

Article

Not peer-reviewed version

Evaluation of Facebook as a Longitudinal Data Source for Parkinson's Disease Insights

[Jeanne M Powell](#)^{*}, [Charles Cao](#), Kayla Means, Sahithi Lakamana, Abeed Sarker, [J Lucas Mckay](#)

Posted Date: 6 May 2025

doi: 10.20944/preprints202505.0280.v1

Keywords: Parkinson's disease; social media; text classification



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Evaluation of Facebook as a Longitudinal Data Source for Parkinson's Disease Insights

Jeanne M. Powell ^{1,*}, Charles Cao ², Kayla Means ², Sahithi Lakamana ¹, Abeed Sarker ¹ and J. Lucas Mckay ¹

¹ Department of Biomedical Informatics, Emory School of Medicine, Atlanta, GA 30030, USA; sahithi.krishnaveni.lakamana@emory.edu; abeed.sarker@emory.edu; j.lucas.mckay@emory.edu

² Rollins School of Public Health, Emory University, Atlanta, GA 30030, USA; charles.cao@emory.edu; kayla.means@emory.edu

* Correspondence: jeanne.marie.powell@emory.edu

Abstract: Background/Objectives: Parkinson's disease (PD) is a neurodegenerative disorder with a prolonged prodromal phase and progressive symptom burden. Traditional monitoring relies on clinical visits post-diagnosis, limiting the ability to capture early symptoms and real-world disease progression outside structured assessments. Social media provides an alternative source of longitudinal, patient-driven data, offering an opportunity to analyze both pre-diagnostic experiences and later disease manifestations. This study evaluates the feasibility of using Facebook to analyze PD-related discourse and disease timelines. **Methods:** Participants (N=60) diagnosed with PD, essential tremor, or atypical parkinsonism, along with caregivers, were recruited. Demographic and clinical data were collected during structured interviews. Participants with Facebook accounts shared their account data. PD-related posts were identified using a Naïve Bayes classifier (recall: 0.86, 95% CI: 0.84–0.88, AUC=0.94) trained on a ground-truth dataset of 6,750 manually labeled posts. **Results:** Among participants with PD (PwPD), Facebook users were significantly younger but had similar MDS-UPDRS scores and disease duration compared to non-users. Among Facebook users with PD, 90% had accounts before diagnosis, enabling retrospective analysis of pre-diagnostic content. PwPD maintained 14 ± 3 years of Facebook history, including 5 ± 6 years pre-diagnosis. They shared an average of 99 ± 172 PD-related posts, with 24 ± 58 occurring pre-diagnosis. Overall, 69% explicitly referenced PD, and 93% posted about PD-related themes. **Conclusions:** Facebook is a viable platform for studying PD progression, capturing both early content from the premorbid period and later-stage symptoms. These findings support its potential for disease monitoring at scale.

Keywords: Parkinson's disease; social media; text classification

1. Introduction

Social media platforms are used by nearly 60% of the global population [1], with Facebook alone reporting 2.89 billion active users in 2024 [2]. Beyond maintaining social connections [3], older adults use social media to share aspects of their daily lives, including health-related experiences [4]. Prior research has demonstrated that social media data, including explicit health disclosures and behavioral indicators (e.g., changes in posting frequency, sentiment shifts), can be leveraged for medical applications [5,6]. However, the extent to which social media serves as a meaningful data source for specific neurological conditions, including Parkinson's disease (PD), remains underexplored.

PD, the second-most common neurodegenerative disorder, affects over 6.2 million people worldwide [7]. Its clinical heterogeneity, spanning both motor and non-motor symptoms, complicates diagnosis and progression modeling [8]. PD is diagnosed based on motor symptoms that emerge after extensive dopaminergic neurodegeneration, often estimated to exceed 50% neuronal

loss [9]. The prodromal phase, during which neurodegeneration occurs before clinical diagnosis, remains difficult to study prospectively, as most PD cases are idiopathic [10]. Leveraging retrospective data sources that document lived experiences before and after diagnosis may provide insights into the premorbid period and early disease manifestations.

This study evaluates the feasibility of using Facebook as a data source for PD research. We assess whether individuals with PD and related disorders are represented on the platform, the extent of their health-related disclosures, and whether their longitudinal social media activity can provide insights into the prodromal phase. Because PD shares clinical features with and is sometimes misdiagnosed as Essential Tremor (ET) and Atypical Parkinsonism (AP) [11], individuals with these conditions were also included. Caregivers, who frequently seek and share PD-related information online [12], were included as well.

This work establishes social media as a viable source of PD-related information and lays the foundation for future studies leveraging large-scale social media data to identify disease-relevant disclosures and track self-reported experiences over time.

2. Materials and Methods

2.1. *Participants and Data Collection*

Participants were recruited through the Emory Brain Health Center, research recontact lists, and PD-specific community events. Eligible participants included individuals diagnosed with PD, ET, or AP, as well as caregivers of individuals with those conditions. Participants had to communicate in English. No additional exclusion criteria were applied. Written informed consent was obtained in person or via teleconferencing, following protocols approved by Emory University's Institutional Review Board (STUDY00005722) and the Declaration of Helsinki.

During structured interviews, demographic and clinical information were collected. Disease severity and symptom burden in PD and AP participants were assessed using the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Parts I (non-motor symptoms), II (motor symptoms), and IV (motor complications) [13], where higher scores indicate greater impairment. ET participants were assessed using MDS-UPDRS Parts I and II, as motor complications (Part IV) are not relevant to ET. MDS-UPDRS Part III (motor examination) was not collected, as physical examinations were outside the scope of data collection.

Participants who opted to share their Facebook data were guided through downloading their Posts, Comments and Reactions, Pages, and Groups in JSON format. They were instructed to download their complete activity history, capturing data from account creation through study participation. All text-based content—posts, comments, media captions, and other written engagement—was authored by the account owner and collectively referred to as "posts." Private messages were neither collected nor analyzed. Study staff provided technical assistance as needed. All data were securely uploaded to a HIPAA-compliant REDCap database [14].

2.2. *PD-Relevant Post Identification*

2.2.1. Text Extraction from JSON Files

Participant-generated text was extracted from Facebook data exports. Given that individual JSON files were not mutually exclusive, deduplication was necessary. Exact duplicates were removed based on identical timestamps and text, and near-duplicates were filtered by retaining only the earliest instance unless at least 180 seconds had elapsed between occurrences, ensuring that only intentionally reshared content was preserved.

Posts flagged as shared memories were retained, as they represented intentional reposts, but associated in-text timestamps (e.g., "X years ago") were removed to prevent redundancy. Text encoding was standardized, and the final dataset was stored in a HIPAA-compliant REDCap database [14].

2.2.2. PD-Related Term Dictionary and Search Strategy

A PD-specific term dictionary with over 1,000 terms was developed from academic literature [8], WebMD, and our prior work on fall-related disclosures in PD [15]. To enhance coverage, ChatGPT-4 generated medical and colloquial synonyms, and the dictionary was refined through manual review. The dictionary spanned motor and non-motor symptoms, treatments, assistive devices, caregiver roles, community engagement, genetics, advocacy, and research participation. See Appendix A for the full dictionary.

The text was lowercased, punctuation was removed, stopwords were filtered, and words were stemmed [16]. A keyword search flagged posts containing either stemmed or unstemmed terms.

2.2.3. Ground-truth Dataset

A subset of 6,750 posts from 14 individuals with PD and 5 caregivers was manually reviewed. Each post contained at least one keyword from the PD-related term dictionary and was independently assessed by two reviewers for relevance to PD using a high-sensitivity approach.

Posts were included if they explicitly mentioned PD, related disorders, treatments, or lacked sufficient context to rule out a PD connection (e.g., "I had a follow-up with my doctor today"). Posts were excluded only if sufficient context indicated they were unrelated (e.g., "I am so fatigued from COVID").

Given that both caregivers and individuals with PD may discuss others with PD, posts were evaluated for general relevance to PD, not just the poster's personal experience. Conflicts were resolved through discussion.

2.2.4. Classifier Development

The dataset was randomly split into 80% training and 20% testing. All feature transformations, including vectorization, scaling, and encoding, were derived from the training set and applied to the test set. Posts were cleaned by removing URLs, tags, hashtags, special characters, punctuation, extra spaces, and stopwords. The processed text was then represented in two ways. First, lemmatization was performed using the NLTK WordNet lemmatizer [17] without explicit part-of-speech tagging, meaning words were lemmatized assuming noun forms by default. Lemmatized text was vectorized using TF-IDF, capturing single words, bigrams, and trigrams while excluding terms appearing in fewer than two documents to reduce noise. Second, tokens were mapped to cluster-based embeddings using a word-clustering approach derived from a Twitter-based corpus [18]. Clusters were vectorized using TF-IDF, capturing only single words and applying the same frequency threshold.

Model features included vectorized lemma and cluster representations, normalized age at time of posting, and one-hot encoded categorical variables (gender, PD diagnosis status). Multiple classifiers, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, AdaBoost, Naïve Bayes, Decision Trees, and XGBoost, were trained using five-fold cross-validation. GridSearchCV [19] was used to optimize macro-averaged recall, ensuring balanced sensitivity across PD-relevant and non-PD content. We also experimented with a soft-voting ensemble that combined all classifiers above.

Performance was evaluated using macro-averaged recall, precision, and F₁-score to account for class imbalance. 95% confidence intervals (CIs) were estimated via bootstrapping, and model discrimination was assessed using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores. The classifier with the highest recall was applied to the full dataset.

2.3. Analyses

Demographic and clinical characteristics were summarized using descriptive statistics. Measures of central tendency are reported as Mean \pm SD unless otherwise noted. Group differences

between Facebook users and non-users were assessed using t-tests for continuous variables and Fisher's exact tests for categorical comparisons.

Facebook-derived text data were analyzed from 45 participants, including a PD-specific subset (N=29) analyzed separately to examine pre- and post-diagnosis content. Facebook engagement was measured by calculating the average duration of use (time between the oldest and newest post), total posts, and time on Facebook before PD diagnosis.

PD-related posts were identified using the best-performing classifier, applied to posts containing at least one term from the custom PD dictionary. The volume of PD-related content was summarized descriptively. Additional analyses examined posting frequency before and after diagnosis and excluded exercise-related content based on predefined terms. A rule-based analysis identified participants who explicitly mentioned 'Parkinson' or '\bPD\b' in any post.

3. Results

3.1. Dataset Characteristics and Participant Engagement

3.1.1. Recruitment and Demographics

We attempted to contact 369 individuals and successfully reached 139. During phone screenings, participants were asked whether they had a Facebook account. Of those contacted, 59% reported having a Facebook account, while 20% did not specify their Facebook status. Among individuals who confirmed having a Facebook account, 85% agreed to participate (see Figure 1A).



Figure 1. Facebook Use and PD-Related Content Among Participants with PD. (a) Facebook use among participants: 59% had accounts, and 85% of those consented to share data. (b) Facebook timelines for participants with PD: blue lines represent posting history, and red dots indicate age at diagnosis. Summary statistics show an average account duration of 14 ± 3 years, 5 ± 6 years of premorbid data, and $4,018 \pm 5,570$ unique posts. (c) PD-related content: 69% explicitly mentioned PD, and 93% referenced PD-related terms.

Of the 139 individuals reached, 62 enrolled, and 60 completed the interview. The final dataset included 38 individuals with PD, 4 with ET, 3 with AP, and 15 caregivers. The sample was 57% male, with an average age of 66.53 ± 12.99 years. Most participants were white (80%), non-Hispanic or Latino (93%), and college-educated (78%). Among participants with PD (N = 38), the average age was 69.21 ± 10.11 years. Disease characteristics included an average disease duration of 8.62 ± 4.73 years, an MDS-UPDRS Part I score of 14.03 ± 7.48 , an MDS-UPDRS Part II score of 13.81 ± 9.68 , and an MDS-UPDRS Part IV score of 6.25 ± 4 .

Among the 60 participants, 46 shared their Facebook data. Eleven participants did not have Facebook accounts and three had Facebook accounts but did not complete the data-sharing process. Of those who shared Facebook data, 30 had PD, 3 had ET, 1 had AP, and 12 were caregivers.

3.1.2. Comparison of Facebook Users and Non-Users

We compared Facebook users and non-users based on demographics and clinical status. Participants with Facebook accounts were significantly younger than those without accounts (64.46 ± 13.24 vs. 75.72 ± 6.41 years; $t(32.14) = 4.16, p < .001, d = -0.91$). However, there were no significant differences between Facebook users and non-users in gender ($p = .09$), race ($p = .49$), ethnicity ($p = 1.00$), or education level ($p = .34$).

Among participants with PD ($N = 30$), those with Facebook accounts were also significantly younger than those without (67.24 ± 9.89 vs. 77.96 ± 5.69 years; $t(15.52) = 3.84, p = .002, d = -0.46$). Facebook users and non-users with PD did not differ in MDS-UPDRS Part I score ($p = .86$), Part II score ($p = .32$), Part IV score ($p = .39$), or disease duration ($p = .81$).

3.1.3. Facebook Account History

Among the 46 participants who shared Facebook data, one had an account but never generated or shared any content, resulting in empty JSON files. This participant was included in overall Facebook user counts but excluded from subsequent analyses. All reported results are based on the remaining 45 participants.

On average, participants had used Facebook for 13 ± 3 years and posted $4,491 \pm 5,637$ times. Among those with PD ($N=29$), the average length of Facebook use was 14 ± 3 years, with an average of $4,018 \pm 5,570$ posts. On average, participants with PD had used Facebook for 5 ± 6 years before their diagnosis, with 90% creating their account before diagnosis (see Figure 1B).

3.2. Identification of PD-relevant Posts

3.2.1. Ground-Truth Dataset

Reviewers achieved substantial interrater reliability (Cohen’s Kappa = 0.79) [20]. Among the 6,750 reviewed posts, 2,400 (35.6%) were classified as PD-relevant, while 4,350 (64.4%) were deemed irrelevant, reflecting the broad range of topics discussed. This imbalance was partly due to keyword ambiguity, where certain terms appeared in non-PD contexts.

3.2.2. Classifier Performance

Table 1 presents the macro-averaged classification metrics for each model. The Naïve Bayes classifier achieved the highest recall (0.86 95% CI: [0.84-0.88]). Although the soft-voting ensemble was implemented, the Naïve Bayes classifier was selected for final PD-related post identification due to its superior recall, ensuring sensitivity to PD disclosures. Please see Appendix B for model hyperparameters.

Table 1. Classifier performance: Macro-averaged metrics.

Model	Recall (95% CI)	Precision (95% CI)	F1-Score (95% CI)
Naïve Bayes	0.86 (0.84-0.88)	0.89 (0.87-0.90)	0.87 (0.85-0.89)
Ensemble	0.84 (0.82-0.86)	0.90 (0.88-0.91)	0.86 (0.84-0.88)
SVM	0.84 (0.82-0.86)	0.86 (0.84-0.88)	0.85 (0.83-0.87)
Decision Tree	0.81 (0.78-0.83)	0.83 (0.81-0.85)	0.81 (0.79-0.84)
Random Forest	0.79 (0.77-0.81)	0.87 (0.85-0.89)	0.81 (0.78-0.83)
XGBoost	0.79 (0.77-0.81)	0.87 (0.85-0.89)	0.81 (0.78-0.83)
KNN	0.77 (0.75-0.79)	0.83 (0.81-0.85)	0.79 (0.76-0.81)
AdaBoost	0.72 (0.70-0.74)	0.85 (0.83-0.87)	0.73 (0.70-0.75)

Overall model discrimination was strong, with an AUC of 0.94 (Figure 2), indicating high separability between PD-relevant and irrelevant posts.

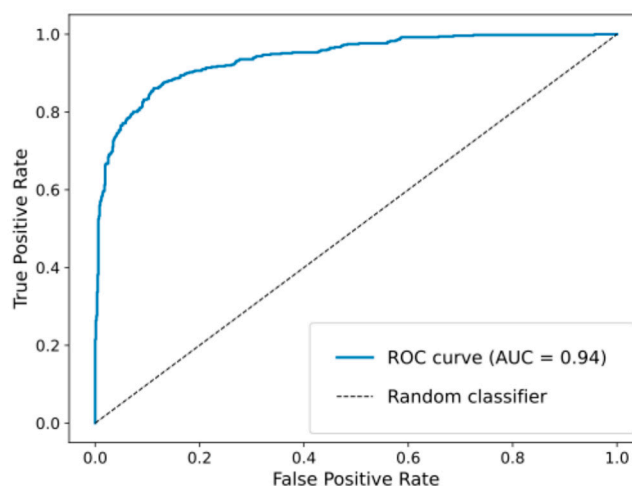


Figure 2. ROC curve of Naïve Bayes classifier to identify PD-relevant posts.

3.3. PD-Relevant Posts

The Naïve Bayes classifier was applied to detect PD-related content. On average, participants made 104 ± 176 PD-related posts. 96% of participants had at least one post flagged as PD-relevant. Additionally, 67% of all participants explicitly referenced PD at least once.

Among individuals with PD ($N = 29$), each participant made an average of 99 ± 172 PD-related posts. Of these, an average of 24 ± 58 posts occurred before diagnosis, and 75 ± 156 occurred after diagnosis. 69% of individuals with PD explicitly mentioned PD at least once, and 93% had at least one post flagged as PD-relevant (see Figure 1C).

Exercise is an important activity for maintaining quality of life in PD [8], though it may be less relevant to detect prodromal symptoms. Since exercise-related terms were included in the term dictionary, we conducted a sensitivity analysis excluding exercise-related posts. Excluding exercise-related posts, the average number of PD-related posts per person before diagnosis decreased from 24 ± 58 to 15 ± 34 , suggesting that some pre-diagnosis PD-related content was driven by exercise but not entirely. After diagnosis, the decrease was smaller (75 to 49), suggesting that exercise-related content accounted for a substantial portion of PD-related discourse but was not solely responsible for the observed differences.

4. Discussion

4.1. Principal Findings

This study establishes Facebook as a feasible and informative data source for PD research, demonstrating that individuals with PD maintain long-term social media histories spanning both pre- and post-diagnosis periods. PD-related content surfaced before formal diagnosis, suggesting that early symptoms and concerns may emerge online. While some pre-diagnosis posts referenced exercise, a non-specific aspect of PD management [8], excluding these did not eliminate PD-related content, reinforcing that prodromal disclosures extend beyond fitness discussions.

Beyond individuals with PD, caregivers and individuals with other movement disorders also contributed to PD-related discussions, reinforcing that insights can be gleaned from a broader community. The high prevalence of caregiver engagement aligns with findings from Chu and Jang [12], who reported that caregivers play a central role in online PD discussions. This finding suggests that social media reflects not only patient experiences but also the perspectives of those supporting them, which may be valuable for understanding disease impact beyond clinical reports.

While keyword-matching and traditional machine learning approaches likely underestimated true PD-related discourse, the study demonstrates the potential of social media data for identifying

early disease signals. Future work should integrate context-aware pipelines, including large language models (LLMs), to enhance detection and interpretation.

4.2. Ethical Considerations

This study highlights the ethical advantages of our Facebook-based approach, where participants actively consented to data sharing rather than having their posts passively scraped via an API. Social media research on platforms like Reddit typically involves extracting user-generated content without direct user awareness, even if permitted by the platform's terms and conditions. Following the Cambridge Analytica scandal, Facebook restricted API access [21], leading us to collect data directly from account owners instead. While this method enhances ethical integrity, it limits scalability and restricts datasets to unidirectional conversations, as only account-owner content is retrievable. Future work should explore ways to balance participant agency with efficient data collection.

4.3. Limitations

Digital literacy and internet access vary widely by demographic [22], influencing the generalizability of social media-based health research. Our sample was predominantly white and college-educated, with Facebook users significantly younger than non-users. While PD severity and duration did not differ between Facebook users and non-users, this finding may reflect selection bias, as individuals with milder disease may be more likely to participate in research. Additionally, the use of keyword-matching and traditional machine learning methods likely led to under detection of PD-related content. Future studies should investigate how digital behaviors shape participant engagement and consider approaches to enhance sample diversity and detection accuracy.

5. Conclusions

This study establishes Facebook as a feasible and ethically sound data source for PD research, demonstrating that individuals share PD-related information both before and after diagnosis. These findings highlight social media's potential for disease monitoring and early detection. Further refinement of computational methods and integration with clinical data could enhance the utility of social media-derived insights.

Author Contributions: Conceptualization, J.M.P., A.S., and J.L.M.; methodology, J.M.P., A.S., and J.L.M.; software, J.M.P.; validation, J.M.P, C.C. and K.M.; formal analysis, J.M.P., A.S., and J.L.M.; investigation, J.M.P, C.C., K.M, S.L.; resources, A.S., J.L.M.; data curation, J.M.P., A.S., and J.L.M.; writing—original draft preparation, J.M.P., A.S., and J.L.M.; writing—review and editing, J.M.P., C.C., K.M., S.L., A.S., and J.L.M.; visualization, J.M.P, J.L.M; supervision, J.M.P, A.S., J.L.M; project administration, J.M.P; funding acquisition, J.M.P, A.S., J.L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a Professional Development Support Funds Competitive Research Grant from Emory University's Laney Graduate School.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Emory University (STUDY00005722; initial approval 07/28/2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Code supporting the project, including text extraction from Facebook exports, keyword flagging, classifier development, application, and analyses, is available at https://github.com/jeannempowell/JCM_pd-on-facebook. However, the dataset and trained models cannot be shared due to the risk of retaining identifiable information, making full de-identification infeasible.

Acknowledgments: We extend our heartfelt gratitude to our participants who generously donated their time and data for this research project. Your contributions have been invaluable, and your willingness to share your personal experiences is deeply appreciated.

During the preparation of this manuscript/study, the authors used ChatGPT4o for the purposes of copyediting. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

API	Application Programming Interface
AP	Atypical Parkinsonism
AUC	Area Under the Curve
CI	Confidence Interval
ET	Essential Tremor
HIPAA	Health Insurance Portability and Accountability Act
JSON	JavaScript Object Notation
KNN	K-Nearest Neighbors
LLMs	Large Language Models
MDS-UPDRS	Movement Disorder Society-Unified Parkinson's Disease Rating Scale
PD	Parkinson's disease
PwPD	People with Parkinson's disease
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine

Appendix A

{"constipation": ["poor bowel habits", "digestive blockage", "need more fiber", "bowel issues", "blocked up", "stool inconsistency", "constipation", "stomach cramps", "bowel discomfort", "irregularity", "intestinal sluggishness", "laxative", "cramping", "irregular bowel movements", "digestive upset", "digestive issues", "digestive slowdown", "gut health concerns", "gastrointestinal issues", "bowel trouble", "bowel difficulty", "stool", "stomach pain", "digestive problems", "abdominal bloating", "infrequent defecation", "straining to defecate", "bowel obstruction", "irregular bowels", "bowel", "poop", "colon concerns", "infrequent bowel movements", "constipated", "bowel movement trouble", "hard stools", "slow bowel transit", "stool softener", "laxative use", "infrequent stools", "intestinal discomfort", "difficulty passing stool", "digestive distress", "lack of fiber", "stool problems", "slow digestion", "stool issues", "fecal impaction", "struggle in the bathroom"],

"REM sleep behavior disorder": ["violent dreaming", "bed partner disturbance", "hitting during sleep", "restless nighttime behavior", "sleep aggression", "nighttime agitation", "sleep disruptions", "violent dreams", "dream reenactment", "unusual sleep actions", "vocalization in sleep", "sleep behavior disorder", "punching in sleep", "nighttime episodes", "sleep walking", "physical dream expression", "REM sleep behavior disorder", "dream enactment", "sleepwalking", "sleep movement", "sleep disturbance", "physical activity during sleep", "acting out while asleep", "violent sleep behavior", "restless sleep", "dream acting", "acting out dreams", "sleep-related violence", "RBD", "night terrors", "sleep interruptions", "thrashing in bed", "vivid dreaming", "dream aggression", "sleep activity", "disruptive sleep behavior", "active dreaming", "talking in sleep", "nocturnal activity", "sleep talking", "kicking in sleep", "abnormal sleep behavior", "aggressive nocturnal behavior", "nighttime restlessness", "agitated sleep"],

"Hyposmia": ["smell impairment", "reduced smell", "Hyposmia", "smelling issues", "weak sense of smell", "loss of smell", "can't smell", "faint odors"],

"Asymmetric vague shoulder pain": ["shoulder pain", "unequal shoulder pain", "unexplained shoulder soreness", "asymmetrical shoulder pain", "shoulder tenderness", "intermittent shoulder pain", "mild shoulder ache", "one-sided shoulder discomfort", "vague pain in shoulder", "shoulder pain without cause", "shoulder pain without injury", "shoulder pain on one side", "shoulder soreness", "shoulder pain without explanation", "uneven shoulder ache", "Asymmetric vague shoulder pain", "one-sided shoulder pain", "one shoulder pain", "shoulder discomfort", "occasional shoulder ache", "shoulder stiffness", "random shoulder pain", "shoulder pain that comes and goes", "irregular shoulder pain", "shoulder hurts", "shoulder ache"],

"Depression": ["feeling down", "low mood", "depressed", "hopeless", "lack of interest", "persistent sadness", "melancholy", "mental", "emotional numbness", "feeling empty", "loss of interest", "Depression", "feeling worthless", "hopelessness", "sad", "low spirits", "lack of pleasure", "chronic unhappiness", "feeling blue", "funk", "sadness"],

"Impaired color vision": ["can't see colors", "color vision deficiency", "color blindness", "dull colors", "Impaired color vision", "colors seem faded", "trouble with colors"],

"Erectile dysfunction": ["ED", "trouble with erection", "Erectile dysfunction", "sexual dysfunction", "intimacy issues", "can't get hard", "sexual issues", "impotence"],

"Reduced arm swing": ["less arm movement", "one arm swing", "arm stiffness", "arm rigidity", "arm doesn't swing", "Reduced arm swing", "stiff arm walk", "arm swing imbalance"],

"Increased stride time variability": ["change in walk pattern", "uneven walking", "gait variability", "irregular stride", "stride irregularities", "walking issues", "Increased stride time variability"],

"Urinary dysfunction": ["urination problems", "urine", "frequent urination", "bladder control problems", "urinary issues", "bladder dysfunction", "leaking urine", "Bladder problems", "Urinary dysfunction", "incontinence", "urinary incontinence", "bladder control issues", "bladder issues", "pee", "peeing issues"],

"Pain": ["aching joints", "tender pain", "radiating pain", "debilitating pain", "excruciating pain", "muscular discomfort", "visceral pain", "pain", "unrelenting pain", "tenderness", "throbbing pain", "nerve pain", "cramping", "ache", "back pain", "twinge of pain", "soreness", "severe pain", "dull ache", "pain management", "pain treatment", "hurt", "pain flare-up", "nagging pain", "painful swelling", "arthritis pain", "subacute pain", "body aches", "mild pain", "pain therapy", "piercing pain", "chronic pain", "pain relief", "discomfort", "burning sensation", "widespread pain", "neck pain", "constant pain", "moderate pain", "painful sensation", "pain control", "acute pain", "unbearable pain", "pain episodes", "neuropathic pain", "migraine", "sharp pain", "inflammation pain", "stabbing pain", "headache", "breakthrough pain", "painful symptoms", "shooting pain", "muscle pain", "post-surgical pain", "stiffness", "numbness and pain", "spasms", "joint pain"],

"Insomnia": ["wide awake", "restless sleep", "insomnia", "can't sleep", "trouble sleeping", "nighttime awakenings", "frequent waking", "sleepless", "Insomnia", "Sleep Disturbances", "sleep problems", "sleep disorder", "sleep disruption", "sleep issues", "wakeful nights", "sleep"],

"Anxiety": ["nervousness", "stress", "anxious", "Anxiety", "nervous", "overthinking", "restlessness", "anxious feelings", "worrying", "tense"],

"Cognitive impairment": ["forgetfulness", "memory loss", "confusion", "Cognitive impairment", "cognitive decline", "difficulty concentrating", "mental fog", "brain fog"],

"Fatigue": ["weary", "exhaustion", "knocked out", "tiredness", "Fatigue", "cranky", "sleepy", "lack of energy", "low energy", "bed", "exhausted", "worn out"],

"Bradykinesia": ["slow to move", "reduced speed of movement", "Bradykinesia", "delayed movements", "sluggish motion", "movement slowness", "slow movement"],

"Rigidity": ["Rigidity", "muscular rigidity", "rigid muscles", "hard to move", "stiff muscles", "inflexible joints", "muscle stiffness", "tight muscles"],

"Tremors": ["shaky movements", "involuntary shaking", "Tremors", "trembling hands", "hand tremble", "body shakes", "shaking", "tremor"],

"bulbar_dysfunction": ["bulbar dysfunction"],

"Dysarthria": ["speech disorder", "strained voice", "slow speech", "talking difficulties", "hard to verbalize", "voice changes", "weak voice", "speaking fatigue", "stammering", "speech impairment", "unclear speech", "hard to understand", "trouble pronouncing", "difficulty articulating", "struggling to talk", "choppy speech", "slurred speech", "mumbling", "garbled speech", "incoherent speech", "difficulty speaking", "stuttering", "altered speech", "speech problem", "hard to speak", "speech difficulties", "slurring words", "nasal speech", "speaking problems", "jumbled speech", "speech changes"],

"Dysphagia": ["can't swallow", "issues with swallowing", "difficulty eating", "choking on food", "chewing problems", "painful swallowing", "throat discomfort", "food getting stuck", "swallowing discomfort", "hard to swallow", "swallowing trouble", "food sticking in throat", "difficulty chewing", "eating difficulties", "trouble eating", "food aspiration", "pain when swallowing", "swallowing problems", "hard to eat", "fear of choking", "coughing when eating", "gagging when eating", "hard swallowing", "sore throat while eating", "swallowing pain", "trouble swallowing", "difficulty swallowing", "swallowing disorder", "feeling of food stuck", "swallowing difficulty"],

"On-Off Periods": ["motor fluctuations", "on periods", "off periods", "medication not working", "variable symptom control", "early morning off", "periods of poor mobility", "drug-induced motor complications", "unpredictable symptom control", "medication off time", "inconsistent medication effect", "wearing-off phenomenon", "fluctuating response", "symptom fluctuations", "sudden OFF", "dose failure", "medication on time", "peak dose dyskinesia", "medication wearing off", "dyskinesia", "motor symptom variability", "increased symptoms", "medication response fluctuations", "medication cycle fluctuations", "good ON time", "deteriorating medication effect", "levodopa wearing off", "medication-related mobility changes", "medication-related motor issues", "erratic symptom relief", "levodopa-induced dyskinesia", "end-of-dose wearing off", "dose wearing off", "delayed ON", "bad OFF time"],

"Postural instability": ["staggering", "balance issues", "leaning", "difficulty standing upright", "wobbling", "swaying", "unsteady", "balance problems", "unsteady standing", "balance difficulties", "poor balance", "dizziness", "loss of balance", "instability standing", "standing issues", "falling easily", "balance", "imbalance", "difficulty standing", "Postural instability"],

"Falls": ["ground", "falls", "help", "serious fall", "slipped", "hurt myself falling", "slipping", "tumbled", "frequent falling", "stumble", "dropped", "stumbled", "sudden fall", "falling", "railing", "plummeted", "caught myself falling", "lost balance", "near fall", "walking", "unexpected fall", "hit the ground", "tumble down", "staggered", "knocked down", "almost fell", "shuffle", "rolled over", "collapse", "pitched forward", "accident", "leaning", "shower", "bruise", "care", "fell", "unsteady", "hurt", "lost my balance", "crashed", "wiped out", "floor", "toppled", "falling down", "took a spill", "imbalance", "seat", "surface", "catch", "stairs", "collapsed", "slip", "recovery", "grabbing", "tripping", "fall scare", "stumbling", "dizziness", "loss of balance", "terrain", "trip", "injury", "slid", "caught", "face plant", "fallen", "toppled over"],

"Gait difficulties": ["trouble walking", "walking instability", "slow gait", "gait freezing", "gait asymmetry", "unpredictability of my legs", "gait disturbance", "walking difficulty", "festinating gait", "gait difficulties", "shuffling gait", "dragging feet", "abnormal walk", "irregular gait", "walking issues", "freezing of gait", "mobility issues", "uneven gait", "shuffling", "walking impairment", "shuffling steps", "unsteady walk"],

"Assistive Device Use": ["mobility scooter", "use of cane", "supportive devices", "walking stick", "adaptive equipment", "handrails", "scooter", "braces", "orthotic devices", "rollator", "assistive walking devices", "mobility aids", "grab bars", "wheelchair use", "assistive devices", "using walker", "adaptive chair", "walking aids"],

"Freezing of gait": ["immobilized", "sudden stop", "stuck in place", "freezing episode", "feet glued", "gait freeze", "Freezing of gait", "freezing", "can't lift feet", "can't move", "start hesitation", "can't move feet", "movement hesitation", "walking freeze", "sudden stop walking", "frozen gait", "legs won't move", "can't step", "temporary paralysis", "motor block", "momentary freeze"],

"Dyskinesia": ["muscle twitching", "hyperkinesia", "motor restlessness", "unintended muscle movements", "drug-induced movement disorder", "involuntary movements", "fidgeting", "athetosis", "abnormal posturing", "muscular jerks", "spontaneous movements", "unpredictable movements", "levodopa-induced dyskinesia", "uncontrolled movements", "writhing movements", "restless movements", "muscle rigidity", "jerky movements", "chorea", "dyskinesia", "jerking movements", "fluctuating movements"],

"Dystonia": ["muscular tension", "muscular spasms", "abnormal postures", "abnormal muscle tone", "neck spasms", "sustained muscle contractions", "twisting movements", "focal dystonia", "generalized dystonia", "muscle stiffness", "writer's cramp", "involuntary muscular contractions", "attack", "legs wouldn't work", "muscle attack", "abnormal body positions", "task-specific dystonia", "sustained postures", "repetitive movements", "torticollis", "twisting postures", "body distortion", "muscle rigidity", "muscle twisting", "dystonia", "muscle attacks", "muscle cramping", "limb dystonia"],

"Related Disorders": ["normal pressure hydrocephalus", "progressive supranuclear palsy", "parkinsonian gait", "tardive dyskinesia", "bradykinesia", "olivopontocerebellar atrophy", "ataxia", "benign essential tremor", "frontotemporal dementia", "vascular parkinsonism", "paraneoplastic syndromes", "restless legs syndrome", "essential tremor", "secondary parkinsonism", "myoclonus", "striatonigral degeneration", "shy-drager syndrome", "akathisia", "drd", "psychogenic movement disorder", "drug-induced parkinsonism", "td", "lewy body dementia", "spinocerebellar ataxia", "parkinson-plus syndrome", "neuroleptic malignant syndrome", "dystonia", "multiple system atrophy", "postural instability", "parkinsonism"],

"Autonomic Dysfunction": ["dysautonomia", "Autonomic Dysfunction", "temperature regulation problems", "irregular heartbeat", "autonomic issues"],

"Orthostatic hypotension": ["light-headedness on standing", "postural hypotension", "Orthostatic hypotension", "sudden dizziness", "dizzy standing", "fainting spells", "low blood pressure", "blood pressure issues"],

"Altered Sweating": ["sweat profusely", "sweating imbalance", "sweating fluctuation", "overactive sweat glands", "drenching sweats", "lack of sweat", "heavy sweating", "profuse perspiration", "sweating disorder", "failed sweat response", "increased sweating", "sweating difficulty", "sweating irregularities", "sweating dysfunction", "clammy skin", "night sweats", "sweating too much", "sweaty palms", "no sweat", "sweating abnormalities", "anhidrosis", "lack of perspiration", "sweat excessively", "abnormal perspiration", "reduced sweating", "non-sweating", "excessive sweating", "sweat attacks", "difficulty sweating", "sudden sweating", "underactive sweat glands", "sweating disturbance", "sweating problem", "heat intolerance", "uncontrolled sweating", "hyperhidrosis", "sweat gland issues", "excessive perspiration", "sweating episodes", "dry skin"],

"Psychosis": ["paranoid delusions", "psychotic symptoms", "unreal perceptions", "schizophrenia-like symptoms", "psychosis", "distorted reality", "visual hallucinations", "psychotic depression", "psychotic break", "psychotic episode", "psychotic behavior", "delusions", "paranoia", "bizarre delusions", "persecutory delusions", "seeing things", "irrational thoughts", "auditory hallucinations", "psychotic disorder", "delusional thinking", "hallucination", "disorganized thinking", "false beliefs", "reality distortion", "paranoid thinking", "hearing voices", "grandiose delusions", "hallucinations"],

"Mental Health": ["mental fatigue", "psychological distress", "stress", "emotional strain", "mental well-being", "mental strain", "anxiety", "psychological well-being", "emotional problems", "mental health support", "mental resilience", "emotional challenges", "mental health", "emotional well-being", "mental toll", "mental health issues", "emotional support", "depression", "mental

health concerns", "psychological issues", "psychological support", "psychological health", "emotional distress", "mental health struggles", "mental health care", "emotional health"],

"Cognitive decline": ["forgetfulness", "mental confusion", "short-term memory loss", "memory decline", "cognitive slowing", "memory deterioration", "cognitive changes", "cognitive dysfunction", "mental fuzziness", "brain fog", "can't remember", "advanced memory issues", "forgetful", "memory loss", "declining memory", "memory lapses", "memory issues", "major forgetfulness", "slow to respond", "cognitive deterioration", "mental decline", "significant cognitive decline", "mental slowness", "brain slowing", "cognitive difficulties", "cognitive struggles", "forgetting things", "thinking issues", "cognitive loss", "cognitive decline in aging", "cognitive decline", "sluggish thought", "cognitive impairment", "disorientation", "dementia", "severe memory loss", "mental fogging", "confused", "mental deterioration", "losing cognition", "delayed cognition", "mental fog", "thinking delay"],

"addiction issues": ["excessive sexual behavior", "addictive tendencies", "compulsive shopping", "alcohol abuse", "substance dependence", "pathological gambling", "excessive gambling", "alcoholism", "compulsive eating", "overeating", "compulsive behavior", "problem gambling", "addictive behavior", "narcotic abuse", "alcohol addiction", "impulse control disorder", "substance abuse", "addiction problems", "binge eating", "drug addiction", "drug abuse", "habitual overeating", "gambling", "chemical dependency", "compulsive", "prescription drug abuse", "compulsive gambling", "sexual compulsivity", "opioid addiction", "drug dependency", "internet addiction", "addictive habits", "addictive personality", "sex addiction"],

"Drooling": ["excessive saliva", "saliva control problems", "drool", "salivating", "drooling issues", "sialorrhea", "dribbling saliva", "mouth drooling", "uncontrolled saliva", "drooling at night", "saliva management", "constant drooling", "spitting", "saliva accumulation", "salivary control", "drooling", "drooling problems"],

"medication": ["Eldepryl", "tolcapone", "pramipexole", "Symmetrel", "medication", "Cogentin", "INBRIJA", "carbidopa", "Gocovri", "drug", "meds", "Xadago", "Comtan", "rasagiline", "selegiline", "impax", "carbidopa", "Zelapar", "mao b inhibitor", "Nourianz", "azilect", "Requip", "Anticholinergics", "Duopa", "Ongentys", "pill", "Sinemet", "apokyn", "entacapone", "Rytary", "ropinirole", "started taking", "baclofen", "levodopa", "Levodopa", "nuplazid", "med", "Dopamine agonists", "benztropine", "l-dopa", "istradefylline", "safinamide", "Tasmar", "rotigotine", "dopamine", "opicapone", "Artane", "new meds", "COMT inhibitors", "Osmolex ER", "Amantadine", "Mao-B inhibitors", "clonazepam", "trihexyphenidyl", "Mirapex", "Neupro", "levadopa"],

"treatment": ["disease stage", "specialist visit", "therapist", "lsvt", "Exercise physiologist", "speech therapy", "speech therapist", "symptom management", "treatments", "occupational therapist", "physician", "rehab", "mph", "speech issues", "therapy", "physical therapist", "treatment side effects", "holistic approaches", "treatment", "neurologist", "dr", "healing", "nursing facility", "neuro", "hospital", "nurse", "physician's assistant", "md", "symptom", "therapy sessions", "doctors", "doc", "dpt", "doctor", "alternative treatment", "neurologist appointment", "treatment options", "physiotherapy", "pt"],

"DBS": ["parkinson's surgery", "neurosurgery", "surgical treatment for Parkinson's", "dbs outcomes", "dbs", "electrical stimulation brain", "dbs procedure", "brain surgery", "surgical options for Parkinson's", "dbs device", "dbs implant", "dbs benefits", "deep brain stimulation", "brain pacemaker", "brain stimulation therapy", "dbs surgery", "neurological surgery", "surgery", "neurosurgical procedure", "dbs therapy", "brain stimulation surgery", "implanting dbs device", "brain stimulation treatment", "dbs treatment", "neurostimulator", "dbs risks"],

"exercise": ["basketball", "elliptical", "muscle building", "move", "fitness class", "bike", "bodybuilding", "tai chi", "swimming", "mountain biking", "water aerobics", "endurance training", "badminton", "running", "rollerblading", "parcour", "parkour", "circuit training", "dance", "tango", "power walking", "pilates", "physical fitness", "jogging", "tennis", "skating", "yoga", "football", "race", "sprinting", "5k", "skateboarding", "group fitness", "boxing", "interval training"],

"kettlebell workout", "cardio workout", "aerobics", "surfing", "skiing", "marathon training", "outdoor activities", "rock climbing", "spinning", "workout", "10k", "soccer", "ride", "walk", "HIIT", "moving", "balance exercises", "sports", "kickboxing", "weight lifting", "cycling", "strength training", "gym", "zumba", "rowing", "aqua aerobics", "volleyball", "crossfit", "barre", "exercise routine", "snowboarding", "spin class", "calisthenics", "functional training", "personal training", "hiking", "boot camp", "trail running", "stretching", "mobility exercises", "exercise", "bicycling"],

"organizations": ["wilkins parkinson", "Parkinson's foundation symposium", "parkinson's foundation", "brian grant foundation", "Fox", "fundraising", "winning round foundation", "Foundation", "Neuro Challenge", "Michael J. Fox", "apda"],

"Quality of Life": ["going natural", "diet changes", "illness impact", "disease milestones", "disability", "driving issues", "disease progression", "quality of life", "daily challenges", "home adjustments", "living with disease", "work adjustments", "day-to-day", "care routine", "new normal", "social isolation", "recent milestone"],

"community": ["patient forum", "community involvement", "community support", "support", "support network", "moving day", "peer support", "support group", "social support"],

"diagnosis": ["diagnostic journey", "Parkinson's diagnosis", "late-onset Parkinson's", "confirming diagnosis", "diagnosis", "yopd", "diagnosed with Parkinson's", "diagnosed", "early-onset Parkinson's"],

"caregiver terms": ["caregiver", "caregiving challenges", "caring for spouse", "caring for parent", "caregiver support", "family caregiver", "spousal caregiver", "caregiver experience", "caregiver journey"],

"research": ["trial", "research study", "clinical study", "research updates", "medical research", "medical trial", "clinical trial", "Parkinson's research", "research breakthroughs"],

"advocacy": ["health advocacy", "raising awareness", "advocacy efforts", "advocating for patients", "community advocacy", "disease awareness", "Parkinson's awareness", "patient advocacy"],

"Other": ["cure", "help", "fight against Parkinson's", "support dog", "overcoming challenges", "living with handicap", "illness", "handicap", "challenge", "adapting to illness", "fight", "helping", "travel concerns", "support animal", "disease", "personal battle"],

"parkinson": ["parkies", "parkinson", "parkie", "pd", "pwp"],

"gene": ["PARK1", "VPS35", "SNCA", "DJ1", "PARK2", "PARK7", "PARK17", "GBA", "lrrk2", "PARK6", "PINK1", "PRKN", "PARK8"]

Appendix B

The optimized hyperparameters for each classifier were as follows:

- Naïve Bayes: $\alpha=0.1$, class_prior=None, fit_prior=True, force_alpha=True;
- Random Forest: 100 estimators, entropy criterion, max_features=sqrt, min_samples_leaf=2, random_state=42;
- XGBoost: binary logistic objective, learning_rate=0.1, max_depth=3, subsample=0.8, colsample_bytree=0.8, n_estimators=100, tree_method=hist, random_state=42;
- Decision Tree: Gini criterion, max_depth=30, min_samples_leaf=2, splitter=best, random_state=42;
- SVM: linear kernel, C=1, probability=True, decision_function_shape=ovr;
- AdaBoost: SAMME algorithm, 100 estimators, learning_rate=1, random_state=42
- KNN: Euclidean metric, n_neighbors=8, weights=uniform;
- A soft-voting ensemble classifier was implemented, combining KNN, SVM, Random Forest, AdaBoost, Naïve Bayes, Decision Tree, and XGBoost models.

References

1. Kelil, T.; Jaswal, S.; Matalon, S.A. Social Media and Global Health: Promise and Pitfalls. *RadioGraphics* **2022**, *42*, E109–E110, doi:10.1148/rg.220038.
2. Hanslo, S. Facebook Business Report. *SSRN Electron. J.* **2024**, doi:10.2139/ssrn.4815047.
3. Gil-Clavel, S.; Zagheni, E. Demographic Differentials in Facebook Usage around the World. *Proc. Int. AAAI Conf. Web Soc. Media* **2019**, *13*, 647–650, doi:10.1609/icwsm.v13i01.3263.
4. Dudina, V.; Judina, D.; Platonov, K. Personal Illness Experience in Russian Social Media: Between Willingness to Share and Stigmatization. In *Proceedings of the Internet Science*; El Yacoubi, S., Bagnoli, F., Pacini, G., Eds.; Springer International Publishing: Cham, 2019; pp. 47–58.
5. Bodnar, T.; Barclay, V.C.; Ram, N.; Tucker, C.S.; Salathé, M. On the Ground Validation of Online Diagnosis with Twitter and Medical Records. In *Proceedings of the Proceedings of the 23rd International Conference on World Wide Web*; April 7 2014; pp. 651–656.
6. Lejeune, A.; Robaglia, B.-M.; Walter, M.; Berrouguet, S.; Lemey, C. Use of Social Media Data to Diagnose and Monitor Psychotic Disorders: Systematic Review. *J. Med. Internet Res.* **2022**, *24*, e36986, doi:10.2196/36986.
7. Dorsey, E.R.; Bloem, B.R. The Parkinson Pandemic—A Call to Action. *JAMA Neurol.* **2018**, *75*, 9, doi:10.1001/jamaneurol.2017.3299.
8. Bloem, B.R.; Okun, M.S.; Klein, C. Parkinson's Disease. *The Lancet* **2021**, *397*, 2284–2303, doi:10.1016/S0140-6736(21)00218-X.
9. Gibb, W.R.; Lees, A.J. Anatomy, Pigmentation, Ventral and Dorsal Subpopulations of the Substantia Nigra, and Differential Cell Death in Parkinson's Disease. *J. Neurol. Neurosurg. Psychiatry* **1991**, *54*, 388–396, doi:10.1136/jnnp.54.5.388.
10. Blonder, L.X. Historical and Cross-Cultural Perspectives on Parkinson's Disease. *J. Complement. Integr. Med.* **2018**, *15*, 1–15, doi:10.1515/jcim-2016-0065.
11. Algarni, M.; Fasano, A. The Overlap between Essential Tremor and Parkinson Disease. *Parkinsonism Relat. Disord.* **2018**, *46*, S101–S104, doi:10.1016/j.parkreldis.2017.07.006.
12. Chu, H.S.; Jang, H.Y. Exploring Unmet Information Needs of People with Parkinson's Disease and Their Families: Focusing on Information Sharing in an Online Patient Community. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2521, doi:10.3390/ijerph19052521.
13. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martinez-Martin, P.; Poewe, W.; Sampaio, C.; Stern, M.B.; Dodel, R.; et al. Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale Presentation and Clinimetric Testing Results. *Mov. Disord.* **2008**, *23*, 2129–2170, doi:10.1002/mds.22340.
14. Harris, P.A.; Taylor, R.; Minor, B.L.; Elliott, V.; Fernandez, M.; O'Neal, L.; McLeod, L.; Delacqua, G.; Delacqua, F.; Kirby, J.; et al. The REDCap Consortium: Building an International Community of Software Platform Partners. *J. Biomed. Inform.* **2019**, *95*, 103208, doi:10.1016/j.jbi.2019.103208.
15. Powell, J.M.; Guo, Y.; Sarker, A.; McKay, J.L. Classification of Fall Types in Parkinson's Disease from Self-Report Data Using Natural Language Processing. In *Artificial Intelligence in Medicine*; Juarez, J.M., Marcos, M., Stiglic, G., Tucker, A., Eds.; Lecture Notes in Computer Science; Springer Nature Switzerland: Cham, 2023; Vol. 13897, pp. 163–172 ISBN 978-3-031-34343-8.
16. Porter, M.F. An Algorithm for Suffix Stripping. *Program* **1980**, *14*, 130–137, doi:10.1108/eb046814.
17. Loper, E.; Bird, S. NLTK: The Natural Language Toolkit 2002.
18. Owoputi, O.; O'Connor, B.; Dyer, C.; Gimpel, K.; Schneider, N.; Smith, N.A. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters.; June 2013; pp. 380–390.

19. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *Mach. Learn. PYTHON*.
20. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46, doi:10.1177/001316446002000104.
21. Mancosu, M.; Vegetti, F. What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data. *Soc. Media Soc.* **2020**, *6*, 2056305120940703, doi:10.1177/2056305120940703.
22. Esper, C.D.; Valdovinos, B.Y.; Schneider, R.B. The Importance of Digital Health Literacy in an Evolving Parkinson's Disease Care System. *J. Park. Dis.* **2024**, 1–9, doi:10.3233/JPD-230229.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.