Recurrent Neural Network Model of Gait Can Predict Parkinson's Disease

Rishav Selvam

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by motor and non-motor symptoms, with subtle gait alterations often appearing in early stages. Early detection is critical for timely intervention, yet current clinical methods are limited. Wearable sensors offer the ability to capture high-resolution gait data, providing an opportunity to detect PD before pronounced clinical symptoms emerge. In this study, we analyzed gait data from 166 participants, including 93 PD and 73 controls, collected using eight force sensors per foot at 100 Hz during a 120-second walking task. Force measurements were normalized using Min–Max scaling to reduce variability due to body weight and individual differences. Walking trials were segmented into 200-sample sliding windows, and a recurrent neural network (RNN) was trained to classify participants based on temporal patterns in gait dynamics. The RNN accurately (96.4% accuracy) distinguished PD patients from healthy controls, capturing subtle differences in gait dynamics that are not apparent to clinicians. Overall, these findings demonstrate the potential of combining wearable sensors with machine learning for non-invasive, objective detection of Parkinson's disease. Future work should validate the model on independent cohorts and explore its applicability in real-world settings to support earlier interventions and improve patient outcomes.

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects movement but also produces a wide range of non motor symptoms. It is the second most common neurodegenerative disease after Alzheimer's disease. Each year, approximately 90,000 new cases are diagnosed in the United States, and more than 1 million people are currently living with PD in the country [1]. The average age of patients at diagnosis is 60 years old. The risk of developing PD increases with age, and men are about 1.5 to 2 times more likely than women to be affected [2]. Ethnic disparities have also been observed, with a higher prevalence in White populations compared to other populations [3].

The prognosis of PD is highly variable but generally progressive. Motor symptoms include tremors, bradykinesia(slowness of movements and speed), rigidity(when muscles become stiff), and postural instability(the inability to maintain balance and upright posture). Non motor symptoms, such as cognitive decline, mood disorders, sleep disturbances, and autonomic dysfunction(where the autonomic nervous system fails to regulate autonomic body functions) also contribute significantly to the struggles of the disease. Motor fluctuations can become an issue 5-10 years after diagnosis, and postural instability typically occurs over around 10 years after diagnosis, although the exact timeline varies from person to person [1].

PD is caused by the loss of dopamine-producing neurons in the substantia nigra, a part of the midbrain that helps control movement. As these neurons die, dopamine levels drop, which disrupts normal signaling in the brain and leads to symptoms such as tremors, rigidity, and bradykinesia [4, 5]. Although medications such as L-DOPA (a medication that the brain converts into dopamine) and device based treatments like deep brain stimulation can reduce symptoms and improve quality of life, there is currently no cure.

Early treatment with L-DOPA is linked to better outcomes, which highlights the importance of earlier diagnosis [6]. Given this context, early detection of PD has become a major research priority. Researchers have explored several approaches, including biomarkers, MRI imaging, and genetic screening, but none have yet become reliable tools for routine clinical use [7, 8]. Gait analysis, however, is already a standard method for tracking PD progression, and subtle gait changes may appear years before formal diagnosis [9]. With wearable sensors, it is now possible to capture high resolution gait data, offering a promising tool for detecting PD at earlier stages.

Machine learning offers new opportunities to analyze high-frequency gait data. Recurrent neural networks (RNNs) are particularly suited for this task because they can capture temporal dependencies in sequential data. In this paper, we present an RNN capable of accurately and efficiently identifying PD diagnosis from gait sensor data collected during standard medical gait tests.

Methods

Gait Sensor Dataset

We analyzed gait data from a publicly available database [10]. The dataset included 166 subjects: 93 with Parkinson's disease and 73 healthy controls. Each participant performed a 120-second walking task while wearing eight force sensors per foot (16 total sensors). Data was sampled at every 100th of a second, yielding 12,000 time points per sensor per subject. In addition to the vertical ground reaction force signals, the dataset also includes demographic information(age, gender, height, weight, and walking speed), clinical assessments(Unified Parkinson's Disease Rating Scale (UPDRS) scores, Hoehn & Yahr (H&Y) staging, and other disease severity measures), and other data that can be calculated by the given data(stride time, swing time, stance time, etc.).

Data Preprocessing

To account for force magnitude differences between participants caused by factors unrelated to PD, such as body weight, force data from each sensor were normalized using Min-Max scaling, rescaling values between 0 (min) and 100 (max) for each sensor on each foot per participant. This approach enables analysis of gait dynamics while minimizing the influence of individual differences in overall force magnitude.

Recurrent Neural Network

An RNN was implemented using PyTorch (v2.8.0) to classify participants as having Parkinson's disease or not based on their gait sensor data. Normalized force data from all 16 sensors, along with total force per foot over the 120-second walking task (sampled at 100 Hz), were used as model inputs. The dataset was randomly split into 80% training and 20% testing subsets. The model was trained for 10 epochs using *nn.CrossEntropyLoss()* to compute classification error and *torch.optim.Adam()* to update model weights. During training, gradients were cleared with *optimizer.zero_grad()*, calculated with *loss.backward()*, and applied with *optimizer.step()*.

Predictions were extracted using *torch.max()*, and the model was evaluated with *model.train()* and *model.eval()* modes.

Statistics

Independent t-tests were performed to compare baseline gait metrics between PD and control groups. P-values < 0.05 were considered statistically significant. Boxplots were generated to display group means and standard deviations. Python libraries used included NumPy, Pandas, Matplotlib, and PyTorch.

Results

Data Exploration

We analyzed gait data from 166 subjects: 93 with Parkinson's disease (PD) and 73 healthy controls. Summary statistics and visualizations were generated to compare gait metrics across groups and disease stages. Gait velocity was lower in PD patients compared to healthy controls (PD: 0.97 m/s vs Healthy: 1.21 m/s; $p = 0.03$) (**Figure 1a**). Within PD patients, Hoehn & Yahr (H&Y) Stage 3 was associated with lower velocity compared to Stage 2 (Stage 2: 1.02 m/s vs Stage 3: 0.81 m/s; $p < 0.05$) (**Figure 1b**). Stage 2.5 showed no significant difference in velocity compared to the other stages ($p > 0.05$) (**Figure 1b**). These results indicate that gait velocity decreases with disease severity, supporting the use of gait metrics for early PD detection.

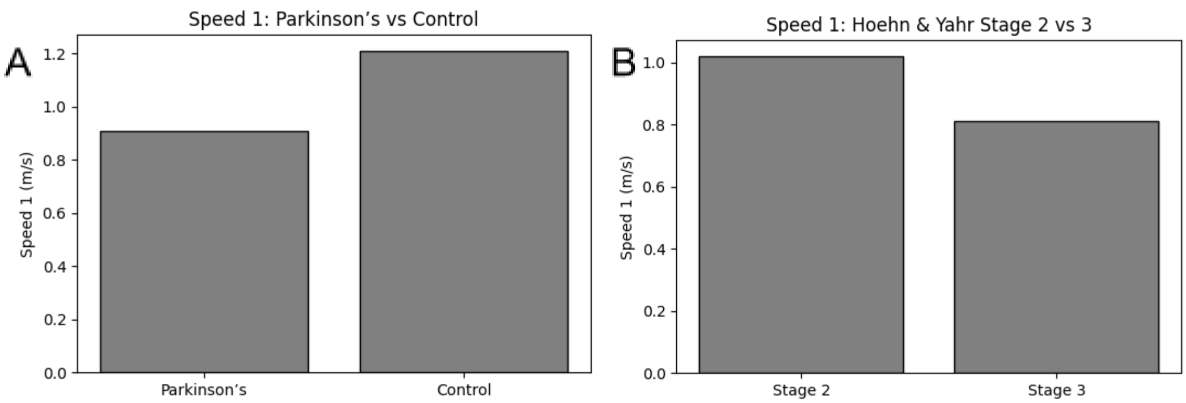

Figure 1: A) Graph of the mean speed for Parkinson's patients vs. control (PD: 0.97 m/s vs Healthy: 1.21 m/s; p = 0.03). **B)** Graph of the mean speed for patients with a Hoehn & Yahr stage of 2 vs. 3 (Stage 2: 1.02 m/s vs Stage 3: 0.81 m/s; $p < 0.05$).

Preprocessing

Normalization was an essential step to ensure that force values reflected gait dynamics rather than body size or weight differences between participants. Since raw vertical ground reaction force (vGRF) measurements are directly influenced by total body mass, Min–Max scaling was applied to each of the 16 foot sensors per subject, rescaling values between 0 and 100. This allowed comparison of gait patterns across participants on a standardized scale. To confirm that normalization reduced individual differences, we compared the mean and standard deviation of average sensor forces before and after preprocessing. Prior to normalization, PD participants exhibited lower average forces compared to controls, a difference largely explained by demographic variation in body weight. After normalization, no significant difference in average normalized force was observed between PD and control groups ($P > 0.05$). **Table 1** summarizes group means and standard deviations before and after preprocessing, averaged across each of 16 sensors. Sensor-by-sensor comparisons also showed that normalization successfully removed baseline force magnitude differences, while retaining temporal gait dynamics. This justified the use of normalized sensor data for subsequent model training and analysis.

Table 1: Gait Force by Sensor pre vs post normalization						
	Pre-Normalization			Post-Normalization		
Sensor	Control	PD	P-value	Control	PD	P-value
1	55.0364 ± 10.0027	53.8398 ± 13.9866	0.398542	20.88 ± 3.79	20.42 ± 5.31	0.398542
2	73.0195 ± 30.4467	66.9888 ± 34.2419	0.127885	9.68 ± 4.04	8.88 ± 4.54	0.127885
3	62.1518 ± 19.4503	56.8844 ± 20.0096	0.032634	9.98 ± 3.12	9.14 ± 3.21	0.032634
4	49.7256 ± 20.5845	47.7888 ± 19.1984	0.442465	8.08 ± 3.34	7.76 ± 3.12	0.442465
5	59.1599 ± 23.3521	63.6273 ± 26.6732	0.143524	9.92 ± 3.92	10.67 ± 4.47	0.143524
6	27.8314 ± 22.9359	36.1110 ± 24.6702	0.005177	6.19 ± 5.10	8.03 ± 5.48	0.005177
7	64.8928 ± 21.1790	69.3770 ± 22.1147	0.095533	13.47 ± 4.40	14.41 ± 4.59	0.095533
8	69.4400 ± 19.8754	81.2448 ± 22.9153	9.59E-06	11.04 ± 3.16	12.91 ± 3.64	9.59E-06
9	36.5557 ± 19.8794	39.0061 ± 23.4574	0.350852	7.29 ± 3.96	7.78 ± 4.68	0.350852
10	71.6570 ± 28.0526	65.7013 ± 36.8570	0.124292	11.11 ± 4.35	10.19 ± 5.71	0.124292
11	59.8245 ± 18.6910	56.3191 ± 19.4014	0.138748	9.80 ± 3.06	9.23 ± 3.18	0.138748
12	50.2973 ± 21.1643	49.6423 ± 22.1196	0.806821	9.16 ± 3.85	9.04 ± 4.03	0.806821
13	61.9705 ± 21.7850	67.6210 ± 25.8153	0.050837	10.80 ± 3.80	11.79 ± 4.50	0.050837
14	27.9660 ± 20.2965	34.8748 ± 23.9073	0.010473	6.98 ± 5.07	8.71 ± 5.97	0.010473
15	63.0427 ±	66.3392 ±	0.191037	12.57 ± 4.06	13.23 ± 3.91	0.191037

	20.3585	19.6258				
16	69.1134 ± 20.7532	79.6281 ± 23.6767	0.000137	6.25 ± 1.88	7.20 ± 2.14	0.000137

Machine Learning Model

We implemented a RNN to classify participants as having Parkinson's disease or being healthy controls based on their gait sensor data. The model inputs consisted of 16 force sensor readings from the left and right feet, captured at 100 Hz during walking. Each walking trial was divided into short segments of 200 consecutive sensor readings. This approach allows the model to focus on short, consistent patterns in gait rather than processing the entire long trial at once, which can be inefficient. Each window inherits the participant's label, Parkinson's disease or healthy control, so the network learns to recognize the temporal gait features associated with each group. By using these overlapping segments, the RNN can capture subtle variations in walking patterns that may indicate early-stage Parkinson's disease. All sensor data were Min–Max normalized to a 0–100 range per sensor, which reduces inter-subject variability due to body weight differences and focuses the model on gait patterns rather than absolute force values. The RNN was trained for 10 epochs using normalized data, achieving strong classification performance (**Figure 2**).

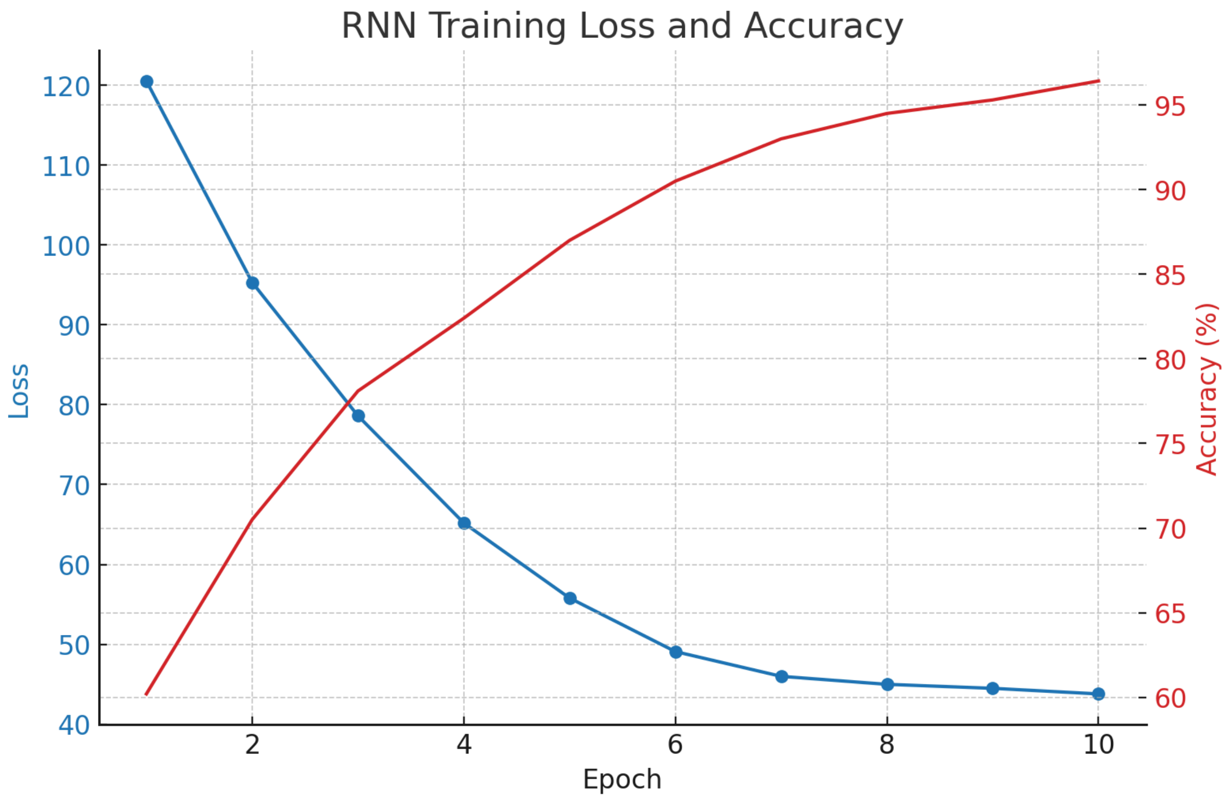

Figure 2: Graph of the accuracy and loss of the RNN vs the epoch. Add more details. Loss-function etc.

Discussion

In this study, we developed a RNN to detect PD from high-resolution gait sensor data. Our goal was to determine whether temporal patterns in walking could serve as reliable indicators of early-stage PD, addressing the need for earlier and more objective detection methods. By preprocessing force sensor data with Min–Max normalization and segmenting walking trials into short, consistent windows, the model focused on gait dynamics rather than differences in body weight or overall force magnitude.

The trained RNN achieved strong classification performance using normalized sensor inputs, demonstrating that subtle temporal patterns in gait can distinguish PD patients from healthy controls. This highlights the potential of wearable sensors combined with machine learning as a non-invasive tool for early detection, which is particularly valuable given the progressive nature of PD and the benefits of early treatment. A unique aspect of our model is its ability to analyze sequential data directly from multiple foot sensors simultaneously, capturing interactions between sensors over time rather than treating each measurement independently.

However, there are several limitations. The model was trained on data from a single publicly available dataset, limiting generalizability. We lacked an independent validation dataset, which means performance metrics could be overestimated. Additionally, the sample size, while moderate, still raises concerns about potential overfitting, especially given the high dimensionality of the input data.

Future work should aim to replicate these findings using independent datasets and explore incorporating additional sensor modalities, longer walking trials, or real-world, free-living gait data. Expanding the dataset and validating the model on diverse populations will strengthen confidence in its clinical applicability. Overall, our findings suggest that RNNs applied to gait sensor data provide a promising approach for earlier and more accurate detection of Parkinson's disease, paving the way for timely interventions and improved patient outcomes.

Conclusion

This study demonstrates that RNNs can effectively classify Parkinson's disease from high-resolution gait sensor data. By leveraging temporal patterns in walking and normalizing for inter-subject variability, the model was able to capture subtle gait changes that distinguish PD patients from healthy controls. Our results highlight the potential of combining wearable sensors with machine learning as a non-invasive, objective tool for early detection of Parkinson's disease. While promising, these findings are based on a single dataset, and further validation on independent and larger cohorts is necessary to confirm the model's generalizability. Future studies should explore applying this approach in diverse populations and real-world settings, which could ultimately support earlier diagnosis and more timely interventions, improving patient outcomes.

References

[1] "Statistics." *Parkinson's Foundation*, www.parkinson.org/understanding-parkinsons/statistics. Accessed 3 Sept. 2025.

[2] Haaxma, Charlotte A, et al. "Gender Differences in Parkinson's Disease." *Journal of Neurology, Neurosurgery, and Psychiatry*, U.S. National Library of Medicine, Aug. 2007, pmc.ncbi.nlm.nih.gov/articles/PMC2117736/.

[3] Wright Willis, Allison, et al. "Geographic and Ethnic Variation in Parkinson Disease: A Population-Based Study of US Medicare Beneficiaries." *Neuroepidemiology*, U.S. National Library of Medicine, 2010, pmc.ncbi.nlm.nih.gov/articles/PMC2865395/.

[4] S;, Dauer W;Przedborski. "Parkinson's Disease: Mechanisms and Models." Neuron, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/12971891/. Accessed 10 Sept. 2025.

[5] Zhou, Zhi Dong, et al. "Role of Dopamine in the Pathophysiology of Parkinson's Disease." Translational Neurodegeneration, U.S. National Library of Medicine, 18 Sept. 2023, pmc.ncbi.nlm.nih.gov/articles/PMC10506345/.

[6] Gandhi, Kavita R. "Levodopa (L-DOPA)." StatPearls [Internet]., U.S. National Library of Medicine, 17 Apr. 2023, www.ncbi.nlm.nih.gov/books/NBK482140/.

[7] Wang, Jian, et al. "Biomarkers of Parkinson's Disease: Current Status and Future Perspectives." Drug Discovery Today, U.S. National Library of Medicine, Feb. 2013, pmc.ncbi.nlm.nih.gov/articles/PMC3557745/.

[8] Lehéricy S;Sharman MA;Dos Santos CL;Paquin R;Gallea C; "Magnetic Resonance Imaging of the Substantia Nigra in Parkinson's Disease." Movement Disorders : Official Journal

of the Movement Disorder Society, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/22649063/. Accessed 12 Sept. 2025.

[9] Morris R;Hickey A;Del Din S;Godfrey A;Lord S;Rochester L; "A Model of Free-Living Gait: A Factor Analysis in Parkinson's Disease." Gait & Posture, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/27883986/. Accessed 13 Sept. 2025.

[10] Hausdorff, Jeffrey. "Gait in Parkinson's Disease." PhysioNet, 25 Feb. 2008, physionet.org/content/gaitpdb/1.0.0/.