Beyond Motor Symptoms: Toward a Comprehensive Grading of Parkinson's Disease Severity

Morteza Rahimi*
Knight Foundation School of Computing and Information Sciences, Florida International University
mrahimi@fiu.edu

Zeina Al Masry
SUPMICROTECH, CNRS, institut FEMTO-ST
zeina.al.masry@ens2m.fr

John Michael Templeton
Knight Foundation School of Computing and Information Sciences, Florida International University
jtemplet@fiu.edu

Sandra Schneider
Department of Communicative Sciences and Disorders, Saint Mary's College
sschneider@saintmarys.edu

Christian Poellabauer
Knight Foundation School of Computing and Information Sciences, Florida International University
cpoellab@fiu.edu

ABSTRACT

This study applies machine learning (ML) feature analysis to an array of multi-functional neurocognitive symptoms specific to individuals with Parkinson's Disease (PD). We provide a framework that can assist with modernizing and objectively individualizing the staging of PD. For that purpose, a hybrid feature score technique is proposed to compute a weighted vector for neurocognitive functions. The methodology is based on Principal Component Analysis and Random Forest for feature selection and extraction purposes. The study enrolled 37 participants who completed various tablet-based functional neurocognitive assessments for motor, memory, speech, executive function, and single versus multi-functional tasks. The study concludes that current assessment and staging schemes exhibit a significant bias toward fine-motor functionalities. Thus, the inclusion of other neurocognitive functions is essential for accurately identifying disease stages. This could be achieved through the integration of multiple functions into a unified score or by adopting function-specific staging. By incorporating ML into disease staging, a more comprehensive understanding of neurocognitive disorders can be obtained, revealing novel insights that affect the design and implementation of staging schemes.

CCS CONCEPTS

• **Computing Methodologies** → **Machine Learning**; • **Applied Computing** → *Health informatics*; Bioinformatics.

KEYWORDS

Machine Learning, Parkinson's Disease, Disease Staging, Digital Health, Neurocognitive Disorder, Mobile Device

*Corresponding author

BCB '23, September 3–6, 2023, Houston, TX, USA

ACM ISBN 979-8-4007-0126-9/23/09…$15.00
https://doi.org/10.1145/3584371.3612988

ACM Reference Format:
Morteza Rahimi, Zeina Al Masry, John Michael Templeton, Sandra Schneider, and Christian Poellabauer. 2023. Beyond Motor Symptoms: Toward a Comprehensive Grading of Parkinson's Disease Severity. In *14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '23), September 3–6, 2023, Houston, TX, USA.* ACM, Houston, TX, USA, 6 pages. https://doi.org/10.1145/3584371.3612988

1 INTRODUCTION

Machine learning (ML) techniques have demonstrated superior performance in various healthcare applications, such as disease staging [12] and diagnosis of conditions such as Alzheimer's and Parkinson's [11]. In order to cover a more comprehensive spectrum of disease disorders and symptoms, ML algorithms employ data from patient records, neuroimaging, and sensor-based monitoring. This enables a more exhaustive evaluation, enhanced prognostic capabilities, and personalized treatment strategies, facilitating more efficient treatment planning and disease management. Moreover, personalized models can account for personal discrepancies in disease progression profiles, which provide a more objective and patient-centered approach to disease staging [4].

Parkinson's Disease (PD), the second most prevalent chronic and progressive neurodegenerative disorder, impacts over 10 million individuals worldwide [8]. In the last couple of years, there has been a substantial increase in demand to modernize and objectify the scaling scheme for movement disorders, such as PD. The existing well-known clinical scaling methods, including the MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and the Hoehn and Yahr (H&Y) scale, classify the severity of motor symptoms and impairments according to clinical observation. In this study, our focus was specifically on the H&Y scale explained below [1]:

- **Stage 1**: Symptoms manifest on one side only (unilateral).
- **Stage 2**: Symptoms present bilaterally without balance impairment.
- **Stage 3**: Balance impairment and moderate disease progression.
- **Stage 4**: Severe disability, but still able to walk or stand unassisted.
- **Stage 5**: Wheelchair-bound or bedridden without assistance.

Although the H&Y and MDS-UPDRS scales are globally known and extensively used, their capability to objectively capture important neurocognitive symptoms is limited, as they heavily focus on motor functionalities. This can potentially result in misclassification and difficulties in tracking disease progression over time. As

a result, there is a need to develop a more objective and accurate scaling system for PD that considers other neurocognitive functions of individuals. The integration of ML into disease staging holds great significance, as it allows for a deeper comprehension of neurocognitive disorders. This, in turn, can lead to novel findings that shape the development of scaling systems.

This study was to acknowledge the motor-centric approach of existing clinical staging scales and to emphasize the need of updating current PD staging methods to include objective measures of other neurocognitive functions. To achieve this, we utilized Random Forest (RF) learning to select significant features and Principal Component Analysis (PCA) to assign a weight to selected features. Subsequently, we introduced a hybrid feature vector score to determine a weighted vector score for each neurocognitive function. Further, by initially examining the effects of PD on different functional areas of neurocognition, this research aims to lay a foundation for future investigations into a wider range of neurological disorders.

2 RELATED WORK

Regarding present staging schemes, Zhao et al. [16] studied the H&Y scale to assess the advancement of PD and analyze variables related to H&Y transition times. The authors employed a large PD database to analyze the time to progress through various H&Y stages and examined the connection between progression to the next stage and baseline variables. They concluded that H&Y transition time is a practical measure of disease progression in PD and may be used in clinical studies. However, the H&Y scale is predominantly motor-based, and further assessments and scaling approaches are required to precisely determine the disease's stage. Moreover, Martinez-Martin et al. [5] investigated the MDS-UPDRS clinical staging scheme, which is a more exhaustive assessment of PD severity than the H&Y staging scale. The study aimed to determine cut-off points for disease severity. It found that these values could be determined and proposed cut-off points to categorize individuals with PD as mild, moderate, or severe according to their MDS-UPDRS scores.

Concerning recent digital biomarkers, Wamelen et al. [14] evaluated non-motor fluctuations (NMF) and non-motor rating scales in the PD population. In order to do so, they studied NMF using both the Movement Disorder Society Non-Motor Rating Scale (MDS-NMS) and wearable sensors. The study revealed that NMF prevalence in the PD population advances with disease duration; however, in a different pattern from motor fluctuations. Ellis and Earhart [2] provided an exhaustive review of digital therapeutics as a means of delivering personalized and evidence-based interventions for individuals with PD via remote platforms. The authors emphasized the increasing need for such interventions and highlight practical applications of digital therapeutic platforms targeting both motor and non-motor symptoms of PD, as well as promoting healthy lifestyle behaviors. In addition, they discuss the latest developments in this field and prospective future applications that could revolutionize patient care and personalized self-management.

Furthermore, Martinez et al. [6] investigated the structural properties of the MDS-NMS using factor analysis and clustering. H&Y staging, demographic and clinical data, and the NMF subscale were also used in this study. They concluded that the majority of subscales on the MDS-NMS are unidimensional. Templeton et al. [12] applied ML to classify PD population data, using self-reported metrics and tablet-based assessments. The study identified significant features, including acceleration, accuracy, and timing, and noted differences between perceived and sensor-based functionality. They concluded that ML in digital health systems can enhance our understanding of neurodegenerative diseases and inform the design of digital health technology.

3 METHODOLOGY

Figure 1 provides an overview of the methodology with five main steps (different colors) and seven detailed steps.

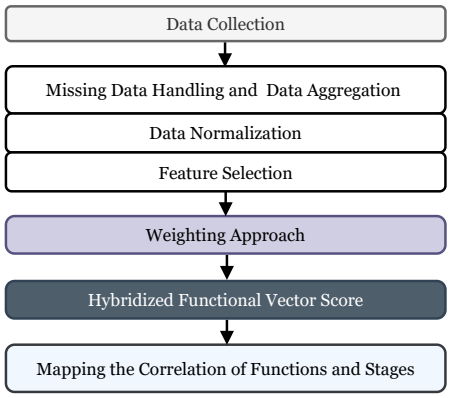


Figure 1: Methodology at a Glance

3.1 Data Collection

A total of 37 individuals, comprising 22 males (59.46%) and 15 females (40.54%) between the ages of 56 and 85, participated in this study. All participants had been diagnosed and staged by a qualified neurologist or specialist in movement disorders. Multiple data samples were collected from each individual, resulting in a total of 74 samples across four stages of the disease. More details regarding the breakdown of the cohort for this research are presented in Table 1. Collected data included various clinical features such as motor and non-motor symptoms, demographic information, as well as cognitive and functional assessments. Participants who didn't understand or speak English were excluded from the study to maintain informed consent. All guidelines from the Institutional Review Board were followed to protect the subjects. All participants of this study were administered a tablet-based neurocognitive assessment designed specifically for individuals with PD [10]. The assessment

Table 1: Cohort Breakdown

Group	Population
Total Individuals with Confirmed PD Diagnosis	37
Number of Confirmed Male PD Cases	22
Number of Confirmed Female PD Cases	15
Total Sample Count from All Individuals	74
Total Number of Samples in Stage 1	28
Total Number of Samples in Stage 2	28
Total Number of Samples in Stage 3/4	18

concentrated on user-device interactions for collecting objective features. A total of 14 neurocognitive functional tests were completed by each participant, encompassing the areas of motor, memory, speech, and executive function. These assessments resulted in 158 digital features pertaining to these areas. The assessment included both single- and multi-functional tests, as individuals with PD exhibit difficulty in performing both task types [13]. Therefore, the analysis of features from both configurations are necessary for understanding how neurocognitive symptoms of PD manifest. All test descriptions are presented in Table 2.

3.2 Data Processing

3.2.1 Data Aggregation and Missing Data Handling. To address this problem, missing data handling and data aggregation techniques, were employed in this study. As the percentage of missing data was moderately low, we opted for the removal of the data method to ensure the completeness of our pilot dataset.

3.2.2 Data Normalization. In order to account for varying scales of data and enhance the comparability of results, the data were normalized. Normalization can help reduce the impact of outliers or extreme values, which may skew the analysis [9]. It scales the data range between -1 and 1 to facilitate comparison and analysis.

3.2.3 Feature Selection. Multiple feature selection methods, including Random Forest, Chi-Squared, Shapley, ANOVA, ReliefF, and Kruskal Wallis were employed to identify significant features. Comparing the outcomes of each technique revealed that RF produced the most meaningful and applicable outcomes. Furthermore, Pearson correlation was used for feature selection in the database. However, the RF method was found to be more accurate and practical for analysis. The RF algorithm's robustness to noisy and correlated features and its ability to handle missing data and outliers are additional benefits that make it a popular choice for feature selection. Thus, the use of RF for feature selection was not only statistically sound but also highly practical and reliable.

3.3 Weighting Approach

PCA is commonly used to extract relevant features and reduce the dimensionality of a dataset. The method works by finding the directions in which the data varies the most, known as the principal components, and representing the data in terms of these components. In this study, PCA was employed to provide the principal components and their corresponding eigenvalues, which can be used to weigh and identify the most important features in our pilot dataset. The first principal component corresponds to the eigenvector with the largest eigenvalue and explains the most variance in the data. Subsequent components explain progressively less variance, with the k-th principal component corresponding to the eigenvector with the k-th largest eigenvalue. By examining the eigenvalues associated with each feature, we were able to assign weights to the most significant features in our dataset and to analyze the mapping between functions and stages.

3.4 Hybrid Functional Vector Score

After selecting the most prominent features using RF and conducting PCA to assign a weight to the selected features, we developed a hybrid RF/PCA score to calculate the weighted sum of each function or sub-function feature for the five neurocognitive functions. RF was chosen for feature selection due to its higher accuracy and practicality, and the eigenvalues produced by PCA were used to create the feature vector score. These scores were then used to generate boxplots for each function, represented in the Results section. To obtain a matrix F containing selected features for a specific cognitive function, this study utilized the RF feature selection approach. Additionally, a vector of eigenvalues $Eigval_{PCA}$ was obtained from PCA. Equation 1 was employed to calculate the weighted sum of each function or sub-function features, denoted as $Fn_{neurocog}$:

$$Fn_{neurocog} = \frac{\sum_{m=1}^{n} F_{:,m} \times Eigval_{PCA}(m)}{\sum_{m=1}^{n} Eigval_{PCA}(m)} \quad (1)$$

Where m iterated through all n features of each neurocognitive function. This equation is used for calculating feature weights based on the importance obtained from RF and the contribution to the overall variance obtained from PCA. The sum of these weighted features is then divided by the sum of all eigenvalues to obtain the final weighted score for each neurocognitive function. $F_{:,m}$ also refers to the features selected by RF. It is important to note that n is a positive integer greater than m, where m could be any number from 1 to $n-1$. Step-by-step details are shown in Algorithm 1.

Algorithm 1 Feature Selection, Weighting Approach, and Hybrid Vector Score

Input: $Metrics_i$, $i = 1, ..., N$, $N = 158$; $Stage_j$, $j = 1, ..., 4$; M: Number of Observations
$d \leftarrow$ load(data)
for $i = 1$ to N **do**
  $F(i) \leftarrow Eigval(d_i)$ `Vector of Eigenvalues Obtained from PCA`
end for
$F' \leftarrow SelectRandomForest(F)$ `Feature Selection`
for $i = 1$ to N **do**
  $Feature\ Vector\ Score(F'_i, RF + Eigenvalue)$ `Functional Vector Calculation`
end for
Weighting Approach
Output: Hybrid Functional Vector Score

4 RESULTS AND DISCUSSION

Figure 2 depicts the ranking of 158 objective digital features across five cognitive functions, i.e. motor, memory, speech, executive, and multi-functional features. The ranking is sorted within each category and ordered by significance. The presented figure indicates that motor-centric features hold the highest significance among the five cognitive functions and multi-functional features, speech, executive function, and memory features follow in descending order.

4.1 Function Specific Results

We proceeded with the following steps to analyze the features of each function. First, using the tests described in Section 3.1 we selected features using the RF ranking algorithm (explained in Section 3.2). These selected features were then weighted using PCA eigenvalues. The purpose of this approach was to assign weights to each feature based on their significance in the model. Finally, these weighted features were applied to Equation 1 to obtain the final results and representations outlined in the following subsections.

Table 2: Each Test and its Description

Test	Description
Fine-motor tracing test	Using the index finger to trace a depicted shape (circle or square).
Gross-motor emulation	"Air-tracing" a depicted shape by manipulating the device in space.
Reaction test	Interacting with targets by tapping on the screen.
Card-matching test	Matching depicted cards by tapping them in pairs until completion.
Trail-making test	Connecting shapes in increasing numerical order using index fingers.
Speech-based test	Participants prompted to read aloud (a sentence and paragraph) and name objects.
Multi-functional test	Combination of a motor and non-automatic speech test.
Executive function/multi-functional	The Stroop Word Color Test requires users to identify color-word matches verbally.
Expanded multi-functional	Participants were instructed to verbally express (through speech), while simultaneously transcribing each word (writing the word being spoken aloud) within the designated area.

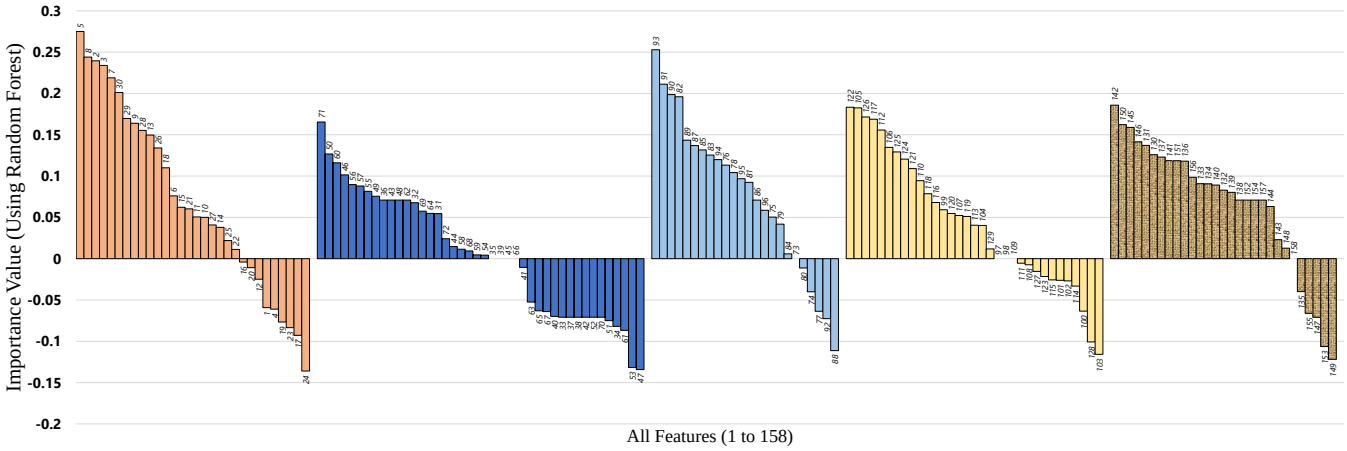
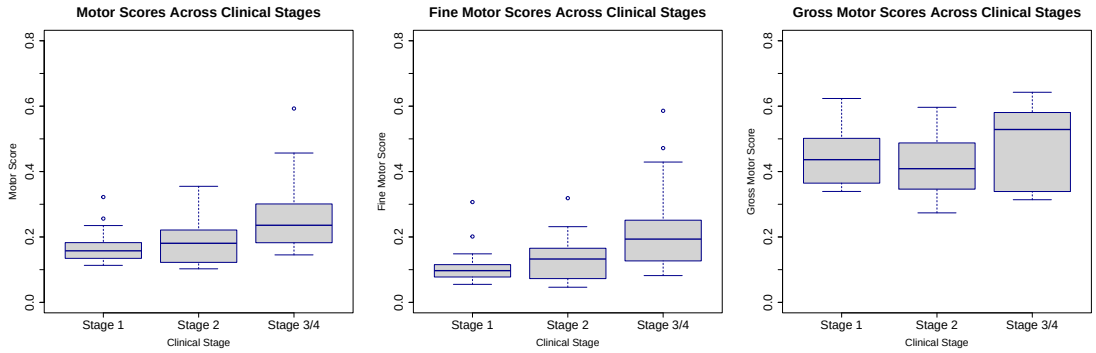


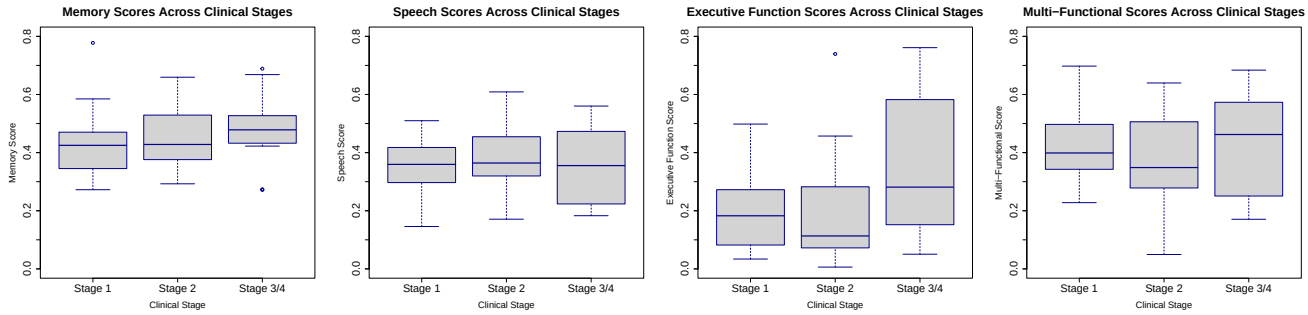
Figure 2: Motor Features (1-30) [Orange] and Non-Motor Features, including Memory (31-72) [Dark Blue], Speech (73-96) [Pale Blue], Executive-Function (97-129) [Yellow], and Multi-Functional Features (97-129, 130-158) [Yellow & Woven Mat Texture]



(a) Motor Function **(b) Fine-Motor** **(c) Gross-Motor**

Figure 3: Motor Function and Its Sub-functions Across Clinical Stages

4.1.1 Motor Function. Given the importance of motor features obtained in the feature selection stage, we selected fourteen fine-motor and six gross-motor features. Fine motor features for both subtasks include time, average distance, total distance drawn, crossings, first point distance to the outline, and first point to the last point drawn distance. Additional fine-motor features for square subtask include average and minimum acceleration. For gross-motor emulation included features are time (circle and square), average acceleration, minimum acceleration, maximum acceleration, and total time. These features were subsequently applied to the weighting approach and Equation 1. Figure 3a presents the comparison of motor



(a) Memory **(b) Speech** **(c) Executive Function** **(d) Multi-Function**

Figure 4: Other Functional Areas of Neurocognition: Memory, Speech, Executive Function, and Multi-Function

function to confirmed clinical stages. As the stage of PD increases, there is a clear trend of decreased motor function overall, where median values for each stage show a step-wise decrease in function. The interquartile range (IQR) also decreases with the increasing stage, indicating that the variability in motor function within each stage becomes smaller. Furthermore, the boxplot shows a greater number of outliers in higher stages (stages 2 and 3/4) compared to stage 1, suggesting that motor impairment becomes more severe as the disease progresses. The importance of the motor function's mapping into the current clinical stages prompted a more in-depth analysis of motor function and sub-functions (e.g., fine and gross).

4.1.2 Fine-Motor. Upon evaluating and comparing results, we perceived that fourteen fine-motor features (mentioned in 4.1.1) created meaningful trends. A depiction of this approach is presented in Figure 3b. As such, it can be observed that as the stage of PD increases, the fine-motor function decreases. The median of the fine-motor tracing function in stage 1 is the highest among the three stages, followed by stage 2 and stage 3/4. This implies that individuals in stage 1 exhibit better fine-motor function compared to those in stage 2 and stage 3/4. Moreover, the IQR of the fine-motor tracing function sharply decreases as the stage of PD increases, indicating less variability of the fine-motor function in individuals with advanced PD. The whiskers of the boxplot also suggest that there are fewer outliers compared to stage 1 and stage 2, further confirming the reduced variability of the fine-motor function in advanced PD. Hence, these findings indicate that there is a substantial association between the stage of PD and fine-motor function, with decreasing fine-motor function as the stage of PD progresses.

4.1.3 Gross-Motor. When it comes to gross-motor, we have observed that incorporating the top six gross-motor features (mentioned in 4.1.1) yielded more meaningful patterns. Figure 3c shows that gross-motor function decreases as the stage of PD increases. The boxplot shows a trend of decreasing median values from stage 1 to stage 3/4, demonstrating that individuals with more advanced stages of PD tend to have poorer gross-motor function. Additionally, IQR appears to widen as the stage increases, suggesting that there is more variability in the advanced stages of PD. The boxplot also shows a number of outliers in each stage, indicating that some individuals with a given stage of PD may have better or worse gross-motor function than expected based on their stage alone. However, the majority of individuals appear to fall within the expected range of gross-motor function for their respective stages.

4.1.4 Memory. For the memory function, the top 4 features selected by RF (including the average time of matched pair, minimum acceleration, card interaction, and minimum acceleration) were applied to Equation 1, resulting in Figure 4a. It indicates a decreasing trend in function scores as the disease progresses from stage 1 to stage 3/4. The median memory function score for stage 1 is around 0.44, while for stage 3/4, it is around 0.5. This highlights the importance of assessing memory, as memory impairment may significantly impact an individual's quality of life. However, it is worth noting that the correlation between memory and current clinical stages of PD is not as robust as the correlation between combined motor and fine-motor functions.

4.1.5 Speech. After doing comparative trials and analysis, it was found that the incorporation of the top four significant features produced the most meaningful, significant trend. These features are the average time for a correct response, total correct, total generated, and number of missed words. The result of the hybrid approach for speech function is demonstrated in Figure 4b. It shows a trend of very moderate decreasing function scores as the disease progresses from stage 1 to stage 2 and very slightly to stage 3/4. The median speech function score for stage 1 is around 0.36 for stage 2 is around 0.37, while for stage 3/4, it is around 0.36. This indicates a slight decline in speech function as the disease progresses, suggesting that the correlation between speech function and clinical stages is not as significant as that of motor function.

4.1.6 Executive Function. Through comparative analysis and evaluation, it was observed that incorporating the top four RF significant features yielded the most notable trend for executive function. These top 4 features are total time, total time acceleration, total correct, and maximum response time. Subsequently, selected features were applied in Equation 1, resulting in Figure 4c. The figure illustrates a declining trend in function scores as the disease progresses, except for stage 2. The median score for stage 1 is approximately 0.18, whereas, for stage 3/4, it is around 2.8. This signifies a significant impairment in executive function during the advanced stage 3/4 of PD. However, the results deviate for stage 2, highlighting the bias of the current clinical staging system towards motor skills,

specifically fine-motor functionalities, and its inability to depict a clear trend in executive functions.

4.1.7 Multi-Function. Finally, it was noted that including the top four RF selected features generated the most meaningful trend pertaining to multi-functional tasks. These four features are average acceleration, total elapsed writing time, average distance, and total distance. Subsequently, these particular features were utilized in Equation 1, resulting in Figure 4d. It displays a similar trend of decreasing function scores as the disease progresses, with a median score of around 0.40 for stage 1 and around 0.46 for stage 3/4. This suggests a slight, fluctuating decline with the progression of PD, indicating the current clinical scheme's inability to provide a clear pattern for the multi-functional abilities of individuals.

4.2 Discussion

Figures 4a – 4d demonstrate that the current clinical staging of PD fails to accurately align with the diverse functions of memory, speech, executive function, and multi-function. Also, Figures 3a and 3b indicate that the current PD scale is biased towards motor functionalities, especially fine-motor. These findings underscore the importance of incorporating other functions beyond motor abilities into PD staging. In line with this, van Wamelen et al. [15] discussed the digital health technology for non-motor symptoms in PD while also highlighting the ongoing digital revolution that sought to objectively measure motor aspects. Thus, a major focus is to now expand this objective assessment toward the "hidden" spectrum of non-motor symptoms. While mobile sensors are increasingly utilized in PD management, their use in addressing non-motor symptoms has been inadequately explored [7]. UPDRS effectively covers motor symptoms, but it lacks detailed scales for non-motor aspects, underscoring the need for incorporating other non-motor features [3]. These findings align with our results and affirm that a comprehensive PD assessment necessitates the inclusion of both motor and non-motor symptoms.

5 CONCLUSION AND FUTURE WORK

This study aimed to address the limitations of the current PD staging systems and highlight the need for updated methods that consider a broader range of neurocognitive functions. To this end, ML feature analysis was applied to various neurocognitive symptoms of the PD population. Specifically, the RF ranking algorithm was used to analyze and assess the significance of features in each function. Moreover, PCA was used to assign a weight to the selected features. we proposed a hybrid feature vector score, which was employed to calculate a weighted vector score for each of the five neurocognitive functions. The results demonstrated a substantial bias on fine-motor abilities in the existing staging system, highlighting the necessity of modernization and inclusion of other functions. The expansion of analysis to include additional sub-functional areas of current neurocognitive areas (e.g., working, skill, and long-term memory) and the development of the proposed approach to other neurodegenerative conditions (e.g., ALS or Alzheimer's disease) would be a perfect path for future work.

REFERENCES

[1] Roongroj Bhidayasiri and Daniel Tarsy. 2012. *Parkinson's Disease: Hoehn and Yahr Scale.* Vol. 36. Humana Press, Totowa, NJ, 4–5. https://doi.org/10.1007/978-1-60327-426-5_2

[2] Terry D. Ellis and Gammon M. Earhart. 2021. Digital Therapeutics in Parkinson's Disease: Practical Applications and Future Potential. *Journal of Parkinson's Disease* 11, s1 (2021), S95–S101. https://doi.org/10.3233/JPD-202407

[3] Christopher C. Goetz. 2003. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders* 18, 7 (2003), 738–750. https://doi.org/10.1002/mds.10473

[4] Heng Huang. 2019. Large-Scale Machine Learning Algorithms for Biomedical Data Science. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '19).* Association for Computing Machinery, New York, NY, USA, 4–4. https://doi.org/10.1145/3307339.3342130

[5] Pablo Martínez-Martín, Carmen Rodríguez-Blázquez, Mario Alvarez, Tomoko Arakaki, Víctor Campos Arillo, Pedro Chaná, William Fernández, Nélida Garretto, Juan Carlos Martínez-Castrillo, Mayela Rodríguez-Violante, Marcos Serrano-Dueñas, Diego Ballesteros, Jose Manuel Rojo-Abuin, Kallol Ray Chaudhuri, and Marcelo Merello. 2015. Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. *Parkinsonism and Related Disorders* 21, 1 (2015), 50–54. https://doi.org/10.1016/j.parkreldis.2014.10.026

[6] Pablo Martinez-Martin, Jose Manuel Rojo-Abuín, Daniel Weintraub, Kallol Ray Chaudhuri, Carmen Rodriguez-Blázquez, Alexandra Rizos, and Anette Schrag. 2020. Factor Analysis and Clustering of the Movement Disorder Society–Non-Motor Rating Scale. *Movement Disorders* 35, 6 (2020), 969–975. https://doi.org/10.1002/mds.28002

[7] Huma Mughal, Abdul Rehman Javed, Muhammad Rizwan, Ahmad S. Almadhor, and Natalia Kryvinska. 2022. Parkinson's Disease Management via Wearable Sensors: A Systematic Review. *IEEE Access* 10 (2022), 35219–35237. https://doi.org/10.1109/ACCESS.2022.3162844

[8] Rana Zia U.R. Rehman, Christopher Buckley, Maria Encarna Micó-Amigo, Cameron Kirk, Michael Dunne-Willows, Claudia Mazzà, Jian Qing Shi, Lisa Alcock, Lynn Rochester, and Silvia Del Din. 2020. Accelerometry-based digital gait characteristics for classification of Parkinson's disease: What counts? *IEEE Open Journal of Engineering in Medicine and Biology* 1 (2020), 65–73. https://doi.org/10.1109/OJEMB.2020.2966295

[9] Dalwinder Singh and Birmohan Singh. 2020. Investigating the impact of data normalization on classification performance. *Applied Soft Computing* 97 (2020), 105524. https://doi.org/10.1016/j.asoc.2019.105524

[10] John Michael Templeton, Christian Poellabauer, and Sandra Schneider. 2021. Design of a Mobile-Based Neurological Assessment Tool for Aging Populations. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST* 362 LNICST (2021), 166–185. https://doi.org/10.1007/978-3-030-70569-5_11

[11] John Michael Templeton, Christian Poellabauer, and Sandra Schneider. 2021. Design of a neurocognitive digital health system (NDHS) for neurodegenerative diseases. In *DigiBiom 2021 - Proceedings of the 2021 Future of Digital Biomarkers (DigiBiom '21).* Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/3469266.3471157

[12] John Michael Templeton, Christian Poellabauer, and Sandra Schneider. 2022. Classification of Parkinson's disease and its stages using machine learning. *Scientific Reports* 12, 1 (aug 2022), 1–11. https://doi.org/10.1038/s41598-022-18015-z

[13] John Michael Templeton, Christian Poellabauer, and Sandra Schneider. 2022. The Case for Symptom-Specific Neurological Digital Biomarkers. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST* 440 LNICST (2022), 235–255. https://doi.org/10.1007/978-3-031-06368-8_16

[14] Daniel Johannes van Wamelen, Silvia Rota, Anette Schrag, Alexandra Rizos, Pablo Martinez-Martin, Daniel Weintraub, and Kallol Ray Chaudhuri. 2022. Characterization of Non-Motor Fluctuations Using the Movement Disorder Society Non-Motor Rating Scale. *Movement Disorders Clinical Practice* 9, 7 (2022), 932–940. https://doi.org/10.1002/mdc3.13520

[15] Daniel J. van Wamelen, Jirada Sringean, Dhaval Trivedi, Camille B. Carroll, Anette E. Schrag, Per Odin, Angelo Antonini, Bastiaan R. Bloem, Roongroj Bhidayasiri, and K. Ray Chaudhuri. 2021. Digital health technology for non-motor symptoms in people with Parkinson's disease: Futile or future? *Parkinsonism and Related Disorders* 89 (2021), 186–194. https://doi.org/10.1016/j.parkreldis.2021.07.032

[16] Ying Jiao Zhao, Hwee Lin Wee, Yiong Huak Chan, Soo Hoon Seah, Wing Lok Au, Puay Ngoh Lau, Emmanuel Camara Pica, Shu Chuen Li, Nan Luo, and Louis C.S. Tan. 2010. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. *Movement Disorders* 25, 6 (2010), 710–716. https://doi.org/10.1002/mds.22875