Beyond Diagnosis: Enhancing Parkinson's Disease Classification and Symptom Profiling Using Wearable Data

Charlotte Wong, Kiarash Kianidehkordi

University of Toronto, Ontario, Canada

* Corresponding author
E-mail: kiarash.kianidehkordi@mail.utoronto.ca (KK)

Word count: [2700]

Funding: The author(s) received no specific funding for this work.
Conflicts of interest: All authors declare no conflict of interest.

Beyond Diagnosis: Enhancing Parkinson's Disease Classification and Symptom Profiling Using Wearable Data

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects aging populations. In Canada, PD impacts approximately 1 in every 500 people.The hallmark motor symptoms of Parkinson's include tremors, muscle rigidity, bradykinesia, and postural instability, which increases the risk of falls (2). As the disease progresses, patients often experience a loss of mobility, independence, and overall quality of life, emphasizing the urgent need for more efficient, innovative, and patient-centered approaches to care. One promising avenue is the application of artificial intelligence (AI) in Parkinson's disease management, particularly through the development of smart drug delivery systems. These technologies aim to modulate medication release in real time based on an individual's fluctuating symptoms. This study explores supervised and unsupervised machine learning techniques for classification of Parkinson's disease (PD) and profiling its motor symptoms using wrist-worn accelerometer data from the Parkinson's Apple Watch Dataset (PADS). We first replicated a published supervised classification pipeline using the Bag-of-Symbolic-Fourier-Approximation-Symbols (BOSS) algorithm combined with a Support Vector Machines. Our results closely matched the original study's performance, validating the reliability and reproducibility of this approach for distinguishing PD from healthy controls and differential diagnoses. For unsupervised analysis, we extracted tremor-related features from filtered acceleration signals, including mean acceleration, amplitude envelope, and power spectral density metrics within the 3–12 Hz tremor band. Gaussian Mixture Models (GMM) were used to identify latent tremor patterns, with the optimal number of clusters selected via Bayesian Information Criterion. Dimensionality reduction using t-distributed Stochastic Neighbor Embedding (t-SNE) enabled visual inspection of clusters.

Keywords
Parkinson's Disease, Motor Symptom Profiling, Wearable Sensors

Introduction

Supervised Disease Classification

The project validates the reliability and reproducibility of supervised machine learning methods for Parkinson's disease classification using wearable movement data. We replicated a published pipeline that combines automatic feature extraction based on the Bag-of-Symbolic-Fourier-Approximation-Symbols (BOSS) algorithm, along with Support Vector Machine (SVM) and nested cross-validation. We chose to replicate the BOSS algorithm as it demonstrated the best performance among all machine learning methods carried out in the original study. By reproducing these techniques and comparing the results, we aimed to compare how reliable the supervised classifications are, which allow us to compare both the supervised

and unsupervised approaches for analyzing motor symptoms in Parkinson's disease.

The Parkinson's Apple Watch Data Set (PADS), which compiles continuous, real-world data from wearable sensors worn by 469 participants, enables researchers to explore these innovations (need reference). By capturing detailed accelerometer data alongside personal and clinical information, PADS provides a foundation for developing responsive treatment solutions. Our research leverages this dataset to apply advanced analytics and machine learning techniques, with the goal of identifying patterns in symptom progression—especially motor fluctuations throughout the day.

Unsupervised Symptoms Profiling.

The objective of this part of the project is to extend the research of the original study by utilizing advanced analytics to uncover and classify motor symptom patterns in individuals with and without Parkinson's disease (PD). This will be achieved using unsupervised machine learning techniques, specifically given the absence of symptom-level labels in the accelerometer data. Unsupervised learning is uniquely suited for this task because its aim is to discover hidden patterns, structures, and relationships within data without requiring predefined output labels. In the context of PD, where symptom presentation can be highly heterogeneous and underlying progression mechanisms are complex, unsupervised methods allow for the learning of meaningful representations of motor behavior directly from raw sensor data, providing an objective, data-driven foundation for symptom profiling.

Accelerometers are micro-electro-mechanical systems (MEMS) that measure the rate of change in velocity, or acceleration, of a body along typically three orthogonal axes (x, y, z). These sensors are widely utilized in wearable devices and are particularly effective for capturing the dynamics of human movement. Their simplicity and cost-effectiveness make them suitable for continuous monitoring in real-world settings. When integrated into wearable inertial measurement units (IMUs), often combined with gyroscopes (which measure angular velocity), they provide comprehensive kinematic information, allowing for a detailed assessment of overall movement changes. For specific motor symptoms in Parkinson's disease, strategic placement of these sensors on limbs, such as fingers, wrists, waist, or lower back, enables effective capture of relevant movement data. The PADs dataset used in this study incorporates accelerometer data collected from Apple Watches worn on both wrists of participants. Accordingly, the data must be processed and analyzed to objectively quantify Parkinson's disease motor symptoms that are observable through this sensor placement.

Motor Symptom Profiling with Accelerometer Data:

- **Tremor:** Parkinson's resting tremor is characterized by a rhythmic shaking, typically occurring at a specific frequency range, often cited as 4-8 Hz or 4-5 Hz. Key metrics

derived from accelerometer data for tremor quantification include tremor amplitude, frequency, and overall intensity. Studies have successfully extracted features such as median frequency, dispersion frequency, fundamental tremor frequency, and power spectral density to assess tremor severity.

- **Bradykinesia (Slowness of movement):** In clinical settings, bradykinesia is assessed through repetitive joint tasks like finger tapping, whole-hand grasping, and pronation-supination movements of the hands. These tasks are particularly amenable to objective tracking using MEMS inertial sensors. Accelerometer and gyroscope data can quantify bradykinesia by measuring various kinematic parameters, including the speed and amplitude of movements, range of motion, and crucially, "fatigability"—the progressive reduction in speed and amplitude over repetitive actions. The smoothness of movement is another important kinematic index that can be extracted.

- **Rigidity/Stiffness:** Rigidity and stiffness are less directly quantifiable by accelerometers alone.

- **Gait Disturbance:** While accelerometers are highly effective in assessing gait quality and disturbances characteristic of Parkinson's disease (PD)—such as shuffling gait, impaired balance, and postural instability—through the use of foot-mounted or waist-mounted accelerometers, the PADs dataset was gathered from wrist-mounted accelerometers, making the data less suitable for profiling gait disturbances.

Given the objective of the project and the characteristics of the PADs dataset (wrist-mounted accelerometer data, absence of symptom-level labels), we will proceed with unsupervised tremor profiling for several key reasons. Firstly, resting tremors are often among the most common and earliest motor symptoms to appear in individuals with Parkinson's disease. Their distinct rhythmic characteristics (4-8 Hz or 4-5 Hz) make them particularly well-suited for identification and quantification through frequency analysis of accelerometer data, even without prior labeling. Secondly, tremor metrics derived from accelerometers (amplitude, frequency, intensity, power spectral density) are robust and well-defined. This allows for the discovery of meaningful patterns using unsupervised techniques, as the inherent structure of tremor data is more readily extractable from raw sensor readings. Furthermore, the ability to objectively quantify tremor in an unsupervised manner can provide a more robust predictor for early Parkinson's detection and progression monitoring. By identifying clusters and patterns related to tremor characteristics, we can potentially differentiate individuals with and without PD, and even stratify those with PD by tremor severity, purely from the sensor data. In contrast, bradykinesia is more complex to profile in an unsupervised manner with wrist-mounted accelerometers, as its assessment typically relies on specific, identifiable tasks. Finally, as previously established, rigidity/stiffness and gait disturbance are significantly less suitable for profiling given the experimental design of the PADs

dataset. Rigidity is not directly quantifiable by accelerometers alone, and gait analysis requires sensor placement on the feet or waist, not the wrists. Therefore, focusing on tremor leverages the strengths of the available data and the unsupervised learning approach most effectively.

Methods

Supervised Disease Classification Replication using BOSS + SVM

For evaluation, we replicated the original study. After preprocessing, the 72-channels accelerometer data with 1789 timesteps was used as an input for the BOSS extraction method. This BOSS algorithm converts each time channel into a symbolic representation, and a SVM classifier was employed for finding the best boundary between classes. Similar to the original study, two classification tasks were evaluated: (1) Parkinson's disease versus healthy controls (PD vs HC), and (2) Parkinson's disease versus differential diagnoses (PD vs DD). We used the nested cross-validation to split data into five folds, and the result is presented as mean ± standard deviation for balanced accuracy, F1 score, precision, and recall.

Unsupervised Symptoms Profiling.

The wrist accelerometer signals were preprocessed to extract meaningful features from the tremor-relevant frequency band. Each subject's wrist data was first transformed from 3-axis (X, Y, Z) to a single-dimensional signal by calculating the Euclidean magnitude of acceleration across axes. To isolate tremor-related components, the magnitude signal underwent bandpass filtering using a 6th-order Butterworth filter in the 3–12 Hz range. This frequency band typically corresponds to physiological tremor, particularly in Parkinson's disease(4-8 Hz frequencies).The filtered signal was further analyzed using Welch's method to compute its Power Spectral Density (PSD). From the PSD, two features were derived: Peak Power: The maximum power value within the 3–12 Hz band. Area Under the Curve : The integral of the PSD within the same band, representing total tremor energy. The envelope of the filtered signal was estimated by identifying local maxima (peaks) and minima (troughs). These extrema points were used to approximate the amplitude envelope, from which the mean envelope value was computed. This feature captures the variability and strength of tremor oscillations.Finally, the absolute mean acceleration over time was computed to represent overall motion intensity, regardless of frequency content. This value helps distinguish between low-movement and high-movement states.Each wrist thus produced a 4-dimensional feature vector :

1. Mean Acceleration
2. Mean Envelope Amplitude
3. Peak Power in Tremor Band
4. PSD AUC in Tremor Band

These vectors were concatenated for both wrists, resulting in an 8-dimensional feature vector per subject.The extracted features were aggregated across all subjects, forming a feature matrix of shape (n_subjects, 8). Before applying clustering algorithms, the feature matrix was standardized using Z-score normalization via StandardScaler, ensuring that each feature contributed equally to the clustering algorithm. **Gaussian Mixture Model (GMM)** was applied. GMM is a probabilistic clustering method that assumes the data is generated from a mixture of Gaussian distributions. The optimal number of clusters was selected based on the Bayesian Information Criterion (BIC). BIC scores were computed for models with 1 to 10 clusters, and the model with the lowest BIC was selected. After fitting the optimal GMM, each subject was assigned to a cluster, representing an unsupervised categorization based on features that resemble tremor characteristics. To evaluate the validity of the clustering, two internal metrics were computed:

- **Silhouette Score**: Measures the separation distance between resulting clusters. Values range from -1 to 1, with higher values indicating better-defined clusters.

- **Calinski-Harabasz Index**: Measures the ratio of between-cluster dispersion to within-cluster dispersion, where higher values indicate better-defined and well-separated clusters.

These metrics provided quantitative support for the quality of the GMM-based clustering.To facilitate visual inspection of the high-dimensional clustering results, **t-distributed Stochastic Neighbor Embedding (t-SNE)** was used for dimensionality reduction. t-SNE projects the 8-dimensional feature space into 2D while preserving local relationships in the data.The resulting 2D coordinates were visualized in scatter plots, colored either by cluster label (from GMM), clinical condition (e.g., Healthy, Parkinson's), or auxiliary patient metadata such as age-at-diagnosis difference. Since parkinson is a progressive disease it was expected that patients with higher difference in age and age at diagnosis would have been in similar clusters. For Parkinson's patients, the absolute difference between age and age_at_diagnosis was computed as a proxy for disease duration. This difference was binned into discrete intervals (e.g., 0–10 years, 10–20 years, etc.), and used as a color code in a separate t-SNE visualization. This helped assess whether movement patterns varied with time since diagnosis enabling a comparison of GMM clusters based on feature vectors with known information about the disease..

Results

Supervised Disease Classification Replication using BOSS + SVM

Our replicated BOSS + SVM pipeline on the PADs dataset (Table #1) achieved results and values that align with the previously published results (Table #2). In the PD vs. HC task, our model

achieved a balanced accuracy of 78.0% ± 5.0%, with an F1 score of 0.75 ± 0.05, precision of 0.75 ± 0.07, and recall of 0.77 ± 0.07; while for the PD vs. DD task, the balanced accuracy was 76.1% ± 3.0%, F1 score was 0.63 ± 0.05, precision was 0.67 ± 0.14, and recall was 0.63 ± 0.08 (Table #1). These metrics are both similar to the performance in the original publication, which achieved a mean balanced accuracy of 81.74% for PD vs HC and 71.57% for PD vs DD.

Table #1: Replication Results of BOSS + SVM Movement Classification on the PADs Movement Dataset (mean ± standard deviation)

Task	Balanced Accuracy (%)	F1	Precision	Recall
PD vs. HC	78.0 ± 5.0	0.75 ± 0.05	0.75 ± 0.07	0.77 ± 0.07
PD vs. DD	76.1 ± 3.0	0.63 ± 0.05	0.67 ± 0.14	0.63 ± 0.08

Table #2: Results of Previously Published Performance of BOSS + SVM on the PADs Movement Dataset (mean) (adapted from (1))

Task	Balanced Accuracy (%)	F1	Precision	Recall
PD vs. HC	81.74	0.9257	0.9124	0.9395
PD vs. DD	71.57	0.589	0.621	0.568

Unsupervised Symptoms Profiling.

The t-SNE projection (Figure 1) reveals distinct spatial groupings corresponding to GMM-derived clusters. These clusters capture meaningful variability in the tremor feature space. The clustering solution demonstrated strong internal validity, with a silhouette score of **0.688** and a Calinski-Harabasz index of **1718.96**, indicating well-separated and compact clusters.To assess alignment between data-driven tremor phenotypes and clinical diagnoses, we colored the same t-SNE projection by patient condition: Healthy (0), Parkinson's Disease (1), and Differential Diagnosis (2) (Figure 2). We observed that all diagnostic categories were distributed across multiple clusters, suggesting that tremor patterns alone do not linearly separate by diagnosis. However, some spatial bias was noted—certain clusters showed higher concentrations of Parkinson's cases, while others were more heterogeneous.Figure 3 shows the t-SNE projection colored by the absolute difference between patient age and age at diagnosis, binned into intervals. The majority of patients exhibited minimal age difference (dark blue), while a few localized regions showed higher deviations (yellow). These outliers may reflect cases of early onset, late diagnosis, or atypical tremor progression.

Figure #1: t-SNE projection of GMM derived clusters

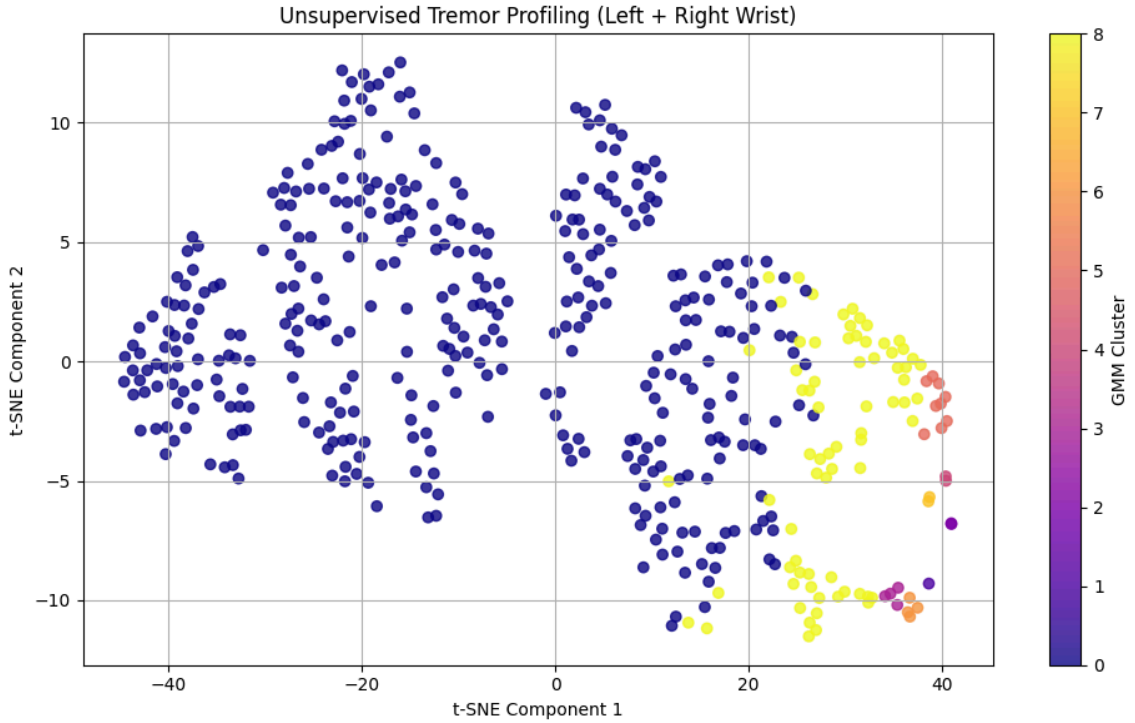


Figure #2: t-SNE projection of GMM derived clusters colored by patient condition

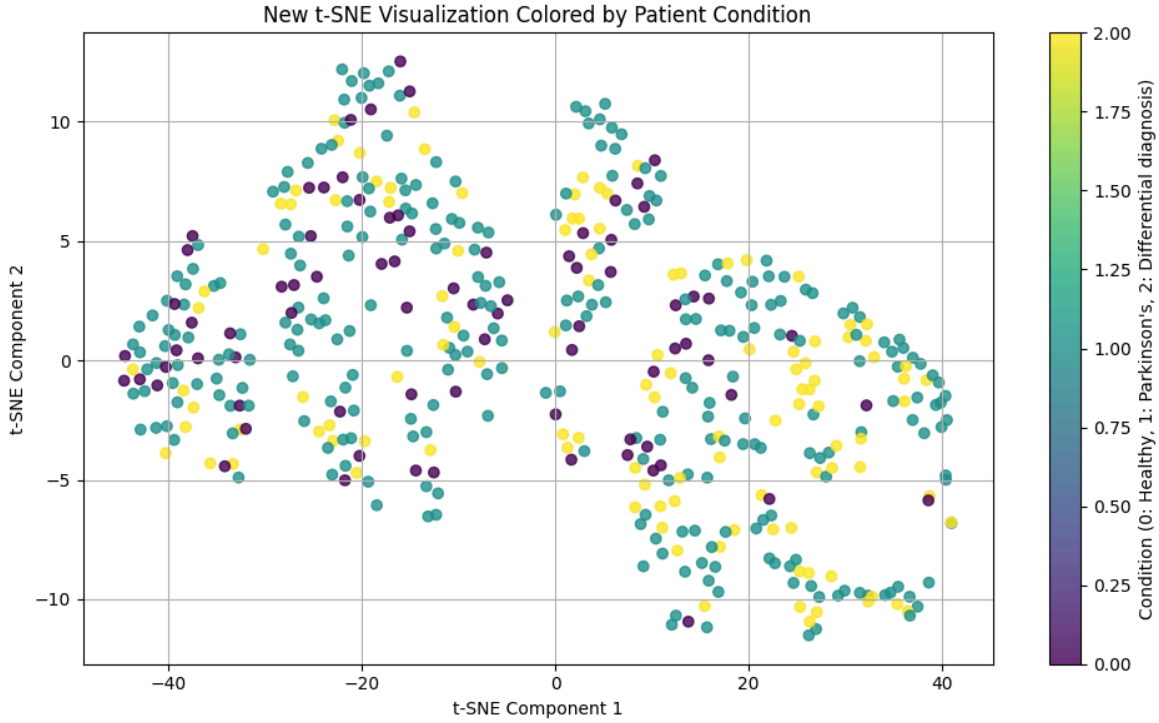
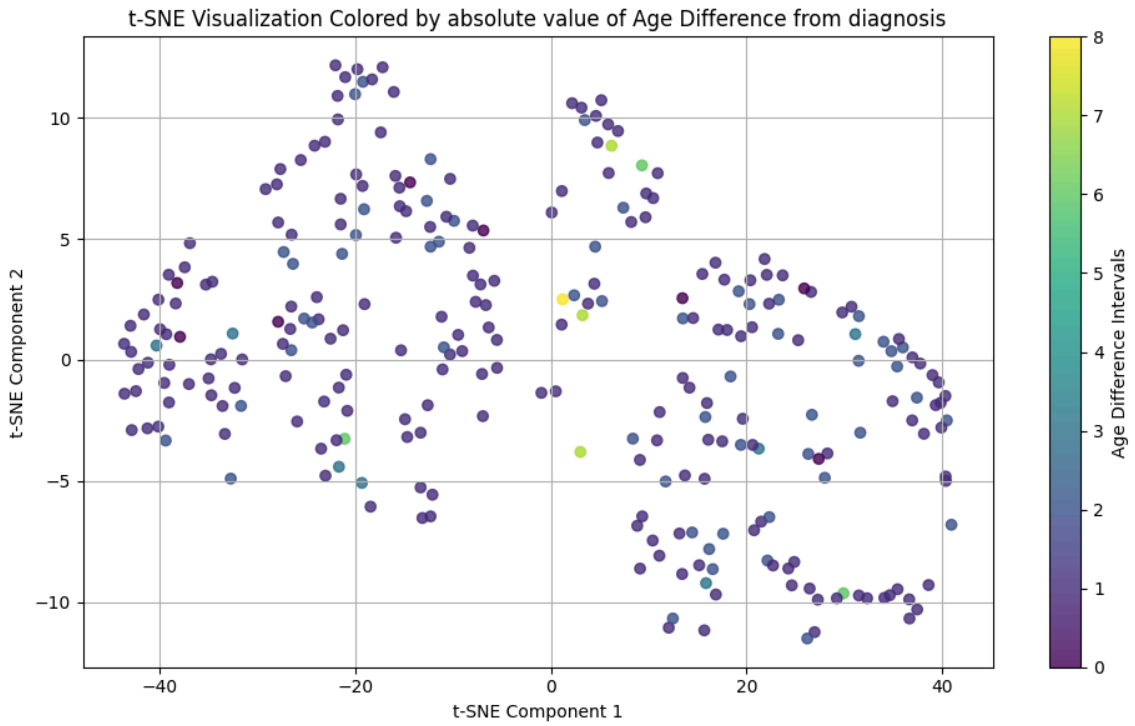


Figure #3: t-SNE projection of GMM derived clusters colored by absolute difference between patient age and age at diagnosis



Discussion

Supervised Disease Classification Replication using BOSS + SVM

The similar results indicate the replication of the BOSS + SVM movement classification was successful. This further validates the use of symbolic time series methods to classify Parkinson's disease from wearable data. The high degree of similarity confirms the overall reliability and generalizability of the BOSS+SVM approach. Therefore, these suggest that BOSS-based feature extraction can serve as a foundation for future digital phenotyping and remote disease monitoring. However, minor differences that are shown in the values between our replication and the original study may be due to differences in random sampling during cross-validation splits and updates in Python library versions (Python 3.11 vs Python 3.13). Despite these technical differences, our results were within 1 SD of the original publication. The close correspondence between the two studies supports the reliability of supervised disease classification, which emphasizes machine learning research is valuable for biomedical sensor data. We hope to integrate these techniques in non-motor symptom data (eg. from questionnaires) in the future to help capture the full spectrum of Parkinson's disease symptoms.

Unsupervised Symptoms Profiling.

Several limitations should be acknowledged. First, the Gaussian Mixture Model (GMM) assumes that the underlying data distribution is a mixture of multivariate Gaussian components. This assumption may not hold for real-world tremor data, which can be non-Gaussian, skewed, or

exhibit complex multimodal characteristics. Consequently, some of the identified clusters may reflect artifacts of model structure rather than true physiological differences.Second, although t-SNE is effective for visualizing high-dimensional data, it is a non-linear technique that is sensitive to initialization and hyperparameter choices such as perplexity. As a result, clusters that appear visually distinct in two-dimensional space may not be as well-separated in the original feature space. These visual patterns should therefore be interpreted with caution. Third, the clustering is inherently constrained by the input features derived from wrist-worn devices. While all participants performed standardized tasks in a clinical environment, tremor is a multifactorial phenomenon influenced by medication, context, and intra-individual motor fluctuations—factors that may not be fully represented in the extracted time-series features.Lastly, the use of an unsupervised learning framework means that cluster labels are not guided by predefined diagnostic categories. While this allows for unbiased discovery of data-driven groupings, it also makes the interpretation of clusters more challenging. The patterns identified by GMM should not be assumed to correspond directly to distinct tremor types or clinical subtypes, despite being informed by time-series characteristics.

Conclusion

This study successfully replicated the use of supervised machine learning techniques, confirming the reliability of using BOSS+SVM in classifying Parkinson's disease on the PADs wearable sensor dataset. At the same time, we demonstrated the potential of unsupervised learning techniques, particularly Gaussian Mixture Models and t-SNE visualization, in uncovering data-driven patterns in wrist-worn sensor recordings of tremor activity. Despite the absence of labeled data, distinct clusters emerged that suggest underlying structure in the tremor feature space. However, caution is necessary in interpreting these groupings, as they are influenced by model assumptions and limitations inherent to both GMM and t-SNE.

Acknowledgments

The paper utilized PADs data set[1] and tried to replicate the result found by [2]. It further extended [2] by implementation of unsupervised profiling of tremors based on [3].

Bibliography

1.Varghese J, Brenner A, Fujarski M, Marie, Plagwitz L, Warnecke T. Author Correction: Machine Learning in the Parkinson's disease smartwatch (PADS) dataset. npj Parkinson's Disease. 2024 Apr 28;10(1).

2.Varghese J, Niewöhner S, Soto-Rey I, Schipmann-Miletić S, Warneke N, Warnecke T, et al. A Smart Device System to Identify New Phenotypical Characteristics in Movement Disorders. Frontiers in Neurology. 2019 Jan 30;10.

3.van der Linden C 1, Berger T 1, Brandt GA 1, Strelow JN 2, Jergas H 1, Baldermann JC 3, et al. Accelerometric Classification of Resting and Postural Tremor Amplitude. ProQuest [Internet]. 2023;8621. Available from: https://www.proquest.com/docview/2882824286?fromopenview=true&parentSessionId=Pk2b1YlB9PHHUTBORwUTKaKrkVucDPOtE8ImtDznbFY%3D&pq-origsite=gscholar&sourcetype=Scholarly%20Journals&parentSessionId=GS94ely6rdDKmUmGuLqi2z2HMeuYY1MzmCvmuhpeHh4%3D

4.Syed Huzaifa Ali, Akhtar A, Meeran Ali Khan, Bilal M. Detection of Parkinson's Tremor in Real Time Using Accelerometers. 2021 Aug 23;

5.National Institute on Aging. Parkinson's Disease: Causes, symptoms, and treatments [Internet]. National Institute of Aging. 2022. Available from: https://www.nia.nih.gov/health/parkinsons-disease/parkinsons-disease-causes-symptoms-and-treatments

6.di Biase L, Summa S, Tosi J, Taffoni F, Marano M, Cascio Rizzo A, et al. Quantitative Analysis of Bradykinesia and Rigidity in Parkinson's Disease. Frontiers in Neurology. 2018 Mar 6;9.

7.Khalil RM, Shulman LM, Gruber-Baldini AL, Reich SG, Savitt JM, Hausdorff JM, et al. Applying Wearable Sensors and Machine Learning to the Diagnostic Challenge of Distinguishing Parkinson's Disease from Other Forms of Parkinsonism. Biomedicines [Internet]. 2025 Feb 25 [cited 2025 Apr 17];13(3):572–2. Available from: https://www.mdpi.com/2227-9059/13/3/572

8.Polvorinos-Fernández C, Sigcha L, Borzì L, Olmo G, Asensio C, López JM, et al. Evaluating Motor Symptoms in Parkinson's Disease Through Wearable Sensors: A Systematic Review of Digital Biomarkers. Applied Sciences. 2024 Nov 6;14(22):10189.

9.Häring V, Selzam V, Martin-Rodriguez JF, Schwingenschuh P, Tamás G, Raethjen J, et al. Phenotypical differentiation of tremor using time series feature extraction and machine learning. 2024 Mar 15;

10.Channa A, Cramariuc O, Memon M, Popescu N, Mammone N, Ruggeri G. Parkinson's disease resting tremor severity classification using machine learning with resampling techniques. Frontiers in Neuroscience. 2022 Oct 28;16.

11.Dattola S, Augusto Ielo, Quartarone A, Cristina M. Integrating Wearable Sensor Signal Processing with Unsupervised Learning Methods for Tremor Classification in Parkinson's Disease. Bioengineering [Internet]. 2025 Jan 6 [cited 2025 Mar 7];12(1):37–7. Available from: https://www.mdpi.com/2306-5354/12/1/37

12.Boutaayamou M, Frédérique Depierreux, Parmentier E, Schwartz C, Olivier Brüls, Verly J, et al. Quantification of refined temporal gait parameters in Parkinson's disease using an accelerometer-based ambulatory system. Frontiers in Neuroscience. 2018 Jan 1;12.

13.Lu R, Xu Y, Li X, Fan Y, Zeng W, Tan Y, et al. Evaluation of Wearable Sensor Devices in Parkinson's Disease: A Review of Current Status and Future Prospects. Parkinson's Disease. 2020 Sep 23;2020:1–8.