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

Felice Dong, Advised by Dr. Hemant Tagare
Department of Statistics and Data Science, Yale University

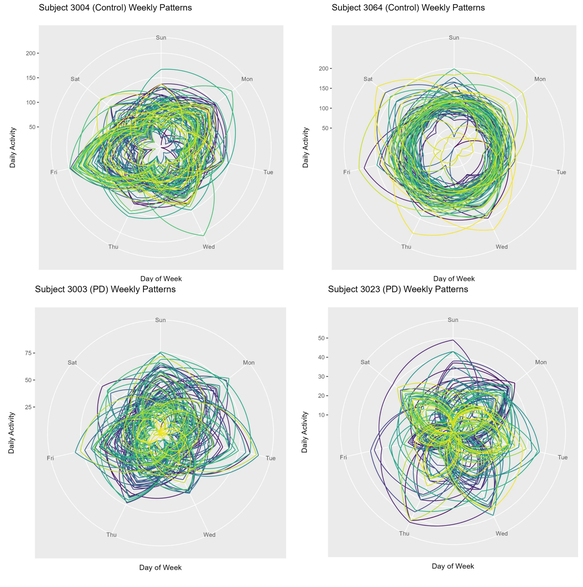
## ABSTRACT


- Parkinson's Disease (PD) is the most common age-related motor disorder, affecting more than 10 million people worldwide.
- We aim to develop a novel progression index for PD monitoring using wearable sensor data and Long Short-Term Memory (LSTM) neural networks.
- We leverage LSTM networks trained on a custom loss function to enable the detection of subtle changes in activity patterns that differentiate PD patients from healthy controls.
- Two feature representations were evaluated: (1) conventional weekday features, Monday through Sunday, and (2) weekly activity levels sorted by intensity.
- Results demonstrate that the progression index consistently shows visual separation between PD and healthy control subjects, 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.


## INTRODUCTION

- Clinical assessments most widely utilize the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS).
- UPDRS has limitations: it is subjective, can pose difficulties for patients without ready access to a doctor, and does not provide a comprehensive score of disease progression.
- Previous research by Verily found that digital biomarkers, especially ambulatory activity, detected clinical treatment effects earlier and with smaller sample sizes than traditional clinical assessments.
- A disease progression index using LSTMs and wearable data provides a more accessible and objective remote method for continuous monitoring, with the ability to identify activity features most predictive of disease state.

## DATA




## METHODS

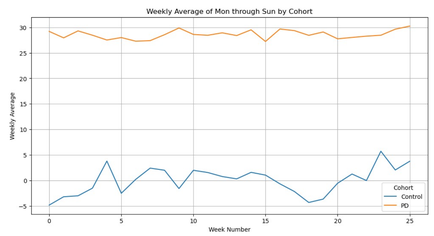



**Figure 1: Feature Transformations**
Since PD patients typically exhibit reduced activity compared to healthy controls due to the disease, we inverted this relationship to create an intuitive index that assigned higher scores to indicate greater disease severity. The transformation can be expressed as:

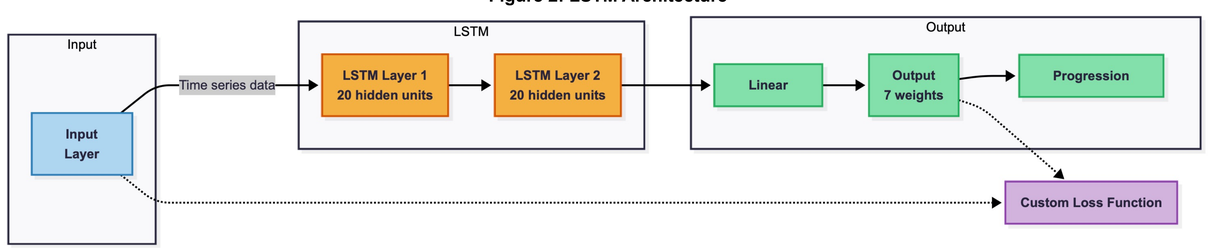
$$transformed\ activity = -original\ activity + mean(HC)$$

**Figure 2:** To train the LSTM model, we implemented a custom loss function designed to optimize separation between PD and healthy control cohorts while accounting for disease progression over time:

$$L = -\sum_{i=1}^{T}\left(\frac{1+t_i}{N_{PD}+\varepsilon}\sum_{j\in PD}\hat{y}_{ij} + \frac{1}{N_{HC}+\varepsilon}\sum_{j\in HC}\hat{y}_{ij}^2\right) + 20w_{std}$$

Progression indices were calculated as the weighted sum of activity features divided by the total weight.

**Figure 2: LSTM Architecture**




## RESULTS

**Figure 3: Progression Indices of PD and HC Test Subjects Across 26 Weeks**

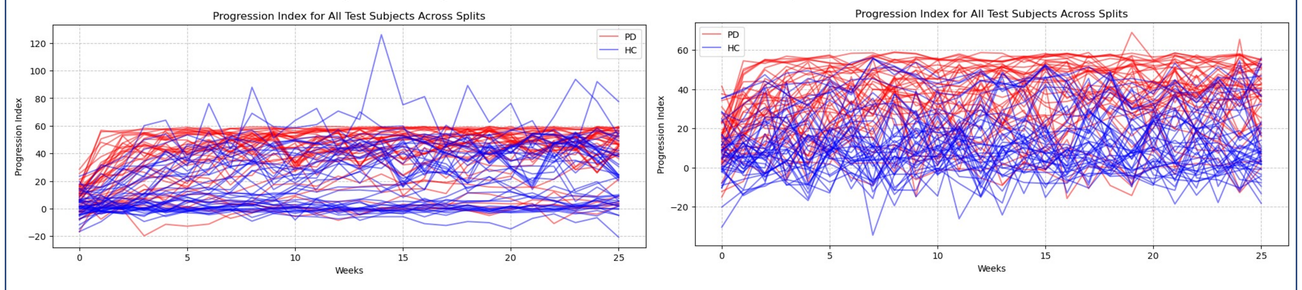



**Figure 3:** Progression indices generated by our LSTM model across 20 independent train-test splits. Red lines represent individual PD subjects, while blue lines represent healthy controls. Despite some variability between splits, PD subjects consistently exhibited higher progression index values compared to HC subjects, demonstrating the model's ability to visually differentiate between cohorts based solely on wearable sensor data. The sorted activity feature representation (left) generally produced clearer separation than conventional weekday features (right).

**Figure 4: Learned Feature Weights Over Time for Different Activity Representations**

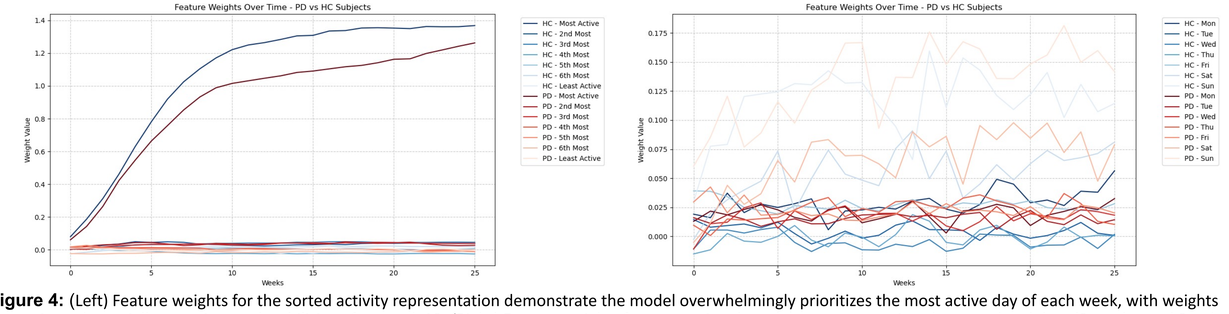
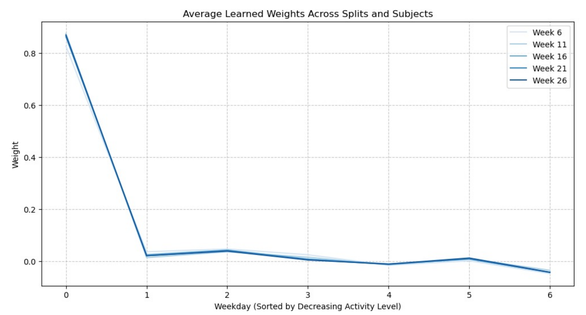



**Figure 4:** (Left) Feature weights for the sorted activity representation demonstrate the model overwhelmingly prioritizes the most active day of each week, with weights increasing substantially over time and stabilizing after week 10. (Right) Feature weights for conventional weekday representation show weekend days (Saturday and Sunday) consistently received higher weights in both PD and HC cohorts, indicating discretionary weekend activities provide stronger signals for distinguishing disease status than more structured weekday routines. This approach distributes importance more evenly across different days of the week, with a preference for weekend days.

## CONCLUSIONS




- The progression index successfully demonstrated consistent visual separation between PD and healthy control subjects across 20 independent train-test splits.
- Using activity features sorted by intensity provided better cohort separation than conventional weekday features, suggesting activity patterns are more informative.
- Learned feature weights revealed the model identified peak activity capability as a key discriminator for disease status.
- For conventional weekday features, weekend days received higher weights than weekdays, indicating discretionary activities on weekends may better reveal disease-related limitations than routine weekday activities.
- While visual separation was achieved, statistical significance was inconsistent across the 26-week period, likely due to our small sample size (73 PD, 22 HC) and heterogeneity in both activities and disease progression.
- Our approach demonstrated the potential of transforming ambulatory activity data from wearables into a meaningful progression index for continuous, remote PD monitoring.

## FUTURE DIRECTIONS

- A refined loss function could that incorporates clinical assessments like UPDRS scores or medication information into the loss function could help anchor the progression index to have it achieve significant separation.
- Implementing masking for handling missing data would increase the model's robustness to real word adherence patterns and enable analysis over longer time, potentially capturing slower progression trends not visible during our current observation period.
- 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.

## ACKNOWLEDGEMENTS

I would like to thank Dr. Hemant Tagare for his continual guidance in advising this project and Yale's S&DS department for their academic support. This work is dedicated to the memory of my aunt, whose courage while living with Parkinson's disease inspired this research. Finally, I thank my family for making my academic pursuit possible.