Freezing of Gait Prognostication in Parkinson's Disease

Disha Parmar[1], Shivam Zala[1] and Madhu Shukla[1][0000-0002-8023-7854]

[1] Department of CE – AI & BDA, Marwadi University, Rajkot, 360003, Gujarat, India

lncs@springer.com

Abstract. Parkinson's disease (PD) affects millions worldwide, and a significant portion experience freezing of gait (FOG), a disabling symptom that impedes mobility and increases fall risk. Despite extensive research, the underlying mechanisms of FOG and effective treatment strategies remain elusive. Objective and precise FOG detection and classification are crucial for advancing our understanding and management of this symptom. Existing FOG detection methods face limitations in accuracy, generalizability, and the ability to distinguish between FOG subtypes. Moreover, current treatment options for FOG are limited and often provide suboptimal outcomes. To address these challenges, we propose a novel approach utilizing machine learning and wearable sensor data to accurately detect and classify FOG episodes. We employ a comprehensive dataset comprising 3D accelerometer data from the lower back of FOG subjects. Using advanced machine learning models, we aim to identify the onset and cessation of FOG episodes and classify them into three distinct types: Start Hesitation, Turn, and Walking. This approach holds the potential to overcome the limitations of existing methods and provide a more comprehensive understanding of FOG. Our research aims to shed light on the intricate mechanisms of FOG, paving the way for the development of more effective treatments and improved quality of life for individuals living with Parkinson's disease.

Keywords: Parkinson's disease, Freezing of Gait, Freezing of Gait prediction, Machine Learning.

1 Introduction

In the context of our comprehensive analysis project—a detailed case study—we focus on Parkinson's disease, a complex neurological condition with significant motor and non-motor symptoms such as freezing of gait and cognitive challenges. Leveraging advanced Python libraries and data analysis tools including phik for uncovering data dependencies, seaborn for informative data visualizations, and lightgbm for powerful predictive modeling, our aim is to dissect the multifaceted factors influencing Parkinson's progression and management. This project seeks to contribute to the understanding and treatment of Parkinson's, ultimately enhancing the quality of life for those affected by this condition. Parkinson's disease is a progressive neurological condition.[7]

2 Aim and Objective

2.1 Aim

This research paper aims to investigate the multifaceted factors impacting individuals with Parkinson's disease, with a focus on less recognized influences. By shedding light on these hidden factors, we seek to enhance our understanding of Parkinson's and contribute to improved care and quality of life for patients and caregivers.

2.2 Objective

The objective of this study is to analyze data from existing Freezing of Gait (FoG) prognostication systems, with a focus on understanding their limitations. Through this data analysis, we aim to bridge the gap between current solutions and a future where we can offer more accurate, non-invasive, and accessible methods for predicting FoG, ultimately enhancing the care and well-being of individuals with Parkinson's disease.

3 Problem Specification

The objective is to identify the start and stop of FOG episodes by detecting the occurrence of three types of FOG events: start hesitation (StartHesitation), turning (Turn), and walking (Walking). For this purpose we use, lower-back 3D accelerometer data from subjects exhibiting FOG episodes. [4]

3.1 Data Description

Three datasets collected in different settings are available for model training:

1) The tDCS FOG (tdcsfog) dataset, collected in the lab, as participants completed a FOG provoking protocol
2) The DeFOG (defog) dataset, collected in the participant’s home, as subjects completed a FOG-provoking protocol
3) The Daily Living (daily) dataset, collected through one week of continuous 24/7 recordings .

The tdcsfog and defog datasets were annotated by expert reviewers that watched videos of the trials and documented the FOG events. Series in the daily dataset were not annotated and it was not used for the development of the presented solution. Each dataset contained three variables related to the acceleration on three axes: V - vertical, ML - mediolateral, AP - anteroposterior. The used sensor data was measured in units of (m/s^2) for tdcsfog data and g (9.81 m/s^2) for defog data. Additionally, the tdcsfog dataset was recorded at 128 Hz, while the defog dataset was recorded using a 100 Hz time resolution.

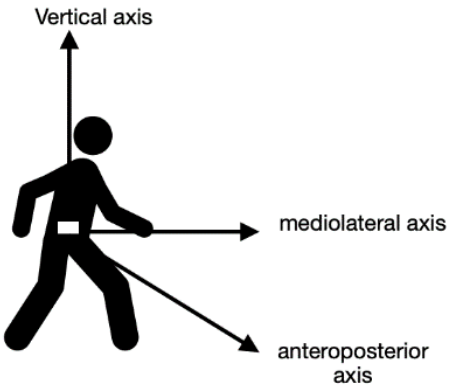

Fig. 1. Vertical, Mediolateral, Anteroposterior Axes

3.2 Literature Review

1. Several studies have explored the development of FoG prognostication systems utilizing various sensors and technologies, such as wearable accelerometers, gyroscopes, and even smartphone applications. These systems often rely on machine learning algorithms to detect and predict FoG episodes based on gait patterns and other relevant data.[1][2][3]

2. While existing FoG prognostication systems have shown promise, they still face significant limitations. Many of these systems struggle with false positives and false negatives, reducing their overall accuracy. Additionally, the effectiveness of these systems can vary among individuals, highlighting the need for personalized approaches. [1][2][3]

3. The placement of sensors and the quality of data collection are critical factors influencing the performance of FoG prognostication systems. Studies have explored optimal sensor placement and data preprocessing techniques to enhance accuracy. [1][2][3]

4. Researchers are increasingly focusing on non-invasive methods for predicting FoG to improve patient comfort and compliance. Moreover, efforts are being made to ensure that these systems are accessible and user-friendly for individuals with Parkinson's, including older adults. [1][2][3]

5. Machine learning and artificial intelligence techniques have played a central role in the development of FoG prognostication systems. These algorithms continue to evolve, aiming to improve predictive accuracy and reduce false alarms. [1][2][3]

6. Challenges in this field include the need for larger and more diverse datasets, validation of predictive models in real-world settings, and addressing the variability in FoG patterns among patients. [1][2][3]

7. "A Machine Learning Approach for Freezing of Gait Prediction in Parkinson's Disease"[10]
    a) Machine learning approach for FOG prediction in Parkinson's Disease.
    b) Utilizes a comprehensive dataset to propose an algorithm for FOG prediction.
    c) Highlights the potential of machine learning in enhancing predictive models
    d)
8. Wearable System for Freezing of Gait Prediction in Parkinson's Disease Patients [11]
    a) Focuses on developing a wearable system for predicting FOG in Parkinson's patients.
    b) Explores degradation of walking patterns preceding FOG episodes using inertial sensors.
    c) Emphasizes the importance of wearable technology in automatic FOG detection and prediction.
    d)
9. Quantifying Freezing of Gait in Parkinson's Disease Using 3D Accelerometry[12]
    a. Advocates for objective quantification of FOG in Parkinson's Disease.
    b. Utilizes 3D accelerometer data to discern onset and cessation of FOG episodes.
    c. Presents a systematic approach to understanding FOG, paving the way for novel therapeutic interventions.
10. Freezing of Gait Detection in Parkinson's Disease Using Inertial Sensors[13]
    a. Focuses on detection of FOG in Parkinson's Disease using inertial sensors.
    b. Employs angular velocity signals and machine learning classifiers.
    c. Achieves excellent performance in FOG detection, contributing to accurate and reliable detection methods
11. A Survey on Freezing of Gait Detection and Prediction in Parkinson's Disease [14]
    a. Provides a comprehensive overview of existing research on FOG detection and prediction in Parkinson's Disease.
    b. Synthesizes information from various studies
    c. Identifies gaps and challenges in the field, laying the groundwork for further advancements

4 Methodology

The data for this analysis project was collected through collaborative efforts involving multiple research groups, including the Center for the Study of Movement, Cognition, and Mobility, the Neurorehabilitation Research Group at Katholieke Universiteit Leuven in Belgium, and the Mobility and Falls Translational Research Center at the Hinda and Arthur Marcus Institute for Aging, affiliated with Harvard Medical School in Boston. This data collection initiative was generously supported by The Michael J. Fox Foundation for Parkinson's Research, making it possible for researchers to access valuable Parkinson's-related data for analysis.[4]

The project's data collection process involved the acquisition of various datasets, including training and testing data for different conditions (defog and tdcsfog), subject information stored in 'subjects.csv', and metadata files ('tdcsfog_metadata.csv' and 'defog_metadata.csv') containing additional context about the data. Moreover, 'events.csv' provided crucial information on individual Freezing of Gait (FOG) episodes.

The collected data served as the foundation for our analysis. Specific utility functions were applied to facilitate data exploration and preprocessing. These functions included 'get_num_cols' for extracting numeric columns from the dataset and 'factorize_column' to handle categorical or object-type data. Additionally, data was organized and accessed through directory structures based on condition (defog or tdcsfog), allowing for systematic data management.

In this section, we outline the comprehensive data preprocessing and exploratory data analysis (EDA) process employed to prepare and investigate the research dataset. The dataset comprises information from clinical studies involving subjects with various attributes. To ensure data quality and to extract meaningful insights, we executed a step-by-step methodology that encompasses data collection, renaming columns for clarity, merging multiple dataframes, conducting data cleaning and feature engineering, calculating descriptive statistics, and employing data visualization techniques.

4.1 Data Collection

The initial phase of our research involved the acquisition of data from relevant sources, which included clinical trials and patient surveys[2]. The dataset incorporates a multitude of variables, with key attributes encompassing subject information, visit details, age, gender, years since diagnosis (years_since_dx), and various clinical assessment scores. This comprehensive dataset serves as the foundation upon which our subsequent data preprocessing and analysis are based.

4.2 Data Preprocessing and Cleaning

To ensure uniformity and enhance interpretability of the dataset, we initiated the preprocessing pipeline by renaming specific columns. This step was executed with meticulous attention to detail, renaming columns such as "Subject" to "subject," "Visit" to "visit," and "Medication" to "medication." Renaming was performed with the goal of maintaining consistency and clarity throughout the dataset.[4]

Our dataset was composed of multiple dataframes, each containing valuable information. To consolidate and leverage these diverse data sources effectively, we employed dataframe merging techniques. We utilized the "inner" join method to merge the subjects_df dataframe with both the defog_metadata and tdcsfog_metadata dataframes[5] on common columns, such as "subject" and "visit."

4.3 Data Exploration and Visualization

	subject	age	sex	years_since_dx	UPDRSIII_On	UPDRSIII_Off	NFOGQ	Id	medication
0	00f674	63	1	27.0	43.0	49.0	24	41bc215f97	0
1	00f674	63	1	27.0	43.0	49.0	24	b4365bba9d	1
2	00f674	63	1	27.0	31.0	30.0	26	3f3b08f78d	1
3	00f674	63	1	27.0	31.0	30.0	26	4c3aa8ea6e	0
4	040587	75	1	26.0	52.0	69.0	21	2cc3c30645	0
...	...	...	...	...	...	...	...	...	...
828	fa8764	60	0	7.0	30.0	NaN	19	8797749a82	0
829	fa8764	60	0	7.0	30.0	NaN	19	98c313f19c	0
830	fa8764	60	0	7.0	30.0	NaN	19	d2382704e0	0
831	fa8764	60	0	7.0	30.0	NaN	19	dbe0a8f2fd	0
832	fa8764	60	0	7.0	30.0	NaN	19	ecd44c6b81	0

970 rows × 9 columns

Fig. 2. Merged Dataset Metadata

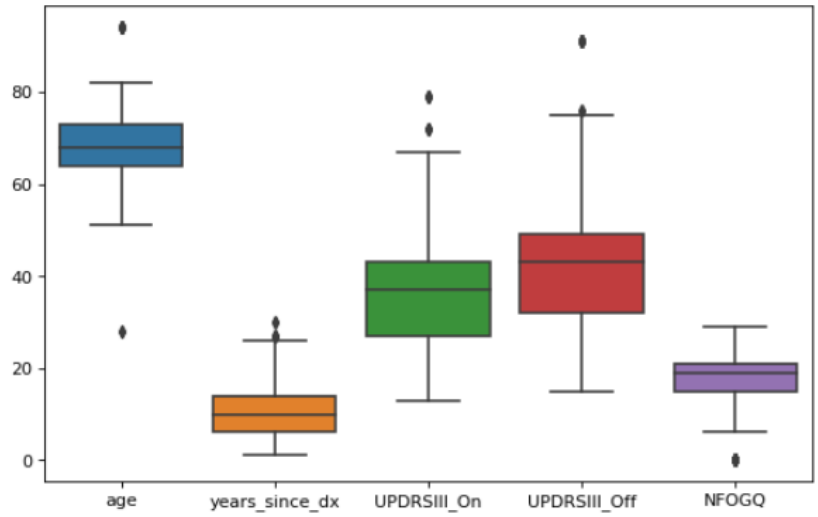

Fig. 3. Box Plot of the full metadata

The box plot figures depicted essential insights about the dataset's numeric attributes, including "age," "years_since_dx," "UPDRSIII_On," "UPDRSIII_Off," and "NFOGQ." These plots efficiently showcased central tendencies, data spreads, and the presence of potential outliers. By visually analyzing[6] these box plots, we gleaned key information about the data's distribution and variability, aiding in the identification of trends and data patterns essential for our research.

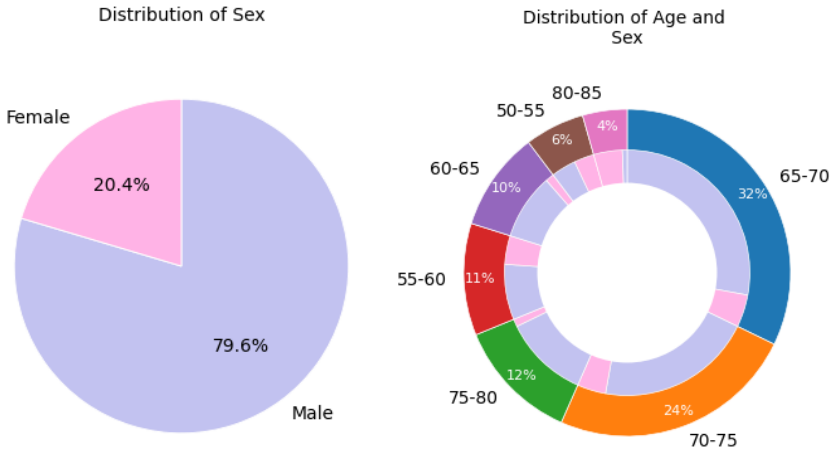

Fig. 4. (*left*) Distribution of sex (*right*) Distribution of Age and Sex

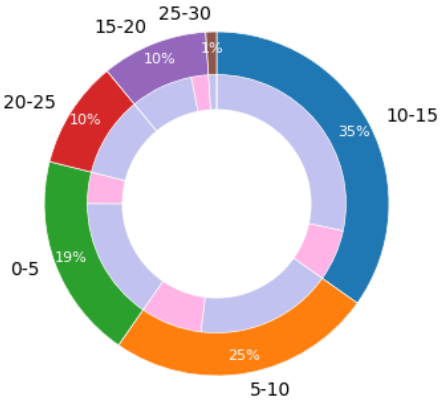

Fig. 5. Distribution of years since Diagnosis and Sex

The visualizations presented here offer an insightful perspective on the dataset, transforming[9] raw data into informative visuals. The initial pie chart provides a clear breakdown of gender distribution ('sex') among the subjects, presenting percentages for each category. Following this, two segmented pie charts depict age-related data, breaking it down into meaningful age ranges while considering gender distribution within each segment. These visualizations collectively provide a comprehensive overview of demographic attributes within the dataset, aiding in the rapid identification of patterns and trends. Through these visuals, researchers gain valuable insights into the composition and distribution of key demographic factors within the study population

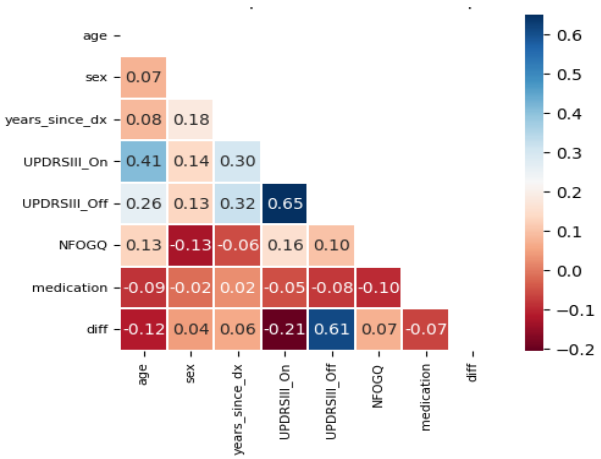

Fig. 6. Full metadata pairwise Correction Heatmap

The correlation heatmap presented here offers a concise visual representation of the relationships between numeric variables within the dataset. This heatmap employs a color-coded grid of squares, where each square signifies the strength and direction of the correlation between two variables. Positive correlations are visually indicated by varying shades of red, while negative correlations are represented by shades of blue, with darker colors representing stronger correlations. The inclusion of numerical annotations within each square provides precise correlation coefficients, facilitating quantitative analysis. This heatmap condenses intricate inter-variable connections into an accessible visual format, making it a valuable tool for identifying noteworthy patterns and dependencies among dataset attributes. Researchers can efficiently identify which variables exhibit substantial correlations, enabling further in-depth analysis and guiding research directions.

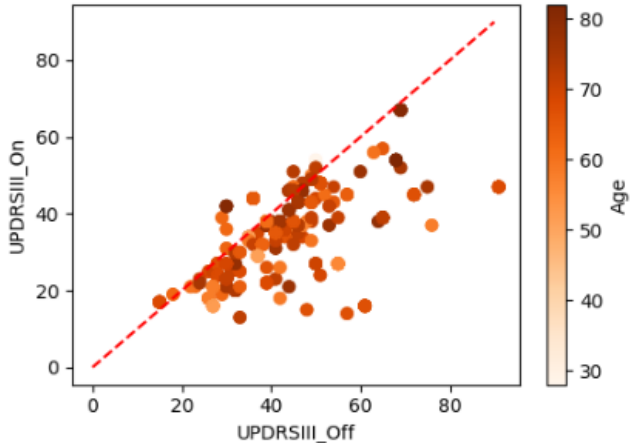

Fig. 7. Scatter plot of UPDRSIII_off vs UPDRSIII_On

Above visualization provides a detailed depiction of the interplay between 'UPDRSIII_Off' and 'UPDRSIII_On' scores. Each data point represents an individual subject, with their 'UPDRSIII_Off' score plotted along the x-axis and 'UPDRSIII_On' score along the y-axis. Notably, the coloration of data points corresponds to the respective age of the subjects, as indicated by the colormap 'Oranges.'

Additionally, a diagonal red dashed line spanning from (0, 0) to (90, 90) serves as a visual guidepost. Points positioned above this line signify instances where 'UPDRSIII_On' scores surpass 'UPDRSIII_Off' scores, while points below it indicate the opposite scenario. This scatter plot offers a nuanced exploration of the relationship between these clinical scores, enriched by the contextual dimension of age, enabling researchers to discern patterns and trends within the data effortlessly.

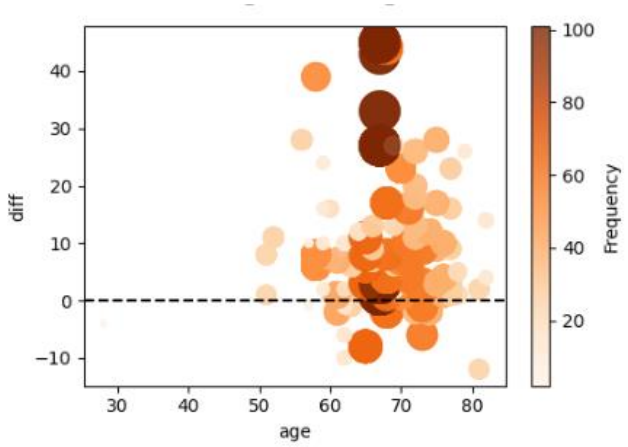

Fig. 8. Relationship between Age and the difference of UPDRSIII_off and UPDRSIII_On istribution of years since Diagnosis and Sex

Above subplot offers a profound exploration of the relationship between age and the difference, denoted as 'diff,' between 'UPDRSIII_On' and 'UPDRSIII_Off' scores. Each data point in this visualization represents an individual subject, with their age determining the coloration of the data point. Furthermore, the size of each data point is directly proportional to the frequency of that specific age value, facilitating a clearer grasp of data distribution patterns. Adding to the interpretability, a horizontal black dashed line positioned at y=0 serves as a crucial reference line, enabling effortless identification of data points where the difference between 'UPDRSIII_On' and 'UPDRSIII_Off' scores equals zero. This subplot offers a comprehensive view of how age correlates with variations in clinical scores, enhancing our understanding of age-related patterns within the dataset, the difference between UPDRSIII, when medication is on and when it is off, is calculated and compared to age.

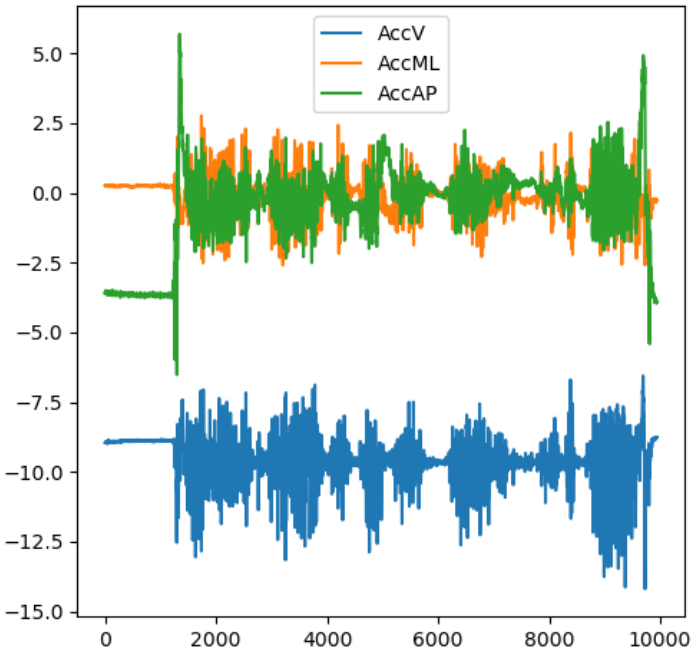

Fig. 9. TDCDFog Series sample

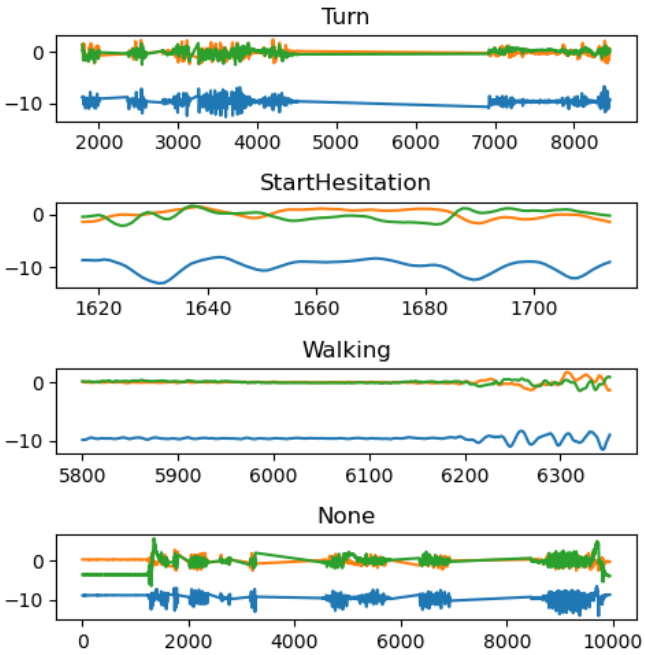

Fig. 10. Variation of acceleration data during various events

In the particular time, series, we observe that the variation of acceleration data during the Turn episode is at its maximum, which is a reasonable expectation. However, what makes the prediction challenging is the observation of non-trivial variations even in the absence of Turn, Walking, or Start Hesitation episodes. This variability in the data adds complexity to the prediction task

5 Conclusion

For ages below 60, medication is almost universally effective. However, for ages above 60, the effects vary greatly in a case-by-case manner. The most diversity is observed in the age range of 65-75 (as we saw before, 66% of patients are in this age group). In this age range, medication can either worsen the symptoms or be even more effective compared to other age ranges

Also, by examining trends with respect to the role of sex we found that approximately 80% of the participants are men. More than 66% of the participants are between 65-75 years old. Around 60% of the participants were diagnosed 5-15 years ago, among participants aged between 80-85, the majority are women.

Women generally exhibit a slightly better condition in Parkinson's disease. On the other hand, medication tends to be trivially more effective in men.

From dataset report and correlation matrix, it can be concluded that:

1. All target variables are highly imbalanced, especially StartHesitation (78.5%) and Walking (80.1%).

2. Based on histograms and skewness values the distributions of the AccAP and AccV columns are moderately left-skewed, the AccML column seems to have a close to normal distribution.

3. The AccAP column have a kurtosis value of less than 3, which indicates that the column is platikurtic. Meanwhile, the AccV column has a kurtosis value of more than 3, which indicates that the column is leptokurtic. And a kurtosis value of theAccML column is close to 3, which is recognized as mesokurtic column.

4. As can be seen from Phik correlation matrices ‘Time’ column have a moderate positive correlation with two target variables Turn, Walking and variable of anteroposterior acceleration measurements. However, between the target variable StartHesitation and other variables there is no strong or moderate correlation

In conclusion, the comprehensive data analysis and visualization undertaken in this study have yielded valuable insights into the relationships and patterns within the

dataset. Our research objectives, focused on understanding the interplay between clinical scores and demographic factors, have been met with noteworthy findings. The scatter plot comparing 'UPDRSIII_Off' and 'UPDRSIII_On' scores, while considering age as a contextual dimension, revealed intriguing trends, with data points both above and below the reference line, indicating variations in these scores. Additionally, the exploration of age in relation to the 'diff' (difference) between 'UPDRSIII_On' and 'UPDRSIII_Off' scores unveiled nuanced associations, further enhancing our comprehension of age-related patterns. These findings hold implications for clinical assessments and research within the field. However, it's important to acknowledge the limitations inherent to this study, such as potential data constraints.

6 Future Scope

Future research endeavors could delve deeper into the uncovered patterns and explore other facets of this multifaceted dataset. Machine Learning Algorithms can be used to predict the onset and cessation of various events. As there are more than one class , it is a multiclass classification problem. Algorithms[8] such as XGboost and Support Vector Classifier could be used as they are efficient in handling complex and non-linear relationships in the such a large data

In summary, this study contributes to our understanding of the intricate interplay between clinical scores and demographic attributes, offering potential avenues for further exploration and application within the domain of medical research and patient care.

References

1. Reches, T., Dagan, M., Herman, T., Gazit, E., Gouskova, N. A., Giladi, N., Manor, B., & Hausdorff, J. M. (2020). Using Wearable Sensors and Machine Learning to Automatically Detect Freezing of Gait during a FOG-Provoking Test. Sensors (Basel, Switzerland), 20(16), 4474. https://doi.org/10.3390/s20164474
2. Manor, B., Dagan, M., Herman, T., Gouskova, N. A., Vanderhorst, V. G., Giladi, N., Travison, T. G., Pascual-Leone, A., Lipsitz, L. A., & Hausdorff, J. M. (2021). Multitarget Transcranial Electrical Stimulation for Freezing of Gait: A Randomized Controlled Trial. Movement disorders : official journal of the Movement Disorder Society, 36(11), 2693–2698. https://doi.org/10.1002/mds.28759
3. Mancini, M., Bloem, B. R., Horak, F. B., Lewis, S. J. G., Nieuwboer, A., & Nonnekes, J. (2019). Clinical and methodological challenges for assessing freezing of gait: Future perspectives. Movement disorders : official journal of the Movement Disorder Society, 34(6), 783–790. https://doi.org/10.1002/mds.27709
4. Kaggle , https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/overview

5. Falla, M., Cossu, G., & Di Fonzo, A. (2022). Freezing of gait: Overview on etiology, treatment, and future directions. Neurological Sciences, 43(3), 1627–1639. https://doi.org/10.1007/s10072-021-05796-w
6. Perez-Lloret, S., Negre-Pages, L., Damier, P., Delval, A., Derkinderen, P., Destée, A., Meissner, W. G., Schelosky, L., Tison, F., & Rascol, O. (2014). Prevalence, Determinants, and Effect on Quality of Life of Freezing of Gait in Parkinson Disease. JAMA Neurology, 71(7), 884–890. https://doi.org/10.1001/jamaneurol.2014.753
7. Schaafsma, J. D., Balash, Y., Gurevich, T., Bartels, A. L., Hausdorff, J. M., & Giladi, N. (2003). Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease. European Journal of Neurology, 10(4), 391–398. https://doi.org/10.1046/j.14681331.2003.00611.x
8. J. Virani, N. Daredi, A. Bhanushali, M. Shukla and P. Shah, "Mental Healthcare Analysis using Power BI & Machine Learning," 2023 4th International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 2023, pp. 73-76, doi: 10.1109/ICSPC57692.2023.10125680
9. K. Pipalia, R. Bhadja and M. Shukla, "Comparative Analysis of Different Transformer Based Architectures Used in Sentiment Analysis," 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2020, pp. 411-415, doi: 10.1109/SMART50582.2020.9337081.
10. Hvingelby VS, Glud AN, Sørensen JCH, Tai Y, Andersen ASM, Johnsen E, Moro E, Pavese N. Interventions to improve gait in Parkinson's disease: a systematic review of randomized controlled trials and network meta-analysis. J Neurol. 2022 Aug;269(8):4068-4079. doi: 10.1007/s00415-022-11091-1. Epub 2022 Apr 5. PMID: 35378605
11. Chow R, Tripp BP, Rzondzinski D, Almeida QJ. Investigating Therapies for Freezing of Gait Targeting the Cognitive, Limbic, and Sensorimotor Domains. Neurorehabil Neural Repair. 2021 Mar;35(3):290-299. doi: 10.1177/1545968321992331. Epub 2021 Feb 9. PMID: 33559531; PMCID: PMC7934156.
12. Bansal SK, Basumatary B, Bansal R, Sahani AK. Techniques for the detection and management of freezing of gait in Parkinson's disease - A systematic review and future perspectives. MethodsX. 2023 Mar 4;10:102106. doi: 10.1016/j.mex.2023.102106. PMID: 36942282; PMCID: PMC10023964.
13. N. K. Orphanidou, A. Hussain, R. Keight, P. Lishoa, J. Hind and H. Al-Askar, "Predicting Freezing of Gait in Parkinsons Disease Patients Using Machine Learning," 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/CEC.2018.8477909.
14. Martinez-Villaseñor, Lourdes & Ponce, Hiram & Miralles Pechuan, Luis. (2020). A Survey on Freezing of Gait Detection and Prediction in Parkinson's Disease. 10.1007/978-3-030-60884-2_13