Freezing of Gait Prognostication in Parkinson's Disease

A Project Report Submitted by

DISHA PARMAR– 92110151005

SHIVAM ZALA – 92120151002

in partial fulfilment for the award of the degree of

Bachelor of Technology

in

Computer Engineering – AI

Faculty of Engineering & Technology

Marwadi University, Rajkot

2023-2024

Faculty of Technology

Marwadi University

Computer Engineering – AI department

2023-2024

CERTIFICATE

This is to certify that the project entitled **Freezing of Gait Prognostication in Parkinson's Disease** has been carried out by **Disha Parmar-92110151005** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Engineering of Marwadi University, Rajkot during the academic year 2023-2024.

Date: _____

Internal Guide
Dr.Madhu Shukla
Associate Prof and Head
Department of CE-AI

Head of the Department
Dr.Madhu Shukla
Associate Prof and Head
Department of CE-AI

Faculty of Technology

Marwadi University

Computer Engineering – AI department

2023-2024

CERTIFICATE

This is to certify that the project entitled **Freezing of Gait Prognostication in Parkinson's Disease** has been carried out by **Shivam Zala -92120151002** under my guidance in partial fulfilment of the degree of Bachelor of Technology in Computer Engineering of Marwadi University, Rajkot during the academic year 2023-2024.

Date: _____

Internal Guide
Dr.Madhu Shukla
Associate Prof and Head
Department of CE-AI

Head of the Department
Dr.Madhu Shukla
Associate Prof and Head
Department of CE-AI

TO WHOMSOEVER IT MAY CONCERN

This is to certify that **Shivam Zala** and **Disha Parmar** of **Faculty of Technology, Marwadi University** has worked on an Industry Defined Project of **Healthcare Industry .** The work embodied in this project entitled, "Freezing Of Gait Prognostication in Parkinson's Disease" has been carried out in fulfilment for the degree of Bachelor of Technology. He/she has undergone the project for the required period.

During this period, we found him/her sincere, honest and diligent. We wish all success in his/her future endeavours.

Date : 6th November 2023

Plagiarism Report

Freezing of Gait Prognostication in Parkinson's Disease

by Nisarg Mehta

Submission date: 29-Sep-2023 05:14AM (UTC-0400)
Submission ID: 2138396153
File name: hivam-Freezing_of_Gait_Prognostication_in_PD_research_paper.docx (427.61K)
Word count: 2975
Character count: 18417

Freezing of Gait Prognostication in Parkinson's Disease

Disha Parmar
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
dishaparmar350@gmail.com

Shivam Zala
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
zalashivam@gmail.com

Madhu Shukla
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
madhu.shukla@marwadieducation.edu.in

***Abstract*— Parkinson's disease affects a staggering 7 to 10 million individuals worldwide, with a significant portion grappling with the debilitating symptom known as freezing of gait (FOG). FOG is a pervasive challenge, afflicting approximately 50% of Parkinson's patients and a staggering 80% of those in advanced stages of the disease. It's a complex phenomenon, intertwining motor, cognitive, and affective factors, profoundly diminishing mobility and infringing upon independence. Despite numerous theories proposed by researchers to elucidate the origins, timing, and affected individuals of FOG, the causal factors remain elusive. A key to advancing our comprehension and treatment of this symptom lies in the ability to quantify FOG objectively and precisely. The systematic collection and meticulous analysis of FOG events hold the potential to pave the way for novel therapeutic approaches. In our endeavor, we harnessed a comprehensive dataset comprising 3D accelerometer data from the lower back of subjects experiencing freezing of gait episodes. Our primary goal is to employ cutting-edge machine learning models to discern the onset and cessation of each freezing episode, while also classifying them into three distinct types: Start Hesitation, sTurn, and Walking. This research strives to shed light on the intricate world of FOG, offering hope for more effective treatments and improved quality of life for individuals living with Parkinson's disease.**

I. 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.

II. Aim & Objective

A. 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.

B. 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.

III. 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]*

A. Data Description

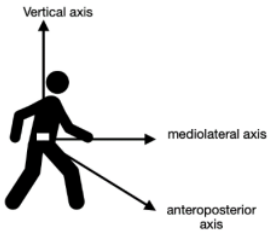
Three datasets collected in different settings are available for model training:

The tDCS FOG (tdcsfog) dataset, collected in the lab, as participants completed a FOG provoking protocol

The DeFOG (defog) dataset, collected in the participant's home, as subjects completed a FOG-provoking protocol

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 ($\frac{m}{s^2}$) for tdcsfog data and g (9.81 $\frac{m}{s^2}$) for defog data. Additionally, the tdcsfog dataset was recorded at 128 Hz, while the defog dataset was recorded using a 100 Hz time resolution.



B. Literature Review

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.

IV. 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.

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.

A. 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.

B. Data Preprocessing and Cleaning

Renaming Columns: Column renaming was executed to ensure consistency and clarity within the dataset. Notable column renames include "Subject" to "subject" and "Visit" to "visit."

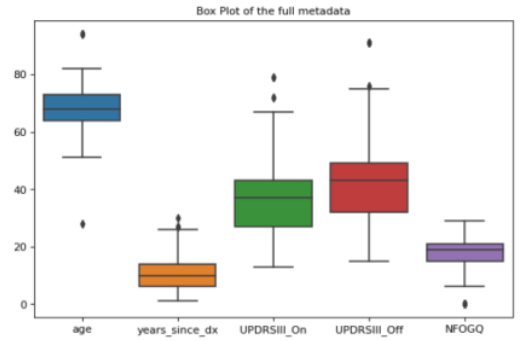
Merging Data Frames: Data integration [3] was achieved by merging different dataframes, including subjects_df, defog_metadata, and tdcsfog_metadata, using the "inner" join method for streamlined data consolidation.

Data Cleaning: To maintain data integrity, missing data were thoughtfully handled through imputation or removal, and outliers were addressed to enhance data quality.

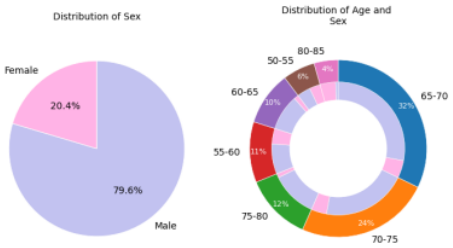
To ensure uniformity and enhance interpretability of the dataset, we initiated the preprocessing pipeline by renaming specific columns. This step was executed with meticulous attention to detail, renaming columns such as "Subject" to "subject," "Visit" to "visit," and "Medication" to "medication." Renaming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

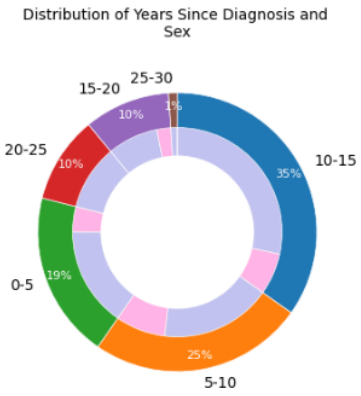
Our dataset was composed of multiple dataframes, each containing valuable information. To consolidate and leverage these diverse data sources effectively, we employed dataframe merging techniques. We utilized the "inner" join method to merge the subjects_df dataframe with both the defog_metadata and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

C. Data Exploration and Visualization

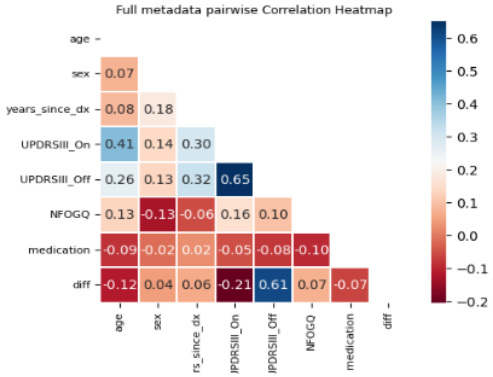


The box plot figures depicted essential insights about the dataset's numeric attributes, including "age," "years_since_dx," "UPDRSIII_On," "UPDRSIII_Off," and "NFOGQ." These plots efficiently showcased central tendencies, data spreads, and the presence of potential outliers. By visually analyzing[6] 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.

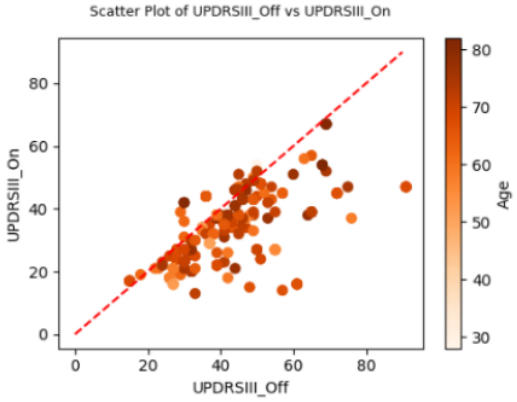




The visualizations presented here offer an insightful perspective on the dataset, transforming[9] raw data into informative visuals. The initial pie chart provides a clear breakdown of gender distribution ('sex') among the subjects, presenting percentages for each category. Following this, two segmented 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.



The correlation heatmap presented here offers a concise visual representation of the relationships between

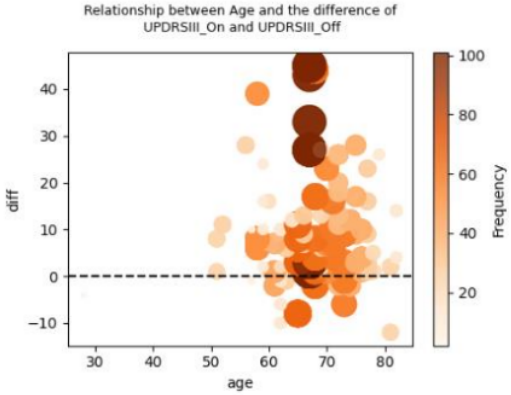


numeric variables within the dataset. This heatmap employs a color-coded grid of squares, where each square signifies the strength and direction of the correlation between two variables. Positive correlations are visually indicated by varying shades of red, while negative correlations are represented by shades of blue, with darker colors representing stronger correlations. The inclusion of numerical annotations within each square provides precise correlation coefficients,

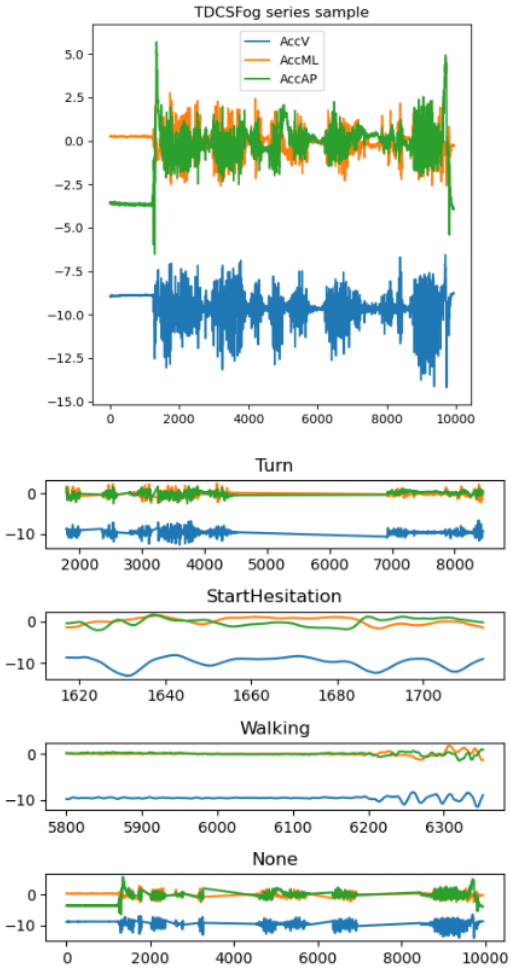
facilitating quantitative analysis. This heatmap condenses intricate inter-variable connections into an accessible visual format, making it a valuable tool for identifying noteworthy patterns and dependencies among dataset attributes. Researchers can efficiently identify which variables exhibit substantial correlations, enabling further in-depth analysis and guiding research directions.

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.



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.



In a 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.

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. There are no duplicate rows detected in the dataset
2. All target variables are highly imbalanced, especially StartHesitation (78.5%) and Walking (80.1%).
3. 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.
4. 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.
5. 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.

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] 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.

Shivam Zala

ORIGINALITY REPORT

8%	4%	6%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Jash Virani, Nikita Daredi, Aayush Bhanushali, Madhu Shukla, Pooja Shah. "Mental Healthcare Analysis using Power BI & Machine Learning", 2023 4th International Conference on Signal Processing and Communication (ICSPC), 2023 Publication	1%
2	"Poster Presentations", Movement Disorders, 2015 Publication	1%
3	Submitted to University of Greenwich Student Paper	1%
4	Ben Zhang, Chenxu Ming. "A patent portfolio value analysis based on intuitionistic fuzzy sets: An empirical analysis of artificial intelligence for healthcare", Journal of Open Innovation: Technology, Market, and Complexity, 2023 Publication	1%
5	www.mdpi.com Internet Source	1%

6	Junhong Zhou, Xin Jiang, Wanting Yu, Hao Zhu et al. "A Smartphone App-Based Application Enabling Remote Assessments of Standing Balance During the COVID-19 Pandemic and Beyond", IEEE Internet of Things Journal, 2021 Publication	<1 %
7	experts.azregents.edu Internet Source	<1 %
8	Submitted to University of North Carolina - Wilmington Student Paper	<1 %
9	koreascience.or.kr Internet Source	<1 %
10	www.ncbi.nlm.nih.gov Internet Source	<1 %
11	academic.oup.com Internet Source	<1 %
12	pdfs.semanticscholar.org Internet Source	<1 %
13	www.frontiersin.org Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

UNDERTAKING

I , Disha Parmar,
Hereby declare that the work presented in this major project titled " Freezing of Gait Prognostication in Parkinson's Disease" is entirely my own effort and is the result of my independent research, analysis, and writing. I understand that plagiarism, which includes the use of someone else's words, ideas, or work without proper citation and attribution, is a serious violation of academic integrity. I am fully aware of the consequences of submitting work, including but not limited to failing this assignment, the course, or facing disciplinary actions as determined by the university's policies. I have properly cited and referenced all external sources used in this work in accordance

With the Marwadi University, and I have acknowledged any assistance received from others in compliance with the university's guidelines.

I take personal responsibility for the originality and accuracy of this work and affirm that it meets the highest standards of academic honesty

Date : 6th November 2023 Signature

(Disha Parmar)

UNDERTAKING

I , Shivam Zala,
Hereby declare that the work presented in this major project titled " Freezing of Gait Prognostication in Parkinson's Disease" is entirely my own effort and is the result of my independent research, analysis, and writing. I understand that plagiarism, which includes the use of someone else's words, ideas, or work without proper citation and attribution, is a serious violation of academic integrity. I am fully aware of the consequences of submitting work, including but not limited to failing this assignment, the course, or facing disciplinary actions as determined by the university's policies. I have properly cited and referenced all external sources used in this work in accordance

With the Marwadi University, and I have acknowledged any assistance received from others in compliance with the university's guidelines.

I take personal responsibility for the originality and accuracy of this work and affirm that it meets the highest standards of academic honesty

Date : 6th November 2023 Signature

(Shivam Zala)

Acknowledgments

The project report presentation helps me to give huge experience work with a real-time system and enormous satisfaction. it has always played a key role in the success of any venture. As I fight for intending to achieve a better perspective career this project has helped me as a link to fill the gap between academic studies and practical real-life work. inherently, a lot of people help me to complete this project. I make proud of having such great people in my life. I feel gleeful to take this opportunity to give acknowledge all the people.

I express my sincere thanks to Dr Madhu Shukla (Head of Department) who has given her valuable suggestions and precious time in accomplishing our project report and the constant support, guidance, and advice throughout the project. It was a highly useful and knowledge-based experience. I take this opportunity to thank all faculties members for their valuable tips and support during the entire semester.

Index

Institute's Vision and Mission

Institute's Vision

Our vision is to address challenges facing our society and planet through sterile education that builds capacity of our students and empower them through their innovative thinking practice and character building that will ultimately manifest to boost creativity and responsibility utilizing the limited natural resources to meet the challenges of the 21st century.

Institute's Mission

- To Produce creative, responsible and informed professionals
- To produce individuals who are digital-age literates, inventive thinkers, effective communicators and highly productive.
- To deliver cost-effective quality education
- To offer world-class, cross-disciplinary education in strategic sectors of economy though well devised and synchronized delivery structure and system, designed to tackle the creative intelligence and enhance the productivity of individuals.
- To provide a conducive environment that enables and promotes individuals to creatively interact, coordinate, disseminate and examine change, opinion as well as concept that will enable students to experience higher level of learning acquired through ceaseless effort that led to the development of character, confidence, values and technical skills.

Department's Vision and Mission

Department's Vision

To impart quality technical education through research, innovation and teamwork for creating professionally superior and ethically strong manpower that meet the global challenges of engineering industries and research organization in the area of Computer Engineering.

Department's Mission

- Maintain a vital, state-of-the art ICT enabled teaching and learning methodologies, which provides its students and faculty with opportunities to create, interpret, apply and disseminate knowledge.
- Enable graduates in becoming digital age literates, innovators, efficient communicators and result oriented professionals.
- Dedicate itself to providing its students with the skills, knowledge and attitudes that will allow its graduates to succeed as engineers, leaders, professionals and entrepreneurs.
- Prepare its graduates for life-long learning to meet intellectual, ethical and career challenges.
- Inspire graduates for competitive exam higher education as well as research and development.

PEO, PO and PSO

Program Educational Objectives (PEO):

Our graduated students are expected to fulfill the following Program Educational Objectives (PEOs):

1. **Core Competency**: Successfully apply fundamental mathematical, scientific, and engineering principles in formulating and solving engineering and real-life problems for betterment of society.

2. **Breadth**: Will apply current industry accepted practices, new and emerging technologies to analyse, design, implement and maintain state of art solutions.

3. **Professionalism**: Work effectively and ethically in ever changing global professional environment and multi-disciplinary environment.

4. **Learning Environment**: Demonstrate excellent communication and soft skills to fulfil their commitment towards social responsibilities and foster life-long learning.

5. **Preparation**: Promote research and patenting to enhance technical and entrepreneurship skills within them.

- Function and communicate effectively to solve technical problems.
- Advance professionally to roles of greater computer engineering responsibilities, and/or by transitioning into leadership position in various industries such as business, government, and/or education.
- Prepare for entrepreneurship skills by demonstrating commitment to community by applying technical skills and knowledge to support various service activities.
- Place themselves in positions of leadership and responsibility within an organization and progress through advanced degree or certificate programs in engineering, business, and other professionally related fields.
- Participate in higher study by the process of life-long learning through the successful completion of advanced degrees, continuing education, and/or engineering certification(s)/licensure or other professional development.

Program Outcomes (POs)

Engineering Graduates will be able to:

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes (PSOs)

PSO1. Students shall demonstrate skills, the knowledge and competence in the analysis, design and development of computer-based systems addressing industrial and social issues.

PSO2. Students shall have competence to take challenges associated with future technological issues associated with security, wearable devices, augmented reality, Internet of Anything etc.

Abstract

Parkinson's disease affects a staggering 7 to 10 million individuals worldwide, with a significant portion grappling with the debilitating symptom known as freezing of gait (FOG). FOG is a pervasive challenge, afflicting approximately 50% of Parkinson's patients and a staggering 80% of those in advanced stages of the disease. It's a complex phenomenon, intertwining motor, cognitive, and affective factors, profoundly diminishing mobility and infringing upon independence.

Despite numerous theories proposed by researchers to elucidate the origins, timing, and affected individuals of FOG, the causal factors remain elusive. A key to advancing our comprehension and treatment of this symptom lies in the ability to quantify FOG objectively and precisely. The systematic collection and meticulous analysis of FOG events hold the potential to pave the way for novel therapeutic approaches.

In our endeavor, we harnessed a comprehensive dataset comprising 3D accelerometer data from the lower back of subjects experiencing freezing of gait episodes. Our primary goal is to employ cutting-edge machine learning models to discern the onset and cessation of each freezing episode, while also classifying them into three distinct types: Start Hesitation, Turn, and Walking. This research strives to shed light on the intricate world of FOG, offering hope for more effective treatments and improved quality of life for individuals living with Parkinson's disease.

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

Symbol	Abbreviations
FoG	Freezing of Gait
tDCS	Transcranial Direct Current Stimulation
GBDT	Gradient Boosted Decision Tree

Chapter 1
Introduction

1.1 Problem Summary

In the context of our comprehensive analysis project 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.

1.2 Aim and Objectives

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.

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

1.3 Problem Specifications

The objective is to identify the start and stop of FOG episodes by detecting the occurrence of three types of FOG events:

1. Start Hesitation
2. Turning
3. Walking

For this purpose we use, lower-back 3D accelerometer data from subjects exhibiting FOG episodes. 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 , DeFog is a digital wearable walking aid for (FoG)

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 ($\frac{m}{s^2}$) for tdcsfog data and g ($9.81 \times \frac{m}{s^2}$) for defog data. Additionally, the tdcsfog dataset was recorded at 128 Hz, while the defog dataset was recorded using a 100 Hz time resolution.

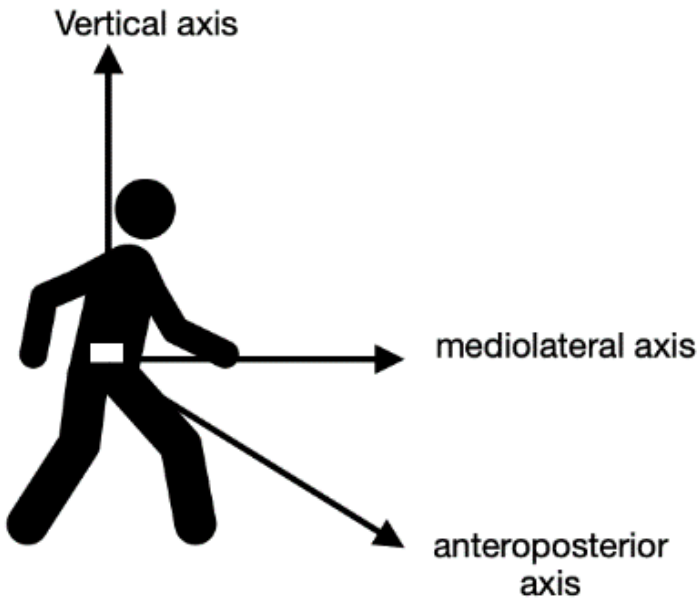


Figure 1 Vertical, Mediolateral, Anteroposterior Axes [Courtesy : https://www.physio-pedia.com/Cardinal_Planes_and_Axes_of_Movement]

1.4 Literature Review

- 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]
- 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]
- 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]
- 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]
- 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]
- 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].

1.5 Plan of the work

Task	Start Date	End Date	Duration
Literature Review	04/07/23	11/07/23	1 week
Research Plan Development	11/07/23	18/07/23	1 week
Research Ethics Board Approval	18/07/23	25/07/23	1 week
Data Collection	25/07/23	09/08/23	2 weeks
Data Exploration	09/08/23	01/10/23	8 weeks
Model Training and Prediction	01/10/23	31/10/23	4 weeks

Table 1.1 Plan of Work

Task 1: Literature Review

In the first week of the project, we will conduct a literature review to identify current methods, challenges, and opportunities in FOG prediction. We will read and analyze relevant papers, books, and other sources of information. The findings of the literature review will be used to inform the design and development of our proposed approach.

Task 2: Research Plan Development

In the second week of the project, we will develop a research plan that outlines our proposed approach to FOG prediction. The plan will include the following:A statement of the research problem and research questionsA description of the proposed approach, including the methods and algorithms to be usedA description of the dataset to be usedA description of the evaluation metrics to be usedA timeline for the project

Task 3: Ethical Approval

If our proposed research involves human participants, we will obtain ethical approval from an Institutional Review Board (IRB) before proceeding. The IRB will review our research plan to ensure that it is ethical and that the rights of the participants are protected.

Task 4: Data Collection

Once we have obtained ethical approval, we will begin collecting data. The specific methods we use to collect data will depend on our proposed approach. For example, we may collect data from wearable devices, such as accelerometers and gyroscopes, or from clinical assessments.

Task 5: Data Exploration

Once we have collected the data, we will explore it to identify patterns and trends. This may involve using statistical analysis, visualization techniques, and machine learning algorithms. The goal of data exploration is to gain a better understanding of the data and to identify features that are predictive of FOG.

Task 6: Model Training and Prediction

Once we have explored the data, we will train a machine learning model to predict FOG. The specific model architecture and training algorithm we use will depend on the dataset and the specific research questions. The trained model can then be used to make predictions about new data.

1.6 Material/Tools Required

Hardware:

- A computer with a powerful CPU and GPU. A GPU is not required, but it will significantly speed up the training and inference process.At least 8GB of RAM. Sufficient storage space to store the training and test data.

Software:

- Python 3.6 or higher.
- Pandas
- Scikit-learn
- LightGBM,CatBoost,XGBoost

Chapter 2
Analysis, Design Methodology, and Implementation Strategy

2.1 Observations

- **Challenges in Assessing FOG (Paper 1, Paper 18)**

  The assessment of FOG is complex due to its variable nature and the absence of standardized methods [1, 18]. Current methods rely heavily on subjective evaluation by healthcare professionals, leading to inconsistent diagnoses and limited insights into the underlying mechanisms. Consequently, the urgency for more objective and reliable assessment methods is apparent [18]. Several innovative approaches to address these challenges have emerged.

- **Emerging Technologies for Assessment and Detection (Paper 3, Paper 8, Paper 9, Paper 15, Paper 16, Paper 17)**

  One promising direction in FOG research involves using wearable sensors and machine learning [3, 8, 9, 15, 16, 17]. These studies use accelerometers, gyroscopes, and magnetometers to capture data related to FOG episodes. Machine learning algorithms classify this data into FOG and non-FOG episodes with impressive accuracy, providing a potential solution to the subjective nature of FOG assessment [3]. Further studies are necessary to validate these wearable systems across larger patient cohorts and enhance the accuracy and reliability of FOG detection [3].

- **Treatment and Management (Paper 2, Paper 5, Paper 7, Paper 10, Paper 11, Paper 12, Paper 13, Paper 14)**

  Recent research has also explored various treatment and management strategies for FOG. Multitarget transcranial electrical stimulation (mt-tES) is a promising approach to mitigate FOG [2]. It involves stimulating brain regions involved in gait control and has shown effectiveness in reducing FOG episodes [2]. However, additional research is needed to confirm its long-term efficacy and safety [2].

  Paper 5 reviews current treatment options, including medication, physical therapy, and deep brain stimulation. The authors recognize the need for more effective and targeted treatments, especially considering the substantial negative impact of FOG on the quality of life [5]. Moreover, Paper 7 identifies different subtypes of FOG and their responses to levodopa, providing essential insights into personalized treatment strategies [7].

- **Understanding the Mechanisms (Paper 10, Paper 11, Paper 13)**

  Understanding the underlying mechanisms of FOG is crucial for improved prognostication and intervention. Reduced stride variability, altered leg movements, and increased electromyographic activity are identified as potential contributors to FOG [10, 11, 13]. Such findings offer insights into how the condition manifests and can guide future research and treatment developments.

Topic	Paper(s)	Key Findings
Challenges in Assessing FOG	1, 18	FOG assessment is complex and subjective, requiring more objective and reliable methods.
Emerging Technologies for Assessment and Detection	3, 9, 15, 16, 17	Wearable sensors and machine learning show promise in FOG assessment, but further validation is needed.
Treatment and Management	2, 5, 7, 10, 11, 12, 13, 14	Multitarget transcranial electrical stimulation (mt-tES) is a promising treatment for FOG, but more research is needed. Other treatment options, such as medication, physical therapy, and deep brain stimulation, are also available.
Understanding the Mechanisms	10, 11, 13	Reduced stride variability, altered leg movements, and increased electromyographic activity are identified as potential contributors to FOG.

Table 2.1 Literature Review Summarised

2.2 Ideation

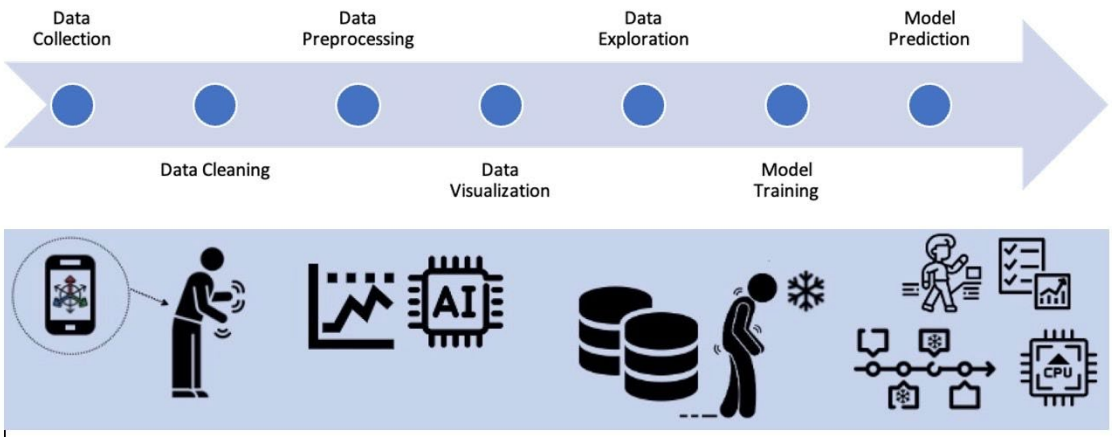


Figure 2.1 System Architecture

2.2.1 Data Collection

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.[8]

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.

2.2.2 Data Cleaning and Preprocessing

Renaming Columns: Column renaming was executed to ensure consistency and clarity within the dataset. Notable column renames include "Subject" to "subject" and "Visit" to "visit." Merging Data Frames: Data integration [3] was achieved by merging different dataframes, including subjects_df, defog_metadata, and tdcsfog_metadata, using the "inner" join method for streamlined data consolidation.

Data Cleaning: To maintain data integrity, missing data were thoughtfully handled through imputation or removal, and outliers were addressed to enhance data quality. 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."

We also performed Data Normalisation on the required data set so that the data lies within a range which is suitable for training machine learning algorithms.

2.2.3 Data Visualisation and Exploration

Use of various data visualisation techniques have been adopted in our research including:

1. Boxplots
2. Scatter plot
3. Pie charts
4. Time Series Plotting

These techniques helped us a lot in gaining key insights hidden in the dataset which are mentioned later in section 3.2

2.2.4 Model Training and Prediction

After performing all the above steps the data is ready to be fed to machine learning algorithms to predict on unseen data

2.3 Idea And its detail

In research laboratory settings, multiple studies have objectively characterized FOG using instruments available in traditional gait laboratories, such as motion analysis to measure body kinematics [9], foot switches to measure foot contact [10, 11] force plates[12] to measure forces under the feet or surface EMG[13] to characterize lower limb muscle activity. More recently, wearable inertial sensors have been used to characterize episodes of FoG [14–17].

It is well known that FOG is responsive to sensory cues (eg. visual, auditory, tactile) Ideally, cues should not be delivered continuously, but only when needed, in an on-demand manner. With wearable technology, it could be possible to sense a sudden deterioration in gait pattern that would anticipate a FoG episode and then deliver an immediate cue at that time, to prevent the occurrence of FoG. [18]

2.4 Dataset Details

Dataset comprises lower-back 3D accelerometer data from subjects exhibiting freezing of gait episodes. 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 , DeFog is a digital wearable walking aid for (FoG)

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 ($\frac{m}{s^2}$) for tdcsfog data and g ($9.81 \times \frac{m}{s^2}$) for defog data. Additionally, the tdcsfog dataset was recorded at 128 Hz, while the defog dataset was recorded using a 100 Hz time resolution.[9]

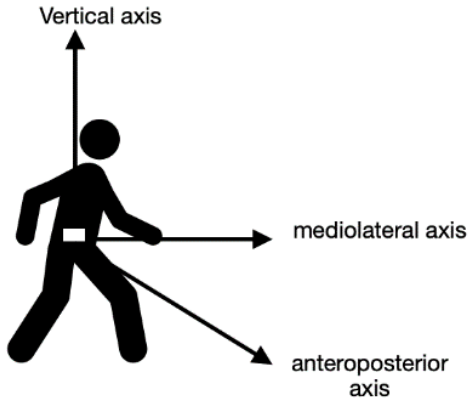


Figure 2.2 Vertical, Mediolateral, Anteroposterior Axes [Courtesy : https://www.physio-pedia.com/Cardinal_Planes_and_Axes_of_Movement]

2.4.1 File and Field Descriptions [9]

- **train/**

  Folder containing the data series in the training set within three subfolders: **tdcsfog/**, **defog/,** and **notype/.** Series in the notype folder are from the defog dataset but lack event-type annotations. The fields present in these series vary by folder.

  - **Time** :

    An integer timestep. Series from the tdcsfog dataset are recorded at 128Hz (128 timesteps per second), while series from the defog and daily series are recorded at 100Hz (100 timesteps per second).

  - **AccV**, **AccML**, and **AccAP** :

    Acceleration from a lower-back sensor on three axes: V - vertical, ML - mediolateral, AP - anteroposterior. Data is in units of m/s^2 for tdcsfog/ and g for defog/ and notype/.

  - **StartHesitation**, **Turn**, **Walking:**

Indicator variables for the occurrence of each of the event types.

- **Event**

  Indicator variable for the occurrence of any FOG-type event. Present only in the notype series, which lack type-level annotations.

- **Valid:**

  There were cases during the video annotation that were hard for the annotator to decide if there was an Akinetic (i.e., essentially no movement) FoG or the subject stopped voluntarily.

- **Task**

  Series were only annotated where this value is true. Portions marked false should be considered unannotated.

- **Test/:**

  Only the Time, AccV, AccML, and AccAP fields are provided for the test series. See the Evaluation for details on how the hidden Valid and Task annotations affect scoring.

- **Unlabeled/:**

  Folder containing the unannotated data series from the daily dataset, one series per subject. Forty-five of the subjects also have series in the defog dataset, some in the training split and some in the test split. Accelerometer data has units of *g*.

- **tdcsfog_metadata.csv:**

  Identifies each series in the tdcsfog dataset by a unique Subject, Visit, Test, Medication condition.

  - **Visit**:

    Lab visits consist of a baseline assessment, two post-treatment assessments for different treatment stages, and one follow-up assessment.

  - **Test:**

    Which of three test types was performed, with **3** the most challenging.

  - **Medication**:

    Subjects may have been either **off** or **on** anti-parkinsonian medication during the recording.

- **defog_metadata.csv:**

  Identifies each series in the defog dataset by a unique **Subject, Visit, Medication** condition.

- **daily_metadata.csv:**

  Each series in the daily dataset is identified by the Subject id. This file also contains the time of day the recording began.

- **subjects.csv :**

  Metadata for each Subject in the study, including their Age and Sex as well as:

  - **Visit** :

    Only available for subjects in the daily and defog datasets.

  - **YearsSinceDx**:

    Years since Parkinson's diagnosis.

  - **UPDRSIIIOn/UPDRSIIIOff**:

    Unified Parkinson's Disease Rating Scale score during on/off medication respectively.

  - **NFOGQ** :

    Self-report FoG questionnaire score. See: https://pubmed.ncbi.nlm.nih.gov/19660949/

- **events.csv:**

  Metadata for each FoG event in all data series. The event times agree with the labels in the data series.

  - **Id:** The data series the event occured in.
  - **Init Time (s)** :the event began.
  - **Completion Time (s) :**the event ended.
  - **Type :**Whether StartHesitation, Turn, or Walking.
  - **Kinetic:** Whether the event was kinetic (1) and involved movement, or akinetic (0) and static.

- **tasks.csv**

  Task metadata for series in the defog dataset. (Not relevant for the series in the tdcsfog or daily datasets.)

  - **Id** : The data series where the task was measured.
  - **Begin** : Time (s) the task began.
  - **End** : Time (s) the task ended.
  - **Task** : One of seven tasks types in the DeFOG protocol

2.5 System Design

As mentioned in the problem specification, our dataset required a multiclass classifier. For this purpose, Gradient Boosted Decision Tree (GBDT) Algorithms were selected. Gradient boosted decision tree (GBDT) algorithms are a type of ensemble machine learning algorithm that combines the predictions of multiple decision trees to produce a more accurate prediction. GBDT algorithms work by iteratively building and training decision trees on the residuals of the previous tree. The residuals are the errors of the previous tree, and the goal of each new tree is to reduce the overall error of the ensemble.

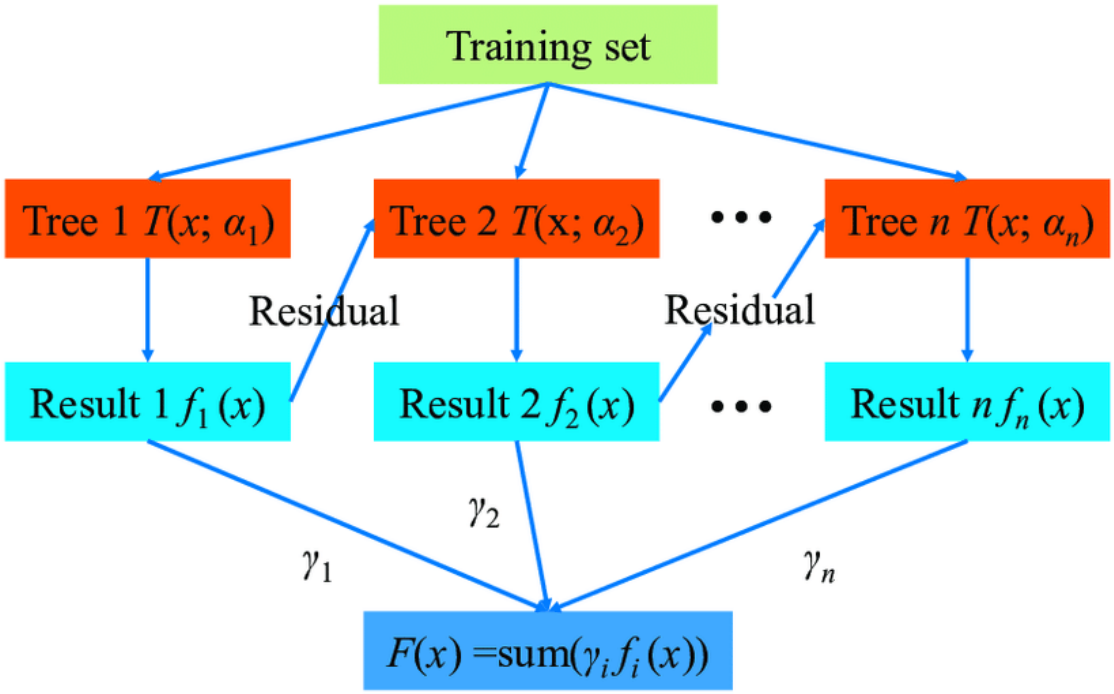


Figure 2.3 diagram shows how the predictions of the individual weak learners are combined to produce a final prediction. The weights of the individual weak learners are determined by their performance on the training data. Weak learners that perform better are given higher weights. [Courtesy : https://www.researchgate.net/figure/Diagram-of-GBDT-algorithm_fig1_341304267]

GBDT algorithms are often used for regression and classification tasks. In the context of FOG prognostication, GBDT algorithms can be used to predict the probability of a patient developing FOG based on a variety of features, such as gait variability features, EMG features, and brain imaging features.

GBDT algorithms have several advantages over other machine learning algorithms for FOG prognostication. First, GBDT algorithms are able to learn complex relationships between the features and the target variable (FOG). Second, GBDT algorithms are relatively robust to overfitting. Third, GBDT algorithms are relatively easy to interpret, which is important for clinical applications.

Three variants of GBDT were used namely:

1. CatBoost Classifier
2. LightGBM Classifier
3. XBBoost Classifier

By training all the three variants using on the accelerometer data we found that CatBoost Classifier gave the most accurate answers and therefore , CatBoost Classifier was used to give the final results

	F1 score	Accuracy
CatBoostClassifier	0.91844	0.97687
XGBClassifier	0.8737	0.96460
LGBMClassifier	0.77167	0.93914

Figure 3 F1score and Accuracy Comparision between three algorithms

Working of CatBoost Classifier

CatBoost Classifier is a machine learning algorithm for classification tasks. It is based on gradient boosting over decision trees, but it has a number of improvements that make it more efficient and accurate than other gradient boosting algorithms.

One of the key features of CatBoost Classifier is that it can handle both categorical and numerical features. Categorical features are features that have a finite number of possible values, such as "red", "green", and "blue". Numerical features are features that can take on any real value, such as height, weight, and age.

CatBoost Classifier also uses a number of other techniques to improve its performance, such as:

- Ordered boosting: CatBoost Classifier uses an ordered boosting scheme, which means that the trees are trained in a specific order. This helps to improve the accuracy of the model.
- Oblivious trees: CatBoost Classifier uses oblivious trees, which are trees that do not depend on the order in which the features are presented. This makes the model more robust to noise in the data.
- Approximate greedy splitting: CatBoost Classifier uses an approximate greedy splitting algorithm to find the best split for each node in the tree. This algorithm is much faster than the exact greedy splitting algorithm, and it produces similar results.

Chapter 3
Implementation

3.1 Implemented Functionality

A machine learning pipeline for predicting Parkinson's freezing of gait (FOG) and transcranial direct current stimulation (tDCS) FOG (tdcsFOG). The pipeline works by first merging data from three different sources:

- **Subject data**: This data contains information about the subjects, such as their age, sex, and years since diagnosis.
- **Defog data**: This data contains information about the subjects' defog scores.
- **TDCSFOG data**: This data contains information about the subjects' tdcsfog scores.

Once the data has been merged, it is preprocessed to convert categorical data to numerical data and to drop any unnecessary columns. The preprocessed data is then analyzed to identify relationships between the different variables.

Finally, a machine learning model is trained on the preprocessed data to predict FOG and tdcsFOG. The model is evaluated using cross-validation to prevent overfitting.

The pipeline was designed to be general and flexible, so that it can be easily adapted to different datasets and prediction tasks. The following are some of the key features of the pipeline:

- **Data merging**: The pipeline uses a powerful data merging function to combine data from multiple sources. This makes it easy to create complex datasets that would be difficult or impossible to create manually.
- **Data preprocessing**: The pipeline includes a variety of data preprocessing functions to convert categorical data to numerical data, to drop unnecessary columns, and to scale the data. This ensures that the data is in a format that is suitable for machine learning algorithms.
- **Data analysis**: The pipeline includes a variety of data analysis functions to identify relationships between the different variables in the data. This information can be used to improve the performance of the machine learning model.
- **Model training and evaluation**: The pipeline uses a cross-validation strategy to train and evaluate the machine learning model. This helps to prevent overfitting and to produce a more reliable model.

The pipeline was implemented using the following Python libraries:

- Pandas: A library for data manipulation and analysis.
- Scikit-learn: A library for machine learning.
- CatBoost: A library for gradient boosting machines.

The pipeline can be used to predict FOG and tdcsFOG for new subjects by passing their subject data, defog data, and tdcsfog data to the model. The model will then predict the probability of FOG and tdcsFOG for the new subjects.

The pipeline is a valuable tool for researchers and clinicians who are working to develop new methods for predicting and preventing FOG and tdcsFOG.

3.2 Results and Reports

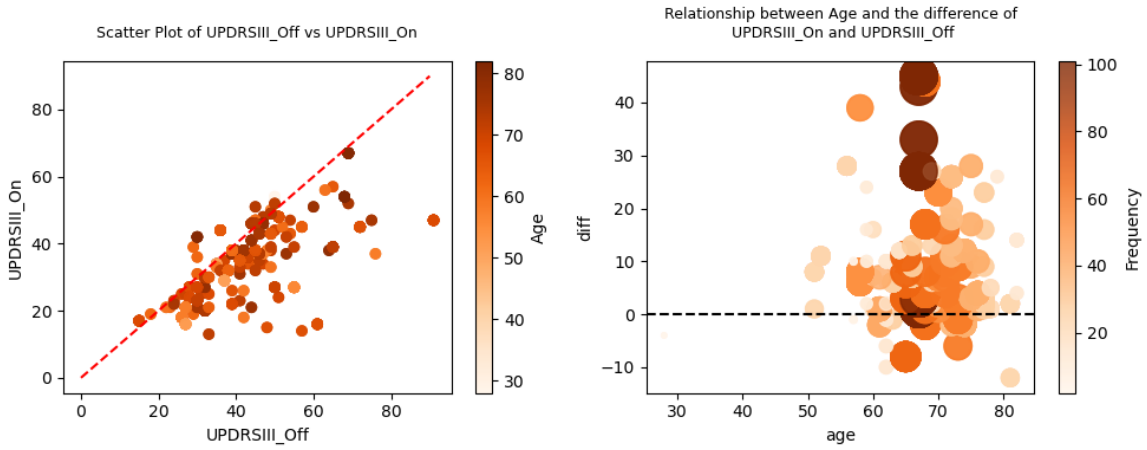


Figure 3.1[left] Scatter plot of UPDRSIII_Off vs UPDRSIII_On [right] Relationship between Age and difference of UPDRSIII_On and UPDRSIII_Off

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.

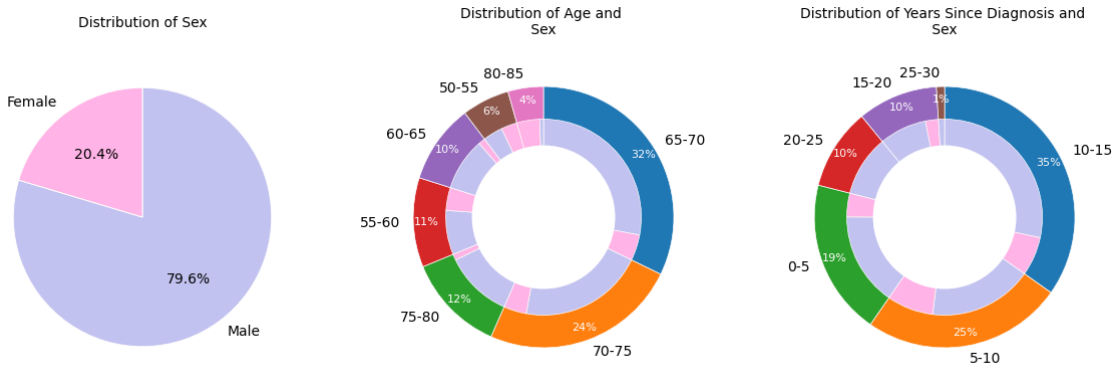


Figure 3.2 [left] Distribution of Sex [center] Distribution of Age and Sex [right] Distribution of Years Since Diagnosis and Sex

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.

```
Average of UPDRSIII_Off in women: 39.55
Average of UPDRSIII_On in women: 31.87
Average of difference between UPDRSIII_On and UPDRSIII_Off in women: 7.72
Ineffective medication cases in women: 12.24
```

Figure 3.3 Details of UPDRSIII values in women

```
Average of UPDRSIII_Off in men: 44.16
Average of UPDRSIII_On in men: 36.03
Average of difference between UPDRSIII_On and UPDRSIII_Off in men: 8.92
Ineffective medication cases in men: 9.17
```

Figure 3.4 UPDRSIII values in Men

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

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

1. There are no duplicate rows detected in the dataset

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

3. 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.

4. 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.

5. 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

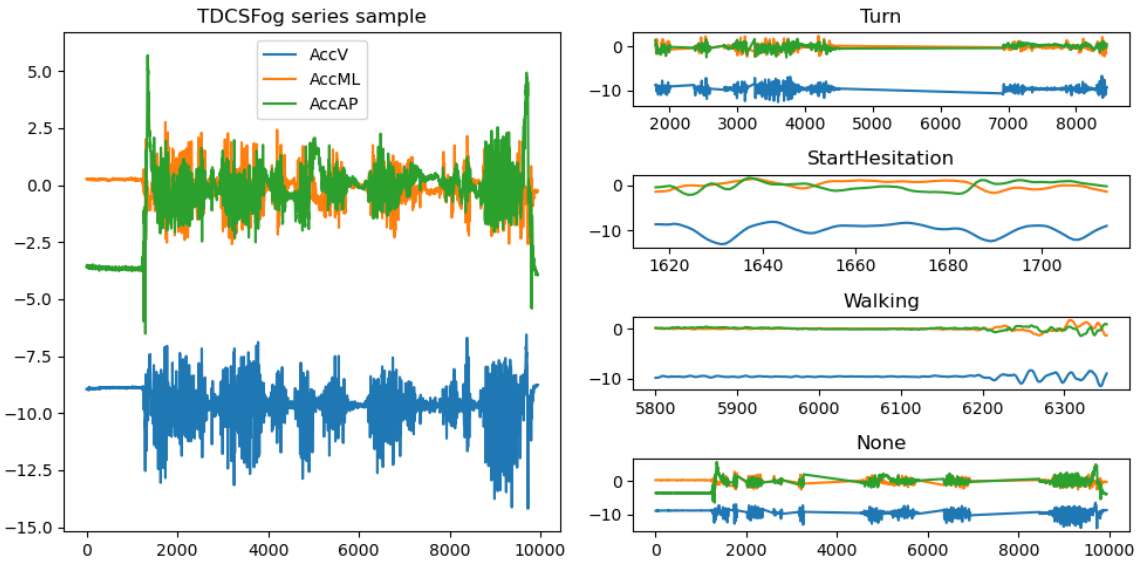


Figure 3.5 Variation of acceleration data during various events

In a 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.

From the comparative study of LightGBM , Catboost and XGBoost it turned out that CatBoost Classifier gave the best accuracy and f1 score in the training phase.

	F1 score	Accuracy
CatBoostClassifier	0.91844	0.97687
XGBClassifier	0.8737	0.96460
LGBMClassifier	0.77167	0.93914

Figure 4 F1score and Accuracy Comparison between the algorithms

Model Evaluation

Since CatBoost Classifier, gave the best accuracy and f1 score in the training phase, it was selected for the training on the entire dataset which gave the following results

```
Average precision of StartHesitation associated with all folds: 0.0001
Average precision of Turn associated with all folds: 0.3865
Average precision of Walking associated with all folds: 0.0450
Overall precision of all folds: 0.1439
```

Figure3.7 Precision of startHesitation,Turn,walking for DeFOg dataset using CatBoost Classifier

```
Average precision of StartHesitation associated with all folds: 0.6476
Average precision of Turn associated with all folds: 0.7098
Average precision of Walking associated with all folds: 0.1117
Overall precision of all folds: 0.4897
```

Figure3.8 Precision of startHesitation,Turn,walking for tdcsfog Dataset using CatBoost Classifier

Final Predictions

	Id	StartHesitation	Turn	Walking
0	003f117e14_0	0.000156	0.000983	0.000064
1	003f117e14_1	0.000156	0.000922	0.000064
2	003f117e14_2	0.000156	0.000922	0.000064
3	003f117e14_3	0.000156	0.000922	0.000064
4	003f117e14_4	0.000156	0.000922	0.000064

Figure 3.9 Final prediction of StartHesitation,Turn & walking on unlabeled data using CatBoost

Chapter 4
Conclusion

4.1 Summary of the results

During model selection phase, CatBoost Classifier gave the most accurate results. This must have been due to its tree symmetry being symmetrical or balanced and being the only one algorithm that uses greedy method as its splitting method.

In model Evaluation , we find that the precision value in the Defog dataset across all the fold is very low compared to that of tDCS dataset owing to the lesser sampling rate with which it was measured since this was measured at the patients home as compared tDCS dataset which is measured at a high sampling rate in the laboratory using advanced machines.

4.2 Advantages of work

Enhanced Quality of Life: FOG prediction can significantly improve the quality of life for individuals with Parkinson's disease by reducing the sudden and debilitating nature of FOG episodes, enabling them to carry out daily activities with greater confidence.

Fall Prevention: Predicting FOG can help prevent falls, which are common and dangerous for Parkinson's patients. This leads to a reduced risk of injuries and increased safety.

Personalized Care: Machine learning models offer personalized recommendations and interventions based on an individual's movement patterns, tailoring care to their specific needs and potentially increasing the effectiveness of treatment.

Early Intervention: Early warnings of impending FOG episodes enable individuals to take preventive actions, such as changing their movement, which is often more effective than responding after FOG has occurred.

Data-Driven Healthcare: The approach showcases the power of data-driven healthcare, leveraging large datasets to identify patterns and insights that may not be apparent through traditional methods, leading to more effective diagnosis and management of Parkinson's disease.

Advancements in Machine Learning: The work challenges machine learning to handle imbalanced and noisy data, pushing the boundaries of the field and fostering innovation in algorithms and techniques applicable to healthcare and beyond.

Accessible Technology: As these predictive models mature, they can become more user-friendly and accessible to individuals with Parkinson's disease, integrating seamlessly into their daily lives and making self-management easier.

Public Health Impact: Reducing the frequency of FOG episodes and related injuries can have a broader public health impact by lowering the healthcare and economic burden associated with Parkinson's disease and potentially improving the lives of a significant patient population.

Increased understanding of Parkinson's disease: By studying the factors that are associated with start hesitation, turn, and walking problems, CatBoost models could help to improve our understanding of the underlying causes of Parkinson's disease. This could lead to the development of new and more effective treatments for the disease.

Overall, CatBoost is a promising algorithm for predicting start hesitation, turn, and walking problems in Parkinson's disease. It has the potential to improve clinical care, reduce healthcare costs, and increase our understanding of the disease.

4.3 Scope of future work.

The future scope of predicting Parkinson's FOG episodes is characterized by a growing integration of wearable devices and mobile apps for real-time monitoring, offering individuals with Parkinson's proactive means to prevent FOG. These technologies may find application in telemedicine, facilitating remote patient monitoring and personalized care delivery. Advances in machine learning, including deep learning and multi-sensor integration, are expected to yield more accurate and robust FOG prediction models. Collaboration and data sharing among researchers and healthcare institutions can accelerate progress and enable a broader global impact, especially in underserved regions. Clinical integration of FOG prediction tools may become more common, aiding healthcare providers in decision-making. Patient empowerment and user-centric design will remain central, enhancing the overall well-being of those with Parkinson's disease and offering a brighter outlook for the management of this condition.

4.4 Unique Features (IDP/UDP)

The unique feature of using CatBoost and predictive models for start, hesitation, turning, and walking in Parkinson's FOG prediction offers improved precision, thanks to CatBoost's

robustness in handling categorical data and imbalanced datasets. It enables multiclass prediction, providing a comprehensive view of an individual's movement patterns during different activities. This, in turn, allows for customized interventions and real-time feedback, empowering individuals to take immediate action tailored to their specific context, ultimately enhancing the effectiveness of FOG prevention during various activities and improving their quality of life.

4.5 Attainment of POs and PSOs

PO / PSO	Attainment Level	Justification
PO1	3	Students demonstrated a strong understanding of the fundamental concepts of engineering through their performance in coursework, projects, and exams
PO2	2	Students showed the ability to identify, analyze, and solve engineering problems of varying complexity. However, there is room for improvement in their ability to apply engineering principles to real-world problems.
PO3	2	Students demonstrated the ability to design and develop solutions to engineering problems, but they often needed guidance from instructors.
PO4	2	Students showed the ability to conduct experiments and collect data, but they often struggled to analyze and interpret the results.
PO5	2	Students were able to use research-based knowledge to solve engineering problems, but they often needed help from instructors in identifying and evaluating relevant research literature.
PO6	3	Students demonstrated a strong ability to apply mathematical and scientific principles to solve engineering problems..
PO7	2	Students showed the ability to design systems, components, or processes, but they often needed guidance from instructors to ensure that their designs were feasible and met all of the requirements.
PO8	2	Students were able to identify, formulate, and solve engineering problems, but they often struggled to develop creative and innovative solutions.
PO9	3	Students demonstrated a strong understanding of professional and ethical responsibility. This was evident in their participation in class discussions,

		their work on team projects, and their responses to ethical case studies.
PO10	2	Students showed an understanding of the impact of engineering solutions in a societal context, but they often needed help from instructors to identify and evaluate the potential impacts of their designs.
PO11	2	Students were able to communicate their ideas effectively in both written and oral form. However, there is room for improvement in their ability to communicate complex technical concepts to non-technical audiences.
PO12	3	Students demonstrated the ability to work effectively in teams. This was evident in their participation in class discussions, their work on team projects, and their feedback to their peers.
PSO1	3	The project encompasses the analysis and advancement of computer-based systems within the realm of adversarial machine learning, while also addressing societal implications.
PSO2	2	The project tackles forthcoming technological challenges, but it does not specifically encompass domains such as wearable devices or augmented reality.

References

[1] 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

[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] 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

[4] https://arxiv.org/abs/2307.03475

[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] https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction/data

[9] Delval, A., Snijders, A. H., Weerdesteyn, V., Duysens, J. E., Defebvre, L., Giladi, N., & Bloem, B. R. (2010). Objective detection of subtle freezing of gait episodes in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, *25*(11), 1684–1693. https://doi.org/10.1002/mds.23159

[10] Hausdorff, J. M., Schaafsma, J. D., Balash, Y., Bartels, A. L., Gurevich, T., & Giladi, N. (2003). Impaired regulation of stride variability in Parkinson's disease subjects with freezing of gait. Experimental brain research, 149(2), 187–194. https://doi.org/10.1007/s00221-002-1354-8

[11] Hausdorff JM, Balash Y, Giladi N. Time series analysis of leg movements during freezing of gait in Parkinson's disease: Akinesia, rhyme or reason? Physica A: Statistical Mechanics and its Applications 2003;321:565–70.

[12]. Nantel, J., de Solages, C., & Bronte-Stewart, H. (2011). Repetitive stepping in place identifies and measures freezing episodes in subjects with Parkinson's disease. Gait & posture, 34(3), 329–333. https://doi.org/10.1016/j.gaitpost.2011.05.020

[13] Nieuwboer, A., Dom, R., De Weerdt, W., Desloovere, K., Janssens, L., & Stijn, V. (2004). Electromyographic profiles of gait prior to onset of freezing episodes in patients with Parkinson's disease. Brain : a journal of neurology, 127(Pt 7), 1650–1660. https://doi.org/10.1093/brain/awh189

[14] Moore, O., Peretz, C., & Giladi, N. (2007). Freezing of gait affects quality of life of peoples with Parkinson's disease beyond its relationships with mobility and gait. Movement disorders : official journal of the Movement Disorder Society, 22(15), 2192–2195. https://doi.org/10.1002/mds.21659

[15] Moore, S. T., Dilda, V., Hakim, B., & Macdougall, H. G. (2011). Validation of 24-hour ambulatory gait assessment in Parkinson's disease with simultaneous video observation. Biomedical engineering online, 10, 82. https://doi.org/10.1186/1475-925X-10-82

[16] Moore, S. T., MacDougall, H. G., & Ondo, W. G. (2008). Ambulatory monitoring of freezing of gait in Parkinson's disease. Journal of neuroscience methods, 167(2), 340–348. https://doi.org/10.1016/j.jneumeth.2007.08.023

[17] Zach, H., Janssen, A. M., Snijders, A. H., Delval, A., Ferraye, M. U., Auff, E., Weerdesteyn, V., Bloem, B. R., & Nonnekes, J. (2015). Identifying freezing of gait in Parkinson's disease during freezing provoking tasks using waist-mounted accelerometry. Parkinsonism & related disorders, 21(11), 1362–1366. https://doi.org/10.1016/j.parkreldis.2015.09.051

[18] 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

Appendix

Review Card I

Marwadi University
Marwadi Chandarana Group

FACULTY OF TECHNOLOGY
Department of CE – AI
Major Project-I

Group Id - 12	Team Size: 2	**Project Type:** IDP/UDP
Student Name	Disha Parmar	
Student Enrollment No.	92110151005	
Name of Internal Guide	Dr. Madhu Shukla	
Title of Project	Freezing of Gait Prognostication in Parkinson's Disease	
Name of Industry (if IDP)	Healthcare	

Performance Evaluation (Poor: 1 | Average: 2 | Good: 3 | Very Good: 4 | Excellent: 5)

Faculty Name		
Dr. Madhu Shukla.	5	Good Work.

Remarks / Suggestions:

Reviewer Faculty Name	Suggestions	Faculty Signature
Vipul. Ladva.	- Try to communicate one paper.	

Marwadi University
Marwadi Chandarana Group

FACULTY OF TECHNOLOGY
Department of CE – AI
Major Project-I

Group Id - 12	Team Size: 2	**Project Type:** IDP/UDP
Student Name	Shivam Zala	
Student Enrollment No.	92120151002	
Name of Internal Guide	Dr. Madhu Shukla	
Title of Project	Freezing of Gait Prognostication in Parkinsonis Disease	
Name of Industry (if IDP)	Healthcare	

Performance Evaluation (Poor: 1 | Average: 2 | Good: 3 | Very Good: 4 | Excellent: 5)

Faculty Name		
Dr. Madhu Shukla	5	Good Work!.

Remarks / Suggestions:

Reviewer Faculty Name	Suggestions	Faculty Signature
Vipul.Ladva	- Try to communicate atleast one research paper.	

Research Paper

Freezing of Gait Prognostication in Parkinson's Disease

Disha Parmar
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
dishaparmar350@gmail.com

Shivam Zala
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
zalashivam@gmail.com

Madhu Shukla
Computer Engineering-Artificial Intelligence Department
Marwadi University
Rajkot, Gujarat, India
madhu.shukla@marwadieducation.edu.in

***Abstract*— Parkinson's disease affects a staggering 7 to 10 million individuals worldwide, with a significant portion grappling with the debilitating symptom known as freezing of gait (FOG). FOG is a pervasive challenge, afflicting approximately 50% of Parkinson's patients and a staggering 80% of those in advanced stages of the disease. It's a complex phenomenon, intertwining motor, cognitive, and affective factors, profoundly diminishing mobility and infringing upon independence. Despite numerous theories proposed by researchers to elucidate the origins, timing, and affected individuals of FOG, the causal factors remain elusive. A key to advancing our comprehension and treatment of this symptom lies in the ability to quantify FOG objectively and precisely. The systematic collection and meticulous analysis of FOG events hold the potential to pave the way for novel therapeutic approaches. In our endeavor, we harnessed a comprehensive dataset comprising 3D accelerometer data from the lower back of subjects experiencing freezing of gait episodes. Our primary goal is to employ cutting-edge machine learning models to discern the onset and cessation of each freezing episode, while also classifying them into three distinct types: Start Hesitation, sTurn, and Walking. This research strives to shed light on the intricate world of FOG, offering hope for more effective treatments and improved quality of life for individuals living with Parkinson's disease.**

I. 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]

II. Aim & Objective

A. 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.

B. 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.

III. 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]*

A. Data Description

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

The tDCS FOG (tdcsfog) dataset, collected in the lab, as participants completed a FOG provoking protocol

The DeFOG (defog) dataset, collected in the participant's home, as subjects completed a FOG-provoking protocol

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 ($\frac{m}{s^2}$) for tdcsfog data and g (9.81 $\frac{m}{s^2}$) for defog data. Additionally, the tdcsfog dataset was recorded at 128 Hz, while the defog dataset was recorded using a 100 Hz time resolution.

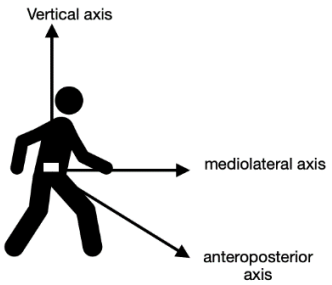


Figure 1 Vertical , Mediolateral , Anteroposterior Axes

B. Literature Review

- 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]
- 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]
- 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]
- 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]
- 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]
- 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]

IV. 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.

A. 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.

B. Data Preprocessing and Cleaning

Renaming Columns: Column renaming was executed to ensure consistency and clarity within the dataset. Notable column renames include "Subject" to "subject" and "Visit" to "visit."

Merging Data Frames: Data integration [3] was achieved by merging different dataframes, including subjects_df, defog_metadata, and tdcsfog_metadata, using the "inner" join method for streamlined data consolidation.

Data Cleaning: To maintain data integrity, missing data were thoughtfully handled through imputation or removal, and outliers were addressed to enhance data quality.

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."

C. 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

Figure 2 Merged Dataset metadata

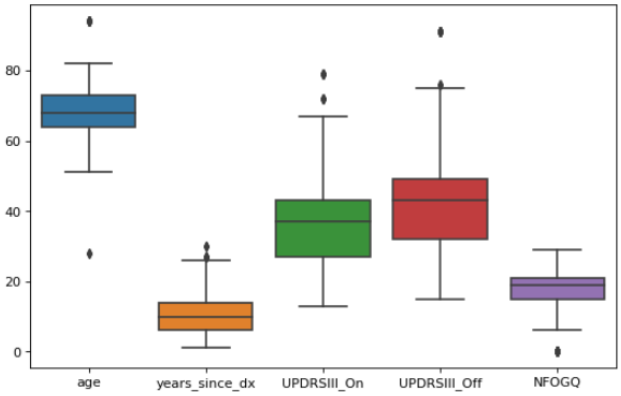


Figure 3 Box Plot of the full metadata

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.

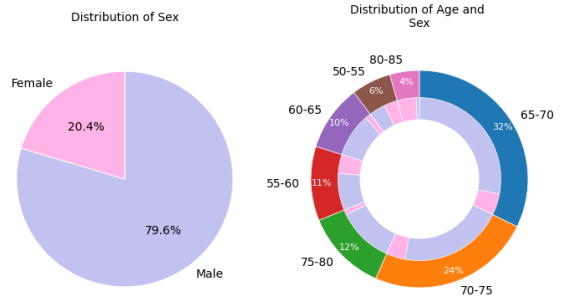


Figure 4 (i) Distribution of sex (ii) Distribution of Age and Sex

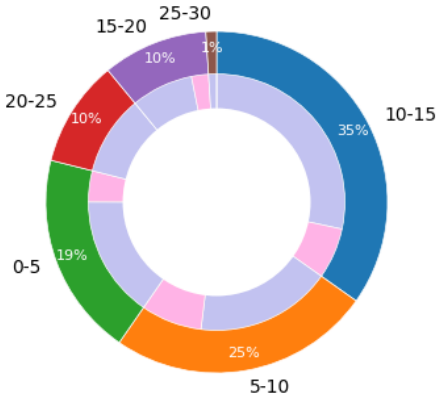


Figure 5 Distribution of years since Diagnosis and Sex

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.

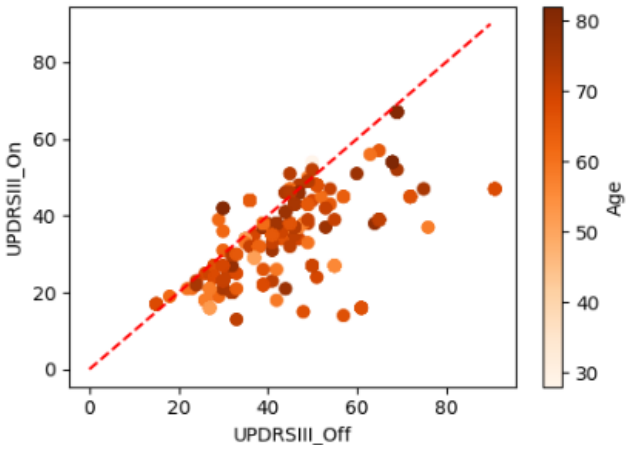


Figure 6 Scatter plot of UPDRSIII_off vs UPDRSIII_On

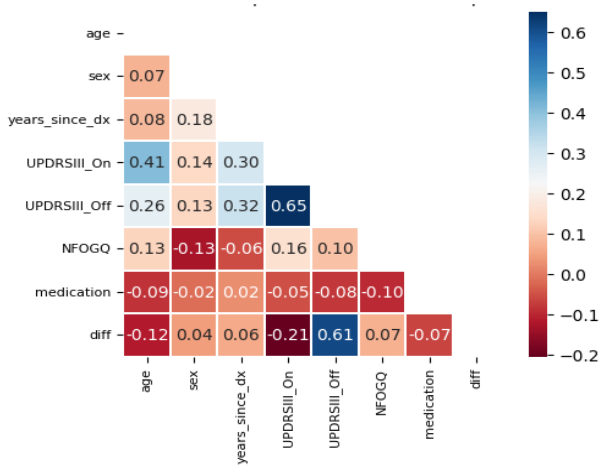


Figure 7 Full metadata pairwise Correction Heatmap

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.

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.

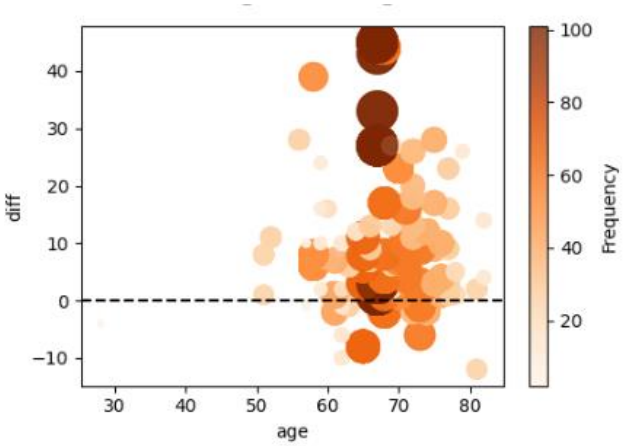


Figure 8 Relationship between Age and the difference of UPDRSIII_off and UPDRSIII_On

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.

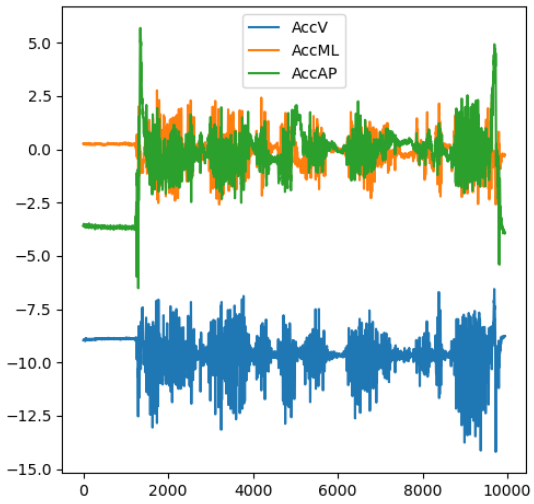


Figure 9 TDCDFog Series sample

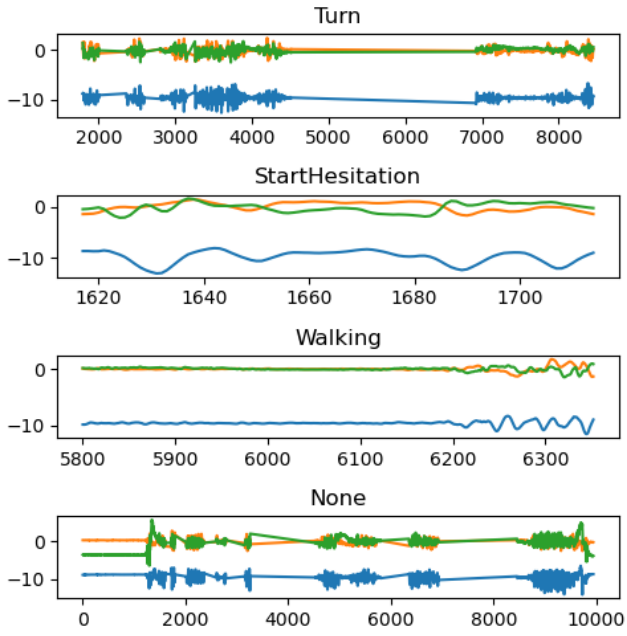


Figure 10 Variation of acceleration data during various events

In a 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.

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. There are no duplicate rows detected in the dataset
2. All target variables are highly imbalanced, especially StartHesitation (78.5%) and Walking (80.1%).
3. 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.
4. 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.
5. 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. s

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] 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.