficulties for patients without ready access to a doctor, and does not provide a comprehensive score of disease progression.

A previous study conducted by Verily, a digital watch company in partnership with the Parkinson's Progressive Markers Initiative (PPMI) found that digital biomarkers, especially ambulatory activity, detected treatment effects earlier and with smaller sample sizes than traditional clinical assessments in a clinical drug trial [3]. Building on these findings, we propose transforming ambulatory activity data into a unified and clinically meaningful disease progression index using Long Short-Term Memory (LSTM) neural networks, which are uniquely suited to model temporal patterns and capture long-term dependencies in time series data. This approach serves as a more accessible and objective remote method for continuous monitoring, with the ability to identify activity features most predictive of disease state.

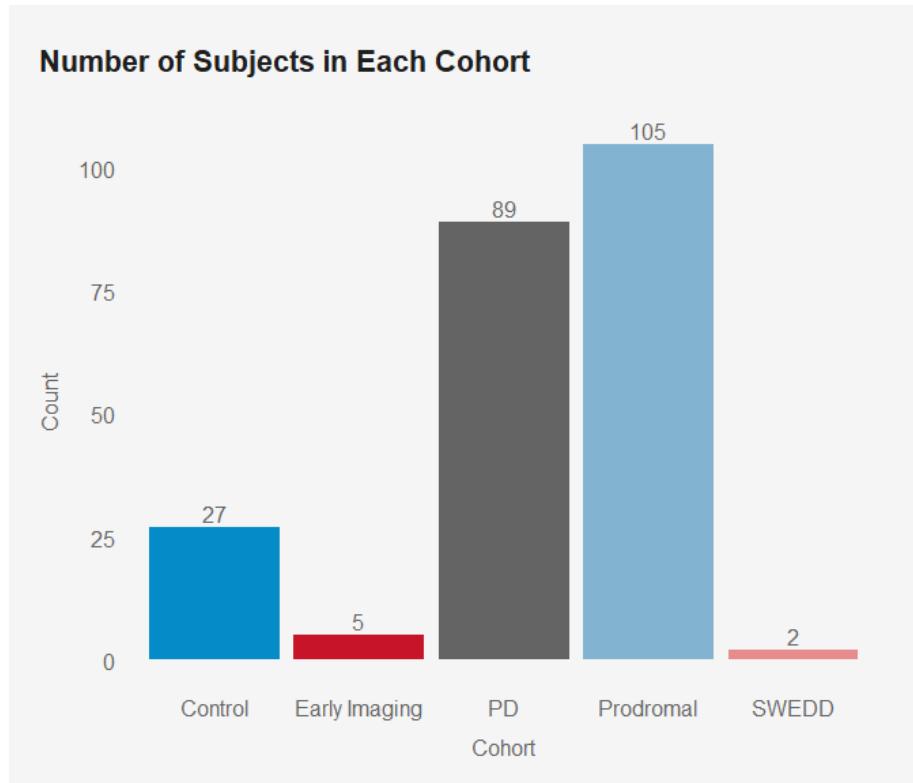
# **Methods**

## **Data and Subjects**

We used data collected from Parkinson's Progressive Markers Initiative (PPMI) participants who consented to wear the Verily Study Watch. The primary data feature used in our analysis was hourly ambulatory minutes, which was derived using a classifier algorithm developed by Verily that predicts ambulatory activity such as walking and running versus other activities based on accelerometer signals from the Study Watch.

The dataset included 229 subjects after initial quality filtering for those with excessive missing data, composed of individuals with Parkinson's disease (PD), healthy

controls (HC), and subjects with other conditions including SWEDD (Scans Without Evidence of Dopaminergic Deficit), prodromal symptoms, and early imaging findings. For our progression index modeling, we focused specifically on PD and HC subjects with at least 26 weeks (6 months) of continuous data, resulting in a final cohort of 73 PD subjects and 22 HC subjects.

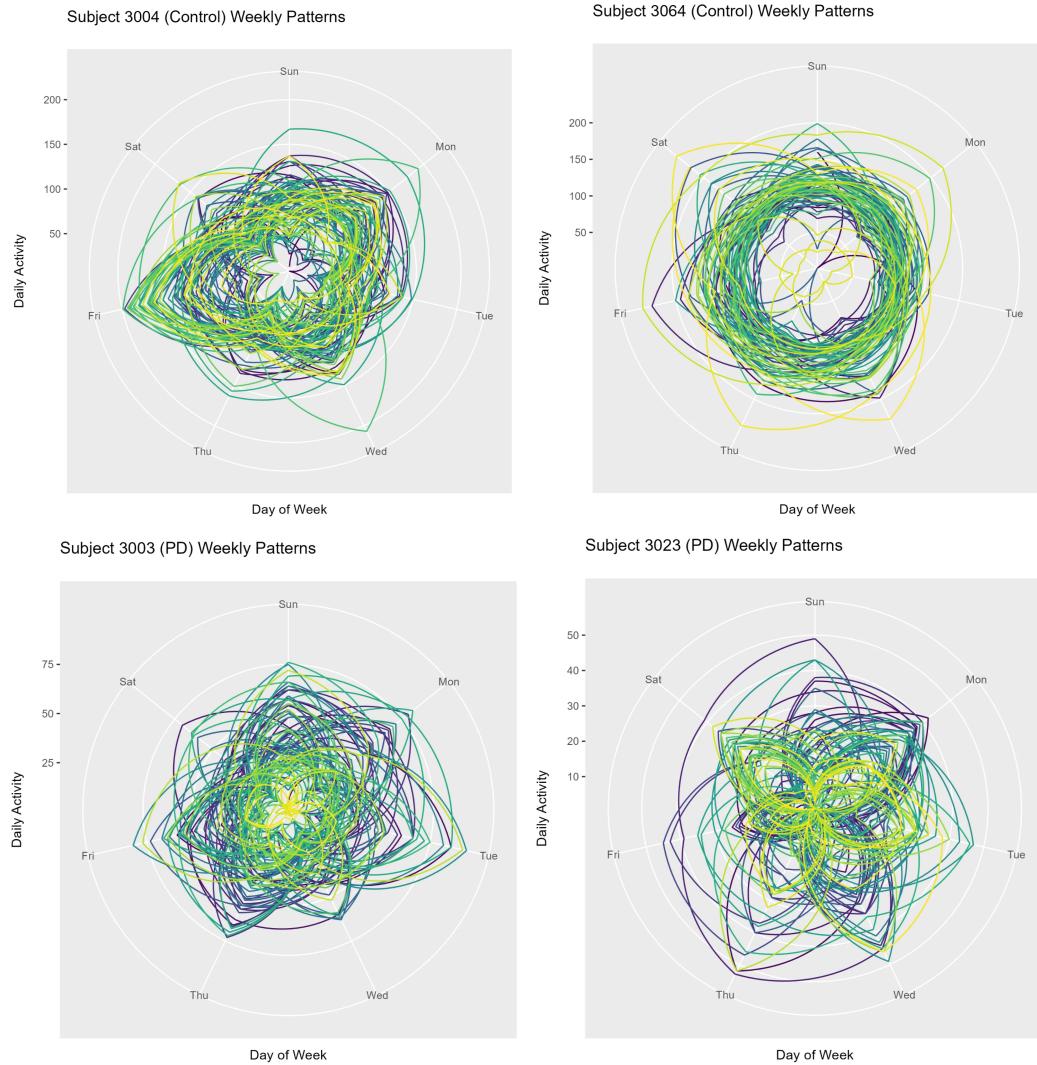


**Figure 1. Distribution of subjects in the PPMI dataset after the initial filtering criteria.** This includes 27 healthy controls, 5 early imaging findings, 89 with clinically diagnosed Parkinson's disease, 105 in the prodromal phase, and 2 in the SWEDD condition.

## Data Preprocessing

The dataset was transformed from hourly ambulatory minutes into daily activity totals, then organized into weekly sequences to capture cyclic patterns observed in polar plots of individual activity levels over time. The resulting dataset structure represented each subject's activity as a time series of weekly activity patterns, with

seven features corresponding to daily activity minutes for each day of the week, Monday through Sunday.



**Figure 2. Polar plots of select subjects in the dataset.** Weekly ambulatory data for each of the subjects were plotted as polar plots with each cycle representing one week, the amplitude representing the number of ambulatory minutes for the corresponding weekday, and color representing time. The passage of time is represented as a gradient from dark purple to light yellow, with dark purple representing earlier weeks. In HC subjects (top), there are a lot of overlaps between weekly activity patterns, as represented by superimposing lines. In PD subjects (bottom), ambulatory activity appears to decline with time yet still maintain weekly patterns—again characterized by superimposed lines.

To ensure data quality and reliability for model training, we implemented the following filtering criteria:

1. *Compliance threshold*: Subjects with less than 50% data compliance were excluded, calculated as the ratio of days with recorded data to the total study duration for that subject.
2. *Minimum duration*: Subjects with fewer than 30 days of data were excluded.
3. *Continuity requirement*: Since the aim is an assessment of progression over time, final analysis focused on subjects with at least 26 weeks (6 months) of continuous data.

For weeks with 1 or 2 missing days of data, linear interpolation was applied to estimate missing values. Weeks with more than 2 missing days were thus excluded from analysis.

## Feature Transformation

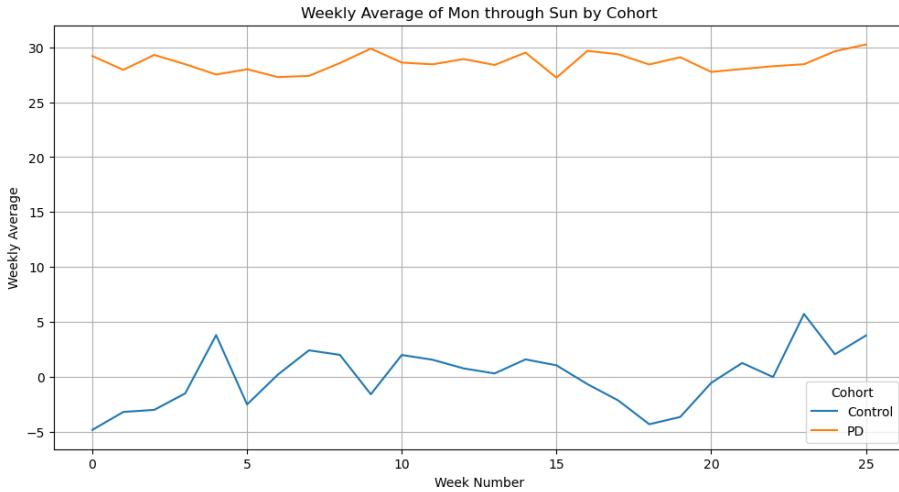
To design a progression index where higher values would appropriately correspond to greater disease progression, we performed a series of transformations to the raw data. Since PD patients typically exhibit reduced activity compared to HC subjects due to the nature of the disease, we inverted this relationship to create an intuitive index that assigned higher scores to indicate greater disease severity. The transformation can be expressed as

$$\text{transformed activity} = -\text{original activity} + \text{mean(HC)} \quad (1)$$

where  $\text{mean(HC)}$  represents the global mean of daily activity minutes across all HC subjects.

This transformation effectively centered HC values around zero while positioning PD ambulatory values higher. In other words, the transformations preserve the relative differences in activity patterns while reorienting the scale such that it aligns with clinical interpretations of disease progression. Figure 3 illustrates the effect of this transformation on the weekly average activity levels for both cohorts.

To enable meaningful comparison across subjects who entered the study at different disease stages, all 95 sequences were aligned such that the 26 weeks of continuous data had consistent weekly markers, 1 through 26, regardless of what weeks the continuous data actually reflected in the study for each subject. For instance, the 26



**Figure 3. Shifted weekly cohort averages.** Prior to transformations, PD subjects were less active on average. Thus, after transformations, the PD subjects (orange) appear higher than HC subjects (blue).

weeks of continuous data could have started at the 10th week of data collection for one subject, and observed for the following 25 weeks. However, after alignment, the 10th week is now represented as week 1—and this is done for every single subject. This approach focuses on capturing relative changes in activity patterns over time rather than absolute disease duration, allowing the model to learn progression patterns despite initial disease state.

## LSTM Model

To develop our progression index, we designed a Long Short-Term Memory (LSTM) neural network architecture to capture temporal trends in the ambulatory activity data. LSTM models are particularly suited for this considering their ability to learn long-term dependencies in sequential data, which enables the effective modeling of disease progression over extended periods.

Our LSTM model consisted of an input layer with 7 features (daily activity minutes), 2 layers with 20 hidden units each, and a linear layer that applied a linear transformation to each hidden state to produce weights. Then weights were normalized and the inner product of the weights and input were taken to produce outputs for progression indices. Normalization, in this case, prevented any single time point from dominating. This normalized weight can be expressed as:

$$\hat{w}_i(t) = \frac{w_i(t)}{\sum_{j=1}^7 |w_{i,j}(t)| + \epsilon} \quad (2)$$

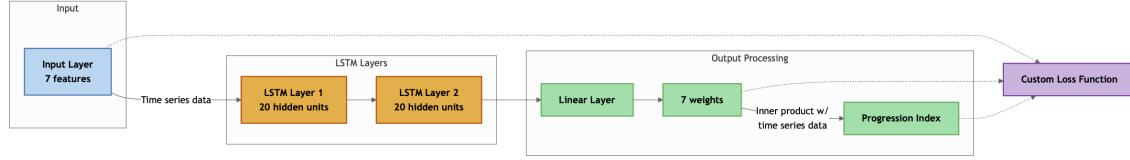
where  $\hat{w}_i(t)$  represents normalized weights, and  $w_{i,j}(t)$  represents the weight for feature  $j$  of subject  $i$  at week  $t$ .

The final progression index  $u_i(t)$  for subject  $i$  at week  $t$  can be expressed as:

$$u_i(t) = \sum_{j=1}^7 \hat{w}_{i,j}(t) \cdot x_{i,j}(t) \quad (3)$$

where  $x_{i,j}(t)$  represents ambulatory data for subject  $i$  for feature  $j$  at time  $t$ .

The progression index thus remained a direct function of activity data, weighed by learned model weights to highlight the most discriminative patterns.



**Figure 4. LSTM architecture**

Furthermore, we investigated two different feature representations. First, we utilized conventional weekday features. Using the seven days of the week—Monday through Sunday—as separate input features, we preserved the natural temporal ordering of the week. Then, we used sorted activity features. Ordering each week’s daily activity levels in descending order from most active to least active day, we aimed to capture varying patterns of activity variation rather than specific weekday schedules. In other words, this approach allowed for the model to potentially identify more generalizable ambulatory patterns independent of personal weekly scheduling preferences.

## Custom Loss Function

To train the LSTM model, we implemented a custom loss function designed to optimize separation between PD and HC cohorts while accounting for disease progression over time.

The loss function can be expressed as

$$\ell = -\text{time weighted PD output} + (\text{HC output})^2 + \text{regularization term} \quad (4)$$

$$\ell = - \sum_{i=1}^T \left( \frac{1+t_i}{N_{PD}+\varepsilon} \sum_{j \in PD} \hat{y}_{ij} + \frac{1}{N_{HC}+\varepsilon} \sum_{j \in HC} \hat{y}_{ij}^2 \right) + 20 \cdot (w_{1\_std} + w_{0\_std}) \quad (5)$$

where

- $T$  is the number of weeks, which in this case is 26
- $t_i$  is week  $i$ , which increasingly emphasizes later weeks for PD subjects
- $N_{PD}$  and  $N_{HC}$  are the number of subjects in the PD and HC cohorts, respectively
- $\varepsilon$  is a small constant added for numerical stability when dividing
- $\hat{y}_{ij}$  is the inner product between the normalized model weights and the ambulatory data for subject  $j$  at week  $i$
- $w_{1\_std}$  and  $w_{0\_std}$  are the standard deviations of the weights for the PD and HC cohorts

The first negative term, representing the PD cohort, encourages loss minimization and drives the model to maximize output values for PD subjects. The time weighting factor places progressively more emphasis on later weeks, aligning with disease expectations that motor functions decline over time, reflecting worsening symptoms.

The second positive term, representing the HC cohort, penalizes the model for producing high values for HC subjects. The squared term imposes a stronger penalty for larger outputs, effectively pushing HC outputs toward zero so that the model learns to assign lower values to these subjects to create separation between cohorts. This does not include a time weighting like the first term because HC subjects are not expected to exhibit significant changes in ambulatory activity over time as their levels should reflect relatively stable conditions.

The third term serves to regularize the learned model weights by penalizing high standard deviations. In effect, this regularization term prevents the model from heavily emphasizing specific time points or features, encouraging more generalized learning to reduce overfitting.

## Model Training and Evaluation

The model was trained using a stochastic optimization approach with data augmentation and early stopping. Then, to evaluate, we implemented a cross-validation framework.

Specifically, for each train-test iteration, random batches of 15 subjects were selected from each cohort. This balanced sampling ensured the model did not bias toward the majority cohort and learn equally from both the PD and HC subjects despite the actual imbalance present in the dataset. Gaussian noise ( $\sigma = 0.02$ ) was added to input features to enhance model robustness against minor differences in activity levels that may be caused by measurement inaccuracies. In other words, the noise helps the model learn actually meaningful signals and changes in activity. Furthermore, 20% of adjacent time points were randomly swapped. Similar to adding noise, time swapping helps the model learn patterns that are robust to slight variations in when certain activities occur. This was crucial in our development for a progression index because it encourages the model to avoid learning weekly schedule quirks or idiosyncrasies of individual subjects and effectively allows for a focus on overall activity patterns.

During each training loop, input data was forward passed through the LSTM to obtain weights and output indices. Loss was then calculated using the custom function, and gradients backpropagated through the network to allow for updates to model parameters. The training process continued until loss convergence was reached. Loss convergence was evaluated every 10 epochs using a 5-epoch sliding window, and reached when the percent change in average loss between the latest and penultimate window fell below 0.01%. If convergence was not reached, training commenced until a maximum of 20,000 epochs. This prevented the model from overfitting to training samples to increase generalizability. After training, the resulting model weights were applied to 4 test subjects, 2 from each cohort, to generate progression indices for each week of the 26-week period.

To assess the model's ability to distinguish between cohorts, we performed independent t-tests to test for significant differences in group means between the test subjects' progression indices at each time point to determine meaningful separation. The earliest week with consistently low p-values ( $p < 0.05$ ) represents the window of time at which the model can first reliably detect differences between cohorts.

This training and testing process was repeated across 20 splits, conducted once for weekday features Monday through Sunday, and once for features sorted by decreasing activity.

Algorithm 1 summarizes the training and evaluation procedure.

---

**Algorithm 1** LSTM-Based Progression Index Training and Evaluation

---

**Input:** Preprocessed ambulatory activity data

**Output:** Validated progression index model and significance windows

**Start:**

**1: Cross-Validation Setup**

    ← Perform 20 independent train-test splits

    ← For each split, randomly select 2 PD and 2 HC subjects for testing

    ← Use remaining subjects for training

**2: Model Initialization**

    ← Input size: 7 (daily activity features)

    ← LSTM layers: 2 with 20 hidden units each

    ← Output size: 7 (matching input dimension)

**3: Training Process**

    ← Initialize Adam optimizer with learning rate 0.0003

    ← Select balanced batches of 15 subjects (from both cohorts)

    ← Apply data augmentation:

        ○ Add Gaussian noise ( $\sigma = 0.02$ )

        ○ Swap 20% of time points with adjacent points

    ← Compute forward pass

    ← Calculate loss using custom function

    ← Backward pass and update parameters

    ← Check convergence every 10 epochs:

        ○ If percent change in loss  $\geq 0.01\%$  over 5 epochs, stop training

        ○ Else continue until max 20,000 epochs

**4: Progression Index Computation**

    ← Apply model to test subjects to generate outputs

    ← Calculate progression index as weighted sum of activity features

**5: Statistical Evaluation**

    ← Perform t-tests between PD and HC indices at each time point

    ← Identify significance windows where p-values  $\leq 0.05$

    ← Record earliest week with consistent significant separation

**6: Aggregate Analysis Across Splits**

    ← Calculate mean progression indices for each cohort

    ← Compute average p-values at each time point

    ← Analyze learned weights to identify most informative features

    ← Compare results between sorted and unsorted feature approaches

**End**

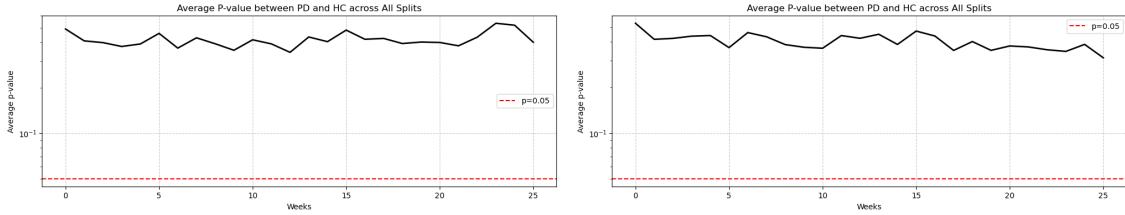
---

## Results and Discussion

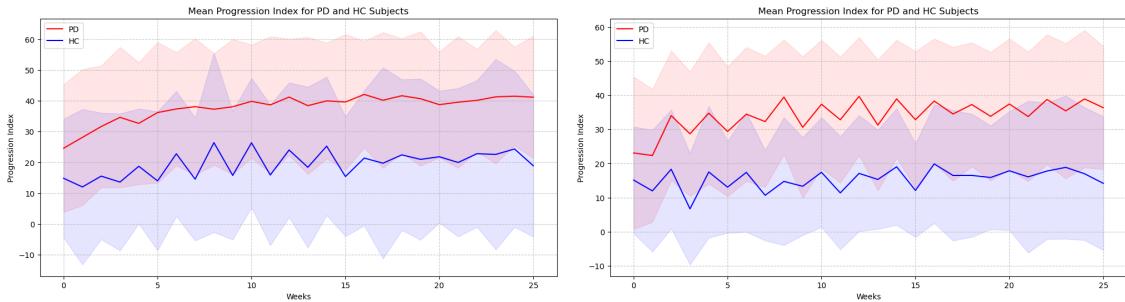
The LSTM-based approach successfully generated a progression index visually capable of differentiating between PD and HC subjects based on ambulatory activity patterns from wearable sensor data. We evaluated both sorted activity features and conventional weekday features across 20 independent train-test splits.

### Statistical Significance of Cohort Separation

The analysis of p-values across the 26-week period showed varying levels of statistical significance between the two cohorts. Although p-values occasionally dropped below 0.05 for individual train-test splits, this was not consistently observed across splits for either the sorted or unsorted feature condition.



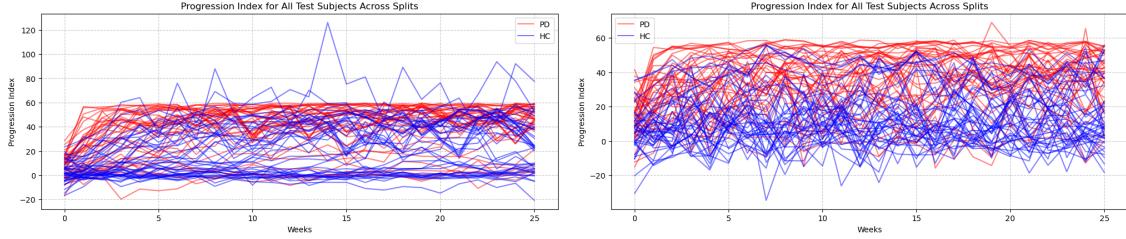
**Figure 5. Average p-values from t-tests assessing the model's ability to distinguish between cohorts.** The left shows the representation sorted by intensity and the right shows the representation using conventional weekday labels.



**Figure 6. Mean progression indices for each cohort plotted with standard deviations.** The left reflects the representation sorted by intensity and the right reflects the representation using conventional weekday labels. There is substantial overlap in the standard deviation bands (shaded areas), which indicates high variability within each cohort.

Despite the fluctuating statistical significance, visual examination of individual progression indices revealed distinct separation between cohorts. Across splits, PD sub-

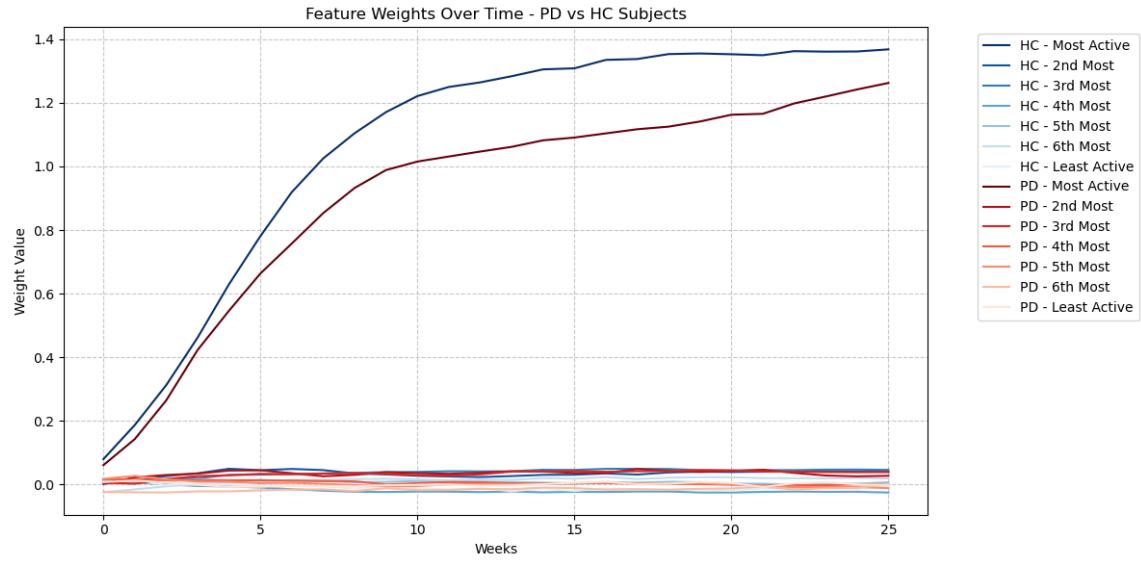
jects consistently exhibited higher progression index values compared to HC subjects, aligning with the goals of the index design. With the exception of a few test subjects, the sorted feature condition generally yielded better separation between cohorts, suggesting the approach potentially provides more robust statistical separation between disease states.



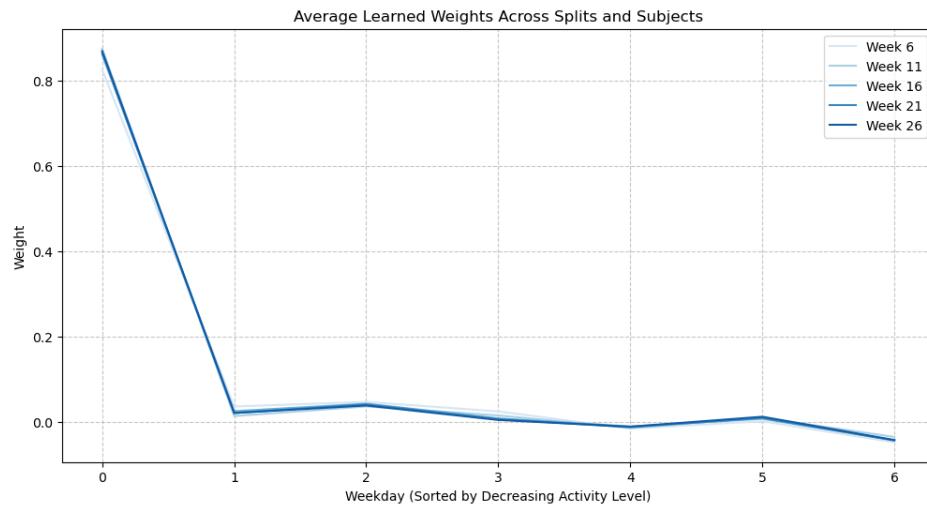
**Figure 7. Plots of individual progression indices for test subjects.** The left reflects the representation sorted by intensity and the right reflects the representation using conventional weekday labels.

## Feature Importance

Nonetheless, analysis of the learned feature weights reveal meaningful patterns across the two cohorts. For sorted features, the model consistently assigned the highest weights to the most active days, suggesting that peak activity capability within each week is a particularly salient discriminator for disease status, consistent with the known impact of PD on mobility.

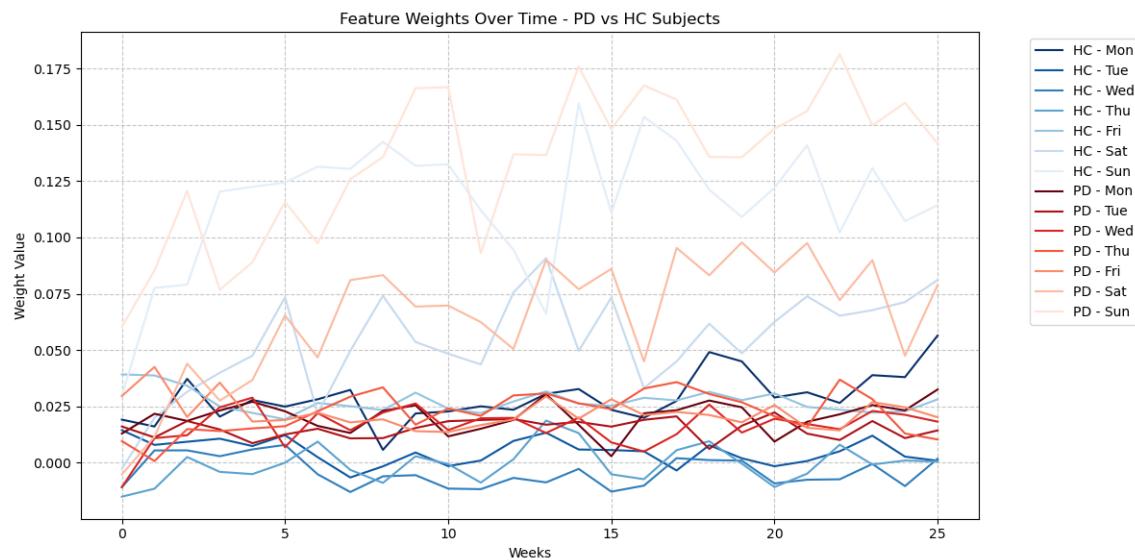


**Figure 8. Feature weights over time for the sorted representation.** Both the PD and HC cohorts assign heavier weights to the most active day of the week, suggesting that peak activity capability may be an informative discriminator for disease status and progression.

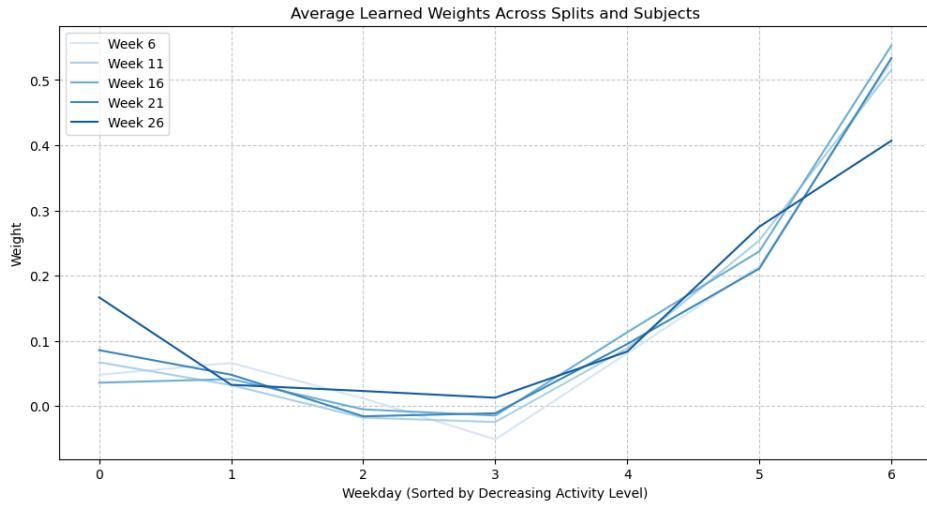


**Figure 9. The average normalized weights assigned to each input feature over time.** Weekdays were sorted by decreasing activity level with 0 representing the most active day and 6 the least active day. Across all time points over the 26-week observation period, the distribution of normalized weights remained consistent with the most weight assigned to the most active day.

For conventional weekday features, weekend days received higher weights compared to weekdays across all time points examined in both cohorts. This pattern likely reflects that weekend days capture more discretionary physical activities, which may be more sensitive to disease-related limitations than the routine, habitual activities that dominate weekdays regardless of disease status. In other words, the higher weights assigned to weekend activity suggest that how individuals spend weekends may be more revealing of disease progression than their performance during typically structured weekday routines, assuming an adherence to a conventional weekly routine. However, the difference in weights assigned to each feature input is not as pronounced as the previous sorted case, indicating that conventional weekday labels offer a less discriminatory and interpretable understanding of activity patterns—plausible given the heterogeneity in lifestyle between subjects.



**Figure 10. Feature weights over time for the unsorted representation.** Both the PD and HC cohorts assigned heavier weights to weekends, reflecting the discretionary nature of those days, which more tellingly distinguishes the cohorts from each other.



**Figure 11. The average normalized weights assigned to each input feature over time.** Weekday labels were not perturbed, with 0 representing Monday, 1 Tuesday, and so on. Although the distribution of normalized weights change slightly over the 26-week observation period, the highest weights are nonetheless still consistently assigned to 5 (Saturday) and 6 (Sunday).

Both the sorted and unsorted features displayed consistent feature importance patterns across different weeks in the observation period, suggesting robust and stable learning by the model, indicating that these activity patterns represent fundamental differences between disease states rather than transient fluctuations.

## Conclusion

This paper presents a novel LSTM-based progression index for Parkinson’s disease monitoring using wearable sensor data. Our approach leverages the temporal learning capabilities of LSTM models to capture subtle changes in activity patterns that offer the possibility of a remote, accessible indicator of disease progression. By transforming ambulatory activity into a clinically meaningful and comprehensive index, we demonstrate the potential of digital biomarkers to distinguish between PD and HC subjects, helping subjects better anticipate their quality of life and aid in evaluating clinical trial effectiveness.

The progression index showed consistent visual separation between PD and HC subjects, particularly when using sorted activity features rather than convention weekday labels, Monday through Sunday. This suggests that the pattern of activity

intensity across the week, specifically the most active day, carries more discriminative information about disease status than calendar-based measurements.

However, our approach faces several limitations that future research should address. While visual separation between cohorts was evident, statistical significance was inconsistent across the 26-week period, in part due to the relatively small sample size, the heterogeneity inherent in activities between subjects, and heterogeneity in PD progression itself. Our current approach uses continuous data over a six-month period, with interpolated short gaps, limiting its applicability to patients with substantial missing data.

## Future Directions

Several promising avenues for future research emerge from this work.

First, a refined loss function could better account for the variable progression rates among PD patients while maintaining separation from HC subjects. Incorporating clinical assessments like UPDRS scores or medication information into the loss function could help anchor the progression index to have it achieve significant separation. The latter, especially, integrated as a weighting factor in the loss function could help account for the confounding effects of treatment on activity patterns, providing a more accurate representation of underlying disease status.

Second, to address the variability in the outputted progression indices, the ambulatory data could be smoothed prior to input into the model and the index calculation. The progression indices are functions of the learned weights from the LSTM and the ambulatory time series data, and since weights stabilize over time (as shown by Fig. 9 and Fig. 11), the noise from the indices therefore result from the noise in the ambulatory data. Smoothing it while keeping the characteristics of underlying walking patterns would allow for more interpretable and significant indices.

Third, implementing masking for handling missing data would increase the model's robustness to real world adherence patterns and enable analysis over longer time periods than the current six-month window, potentially capturing slower progression trends not visible during our current observation period.

Fourth, adopting a validation-based approach to determine optimal training duration would improve model performance and generalizability by more robustly preventing both underfitting and overfitting to the training data than the current convergence-based approach.

Fifth, our current approach focuses on weekly scales as a result of patterns observed in exploratory polar plots. However, future research should consider different time scales to capture different aspects or potentially more informative features of the disease.

In conclusion, our LSTM-powered progression index demonstrates the promise of wearable-based digital biomarkers for objective, continuous PD monitoring. While refinement is needed, this approach represents a meaningful step toward more sensitive, accessible disease monitoring tools that could transform both clinical research and patient care in Parkinson's disease.

# References

- [1] Parkinson's Foundation. (n.d.). *Statistics*. Retrieved April 30, 2025, from <https://www.parkinson.org/understanding-parkinsons/statistics>
- [2] Parkinson's Foundation. (n.d.). *Stages of Parkinson's*. Retrieved April 30, 2025, from <https://www.parkinson.org/understanding-parkinsons/what-is-parkinsons/stages>
- [3] Chen, C., Kowahl, N. R., Rainaldi, E., Burq, M., Munsie, L. M., Battioui, C., Wang, J., Biglan, K., Marks, W. J., Jr., & Kapur, R. (2023). Wrist-worn sensor-based measurements for drug effect detection with small samples in people with Lewy Body Dementia. *Parkinsonism & Related Disorders*, 109, 105355. <https://doi.org/10.1016/j.parkreldis.2023.105355>