Structural MRI-based Brain Connectivity Biomarkers for Parkinson's Disease

A Thesis

Submitted in Partial Fulfillment of the Requirements for the award of Degree of

DOCTOR OF PHILOSOPHY

by

TANMAYEE SAMANTARAY
(Registration No: 176106113)

Under the supervision of

DR. COTA NAVIN GUPTA

Department of Biosciences and Bioengineering
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
Guwahati – 781039, Assam, India
September, 2023

Dedicated to

My parents

Mr. Pradipta Kishore Samantaray & **Mrs. Harapriya Samantaray**

for being my pillar of strength

My brother
Mr. Tanmay Samantaray

&

My huband
Mr. Manoranjan Mantri

for their love and support

&

My cute little daughter

Ms. Kritika Samaira

for her endless sacrifices

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

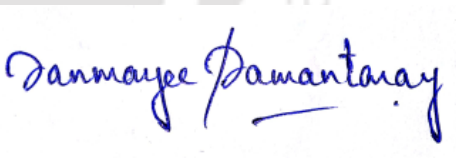
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI- 781039, ASSAM, INDIA

STATEMENT

I do hereby declare that the research findings of this thesis is the original result of research work carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India, under the supervision of Dr. Cota Navin Gupta.

As per the general norms of reporting research findings, due acknowledgements have been made, wherever the research findings of other researchers have been cited in this thesis.



Date: 08th Sept 2023

TANMAYEE SAMANTARAY
(Registration No. 176106113)

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI- 781039, ASSAM, INDIA

CERTIFICATE

It is certified that the work described in this thesis entitled "**Structural MRI-based Brain Connectivity Biomarkers for Parkinson's Disease**" by **Ms. Tanmayee Samantaray** for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision and guidance in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Assam, India. This thesis has fulfilled all requirements as per regulations of the institute and in my opinion has reached the standard needed for submission to Indian Institute of Technology Guwahati, Assam, India. The findings of this research work has not been submitted elsewhere for the award of any degree.

Date: 08th Sept 2023

CERTIFIED

DR. COTA NAVIN GUPTA

(Thesis Supervisor)
Assistant Professor
Department of Biosciences and Bioengineering,
Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India

ACKNOWLEDGEMENTS

My PhD is not just a degree, but a precious chapter of my life. A chapter, which taught me a lot many things, besides domain knowledge. It was full of ups (gyrus) and downs (sulcus), but now when I pen down this section of my thesis, I realize it was a worth experience. It's an honour to recall all those whose help has made this scientific journey a successful one.

First and foremost, I express my sincere gratitude and heartfelt thanks to my PhD supervisor **Dr. Cota Navin Gupta**, Assistant Professor, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, for his tremendous support, guidance and encouragement over these years. Thank you very much for your valuable discussions, including manuscript till code verification, which greatly improved my work. You always provided great ideas and valuable suggestions with all your experiences. Apart from focussing on my research in lab, you encouraged me to participate in national and international conferences, workshops and talks. Most importantly, you provided exposure to international hackathons and challenges, where not only did I learn new things, but also ranked among top positions and hence gained confidence in myself. I am truly grateful for the time you invested to ensure that my research was conducted with a high standard and my thesis is best shaped.

I am also grateful to all the members of my doctoral committee: **Prof. Shyamanta Moni Hazarika**, (Professor, Department of Mechanical Engineering, & Chairman of my doctoral committee), **Prof. Bithiah Grace Jaganathan** (Professor, Department of Biosciences and Bioengineering) and **Dr. V. Vijaya Saradhi** (Associate Professor, Department of Computer Science and Engineering), for evaluating my work and providing me with valuable suggestions and feedback, which helped me improve my quality of research over the years.

I would like to thank the present and former Heads of Department of Biosciences and Bioengineering (**Prof. Rakhi Chaturvedi** (present HoD), **Prof. Latha Rangan** and **Prof. Kannan Pakshirajan**) for providing all necessary facilities from the department. I especially acknowledge the computational and library facility provided to me during the entire tenure. I am also thankful to the office staff members of Department of BSBE for

their timely help. I also thank all the course instructors who imparted the best knowledge in different subjects during my coursework.

I would also acknowledge **Ministry of Education** (formerly, Ministry of Human Resource Development), Government of India, for the financial assistance provided over the years.

I also thank **National Institute of Mental Health and Neurosciences** (NIMHANS) and **Parkinson's Progression Markers Initiative** (PPMI) for providing me with the clinical and neuroimaging data of the subjects involved in my study. NIMHANS provided with healthy and Parkinson's data of subjects from Indian, and the data were recorded at NIMHANS, Bangalore, India. PPMI is a publicly-available large international database which provides information of healthy and Parkinson's individuals recorded at multiple centres. I express my heartfelt thanks to the **patients** and their families, who indirectly participated in this study. I genuinely hope your contribution in this study will have a significant impact in identification of Parkinson's subtypes and future treatment. You will have never-ending impact in scientific community and our society as a whole.

I thank **Prof. Jitender Saini** (Professor, Department of Neuroimaging and Interventional Radiology, NIMHANS, India) with whom I worked and learnt over various discussions.

It's my pleasure to thank all the present and previous members of **Neural Engineering Lab** for being really nice and giving a healthy work environment. I would like to thank Doli, Srihari, Shwetank and Abhishek for developing the lab during the early days and being always supportive to me. I specially thank Aditya for organizing the clinical data for further use in my research work. Besides the ones mentioned here, I also thank Navarun, Suraj, Utsav, Manish, Shreyansh, Shivakumar, Nanaki, Vishnu, Rekshand, Manav, Samyak and Arup for their feedback and discussions during various lab meetings.

My heart-felt thanks to my friends Ranjana, Shilpi, Khyati Raina and Deepshikha for supporting and constantly motivating me through the rough phase. I am very fortunate to have great friends who held my hands to overcome all the small or big problems in life. Thank you Akshit, Emte, Yogesh, Shanta, Aditee, Adeline, Shrusti, Praveena, Cathy, Vimal, Rosh, Sandy for your unconditional love and support.

Thanks to all the people who made my life enjoyable during my stay at IIT Guwahati. A big thanks to IIT Guwahati for enhancing my daily life with its natural beauty and stunning views.

Finally, I would like to convey my endless gratitude to my parents (Bou and Bapa) and my brother (**Mr. Tanmay Samantaray**) for your moral support throughout. My mother, **Mrs. Harapriya Samantaray,** showered me with her love and prayers and stood by me all the time. My father, **Mr. Pradipta Kishore Samantaray**, is my role model and my inspiration. Words won't be enough that you all have made numerous personal sacrifices and have taken care of my daughter so well that I can focus and complete my research with dedication. I am forever grateful to my husband, **Mr. Manoranjan Mantri,** who always had my back, through thick and thin. Thank you for clarifying many concepts on signals which still helps me a lot.

My special thanks to my cute little daughter, **Ms. Kritika Samaira**, for sacrificing her childhood without me and living far from me. Thanks for enquiring me on my work update and telling me to successfully complete my PhD, everytime I speak to you (though you don't understand anything). Thanks for being an amazing daughter. I am running in short of words to convey my deep respect and gratitude towards my family. This PhD is yours as much as mine.

Thanks to the **Almighty** for giving me strength and showing me the right path to keep going.

Tanmayee Samantaray

TABLE OF CONTENTS

Page No.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AM:	Adjacency Matrix
ANOVA:	Analysis of Variance
APD:	Atypical Parkinson's Disease
BCT:	Brain Connectivity Toolbox
BICA:	Biclustered Independent Component Analysis
CAT:	Computational Anatomy Toolbox
CBD:	Corticobasal Degeneration
CI:	Confidence interval
CSF:	Cerebro Spinal Fluid
DEC:	Decimal equivalent
DICOM:	Digital Imaging and Communications in Medicine
DLB:	Dementia with Lewy Bodies
DTI:	Diffusion Tensor Imaging
DWI:	Diffusion Weighted Imaging
FDR:	False Discovery Rate
GDS:	Geriatric Depression Scale
GIFT:	Group ICA for fMRI Toolbox
GIG-ICA:	Group Information Guided - Independent Component Analysis
GM:	Grey Matter
H&Y:	Hoehn and Yahr scale
HAM-D:	Hamilton Depression Rating Scale
HC:	Healthy control

ICA:	Independent Component Analysis
IFG:	Inferior Frontal Gyrus
ITG:	Inferior Temporal Gyrus
JICA:	Joint Independent Component Analysis
LNS:	Letter Number Sequencing
LPBA:	LONI Probabilistic Brain Atlas
MFG:	Middle Frontal Gyrus
MKNN:	Mutual K-nearest neighbor
mm:	milimeter
MNI:	Montreal Neurological Institute
MRI:	Magnetic Resonance Imaging
MSA:	Multiple System Atrophy
MTG:	Middle Temporal Gyrus
N-BiC:	N-way Biclustering method
NIfTI:	Neuroimaging Informatics Technology Initiative
NIMHANS:	National Institute of Mental Health and Neurosciences
PCA:	Principal Component Analysis
PD:	Parkinson's Disease
PDD:	Parkinson's Disease with Dementia
PPMI:	Parkinson's Progressive Markers Initiative
PSP:	Progressive Supranuclear Palsy
QUIP:	Questionnaire for Impulsive -Compulsive Disorders
RBD:	Random Eye Movement- Sleep Behaviour Disorder
rGMV:	Regional grey matter volume

ROI: Regions of Interest

rWMV: Regional white matter volume

SBM: Source - Based Morphometry

SCoPA-AUT: Scales for Outcomes in Parkinson's Disease-Autonomic

SDM: Symbol-Digit Modalities score

SFG: Superior Frontal Gyrus

sMRI: Structural Magnetic Resonance Imaging

SPM: Statistical Parametric Mapping

STAIT: State-Trait Anxiety Inventory for Adult Test

STG: Superior Temporal Gyrus

T (1.5/3): Tesla

T1-w: T1 weighted

UBNIN: Unique Brain Network Identification number

UPDRS: Unified Parkinson's Disease Rating Scale

UPSIT: University of Pennsylvania Smell Identification Test

VBM: Voxel Based Morphometry

WM: White Matter

α-Syn: Alpha Synuclein

ABSTRACT

Individuals, whether healthy or disease, have a specific pattern of brain network due to inter-individual differences. So, now the question is- Can a brain network be encoded? To address this, a novel algorithm called Unique Brain Network Identification Number (UBNIN) was proposed for encoding brain networks of individual subject. To realize this objective, each subject's brain volume was parcellated from structural MRI scans and individual adjacency matrix was constructed using correlation between grey matter volume of every pair of regions. The numerical representation (UBNIN) was observed to be distinct for each individual brain network, which may also be applied to other neuroimaging modalities. This model may be implemented as neural signature of an individual's unique brain connectivity, thereby useful for brainprinting applications.

Also, Parkinson's disease (PD) is the second-most common neurodegenerative disorder, causing fatal consequences with severity and duration. Additionally, it is considered as a disconnection syndrome. Hence, first, patterns of disconnectivity were tested in PD patients from the study data, and then, all PD patients were segregated into five age-cohorts to study the variation in network topology over age. For each age-cohort, a decreasing trend was observed in mean clustering coefficient with increasing sparsity. Significantly different clustering coefficient was noted in PD between different age-cohorts. The findings suggest network connectivity patterns change with age, indicating network disruption possibly due to the underlying neuropathology. Varying clustering coefficient for different cohorts indicate that information transfer between neighboring nodes change with age. This provides evidence on age-related brain shrinkage and network degeneration.

PD is often characterized with brain tissue atrophy. Its heterogeneity in clinical symptoms is clearly visible by its phenotypic representation such as overlapping motor and non-motor dysfunctions. Although several attempts have been made in the past to address this heterogeneity, a stable subtyping approach to categorize the patients, is the need of the hour. Thus, successful subtyping will not only help in clinical analysis, but also be useful in precision treatment. Hence, in this thesis, attempts were made for finding subtypes within Parkinson's disease using neuroimaging- based data-driven

approach, especially using structural magnetic resonance imaging (sMRI) data. The present study has analysed brain grey matter information to find distinct subtypes and have correlated them with the clinical features to find association between imaging and clinical features. This study also explores if the deciphered subtypes had differences in connectivity pattern amongst them. To analyse this, graph-theory based network analysis was performed based on connectivity metrics. Three subtypes were found with differences in frontal and temporal gyrus regions of brain. Inter-subtype differences in network metrics were also observed.

Furthermore, realizing the fact that PD is also associated with white matter atrophy in addition to grey matter, subtyping of PD only patients were attempted by fusing both grey and white matter tissue information in a joint ICA framework. The brain networks of these deciphered subtypes were then analysed using association matrices based on correlation between both tissue volumes. The adjacency matrices were obtained by binarizing using mutual K-nearest neighbor based thresholding. Three subtypes were found based on joint loading parameter. Further, upon associating the subtypes with the motor features, they were termed as mild PD, intermediate PD and severe PD.

The findings of this study suggest that specific brain regions may act as biomarkers for subtyping PD, as they have associations with clinical symptoms. The network metrics, significantly varying between the data-driven subtypes, indicate differences in network topology and may hence be used as connectivity biomarkers.

The investigations covered in this thesis were conducted on data obtained from an Indian cohort, *i.e.,* National Institute of Mental Health and Neurosciences (NIMHANS) and a publicly available international database, *i.e.,* Parkinson's Progressive Markers Initiative (PPMI).

Keywords: Parkinson's disease, Structural magnetic resonance imaging (sMRI), Subtyping, Brain connectivity, Unique brain network identification number (UBNIN)

CHAPTER – 1

INTRODUCTION

Chapter 1. INTRODUCTION

1.1. Parkinson's Disease -Epidemiology

Parkinson's disease (PD), first described in 1817 by Dr. James Parkinson [1], is a progressive neurological disorder characterized by loss of dopamine-producing neurons in substantia nigra pars compacta [2]. It was initially termed as "paralysis agitans" in his work "An Essay on the shaking Palsy" in 1871. It is the second-most common neurodegenerative debilitating disorder, after Alzheimer's disease, beginning in the fifth to sixth decade of life and affects a higher proportion of males [3], [4], [5]. The occurrence of Parkinson's disease (PD) has experienced a twofold increase over the last 25 years, and it has resulted in 329,000 fatalities, reflecting a rise of more than 100% since the year 2000 [6]. PD has a global prevalence of 6.1 million individuals in 2016 [7] and 8.5 million individuals in 2019. Prevalence means percentage of the PD population. **Figure 1** showing age-standardised prevalence rates of PD indicates highest rates in high-income North America and lowest rates in sub-Saharan Africa.

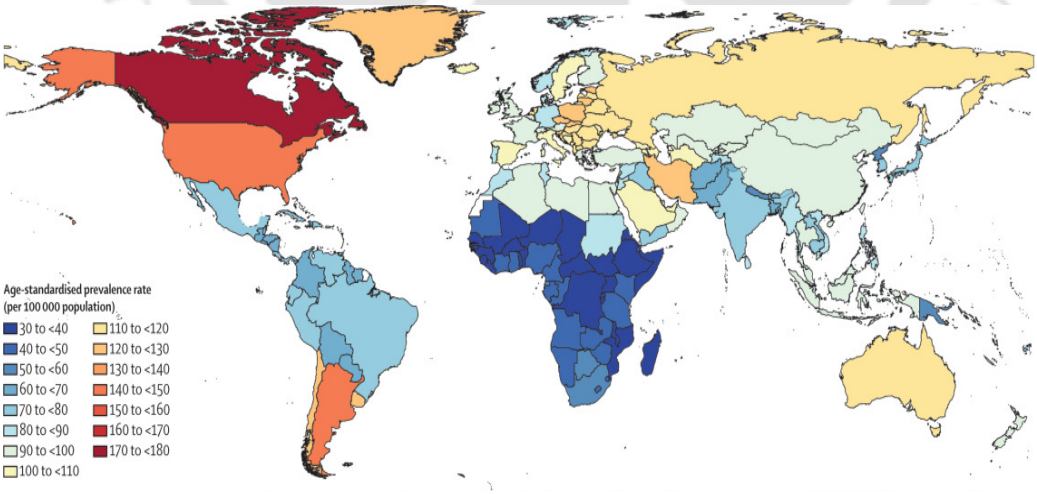


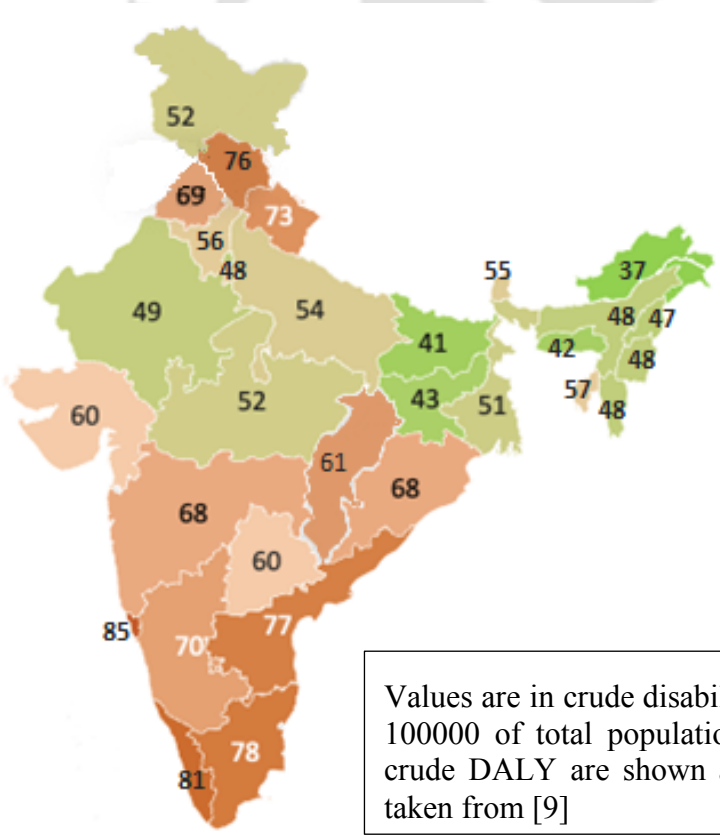
Figure 1. Global prevalence of PD per 1,00,000 populations, 2016 (Figure taken from Dorsey *et al.* [7])

Looking into Indian scenario, PD was found in a total of 770800 (95% CI: 635100-919400) people, with 34900 (95% CI: 287200-416500) females and 421800 (95% CI: 348300-502100) males as of 2019 [8]. The estimated number of deaths were 45300 (95% CI: 38600–52800).

Figure 2 shows crude disability-adjusted life year (DALY) rates in different states of India, given by equation 1, where YLL is years of life lost due to premature mortality and YLD is years of healthy life lost due to disability.

$$DALY = YLL + YLD \quad (1)$$

Additionally, PD prevalence by age and gender is shown in **Figure 3**.



Values are in crude disability-adjusted life year (DALY) per 100000 of total population at 95% CI. Indian states with crude DALY are shown as heatmap in the figure. Data is taken from [9]

Figure 2. Crude DALY rates of neurological disorders in the states of India, 2019

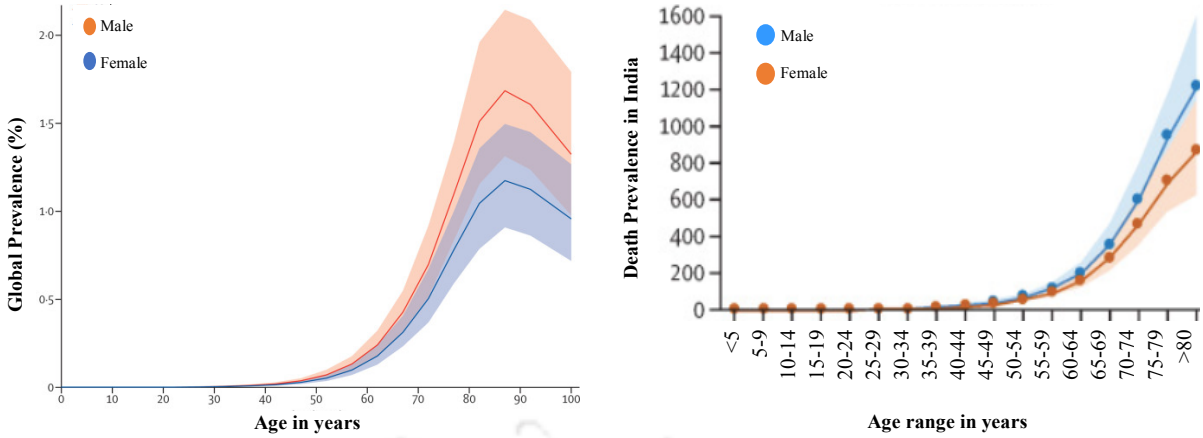


Figure 3. Age and gender-based analysis: global prevalence (Left) [7] & death prevalence in India per 100000 (Right) [8]

1.2. Clinical Symptoms

PD is clinically diagnosed based on motor dysfunctions like tremors, bradykinesia and rigidity. The associated non-motor dysfunctions are depression, rapid eye movement sleep behaviour disorder, constipation, autonomic dysfunction, pain and cognitive decline [9], [10]. PD is associated with Lewy bodies, large aggregates of *α-synuclein* [11] and tau protein [12], [13] formed in the neuronal cell bodies and axonal fibers. However, neurodegeneration in PD begins long before the onset of clinical motor symptoms [14], [15], also termed the prodromal phase. This presymptomatic phase preceding the onset of neuronal loss in Parkinson's disease is approximated to span around five years [14]. The features [16], [17], [18], [19] presented by PD patients are described as below:

1.2.1. Cardinal motor symptoms

Bradykinesia - Bradykinesia is a medical umbrella describing slow motion. Bradykinesia may result in hypomimia (decreased facial expression), hypophonia (soft voice), and swallowing issues in addition to difficulty in walking or rising from a chair. Clinical assessments in which the patient is instructed to repeat repetitive motions as quickly and widely as feasible can identify bradykinesia in milder forms of PD [20].

Rigidity - An elevated muscular tone, comprising both flexor and extensor muscles, is the cause of Parkinsonian rigidity. In contrast to spasticity, parkinsonian rigidity does not worsen with faster movement, and a common clinical finding is "cog wheel rigidity," which develops when rigidity and rest tremor coexist.

Tremor - The typical parkinsonian tremor is a low- to mid-frequency rhythmic involuntary movement that is present at rest but disappears with vigorous activity. The term "re-emergent tremor" is used when the tremor returns in a static position, such as extending out the arms. Rest tremor in Parkinson's disease (PD) is typically most noticeable in the distal portion of an extremity and may also include face (lips, chin, and/or jaw) [20].

Postural and gait impairments - Patients with advanced PD often assume an arched posture, with their neck and torso flexed, as well as a sluggish walk and short, shuffling steps. Due to stiffness, poor axial coordination, and slow reflexes, these deficits exist. The inability to move suddenly and briefly, or "freezing," is one of the most incapacitating symptoms of PD. Patients with Parkinson's disease frequently fall because of this condition, which most frequently affects the legs. To assess the balance in PD patients, the "pull test"—which involves abruptly moving the patient's shoulders backward or forward—is frequently employed in clinical procedures. Balance deficits that are clinically obvious are typically regarded as PD late-stage symptoms [20].

1.2.2. Non-motor symptoms

Autonomic dysfunction - Constipation, orthostatic hypotension, urinary dysfunction, sexual dysfunction, excessive sweating, seborrhea, and sialorrhea [21]**.**

Sensory dysfunction - Hyposmia, visual deficits (reduced contrast/color discrimination and visual motor perception), paresthesia, and pain are examples of sensory dysfunction.

Cognitive and neuropsychiatric features - Dementia is a major side effect of PD, and many patients eventually have cognitive impairment. Depressive, anxious, and apathetic symptoms are additional typical psychiatric traits [21].

Fatigue - In PD, fatigue may be linked to sleep disorders but it can also happen independent of sleep disturbances.

Sleep disorders - Rapid eye movement (REM) sleep behaviour disorder (RBD), insomnia, restless legs syndrome, periodic limb movements during sleep, and excessive daytime sleepiness are all examples of sleep disorders.

1.2.3. Complications of dopaminergic replacement therapy

Motor fluctuations - Changes between periods of effective motor symptom control (referred to as "on" time) and insufficient motor symptom control (referred to as "off" time).

Impulsive control disorders - Dopamine agonists are most frequently linked to impulsive control issues, which include hypersexuality, compulsive gambling, excessive spending, and binge eating.

Dyskinesia - Involuntary dystonic or choreiform (uncontrollable jerking) movements are known as dyskinesia and are most prevalent during levodopa peak concentrations.

Hallucinations - Both as a side effect of DRT and as a core characteristic of severe disease in PD, hallucinations are possible.

1.3. Neuropathology

Studies [22], [23] on the spread of Lewy body pathology suggested that Parkinson's disease propagates along the neuronal pathways from the brain stem to subcortical areas. This

follows cerebral hemisphere, which was in synchrony with the clinical progress of the disease. Tau protein, under physiological conditions, is highly soluble and unfolded. It plays a vital role in stabilizing microtubules, thereby regulating axonal transport [24], [25], [26], [27]. However, this tau protein hyperphosphorylation leads to their disintegration from microtubules, causing damage to neurons and axonal transport [28]. The hyperphosphorylated and insoluble *α-synuclein* aggregates as an abnormal amyloid fibrillar configuration, destabilizing the synaptic terminal. This interrupts synaptic vesicular transport in contrast to its stable form [29], [30], [31], [32]. Thus, neuropathological hallmarks of PD include aggregation of tau, *i.e. tauopathies* [33], [34], and enrichment of *α-synuclein*, *i.e. synucleinopathies* [12], [35], [36]. Tau pathology is described pictorially in **Figure 4**.

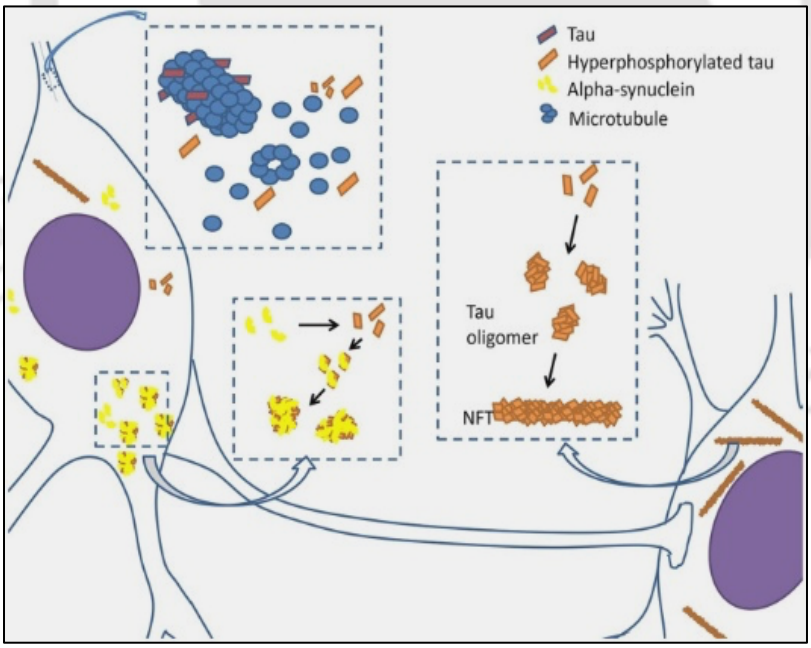


Figure 4. Tau Pathology in Parkinson's disease [27]

1.4. Genes Involved

Although the vast majority of PD instances seem to be sporadic, 10–30% of PD patients have at least one first-degree relative who has parkinsonism, with large population of young onset PD [37]. A few genes involved in the development of Parkinson's disease on different loci of PARK are: SNCA, PRKN, UCH-L1, PINK1, DJ-1, LRRK2, ATP13A2, GIGYF2, HTRA2, PLA2G6, FBXO7, VPS35, EIF4G1, DNAJC6, SYNJ1, DNAJC13, CHCHD2, VPS13C, PSAP, as registered in Online Mendelian Inheritance in Man (OMIM) database. Alpha-synuclein gene (SNCA) is responsible for autosomal dominant form of PD, and is reported since a quarter of a century [38].

1.5. Atypical Parkinson's Disease/ Parkinsonism

Atypical Parkinson's disorders (APD) or Parkinsonisms are the ones with overlapping clinical features with PD. APDs comprise a heterogeneous group of disorders, such as Synucleinopathies (multiple system atrophy (MSA), dementia with Lewy bodies (DLB)) and Tauopathies (progressive supranuclear palsy (PSP), corticobasal degeneration (CBD)) [12], [39], [40]. The predominant GM atrophy regions, reported so far, is represented (**Figure 5**) to understand and relate with the readers' primary finding.

Multiple system atrophy (MSA) [41] disables one or more body systems, and the autonomic nervous system is severely affected in the early course of the disease. DLB [42] causes progressive intellectual and functional deterioration and has a slight tremor. PSP [43] is symptomized by unexpected falls, forgetfulness and personality changes, usually beginning in the 60s. Its most feeble effect involves the inability to move eyeballs because of weakness or paralysis of the muscle. CBD was reported first by Rebeiz et al. [44] but described clinically and pathologically by Gibb and colleagues [45].

It has remained underdiagnosed with symptoms such as loss of function on one side of body, involuntary and jerky movements of limbs and speech problems. Moreover, this classification of APDs into synucleinopathies and tauopathies is based on the intracellular aggregates of the *α-synuclein* or *Tau* protein, respectively [12], [46].

1.6. Neuroimaging

Human brain structure and function can be understood using diverse neuroimaging and electrophysiological modalities like structural magnetic resonance imaging (sMRI), functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), magnetoencephalogram (MEG), diffusion - weighted magnetic resonance imaging (DW-MRI), diffusion tensor imaging (DTI), diffusion weighted imaging (DWI), etc.

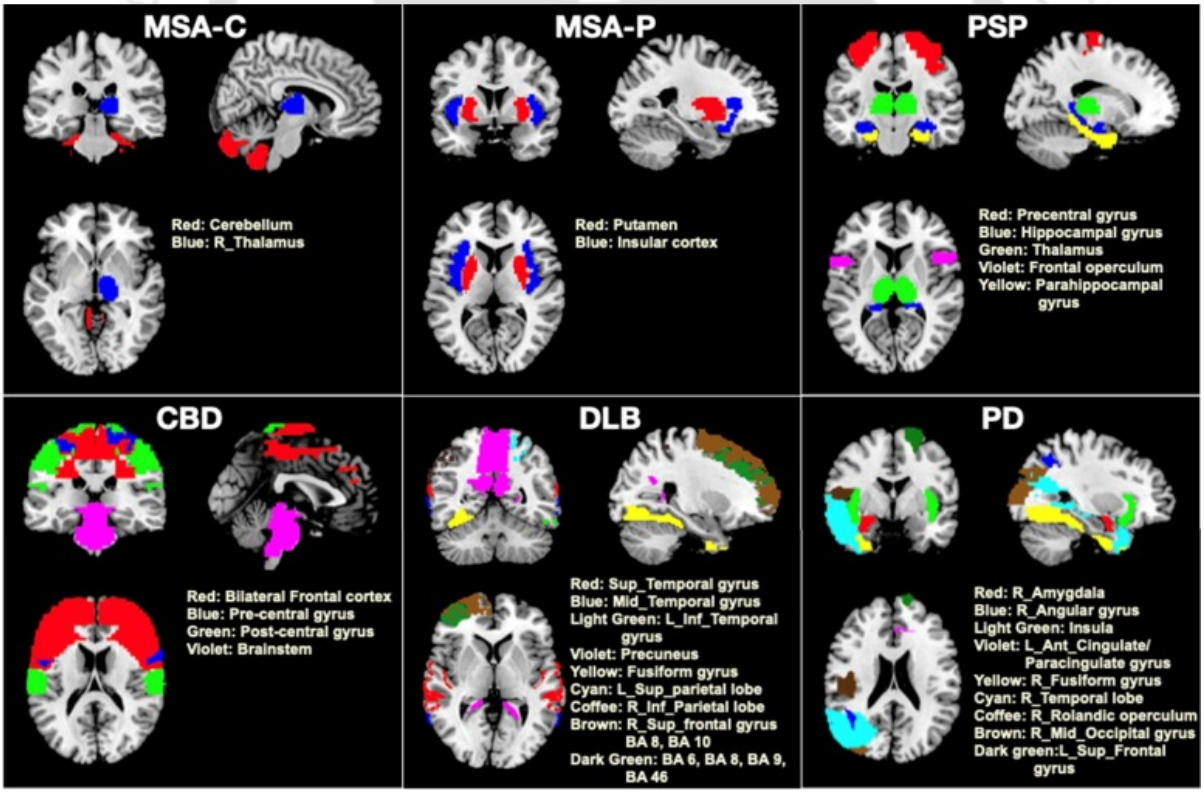


Figure 5. Brain Regions showing Grey Matter atrophy in Parkinson's affected compared to control. Different colors represent distinct atrophy regions. Ant: Anterior; BA: Brodmann Area; CBD: Corticobasal Degeneration; DLB: Dementia with Lewy Bodies; Inf: Inferior;

L: Left; Mid: Middle; MSA-C: Multiple System Atrophy-Cerebellar type; MSA-P: Multiple System Atrophy-Parkinsonian type; PD: Parkinson's Disease; PSP: Progressive Supranuclear Palsy; R: Right; Sup: Superior; This figure is constructed in MRIcron (https://www.nitrc.org/projects/mricron) [47]

1.6.1. Basics of MRI

The subject is positioned within the static magnetic field generated by the MRI scanner's magnet. Within living tissues, a multitude of hydrogen atoms are present, either within water molecules or various other molecules. Hydrogen's nucleus, known as a proton, possesses an inherent magnetization referred to as "spin." This spin magnetization vector undergoes precession (rotation) around the magnetic field, occurring at a frequency called the Larmor frequency, which correlates with the strength of the magnetic field. Consequently, the resultant magnetization of all protons within the tissues aligns in parallel with the magnetic field's direction. This parallel magnetization varies based on the magnetic field's intensity [48].

While acquiring images, the MRI scanner emits a radiofrequency (RF) pulse. When this pulse is adjusted to match the Larmor frequency, it achieves resonance: initiating a synchronized phase in the precession of all proton spins. The duration of this RF pulse is selected so as to orient the spin magnetization perpendicular to the magnetic field. Upon placing a receiving coil (an electrical conductor) near the tissue, the transverse magnetization – which continues its rotation in alignment with Larmor precession – induces an electric current within the coil. This phenomenon is recognized as the nuclear magnetic resonance (NMR) signal. The NMR signal experiences reduction due to the simultaneous operation of two relaxation processes. One of these is the loss of coherence within the spin system, leading to signal attenuation with a specific time constant, known as the transverse relaxation time (T2). Concurrently, the magnetization vector gradually returns to its

equilibrium alignment, parallel to the magnetic field; this process is governed by a time constant referred to as the spin-lattice relaxation time (T1). The distinguishing factor in MR images emerges from the reality that diverse tissues generally possess distinct T1 and T2 relaxation times. This contrast is particularly evident in soft tissues, accounting for the exceptional differentiation of soft tissue information in MRI images [48].

The spatial representation of the MRI signal is achieved by employing magnetic field gradients, which are smaller supplementary magnetic fields characterized by intensity that varies linearly according to their spatial positions. As a result, the spins of protons situated in distinct locations undergo precession at marginally varying rates. The component of the gradient coils and their related current that stands perpendicular to the primary magnetic field generates a force (known as the Lorentz force) on the coils. In this procedure, the gradients are rapidly activated and deactivated, causing them to oscillate and the generate the majority of the auditory noise encountered during the acquisition of MR images. When magnetic field gradients are applied, the resultant NMR signal encompasses diverse frequencies corresponding to distinct positions of tissue spins. This amalgamated signal is referred to as MRI signal. Further, the analog MRI signal is converted into digital format by sampling and then preserved for subsequent processing. This processing involves segregation of signal components stemming from distinct spatial positions, which are depicted as pixels within the ultimate image, through Fourier transform [48].

Thus, the primary origin used to generate almost all MRI images comes from hydrogen nuclei because of the presence of a single proton that carries a positive electrical charge. When a group of protons is exposed to an external static magnetic field, it precesses. Protons precessing parallel to the external magnetic field, sums up to a net magnetization

called as longitudinal magnetization (LM) [49]. When the subject possesses LM, RF pulses are switched on and off to disturb the protons so as to misalign the protons from external magnetic field. This occurs when the RF pulse has the same frequency as the precessional frequency of the protons, and this phenomenon is called resonance; hence the term magnetic resonance imaging [50]. Transverse magnetization (TM) is a rotating magnetic field which generates an electrical current. As soon as the RF pulse is switched off, the protons begin to fall out of phase with each other and return to a lower energy state, i.e. protons relax. So, when the LM begin to return to its original value, it's termed as longitudinal/ Time 1 (T1) relaxation and when the TM begins to disappear, it's termed as transverse/ Time 2 (T2) relaxation [49]. Adjacent structures are differentiated from one another by contrast, which is determined by signal intensities, which in turn are governed by the T1 and T2 relaxation times of the tissues within an image. The image in which the difference in signal intensity between tissues is predominantly due to differences in tissue T1 relaxation time is called T1 weighted image (**Figure 6**) and that due to T2 relaxation time is called as T2 weighted image (**Figure 7**). Moreover, T1-weighted sequences offer high spatial resolution, high contrast between grey and white matter, and the capability to be reoriented in 3D to comprehensively evaluate brain atrophy [51].

1.6.2. Structural MRI

The structural MRI (sMRI) technique, using T1 and T2 contrasts, is a simple, robust and non-invasive modality for brain imaging, and has become the most accepted standard for routine examination of brain offering visualization, detection and monitoring of morphological changes in the brain along with exquisite anatomical detail and high sensitivity to pathological changes [52]. The conventional sMRI provides both quantitati

ve and qualitative information in terms of shape, size and tissue information (grey matter (GM), white matter (WM), cerebro-spinal fluid (CSF), etc) present in the brain. MRI signal varies across tissue types because of the varied concentration of cell bodies. The GM in the brain contains more cell bodies (eg. neurons and glial cells) than WM which comprises of long-range nerve fibres/ myelinated axons along with supporting glial cells. The sMRI images can be viewed as sagittal, coronal or axial view (**Figure 6** and **Figure 7**). The detection and differential diagnosis of PD has been successful in several studies by using sMRI as the imaging modality [53], [54], [55]. Moreover, MRI technique has enabled the confirmation of mild cognitive impairments prodromal for dementia due to Lewy bodies in case of PD [56].

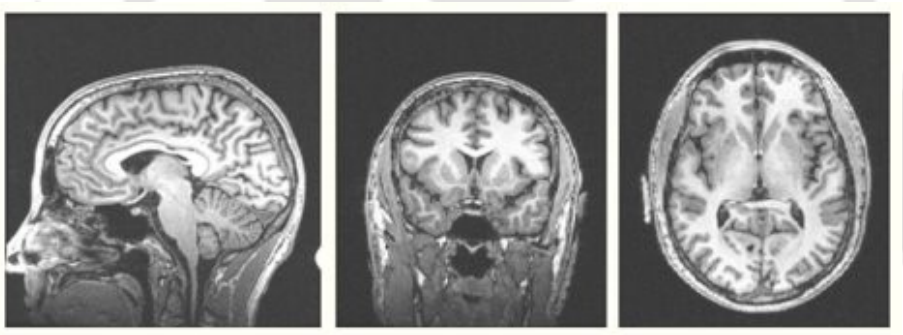


Figure 6. T1 weighted MRI image: Sagittal, Coronal, Axial View (Left to Right); Good contrast between grey matter (dark grey) and white matter (lighter grey) tissues. CSF is void of signal (black). Fat, such as lipids in myelinated white matter, appears bright [57]

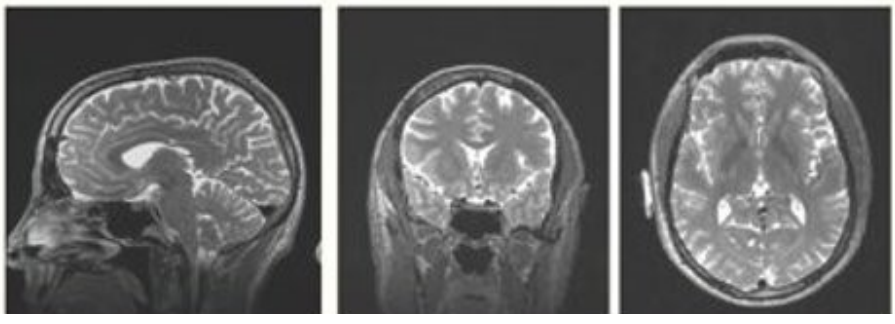


Figure 7. T2 weighted MRI image: Sagittal, Coronal, Axial View (Left to Right); Good contrast between CSF (bright) and brain tissue (dark). Some T2 sequences demonstrate additional contrast between grey matter (lighter grey) and white matter (darker grey). Fat, such as lipids in the white matter, appears dark [57]

1.7. Brain Connectivity

Another area that requires investigation in the context of Parkinson's disease is the brain network of individuals with PD.

1.7.1. Human brain as a complex network

Brain connectivity elucidates how neurons and neuronal networks process information. And the connectivity patterns play a crucial role in determining the structural and functional properties of neurons and neural systems. The underlying structural substrate and mechanistic insights need to be understood to explore the functioning of a network [58]. Although several studies have investigated functional connectivity, little is known about the network of anatomical connections. Due to limited study on subgroups of neuropsychiatric disorders, characterizing the architecture of the human brain network is essential. Construction of brain networks needs a comprehensive and detailed structural model of the neuronal units and their connections.

Structural maps indicate that each neuronal element maintains a specific pattern of connection with a few other similar and/or different neuronal elements. The complexity of the human brain is that it contains billions of neurons connected by trillions of links. The essential circuits and their functions depend on these neurons and their inter-connections. The brain functions are dependent on individual and ensemble of neurons, the neural circuits, a single brain region, and the various interactions between brain regions [47], [58].

1.7.2. Scales and levels of structural description

A brain network comprises neurons at the microscale, neural population at the mesoscale and brain regions at the macroscale, and connectivity refers to the association between these structural components. An advantage of single neurons (considered at microscale) is that

the elements are easily demarcated and well defined. These maps represent anatomical (also named morphological or structural), functional or effective connections depending on the dataset. In contrast, brain areas and neuronal populations are more difficult to delineate. Structural networks or anatomical connections typically correspond to correlation between pairs of brain regions or neuronal fibers connecting distinct regions, while functional connections correspond to statistical dependence between events of neuronal elements, and effective connections indicate the influence of one neuronal system over another at either synapses or an ensemble of neurons [59].

1.7.3. Graph theory in human brain

Graph theoretical approaches have been the basis for complex network analysis of structural and functional brain networks and quantifying the healthy and patients' neuronal systems. Graph theory represents any real-world network mathematically as a graph G = (V, E) where V is set of vertices (or nodes) and E is the set of edges (or links or connections) between the nodes. A human connectome is a map of the structural networks of the human brain, which has been attempted in several studies on brain regions as nodes [60], [61].

1.7.4. Nodes and parcellation schemes

Nodes are the basic units of a brain network and could be described as neurons or cortical areas or spatially non-overlapping regions. These anatomically-defined nodes of brain networks can be defined in many ways [62]. The simplest one is the smallest unit in the data which could be individual voxels in sMRI scans. The advantage includes no additional processing or averaging to analyze the data. Nevertheless, the disadvantage is that it may lead to an extensive network and hence is difficult to analyze. Another disadvantage is adjacent voxels may be redundant when compared with the functional aspect.

Table 1. Standard atlas in brain network studies

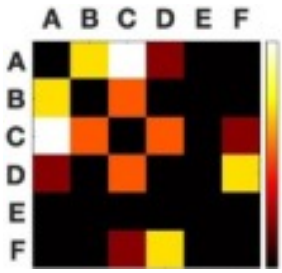
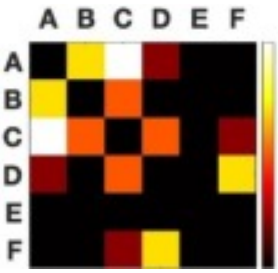
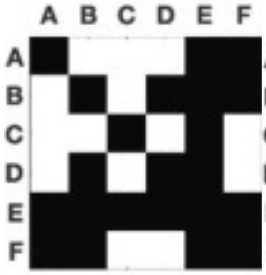
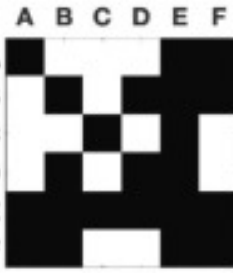
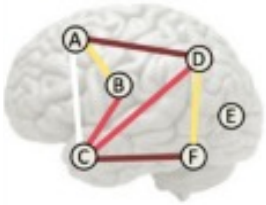
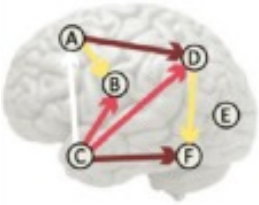
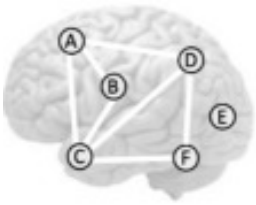
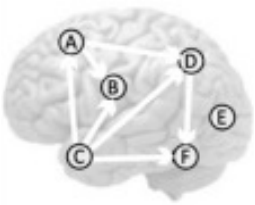
Atlas Name	Coverage	Regions	Links	Reference
Anatomical Automatic Labeling Atlas (AAL)	Whole brain	116	https://www.cyceron.fr/index.php/en/plateforme-en/freeware	[63]
Destrieux Atlas	Cerebral Cortex	148	https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation	[64]
Brainnetome Atlas	Cerebral cortex, Subcortical structures	246	https://atlas.brainnetome.org/	[65]
Brodmann Area	Cerebral cortex	104	http://www.fmriconsulting.com/brodmann/Interact.html	[66]
LONI Probabilistic Brain Atlas (LPBA)	Whole brain	56	https://resource.loni.usc.edu/resources/atlases/	[67]
Harvard Oxford Atlas (HO)	Cerebrum	69	http://neuro.debian.net/pkgs/fsl-harvard-oxford-atlases.html	[68], [69], [70], [71]
HCP Multimodal Parcellation (HCP MMP1.0)	Cerebral Cortex	360	https://balsa.wustl.edu/study/show/RVVG	[72]
Desikan-Kiliany	Cerebral Cortex	68	https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation	[68]
MarsAtlas	Cerebrum	82	http://brainvisa.info/web/index.html	[73]
Desikan-Killiany-Tourville	Cerebral Cortex	62	http://mindboggle.info/data	[74]
Indian Brain Template (IBT)	Cortical and Subcortical structures	[68], [64]	https://hollabharath.github.io/IndiaBrainTemplates/	[75]

The second method is to express nodes as brain regions based on a meta-analysis. This brings the idea of considering nodes by constructing a disease-specific atlas. The third method uses connectivity to define nodes, where linkage of each voxel to the rest of the voxels is measured and similar functioning voxels are grouped as a single node. The most widely accepted method establishes nodes from experimental data by registering them to a standard parcellation atlas. Automated Anatomical parcellation atlas [63] is the most popularly used map for cortical parcellation among the standard atlases available. **Table 1** provides a list of web links, number of brain regions and brain coverage of commonly used open-source atlases. However, no single universally adopted parcellation scheme exists for partitioning human brain regions, posing a significant obstacle to creating a unified resource such as connectome [58]. However, use of different parcellation atlases gives scope to analyze various network properties in different studies [76], [77].

1.7.5. Edges

In brain networks, edges represent connectivity between two neuronal systems or brain regions. The set of nodes and edges primarily affect the neurobiological interpretation of network topology and analysis. Interestingly, brain network analyses have deciphered that the nodes with similar connection patterns exhibit similar functionality [78]. Structural brain connectivity, using sMRI, can be interpreted from the interregional correlation of GM volume, CT or other similar morphometric measures [60], [79], [80]. The correlation is calculated between two brain regions after extracting the required measures from each region, further representing the edges of the brain network of interest. For instance, a connection or edge between two nodes would exist if there exists a strong correlation between them.

Table 2. Types of brain network topology

Types	Weighted undirected network	Weighted directed network	Binary undirected network	Binary Directed network
Matrix				
Network				

Note: A, B, C, D, E and F are brain regions or nodes. The colors represent the edge correlation coefficient in the matrix and strength in the network; black being the lowest and white being the highest.

Edges can be broadly classified based on two features- Strength (weighted or binary) and Directionality (undirected or directed), producing different brain network topologies (**Table 2**) [81], [82]. The weighted edges preserve the information of strength, which might represent the size or density of anatomical links, thereby reflecting axonal fibers in the structural networks. Thresholding methods such as proportional or sparsity thresholds are used to remove ostensibly false connections, resulting in an adjacency matrix. A threshold of a much lower value would introduce noisy connections, whereas a threshold of a much higher value would lead to the loss of significant links [83]. Thresholding often results in binary links, indicating the presence or absence of connections. Weighted-directed matrix has the edge weight information along with directional information, and the respective network is accordingly termed a *weighted-directed network*. *Weighted-undirected* matrix

and network have only the weight information of the edge. However, when the links are weak and non-significant, the spurious connections need to be removed by thresholding. The resultant network hence obtained is *binary-directed* when directional information is available, else *binary-undirected* (**Table 2**).

1.7.6. Structural networks from different imaging modalities

Anatomical and neuroimaging studies have revealed significant inter-individual variability not only in the size and location of brain areas, but also in the relationship between macrostructures and microstructures [58]. The presence of substantial inter-individual variability in the structural connection patterns studied at macroscale gives better interpretability on fundamental dynamics of the human brain. Recent studies have described structural networks are physical wiring between the different brain regions and are mapped using modalities such as sMRI [84], [85], [86] and diffusion tensor imaging [87], [88]. Sex differences in the structural connectome have been studied in the human brain with age and inferred significantly different connectivity patterns. Brain complexity increases from childhood age (8-13.3 years) to young adulthood age (17.1 – 22 years) through adolescent age (13.4-17 years) [89]. The structural networks are often constructed based on either of the two measurements: regional GM volume and regional cortical thickness [84], [85], [86]. The pipeline for brain network construction using various modalities are illustrated in **Figure 8**.

1.7.7. Open repositories and brain connectivity analysis software

A few open access repositories for structural MRI of Parkinson's patients are made available to investigators in the field, such as Parkinson's Progressive Markers Initiative (PPMI) [90], Parkinson's Disease Biomarkers Program (PDBP) [91], Dementias Platform UK [92] and TaoWu and Neurocon by ICI [93].

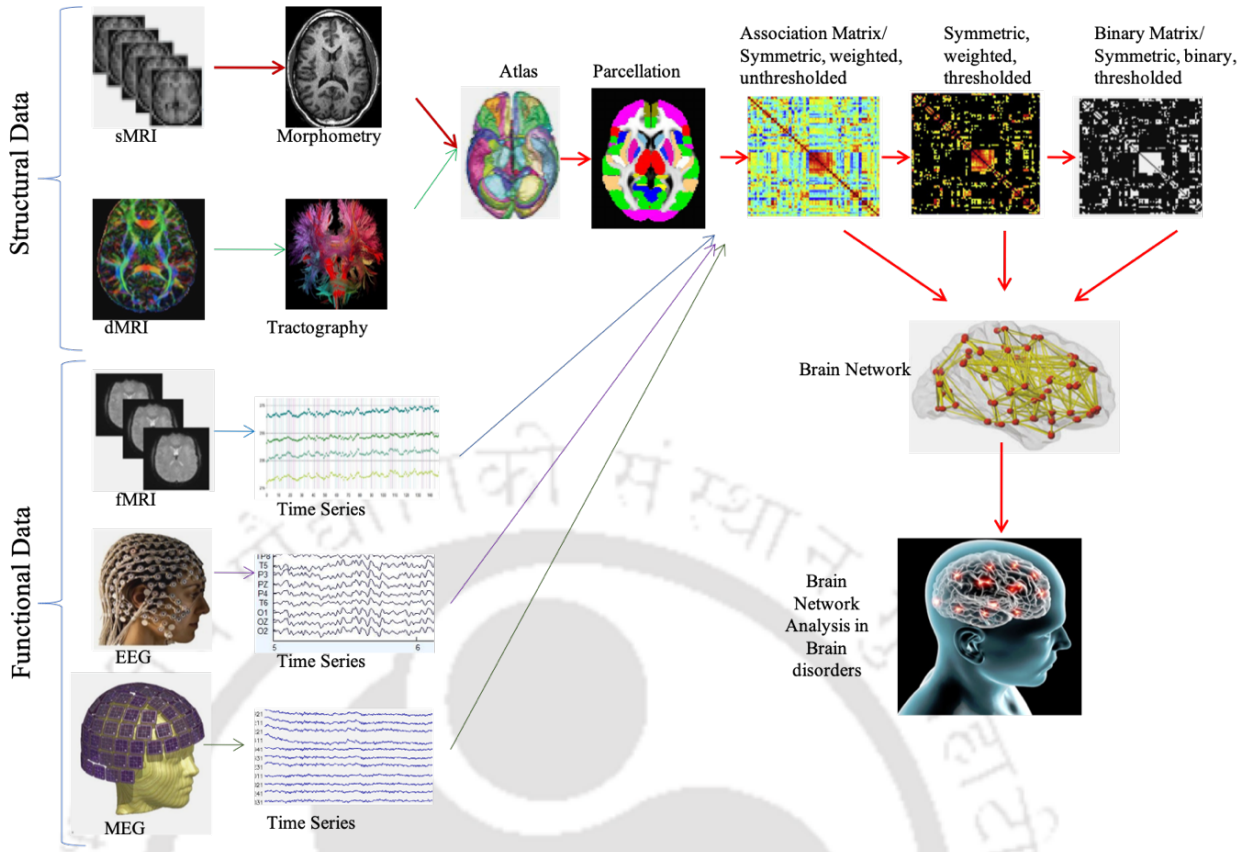


Figure 8. Pipeline for brain network construction

Software packages available for brain network analysis assist to construct and visualize a complex network. **Table 3** includes a list of the software, along with web links and connectivity techniques for the researcher's study. The brain connectivity toolbox (BCT, **Table 3**) [94] is the most commonly used open-source MATLAB toolbox that provides functions for network manipulations (thresholding). This toolbox includes algorithms for network topology analysis and contains datasets for large-scale neuroanatomical networks of several species. It can be utilized to analyze both structural and functional brain networks. Graph theoretical network analysis (GRETNA, **Table 3**) [95] is helpful for imaging connectomics with added advantages like open-source, MATLAB-based software with a graphical user interface. It can be used for topological analysis across modalities.

Graph-Theoretical Analysis Toolbox (GAT, **Table 3**) [96] provides a graphical user interface that allows both structural and functional brain network construction and analysis. It facilitates inter-network comparison of regional and global topological properties. Multimodal Imaging Brain Connectivity Analysis (MIBCA, **Table 3**) [97] is a fully automated toolbox that allows batch-wise multiple tasks like preprocessing, connectivity and graph theoretical analysis of multimodal image data. It also allows visualizing matrices and brain networks as three-dimensional brain graphs or connectograms. Brain Graph (BRAPH, **Table 3**) [98] is a MATLAB toolbox with a graphical user interface for analyzing networks from multiple modalities, including sMRI. Network-based statistic (NBS, **Table 3**) [99] is also a MATLAB- based toolbox used to identify connections and networks and allows analysis of both structural and functional brain networks. BrainNet Viewer [100] is often used for network visualization. A recently developed toolbox, Intelligent Analysis of Brain Connectivity (IABC) for functional networks, may assist in differentiating brain disorders.

1.7.8. Brain network analysis

Brain network analysis [78], [94], [101] has been extensively applied to identify detailed abnormalities of structural network topologies associated with brain disorders. Computational analyses of neuroimaging datasets have revealed a broad range of network characteristics ("metrics"). **Table 4** summarizes connectivity metrics, viz. path length, efficiency, connectivity strength, clustering coefficient, etc. along with a sample network and the respective calculations.

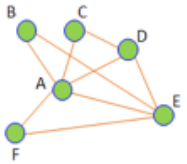
Table 3. Toolbox for Brain network construction and connectivity analysis

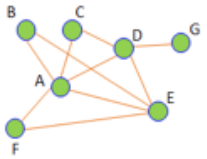
Sl.	Toolbox names	Purpose	Links
1	Brain Connectivity Toolbox (BCT)	Structural and Functional network analysis	http://www.brain-connectivity-toolbox.net
2	Graph Theoretical Network Analysis (GRETNA)	Topology analysis of brain connectome	https://www.nitrc.org/projects/gretna
3	Biomarker and Neural Connectivity Toolbox (BioNeCT)	Analyzing connectivity in EEG recordings	https://sites.google.com/site/bionectweb
4	CONN	Functional connectivity toolbox	https://web.conn-toolbox.org/
5	Graph Theoretical Analysis (GAT)	For structural and functional brain networks	https://www.nitrc.org/projects/gat/
6	Multimodal Imaging Brain Connectivity Analysis (MIBCA)	Automated all-in-one network analysis with batch processing	http://www.mibca.com/
7	GraphVar	GUI-based toolbox for brain graph analysis	https://www.nitrc.org/projects/graphvar/
8	Brain Graph (BRAPH)	For structural and functional network analysis	http://braph.org/
9	Network-Based Statistic (NBS)	Functional and structural brain connectivity	https://www.nitrc.org/projects/nbs/
10	Intelligent Analysis of Brain Connectivity (IABC)	For functional connectivity	https://www.nitrc.org/projects/iabc

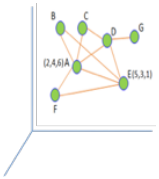
1.7.9. Biomarker

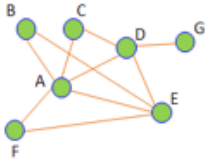
Biomarker is an indicator of biological phenomenon, whether normal or pathogenic. In psychiatry, the discovery of biomarkers is crucial for diagnosis, care and treatment of each patient. In contrast to other diseases, mental illnesses have the uniqueness of being categorised into different diagnostic categories and heterogeneous symptoms. With the development of neuroimaging modalities, a thorough knowledge of neuropsychiatric illnesses is now achievable. This also offers reliable data-driven approaches for investigating both structural and functional aspects of brain. In this regard, the discovery of psychiatric biomarkers enables healthcare professions not only to differentiate health and disease, but also to identify subtypes within a disease. This, in turn, may result in more targeted therapy options or precise treatment.

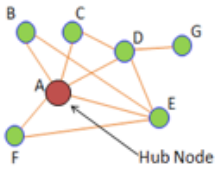
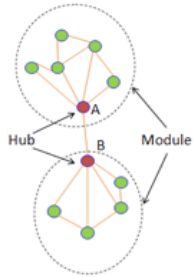
Table 4. Mathematical definitions and calculations of Brain Connectivity Metrics

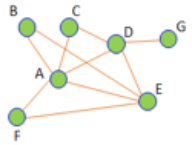
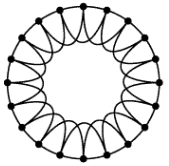
Sl.	Metrics	Definition	Significance	Formula	Calculation
1.	Degree				
	a. Node degree	Number of connecting links it has to other nodes.  Node Edge	A higher degree reflects connectedness and greater significance in the brain network.	The degree, k_i of a node i is: $k_i = \sum_{\substack{j \neq i \\ i,j \in N}} a_{ij}$ where a_{ij} is the connection between i and j; $a_{ij} = 1$; if connection exists, else 0 N : Total number of nodes, i, j : Nodes	 $(\text{Node degree})_A = k_A = 5$
	b. Network Degree	Average of the degrees across all nodes in the network.	Indicates strength of the entire network	The network degree (S) is: $S = \frac{1}{N}\sum_{i \in N} k_i$ where k_i is the degree of node i, N : Total number of nodes	$S = \frac{k_A + k_B + k_C + k_D + k_E + k_F}{6}$ $= \frac{5+2+2+3+4+2}{6} = 3$
2.	Assortativity	Correlation between the degrees of connected nodes.	Assortative network indicates that high degree nodes tend to connect to each other. Disassortative	Perfect assortative network: $r = 1$ Non- assortative network: $r = 0$ Perfect disassortative network: $r = -1$	

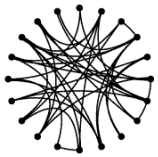
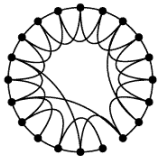
		network implies that network hubs are not connected.	where r is the correlation coefficient	
3.	Path Length			
	a. Shortest Path Length	It is the minimum number of edges between two nodes.	Essential in information transmission of the brain network. It can efficiently transmit information faster and reduce brain consumption.	Path length, l_{ij} between two nodes i and j is: $$l_{ij} = \sum_{a_{st} \in l_{i \leftrightarrow j}} a_{st}$$ where $l_{i \leftrightarrow j}$ is the shortest path between nodes i and j, $a_{st} = 1$, if path exists, else 0	l_{AG} =2 where path is $A \rightarrow D \rightarrow G$
	b. Average shortest path length	Average of the shortest path length from a specific node to the remaining nodes.		Average shortest path length l_i between node i and other nodes is: $$l_i = \frac{1}{(N-1)} \sum_{i \neq j} l_{ij}$$ where N : Total number of nodes; l_{ij}: the shortest path length between nodes i & j	$l_A = \frac{l_{AB} + l_{AC} + l_{AD} + l_{AE} + l_{AF} + l_{AG}}{6}$ = (1+1+1+1+1+2)/6 = 1.17
	c. Characteristic path length	Average of the shortest path lengths between every pair of nodes	Determines how fast information gets processed from distinct brain regions	Characteristic path length, L is: $$L = \frac{1}{N(N-1)} \sum_{\substack{i,j \in G \\ i \neq j}} l_{ij}$$ where N: Total number of nodes, l_{ij} : shortest path length between nodes i and j of graph G.	$L = \frac{l_A + l_B + l_C + l_D + l_E + l_F + l_G}{(6 * 7)}$ = (7+11+10+8+8+11+13) / 42 = 1.62

4.	Anatomical distance	Anatomical distance between two regions is the Euclidean distance between regional centroids.	It gives an approximate reflection on the true physical length (axonal length) of connections between regions	The anatomical distance (D_{ij}) between the two regions is given by: $D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ Where x_i, y_i, z_i: centroid coordinates of region i, and x_j , y_j, z_j for region *j*	 $D_{AE} = \sqrt{(2-5)^2 + (4-3)^2 + (6-1)^2}$ = 5.92
5.	Clustering coefficient				
	a. Local clustering coefficient	Ratio of number of edges between neighboring nodes to the maximum possible number of edges in the entire network.	Determines the possibility of connectedness between any two neighboring node	The local clustering coefficient, *C(i)* of a node *i* is: $C(i) = \frac{t_i}{k_i(k_i-1)/2}$ where, k_i : total number of nodes; t_i is the number of edges between neighbors of *i*	*C(A)* = 4/ (7*6/2) = 0.19 where, t_A = 4 (i.e. number of edges between neighboring nodes of A) and k_i = 7
	b. Global clustering coefficient	Average of the local clustering coefficients of all nodes in a network.		The global clustering coefficient, *C* is: $C = \frac{1}{N}\sum_i C(i) = \frac{1}{N}\sum_i \frac{2t_i}{k_i(k_i-1)}$ where N : Total number of nodes, *C(i)*: local clustering coefficient of node *i*	$C = \frac{2*(4+1+1+2+3+1+0)}{7*7*6}$ = 0.082

6.	Centrality	Measure of importance of nodes of a brain network. Also estimates the efficiency of a network.	Centrality of a node directly implies its information transmission efficiency in brain network.		
	a. Degree centrality	Equivalent to the degree of a node. Employs node degree to describe its role in the network.	Determines the direct influence of a node on its neighboring nodes, i.e., regional influence.	The degree centrality, $C_d(i)$ of a node i is equivalent to the degree (k_i) of the node in a binary undirected network: $C_d(i) = k_i = \sum_{j \neq i} a_{ij}$	$C_d(A) = k_A = 5$
	b. Closeness centrality	Inverse of average shortest path between a region to the remaining region. Reflects the closeness between nodes in a brain network.	Detects nodes that can efficiently spread information throughout the graph. Scales indirect influence of a node on remaining nodes.	The closeness centrality, $C_c(i)$ of a node i in an undirected binary network is: $C_c(i) = \frac{N-1}{\sum_{j \neq i} l_{ij}}$ where N-1 : remaining number of nodes to standardize; l_{ij} : shortest path length between nodes i and j	$C_c(A) = \frac{7-1}{1+1+1+1+1+2} = 0.86$
	c. Betweenness centrality	The proportion of the shortest path between any other two nodes that pass through a node. Counts the connecting node along the shortest path between two other nodes.	Scales the influence of a node on information transmission through the brain network.	The betweenness centrality, $C_b(i)$ of a node i is: $C_b(i) = \frac{2}{(N-1)(N-2)} * \sum_{j \neq h \neq i} \frac{n_{hj}(i)}{n_{hj}}$ where $n_{hj}(i)$: count of shortest paths between h and j passing through I; n_{hj}: count of shortest paths between h, j; N : Total number of nodes	$C_b(A) = \frac{2}{(N-1)(N-2)} * \sum_{F \neq G \neq A} \frac{n_{FG}(A)}{n_{FG}}$ $= \frac{2}{(7-1)*(7-2)} * \frac{1}{2} = 0.03$ where, $\frac{(N-1)(N-2)}{2}$: number of node pairs excluding node i and it is for standardization

7.	Hub Node	Node with degree greatly exceeding the average node degree of the network. Holds the central position in the brain network.	Employed to recognize central brain regions.		
8.	Modularity	Modules refer to a group of nodes showing higher connectivity within the group and sparse interconnections with other nodes.	Modularity measures the strength of clustering a network into modules. Breaks larger network into smaller and simpler modules, making them easier to understand.	The modularity, Q can be found by two methods: $$Q = \sum_{u \in M} \left[P_{uu} - \left(\sum_{v \in M} P_{uv} \right)^2 \right]$$ P_{uu}: Proportion of actual edges in module u, P_{uv}: Proportion of actual edges between modules u and v	
	a. Participation Index	Differentiates regional or provincial hub from a connector hub.	A node with a participation index higher than the threshold indicates it as a connector hub, else a provincial hub.	The participation index, P_i of a node i is: $$P_i = 1 - \sum_{m=1}^{N_m} \left(\frac{k_{im}}{k_i} \right)^2$$ where N_m: count of modules, k_{im}: count of edges from node i to module m and k_i : the degree of node i	$$P_A = 1 - \left(\frac{1}{5} \right)^2 = 0.96$$
	b. Rich club	Nodes with a larger number of degrees and forming closely interconnected subgraphs (i.e., clubs).	Significant role is global brain information transmission.	Rich club coefficient, $\emptyset(k)$ of degree k is: $$\emptyset(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)}$$ where $N_{>k}$: count of nodes with degree larger than k, $E_{>k}$: count of edges connecting $N_{>k}$ nodes	$$\emptyset(2) = \frac{2E_{>2}}{N_{>2}(N_{>2} - 1)}$$ = 2*17/ (6*(6-1)) = 1.13 Where $E_{>2}$ = 17 and $N_{>2}$ = 6

9.	Efficiency	Indicates the potential of information exchange across a network. Efficiency is inversely related to path length.	The efficiency of a network reflects its ability to exchange information.	The importance of an individual node in network efficiency can be assessed by deleting it and estimating the efficiency of the lesioned network.	
	a. Nodal efficiency	Reciprocal of average shortest path lengths from a node to the remaining nodes of the network.	Measures the ability of a node to get integrated within the network through the shortest path. Regions with high values indicate better information distribution, termed as hubs.	The Nodal efficiency, $E_{nodal}(i)$ of a node i is: $$E_{nodal}(i) = \frac{1}{N-1}\sum_{i \neq j}\frac{1}{l_{ij}}$$ where N : Total number of nodes; l_{ij} : shortest path length between nodes i and j	$$E_{nodal}(A) = \frac{1}{N-1}\sum_{A \neq j}\frac{1}{l_{Aj}}$$ $$= \frac{1}{7-1}\left(\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{2}\right)$$ = 5.5/6 = 0.92
	b. Global efficiency	The average nodal efficiencies of all nodes in a network. The higher, the better.	Scales the potential of information exchange across the entire network.	The Global efficiency E_{glob} of a network is: $$E_{glob} = \frac{1}{N}\sum_{i} E_{nodal}(i)$$ where N : Total number of nodes	$$E_{glob} = \frac{1}{7*6}(5.5 + 3.83 + 4 + 5 + 5 + 3.83 + 3.17)$$ = 0.72
10.	Small world network	Ensures efficient local and global information transmission. A random network is a network with the same number of nodes, total edges and degree	Have both • high clustering characteristics (compatible with segregated processing) similar to regular networks and • shorter shortest path lengths (compatible with integrated processing)	Three criteria to ensure small-worldness: 1. $\gamma \gg 1$ and $\gamma = \frac{C}{C_{rand}}$ 2. $\lambda \approx 1$ and $\lambda = \frac{L}{L_{rand}}$ 3. $\sigma > 1$ and $\sigma = \frac{\gamma}{\lambda}$ where γ: normalized characteristic path length of the network,	Regular Network  High clustering coefficient and long characteristic path length

	distribution as that of the network under study.	identical to random networks If the brain network has small-world property, it shows better information processing performance without significant disruption.	C_{rand}: average clustering coefficient of matched random networks, λ : normalized clustering coefficient of the network, L_{rand}: characteristic path length of matched random networks, σ: small world index of the network	Random network  Small world network 	Low clustering coefficient and short characteristic path length High clustering coefficient and short characteristic path length

Network images for the small-world network have been adopted from Watts and Strogatz [102].

The toy-network depicted in this table is a binary-undirected network

CHAPTER - 2

REVIEW OF LITERATURE

Chapter 2. REVIEW OF LITERATURE

In this chapter, the state-of-the-art results are presented for clinical-based as well as data-driven approaches for subgrouping. PD patients are subtyped into juvenile parkinsonism, young-onset PD (YOPD), late-onset PD [21]. Juvenile parkinsonism pertains to individuals who exhibit clinical characteristics of PD before the age of 21 years. This form of parkinsonism is often familial, displaying distinctive clinical traits, and demonstrating diverse pathological discoveries during postmortem examinations. YOPD are the patients between 21 to 40 years old with slower disease progression, less cognitive deterioration than late-onset PD. Although distinctions have been established among juvenile parkinsonism, YOPD and late-onset PD, the attempt to subtype PD in this manner is not efficient. This is due to the fact that patients whose symptoms begin at 50 years of age or less make up only a small portion (5-10%) of the Parkinson's disease population. The subtype characterized by later onset remains notably diverse in its presentation and outcomes [21]. Moreover, in numerous studies, the division between YOPD and later onset PD has been based on various age thresholds such as 50, 55, and 60, underscoring the arbitrary nature of this classification approach [103]. Additionally, previous review articles [104], [105] have listed studies using data mining and cluster analysis in PD.

Cclinical symptoms-based studies (**Section 2.1.1**) in PD as well as studies on GM data from structural MRI scans of PD patients (**Section 2.1.2**) are explained. **Table 5** summarizes the subgrouping studies with respective sample sizes and classification algorithms. In addition, the clinical features of the subjects, along with the resultant subgroup-specific feature, are described. **Table 6** summarizes the data-driven classification studies containing MRI scanner detail, sample size, subject profile (Age, disease duration), and classification performance. Finally, this thesis can guide through the steps to perform network analysis.

Connectivity pipeline, different network topologies and metrics described here, could provide better understanding of neurological disorders like Parkinson. The various network analysis performed on PD is summarized in **Table 7**, and metric calculations are performed on the illustrated toy network in **Table 4**. Lastly, the research gaps are identified and possible opportunities are suggested for further research.

2.1. Subgrouping

PD is a complex disease with various clinical, cognitive, task-based and motor studies differentiating the neuropathological or clinical subgroups [106], [107], [108], [109], [110]. Despite insights about the disease, researchers and clinicians are working towards well-defined subgroups for possible personalized treatment [47], [111]. Hence, due to heterogeneity in PD, the International Parkinson and Movement Disorders Society (MDS) [112] and the National Institute of Neurological Disorders and Stroke (NINDS) [113] have highlighted the priority-based need for investigating stable PD subgroups. This section will give an overview of different clinical scales and methods adopted in various subgrouping studies. In this thesis, subgrouping and subtyping are used alternately.

2.1.1. Based on clinical symptoms

As PD is a progressive neurologic condition, its state and severity have been rated according to the below rating scales [114], [115], which are subjective and heterogeneous.

a. Geriatric Depression Scale (GDS),

b. Hospital anxiety and depression scale (HAD),

c. State-Trait Anxiety Inventory for Adults (STAI),

d. University of Pennsylvania Smell Identification Test (UPSIT),

e. Random Eye Movement Sleep Behavior Disorder (RBD) Screening Questionnaire,

f. Scales for Outcomes in Parkinson's Disease-Autonomic (SCOPA-AUT),

g. Postural Instability Gait Disorder (PIGD) score,

h. Hoehn and Yahr scale (H&Y),

i. Hamilton Depression Rating Scale (HAM-D),

j. Unified Parkinson's Disease Rating Scale (UPDRS),

k. Schwab and England Scale,

l. Mini-Mental Score examination (MMSE), and

m. Montreal Cognitive Assessment (MoCA)

In addition, several efforts were made toward diagnosing and subgrouping PD based on non-motor features [116], [117], [118], [119], motor features [120], age of onset [121] and a variety of clinical manifestations. Based on the combination of clinical data and etiology, Hoehn and Yahr [122] proposed three significant subgroups of PD patients, further representing five stages. Rooden *et al.* [123], Gasparoli *et al.* [124] and a few other studies [110], [125], [126], [127], [128], [129], [130] have performed cluster analysis for differentiating Parkinson's patients using symptoms-based clinical features. Cluster analysis can classify subjects into subgroups so that intra-class differences are less and inter-class differences are more. Hence, this method has been proved to be an unsupervised classification where the variable to be selected for analysis, clustering technique and number of clusters are the crucial factors on which the subgroups are highly dependent. Rooden et al. [120] interpreted the classification of PD patients into subgroups based on cluster analysis (CA) on motor symptoms, cognition, depression and age of onset. This fully-automated method was much advantageous as the a-priori structure of the data was unknown, and each cluster had internal cohesion and external isolation. A few other studies have also identified Parkinson's subgroups by employing the hierarchical cluster analysis method on clinical features [109], [131], as shown in **Table 5**.

The hierarchical clustering method does not directly cluster data but groups or divides entire data to progressively assemble (agglomerative or bottom-up method) or disassemble (divisive or top-down method) the data into clusters [132]. In the end, a dendrogram is constructed to connect all the data points to understand the hierarchical arrangement. In addition, the classification of Parkinson's patients has also been attempted based on various clinical features by using k-means cluster analysis, a non-hierarchical method [106], [107], [108], [131], [133], [134], [135], [136]. K- means clustering measures centroid from the pre-defined k clusters. It further delineates the total samples into k clusters such that each sample belongs to the nearest cluster. Consequently, the cluster mean becomes the new centroid, and the centroid is recalculated after every iteration until the centroid remains unchanged. Thus, it can efficiently differentiate large samples with lesser time for which it has been widely adopted [132], [137]. The highly cited studies conducted on subgrouping based on clinical symptom scores are summarized in **Table 5**, and other previous studies are listed in Supplementary material of Samantaray *et al.,* 2022 [47]. Due to the subjective and overlapping nature of clinical scales, recent studies have considered neuroimaging-based data-driven approaches along with clinical data to classify Parkinson's patients into stable and reliable subgroups.

Table 5. Demographic details and subgrouping results of state-of-the-art studies on clinical data-based subtyping in Parkinson's Disease

Study	Clinical Features Underlined: feature type	Subjects	Classification Algorithm	Results
[138]	Motor: Tremor, rigidity, bradykinesia, dyskinesias, motor fluctuations, axial symptoms, PIGD Non-motor: Depression, anxiety, apathy, psychosis, impulse control and related disorders, cognition, orthostatic hypotension, urinary, gastrointestinal (GIT), sexual, sleep and wakefulness, pain Others: H&Y	402 PD patients	Conditional linear Gaussian Bayesian network model-based multi-partition clustering method	Subgroup A: No problem with Impulse controlling, n = 335; slight problem, n = 67; Subgroup B: no apathy, slight (cognitive changes, pain, GIT, sleep, urinary) issues, n = 199; slight all issues, n = 203 Subgroup C: no dyskinesias or psychosis, n = 242; slight both, n = 112; slight dyskinesias & acute psychosis, n = 48 Subgroup D: No mental & slight physical fatigue, n = 272; Mild mental and mild physical fatigue, n = 130 Subgroup E: slight axial symptoms, slight bradykinesia, mild loss of smell but no motor fluctuations, n = 199; mils axial symptoms, bradykinesia, motor fluctuation and moderate loss of smell, n = 203 Subgroup F: No orthostatic hypotension and slight sexual problems, n = 214; Slight orthostatic hypotension and mild sexual problems, n = 188; Subgroup G: No weight loss and slight depression, n = 317; mild weight loss ($\mu\mu$ = 0.00, $\sigma\sigma$ = 0.01) and slight depression, n = 85 Subgroup H: no degree of excessive sweating and slight anxiety, n = 299; mild degree of excessive sweating and slight anxiety, n = 103

[139]	Demographic: Disease duration, Motor: Fine motor assessment (FTT), (H&Y) Non-motor: L-Dopa Cognitive: Mattis Dementia Rating Scale, GDS, State-Trait Anxiety Scale, Modified fatigue impact Functional disability: Lawton and Brody Activities of Daily Living scale's modified version	157 PD patients	Cluster analysis : Hierarchical (Ward's method) and k-means	Subgroup 1: Non-Apathetic group, i.e., low levels of apathy symptoms, n = 48; Subgroup 2: Low Interest/ Energy group, i.e. worse depression, fatigue, anxiety, health-related quality of life, and caregiver burden than Subgroup 1, n = 76; Subgroup 3: Low Initiation Group, ie. Feature of subgroup 2 along with worse overall cognition, emotional well-being, state anxiety, communicative ability, and functional ability than Subgroup 2, n = 33
[109]	Demographic: Age at diagnosis, gender, race, education history, symptom duration Motor: UPDRS-II, UPDRS-III, Shwab and England score, PIGD, tremor score Non-motor: RBD, SCOPA-AUT Cognitive: MoCA	421 PD patients	Hierarchical clustering method	Subgroup 1: mild motor predominant, n = 223; Subgroup 2: intermediate type, n = 146; Subgroup 3: diffuse malignant type, n = 52
[131]	Demographic:Onset & diagnosis age,disease duration Motor: UPDRS- II, III, PIGD, FOSS score, dyskinesia, Rigidity, Tremor, falls Non-motor: Depression, anxiety, apathy, hallucination, UPDRS-I, RBD, Autonomic score, orthostatic hypotension Cognitive: MMSE score	113 idiopathic PD patients	k- means clustering + Hierarchical clustering	Subgroup 1: Motor/ slow progression, n = 43; Subgroup 2: Diffuse/ Malignant, n = 40; Subgroup 3: Intermediate, n = 30
[133]	Non-motor: Sensory, dysautonomic, sleep-related, Neuropsychiatric	109 PD patients 107 Healthy Control	k- means clustering method	Subgroup 1: RBD symptoms-constipation related; Subgroup 2: cognition-related; Subgroup 3: mood-related;

				Subgroup 4: sensory-loss related
[123]	Motor: Tremor, Bradykinesia, PIGD, FOSS, Dyskinesia Non-motor: Daytime sleepiness, nighttime sleep problems, autonomic dysfunctions, psychotic symptoms, depressive symptoms	344 PD patients	Cluster Analysis	Subgroup 1: n = 169; Subgroup 2: n = 45; Subgroup 3: n = 104; Subgroup 4: n = 26
[108]	Demographic: Age of onset, at diagnosis, disease progression Motor: tremor, hypokinesia, rigidity, PIGD Non-motor: depression, apathy, L-Dopa complications, hallucination	173 PD patients	k-means cluster analysis	Subgroup1: Rapid disease progression, n =11; Subgroup 2: Young-onset, n = 51; Subgroup 3: Non-tremor dominant and psychopathology; n = 29; Subgroup 4: Tremor-dominant, n = 82
[136]	Demographic: Gender, alcohol intake, daily levodopa dose, dopamine agonist daily dose Motor: UPDRS, alternate finger tapping, dyskinesia, falls Cognitive: dementia, digit ordering, letter fluency, CANTAB	176 idiopathic PD patients	k-means cluster analysis	Subgroup 1: n = 60; Subgroup 2: n = 37; Subgroup 3: n = 37; Subgroup 4: n = 23; Subgroup 5: n = 19
[122]	Demographic: Gender, age of onset, age at diagnosis, disease duration, progression, degree of disability, associated diseases Motor: Tremor, rigidity, delayed movement, slowness, impaired postural and righting reflexes	802 patients with paralysis agitans or PD or parkinsonism	Based on the combination of clinical data and etiology	Subgroup 1: Primary parkinsonism, n = 672; Subgroup 2: Secondary parkinsonism with Definite encephalitis lethargica (n = 82) or probable encephalitis lethargica (n = 14) or Others (n = 22); Subgroup 3: Indeterminate parkinsonism with encephalitis lethargica (n = 12)

n: number of subjects in the subgroup

2.1.2. Based on structural neuroimaging data-driven approaches

The structure and function of the human brain can be analyzed non-invasively using structural magnetic resonance imaging (sMRI), which has great potential in the differential diagnosis of PD [140], [141], [142], [143]. PD is associated with Grey Matter (GM) and White Matter (WM) atrophy, often detected as morphological changes in sMRI images. These changes have been well interpreted and understood by voxel-based morphometry (VBM) in terms of volumetric and intensity changes in different brain tissues in various regions. A voxel is a 3-dimensional pixel or a volumetric pixel that refers to a small volume of space, used in imaging or three- dimensional modeling. The raw medical image of brain from sMRI is usually obtained in Digital Imaging and Communications in Medicine (DICOM) format and further converted to Neuroimaging Informatics Technology Initiative (NIfTI) format for further channelizing through the preprocessing pipeline of VBM. VBM is an unbiased, operator-independent and automated morphometric technique allowing detection of significant differences in various brain tissue between groups of the population without considering the *a-priori* regions of interest. It has been reported to a greater extent for the growing role of sMRI in the differential diagnosis [140], [141] of APD and PD by enabling a whole-brain voxel-wise comparison of GM and WM using MRI images [144]. **Figure 5** shows the various GM atrophy regions in PD [145] and APDs such as MSA-cerebellar type [146], MSA-Parkinsonian type [146], PSP [147], CBD [148] and DLB [149] as compared to healthy control subjects. Interestingly, MRI could also connect the neuropathological expansion of the disease with the clinical phases [150]. In addition, novel diagnostic approaches like the automated Magnetic Resonance Parkinsonism Index have been developed to differentiate PSP from PD [151].

There is widespread axonal pathology from the earliest stages of PD, suggesting an impaired axonal function in this disorder [152]. In addition, WM disintegrity in PD patients

leads to increased disease severity and systemic inflammation [153]. The alterations of WM in PD are associated with cognitive impairment, and such alterations may help assess the onset of dementia in PD patients [154]. The prefrontal WM alteration was also reported to be a valuable biomarker of dementia in PD. Additionally, WM atrophy has been reported in specific brain regions of PD patients compared to healthy control subjects [155], [156].

Exploring the heterogeneity in PD with limited consideration of the clinical variability has remained a greater challenge. This has been attempted by involving machine learning approaches, the highly cited studies are in **Table 6** and the other studies may be found in Supplementary material of Samantaray *et al.,* 2022 [47]. These mentioned studies have performed a classification approach using support vector machine (SVM) on three-dimensional T1-weighted sMRI storing information of every voxel. SVM has not only been applied in other neurological issues [157], [158], [159] but is widely accepted in Parkinson's disease studies too. SVM is a classification algorithm in machine learning that trains the subjects with known (labeled) classes and estimates a hyperplane for maximally and distinctly separating these classes. The classification was further cross-validated using leave-one-out approach (LOO) approach. It validates the classification as many times as the number of subjects in the data set. It considers all but one subject for testing and tests the remaining subjects. This approach assesses classification model's accuracy, specificity, and sensitivity by presenting the testing or unknown (unlabeled) samples to the model.

A study was performed by Park et al. [160] to discriminate PD and APD from HC, PD from APD, and MSA from PSP. The T1-weighted 3 Tesla sMRI scans were preprocessed using the SPM12 pipeline, and the grey matter (GM) volume measures were further considered. The GM regions were parcellated into 120 regions based on different atlas separately, and for each of these regions, the GM volume was divided by total intracranial volume. SVM was implemented as a machine learning method, followed by LOO for cross-validation.

However, they also fused functional MRI measures in their study to improve the performance of the diagnostic classification model, thereby encouraging the use of multimodal neuroimaging measures.

Rana et al. [161] classified PD patients based on the 3D local binary pattern (LBP) feature extracted from brain regions using sMRI scans. The 3D sMRI scans were preprocessed using SPM8 software, and GM, WM and CSF volumes were taken into account for fused volume construction, followed by parcellating each subject's fused volume based on an atlas with 118 regions. The three-dimensional LBP was used for feature extraction, which stored every pixel's intensity related to its neighboring intensities. Feature selection was carried out to obtain a set of discriminating and uncorrelated features. SVM was adopted as a machine learning classifier and LOO strategy for validation. This approach could successfully classify the PD patients from the healthy controls with 95% accuracy, 96.67 % specificity and 93.33 % sensitivity.

Huppertz and coworkers [54] differentiated neurodegenerative parkinsonian syndromes (PS), i.e., PSP with Richardson's syndrome (PSP-RS), MSA of Parkinsonian type (MSA-P), MSA of cerebellar type (MSA-C) by volumetric MRI analysis and SVM. They preprocessed the T1-weighted sMRI by following the SPM12 pipeline and used atlas-based volumetry from different brain atlases for pair-wise classification. The groups were inversely weighted by their size for training of the SVM. The validation of the classifier was evaluated using the LOO validation method.

Further, Salvatore et al. [162] performed machine learning on individual T1- weighted brain MRI images for differential diagnosis of PD and PSP. This study proposed a supervised machine learning approach that was data-driven and did not require a-priori hypotheses and subsequently tried identifying sensitive medical image biomarkers. The

images were initially preprocessed, and the resulting whole-brain volumes were considered for feature extraction by principal component analysis (PCA). They used an SVM classifier for correctly classifying the test samples based on the associated label and the maximal margin hyper-plane. They cross-validated the classification by the LOO approach, along with another approach with half randomly chosen subjects as training samples and the remaining half as testing samples. Subsequently, they generated voxel-based morphological biomarkers for the diagnosis of PD and PSP using the weighted training set, indicating the importance of the image for binary separation.

Focke et al. [163] conducted a voxel-based subtype prediction of IPD versus APD involving multiple-system atrophy of Parkinsonian variant (MSA-P) and PSP. They preprocessed the 3D-T1 weighted images using Statistical Parametric Mapping (SPM 8) and used SVM as a classification algorithm. The testing samples were then presented to the model for classification, and the accuracy, specificity and sensitivity of the model were assessed using the LOO approach. Interestingly, they investigated tissue volume loss of both grey matter and white matter. Successful classification of PSP versus IPD and MSA versus IPD was observed but failed in the case of IPDs versus control subjects.

Duchesne et al. [164] performed a classification approach on three-dimensional T1 weighted sMRI for automated differential diagnosis of Parkinson's disease (IPD) versus APDs, specifically PSP and MSA, based on tissue composition and deformation features extracted from the hindbrain region. This unicentric study employed PCA for dimensionality reduction of the healthy subjects, followed by SVM to separate the two groups maximally. In this study, over-determination was avoided by the leave-one-out (LOO) classification. Overall, these studies prove SVM for classification with LOO validation to be the most commonly adopted approach and suggest a fused-feature method

for better accuracy. Thus neuroimaging modalities can provide additional viewpoints along with clinical-criteria-based differentiation.

Besides, deep-learning is a machine learning method successfully adopted in differentiating PD patients from non-PD subjects [165], [166], [167]. Deep learning models possess multiple levels of feature representation and train several hidden layers with unsupervised initialization. The features are picked by the neural network from the huge volume of data without human intervention. A deep neural network was adopted for diagnosing PD using regional GM, WM, cerebrospinal fluid (CSF) volumes, cortical thickness (CT), and cortical surface area of different regions based on MRI data [168]. A correlation matrix was constructed based on CT measures giving 3003 features, of which 111 features were selected for the deep neural network [168]. Furthermore, convolutional neural networks have also been proposed as diagnostic biomarkers of PD based on the region around the brain stem obtained from neuromelanin sensitive- MRI data of 45 PD, 20 APD (MSA and PSP, together), and 35 healthy control subjects [169]. A very recent study [170] has conducted VBM on MRI scans to detect the affected regions in PD patients and extracted different features from these regions. The PD patients were then differentiated from healthy controls based on these features, thereby claiming to be used as a diagnostic tool. Due to changes in brain anatomical characteristics and tissue volume, the connection between the various brain regions is assumed to vary amongst the disease subgroups.

Table 6. Demographic details and results of state-of-the-art structural neuroimaging-based subgrouping methods in Parkinson's disease

Study	Modality	Subjects Count(male)	Age in years Mean (SD)	Disease duration in years Mean (SD)	Classification algorithm	Validation Scheme	Accuracy (%)	Sensitivity (%)	Specificity (%)
[160]	sMRI T1 w 3 T	HC 53(25) PD 77(44) PSP 42(23) MSA 44(26)	HC 66.87(8.36) PD 67.62(7.72) PSP 71.04(6.56) MSA 61.73(9.19)	PD 22.49(20.96) PSP 33.33(16.19) MSA 25.31(18.94)	SVM	LOO	HC vs. PD 75.38 HC vs. MSA 89.69 HC vs. PSP 83.16 PD vs. MSA 82.64 PD vs. PSP 75.63 MSA vs. PSP 75.58	NA	NA
[161]	sMRI T1 w 1.5 T	HC 30 PD 30	HC 58.23(6.87) PD 61.83(8.01)	NA	SVM	LOO 10 Fold CV	95 89.67	HC vs. PD 93.33 HC vs. PD 86.67	96.67 92.67
[54]	sMRI T1 w 1.5 T	HC 73 PD 204 PSP-RS 106 MSA-C 21 MSA-P 60	HC 63.3(0.9) PD 64.0(0.8) PSP-RS 69(0.6) MSA-C 63.1(1.6) MSA-P 63.3(1.0)	PD 6.5(0.4) PSP-RS 3.2(0.2) MSA-C 3.8(0.5) MSA-P 3.6(0.3)	SVM	LOO	82.3 92.6 91.1 85.5 89.8	MSA-P vs. HC 83.3 MSA-C vs. HC 81.0 PSP-RS vs. HC 89.6 MSA-P vs. PSP-RS 83.3 MSA-C vs. PSP-RS 71.4 MSA-P vs. MSA-C	81.5 95.9 93.2 86.8 93.4

							75.3	57.1	81.7
[162]	sMRI T1 w 1.5 T	HC 28(15) PD 28(15) PSP 28(18)	HC 67.5(7.1) PD 68.2(5) PSP 69.4(5.7)	PD 8(4.8) PSP 3(1.6)	SVM	LOO		PSP vs. PD	
							88.9	89.5 PD vs. HC	88.5
							85.8	86 PSP vs. HC	86
							89.1	89.5 PSP vs. PD	89.1
						Training: N/2 subjects and Testing: N/2 subjects	84.7	83.8 PD vs. HC	87.5
							83.2	85.4 PSP vs. HC	81.9
							86.2	82.9 MSA-P vs. PSP	92.1
[163]	sMRI T1 w 3 T	HC 22(13) IPD 21(15) PSP 10(7) MSA-P 11(6)	HC 69.3(9.1) IPD 65.2(8.0) PSP 67.0(4.0) MSA-P 62.5(8.0)	IPD 5.5(3.4) PSP 2.4(2.3) MSA 4.4(2.6)	SVM	LOO	33.33 (GM) 76.19 (WM)	-	-
							71.87 (GM) 65.63 (WM)	MSA-P vs. IPD 36.36 (GM) 27.27 (WM)	90.48 (GM) 85.71 (WM)
							87.1 (GM) 96.77 (WM)	PSP vs. IPD 70 (GM) 90 (WM)	95.24 (GM) 100 (WM)
[53]	sMRI T1 w 1.5 T	HC 149 IPD 16(9) PSP 8(5) MSA 8(4)	IPD 57.9 (8.5) PSP 65.1 (9.4) MSA 58.6 (7.1)	IPD 11.8(2.6) PSP 4.1(1.9) MSA 4.3(1.9)	SVM	LOO	90.6	IPD vs. (PSP and MSA) 93.32	88.2

CV: Cross validation; GM: Grey Matter; HC: Healthy Control; IPD: Idiopathic Parkinson's Disease; LOO: Leave-one-out; MSA-C: Multiple System Atrophy −Cerebellar type; MSA-P: Multiple System Atrophy −Parkinsonian type; MSA: Multiple System Atrophy; N: Number of subjects; NA: Not available; PD: Parkinson's Disease patients; PSP-RS: Progressive Supranuclear Palsy - Richardson Syndrome; PSP: Progressive Supranuclear Palsy; SD: Standard Deviation; sMRI: structural Magnetic Resonance Imaging; SVM: Support Vector Machine; T: Tesla; T1 w: T1 weighted; WM: White Matter

2.2. Multivariate Approaches for Data Fusion

Integrating information from different modalities may provide a better picture of how different data are associated with a single phenomenon. In this direction, several studies use data fusion techniques such as multimodal canonical correlation analysis (mCCA) [171], partial least square (PLS) [172], [173], joint independent component analysis (JICA) [174], [175], [176], linked ICA [177], parallel ICA [178], coefficients-constrained ICA [179], spatial CCA (sCCA), combination of mCCA and JICA [180].

Previous studies utilized multimodal information for different purposes. Multimodal group principal component analysis [181] was performed using three different imaging modalities, namely structural MRI, functional MRI, diffusion MRI. Features such as grey matter, mean amplitude low frequency fluctuations and fractional anisotropy were extracted from respective modalities. They found age-related significant reductions in GM density in specific brain regions with covarying reduction in fractional anisotropy and low-frequency fluctuations [181].

Joint decomposition of multiple imaging datasets using independent component analysis (ICA) resulted to independent components which were combined to form subspace of all possible combinations of components [182]. The association between the sources were analyzed and hidden features were extracted from multiple modalities. The proposed approach showed improved performance over contemporary datasets, even at low signal to noise ratio [182].

A recent study [183] analyzed the clinical parameters pertaining to Impulse control behaviour (ICB) in PD patients. They assessed GM volume from sMRI, white matter integrity from diffusion MRI and graph properties from fMRI. They also, identified

different brain regions with GM loss, higher axial diffusivity and abnormal functional connectivity, along with found association of structural alterations with anxiety and depression in ICB [183].

Neuro-imaging modalities such as sMRI, fMRI, dMRI have played a crucial role in diagnosis of neurological, neuro-psychiatric, neuro-degenerative diseases such as Attention deficit hyperactivity disorder (ADHD), Schizophrenia (SZ), brain-tumors, ischemic strokes, Multiple Sclerosis, Alzheimer's disease, etc [184]. There are numerous studies [185], [186], [187], [188], [189] that have fused multiple modalities for analyzing neurological disorders.

Radiomics-based machine learning approach was used for distinguishing multiple system atrophy (MSA) with parkinsonian type from that of cerebellar type [190]. They used 3D T1 sMRI and a resting-state-fMRI to understand the structural and functional features for differentiating MSA subtypes.

Canonical correlation analysis [191] is a technique to jointly discover relationships between multiple modalities. Basically, the correlation between the modalities is maximized, by finding linear combinations of the variables. One of the advantages of CCA is that it doesn't assume any directionality, making it suitable for neuroscience research. Different variants of CCA, such as kernel CCA [192] constrained CCA [193] and multi-set CCA [194] have also been used in brain science research.

A data fusion method, used simultaneously on fMRI and EEG data, studied the co-variation between the modalities [194]. The dataset was decomposed into spatial maps for fMRI and corresponding temporal dependencies for EEG data using multiset CCA (MCCA). This allowed them to map the locations as well as brain connectivity and understand trial- to-trial variations in amplitude modulations over the decomposed data [194].

A combined model, employing multimodal-CCA and joint-ICA, was utilized for fusion of GM data from sMRI and WM data from DTI [195]. The input information was decomposed and the resultant mixing coefficients were correlated with neuropsychological scores of different groups. This multimodal approach enabled understanding of associations among different tissue types and its role in neuropathology.

Wolf et al., 2020 [196] analyzed parkinsonism features in Schizophrenia spectrum disorders from grey matter maps (sMRI) and fractional ALFF (fALFF, from rs-fMRI). They used parallel ICA (p-ICA), which unlike joint-ICA, decomposes the modalities separately, into their independent components and finds association between them.

2.3. Brain connectivity

2.3.1. Construction and application of structural brain networks in PD

Structural network patterns obtained from correlation (usually Pearson's or Partial correlation) in morphometric measures between regions indicate the presence of cortico-cortical pathways [197]. Some of the open-source tools available to measure morphological parameters of the brain regions are FreeSurfer (http://www.freesurfer.net), and Statistical Parametric Mapping (SPM, http://www.fil.ion.ucl,ac.uk/spm/). Structural brain network analysis has contributed to exploring network-based biomarkers and successful pathophysiology studies of neurological disorders. Since PD is a disconnection syndrome [198], brain network analysis has identified abnormalities of network topologies using sMRI, as detailed in **Table 7**. Majorly brain network connectivity was focused at macroscale where nodes are brain regions. Pereira and co-workers [84] showed a significant reduction in connectivity strength between cortical and subcortical regions in PD patients with mild cognitive impairment (PD-MCI) compared to healthy control (HC). Furthermore, reduced global efficiency but larger characteristic path length was observed

in PD-MCI than PD with normal cognition (PD-NC) in frontal and parietal regions. Interestingly, both PD groups (MCI, NC) exhibited reorganization of densely connected regions [84]. This study showed abnormal network topology in patients compared to HC, which corroborates the finding that neurodegenerative disease targets highly correlated brain regions in HC [199]. Zhang et al. analyzed PD patients with tremors [86]. They observed an increased nodal local efficiency in a few regions, while a decreased nodal local efficiency in a few others compared with HC. However, a consistently significant improvement in the global efficiency in PD patients with tremors supported the compensatory mechanism in these patients [86]. A gender-based comparative study was performed between PD patients (43 males, 21 females) and HC (12 males, 34 females). A lower connection strength was observed in male patients due to altered CT. Small-world network topology was noticed in PD patients of both genders, indicating efficient information transmission at both local and global levels. Patients of either gender showed no difference in normalized clustering coefficient and normalized path length. The clustering coefficient reflects the connectedness of a node to its neighboring nodes in the network, and path length estimates the transmission of information and the amount of brain consumption [85]. A recent study has employed regional grey matter volume from 3T sMRI of 70 PD patients and an equal number of age- and gender-matched HC [200]. Sparsity, as a measure of threshold, was applied to the undirected matrix to measure the network metrics. Reduced clustering coefficient and local efficiency implied alteration in brain connectivity in the PD patients' network [200].

Table 7. Demographic details and structural MRI-based Brain Network studies in PD

Study	Modality	Subjects N (male)	Age Mean(SD)	Tissue	Preprocessing	Node definition	Parcellation Scheme	Software Used	Connectivity metrics	Threshold
[200]	sMRI T1 w 3T	PD: 70 (55) HC: 70 (52)	PD: 52.56 (11.03) HC: 49.24 (10.99)	Grey Matter maps	CAT 12	Grey matter volume	LONI Probabilistic Brain Atlas	Brain Connectivity Toolbox BrainNet Viewer	Partial correlation	Sparsity 0.6 to 0.9 Interval: 0.03
[201]	sMRI T1 w 3T	PD: 31 (15) HC: 37 (14)	PD: 57.1 (8.4) HC: 57.7 (7.9)	Grey Matter maps	CAT 12	Grey matter volume	Automatic Anatomic Labeling Atlas	Graph Theoretical Analysis Toolbox	Pearson's correlation	Density 0.01 to 0.05 Interval: 0.02
[202]	sMRI T1 w 3T	PD:48(25) HC:50(24)	PD: 57.64(7.0) HC: 57.74(5.56)	Cortical structures	VBM (SPM 8)	Cortical volume	Automatic Anatomic Labeling Atlas	Graph Theoretical Analysis Toolbox	Pearson's correlation	Sparsity 6 % to 40 %

[85]	sMRI T1 w 1.5 T	PD: 64 (43) HC: 46 (12)	PD male: 71.2(6.3) PD female: 69.5(7.0) HC: 73.0(10)	Grey Matter surface maps	FreeSurfer	Cortical Thickness	Clusters with significant group differences	Graph Theoretical Analysis Toolbox	Pearson's correlation	Connection density 0.42 to 0.5 Interval: 0.01
[86]	sMRI T1 w 1.5 T	PD: 16 (9) HC: 20 (11)	PD: 60.5(11.8) HC: 59.2(8.7)	Grey Matter maps	VBM (SPM8)	Cortical Thickness	Meta-analytic functional areas	GRETNA BrainNet Viewer	Intercortical similarity	Sparsity 0.02 to 0.4 Interval: 0.02
[84]	sMRI T1 w 3 T	PD-MCI: 33 (20) PD-NC: 90 (55) HC: 56 (33)	PD-MCI: 63.4(7.6) PD-NC: 59.4(10) HC: 58(10.4)	Cortical surface & Subcortical structures	FreeSurfer	Cortical Thickness and Subcortical volumes	Destrieux Atlas	Brain Connectivity Toolbox BrainNet Viewer	Pearson's correlation	Sparsity 2 % to 12 % Interval: 0.5 %

CAT: Computational Anatomy Toolbox; GRETNA: Graph theoretical network analysis toolbox; HC: Healthy Control; PD: Parkinson's Disease patients; PD-MCI: Parkinson's Disease with Mild Cognitive Impairment; PD-NC: Parkinson's Disease with normal cognition; SD: Standard Deviation; sMRI: structural Magnetic Resonance Imaging; SPM: Statistical Parametric Mapping; T: Tesla; T1 w: T1 weighted; VBM: Voxel-Based Morphometry; FMRIB: Functional Magnetic Resonance Imaging of Brain

2.4. TAKE AWAY FROM LITERATURE REVIEW

Through detailed literature survey, it is suggested that clinical symptoms-based subgrouping approach may lead to unreliable subgroups and hence data-driven subgrouping could also be considered alongside. Classification algorithms like support vector machine, random forest, density-based spatial clustering of applications with noise (DBSCAN) and deep learning need to be explored as well [21]. Few data-driven algorithms, such as generalized factorization method with non-negative factorization (GFM-NMF) method [203], biclustered independent component analysis (BICA) algorithm [204] and n-way biclustering (n-BIC) algorithm [205], have successfully classified Schizophrenia subjects and hence could be attempted for subgrouping PD. In addition, classification has been attained between healthy controls and PD patients in a few studies, while in other, clinical subgroups of PD (such as PSP, MSA, etc.) and healthy control subjects have been considered. So there is a prospect of classifying PD patients only without labelled clinical targets, i.e., without having knowledge of patients belonging to MSA or PSP or any other clinical subgroup. Though several studies have used both clinical and data-driven approaches to find different subgroups of PD, there seems to be no consensus yet and warrants further research. Another practical issue is data unavailability of elderly patients due to the difficulty in conducting MRI scanning.

Interrelation among different tissues may be exploited to differentiate PD patients into subgroups because of intra- and inter-tissue interactions and connections. Multi-modal imaging fusion seems to provide effective and reliable biomarkers for neuropsychiatric disease diagnosis [206]. Furthermore, the use of multiple datasets in neural imaging may provide complementary information useful for clinical decisions. However, there is limited research on multimodal neuroimaging measures for Parkinson's data. Multivariate

multimodal algorithms like joint ICA and parallel ICA may be attempted to find linked associations between grey and white matter measures for PD datasets [207], [208]. Both of these are based on blind source separation (BSS) methods. Integration of multiple modalities could make computation time-intensive and could come across as an intractable problem. Hence, there exists is a possibility to incorporate evolutionary algorithms which can tackle approximation versions of these problems and achieve the integration goals.

The progressive use of structural brain images has enabled neuroscientists to construct human brain structural connectivity networks. Interestingly, multimodal approaches have proved to be more informative in discovering disconnections in brain networks [209]. Connectivity analysis provides insights into several important network measures, including, but not limited to, strength, degree distribution, clustering coefficient, modularity, and efficiency. Previously, structural brain connectivity in PD has been studied using sMRI based on the correlation between regional cortical thickness. However, because PD has been linked to GM and WM atrophy, further in-depth research is needed with available connection metrics to compare brain networks of PD subgroups. Additionally, the intergroup topographical patterns could be analyzed by first subgrouping Parkinson's disease and then studying the brain network connection of the deciphered subgroups. Longitudinal investigations based on multimodal imaging may aid in the discovery of neurophysiological sequences of events in specific brain regions of the deciphered subgroups. This may aid with a thorough understanding of disease pathology. A major challenge in brain connectivity includes selecting an appropriate brain parcellation atlas. Employing a common atlas during brain parcellation and network construction might lead to a better conclusive result. This would also help deduce a converging result while comparing across studies. Besides, the use of disease-specific nodes should be encouraged

to use as they accurately reflect the regions affected and alongside reduces the redundant nodes and computational complexity.

Thresholding is the fundamental step for obtaining an adjacency matrix from the correlation matrix. A much lower or higher threshold value may possibly falsify the results and mislead the analysis. In this direction, proportional or sparsity threshold is a user-defined density-based thresholding method conventionally used for network analysis. However, a few other estimators like a histogram and mean–standard deviation should also be considered alongside for an optimal threshold. It is noteworthy to mention that generative models like the adaptive thresholding method could be explored to binarize a correlation matrix. It assumes the presence of two peaks in the histogram and attempts to find an optimal value to separate these two peaks. The age-related modular organization of the human brain connectome has been investigated [210] using diffusion MRI data. The metric of modularity showed a U-shaped trend with higher values at the beginning and end of the age range. This phenomenon could be reinvestigated using other imaging modalities (like sMRI) on patients with neurodegenerative disorders. This is due to the fact that they exhibit diminished connections with age due to neuronal loss and severity. The state-of-the-art studies suggest that structural MRI may provide better and more accurate insights into the diseased brain as it avoids the effect of the dynamic behavior of the human brain. Furthermore, better classification success could perhaps be achieved using fused features from structural imaging-based and connectivity-based data and be considered for early diagnosis. Differential classification of Parkinson's only patients is challenging and requires tailoring of accurate, stable, and, most importantly, distinct subgroup-specific biomarkers.

2.5. MOTIVATION AND OBJECTIVES

Inter-individual variations in patients with Parkinson's disease (PD) may be evident in their brain network structure. To realize this, developing an encoding algorithm was proposed to numerically represent any given subject's brain network in the first objective. The neural signature of a person's particular brain connections might be created using this model for possible applications in brainprinting. Moreover, PD is a neurodegenerative disorder, implying loss of neurons with time, thereby leading to loss of connections in brain network of PD patients with age. Hence, this objective also aims to analyse brain connectivity of PD patients with age.

PD is often characterized with grey matter atrophy and overlapping clinical symptoms, but due to limited neuroimaging-based data-driven approaches, subgrouping PD has remained a challenge. Most of the state-of-the-art methods classify PD patients from healthy control, but limited attempt has been made towards finding subgroups within the disease. Differentiating Parkinson's patients based on clinical, cognitive, task-based or motor symptoms involves heterogeneous symptom scores. In addition, there are often varying results because of varying clinician scoring across different locations and is hence biased.

This led to development of data-driven approaches like Biclustered Independent Component Analysis (BICA) and joint Independent Component Analysis (JICA). BICA helps in deciphering subgroups based on the joint distribution of loadings from two components (i.e regions of the brain considering grey matter) to obtain subsets which are then tested for association with clinical symptoms. The notion that Parkinson's disease is linked with abnormal structural connectivity has been evidenced in the recent past. The state-of-the-art also reveals the fact that assessing structural brain connectivity in PD using

structural Magnetic Resonance Imaging (sMRI) data has remained scarce. Furthermore, there is no structural connectivity study reported on data-driven subgroups of PD only to the best of knowledge. Hence, objective two proposes to differentiate PD patients into subgroups by applying BICA on sMRI data and further obtain brain connectivity which may provide a neuroimaging marker for accurate early diagnoses.

Furthermore, PD is concomitant with grey matter and white matter damages. However, the interrelationship of volumetric changes between grey and white matter remains poorly understood in PD. Combining multiple features from brain imaging data provides more information for individual subjects by exploiting the abundant information that exists. Also, the mutual k-nearest neighbors (MKNN) between grey matter and white matter regions may help understand how these tissues are interconnected. Hence, the third objective focusses on applying JICA to identify maps with covariant change in grey matter and white matter volumes in PD patients based on sMRI data and thereby could differentiate PD patients based on fusion of multiple features (i.e. based on integrated information of grey matter and white matter volumes). Futher, this objective shows the inter-tissue connections using MKNN concept. Also, the 4th objective optimizes permutations in biclustering algorithm.

Therefore, taking into account the aforementioned concerns and motivated by these factors, the thesis objectives are listed below.

1. Unique Brain Network Identification Number & Age-based Brain Connectivity analysis of Parkinson's disease (Chapter 3 of thesis)
2. Subgroup identification of Parkinson's Disease by neuroimaging- based data-driven approach using grey matter information & Subgroup-based Brain Connectivity Analysis (Chapter 4 of thesis)

3. Subgroup identification of Parkinson's disease using joint information of grey and white matter & Mutual Nearest Neighbor based Bipartite Graph analysis (Chapter 5 of thesis)
4. Optimizing permutations in biclustering algorithm (Chapter 6 of thesis)

CHAPTER – 3

Unique Brain Network Identification Number & Age-based analysis of Parkinson's disease

Chapter 3. UNIQUE BRAIN NETWORK IDENTIFICATION NUMBER & AGE-BASED ANALYSIS OF PARKINSON'S DISEASE

3.1. INTRODUCTION

Brain networks provide possible biomarkers not only for a disease cohort, but also for individual subjects [211]. Although group-level structural network analysis has been widely explored in PD, there is minimal research in individual-level network [211], [212], [213], [214] and structural connectivity studies in PD seem unexplored. Intriguingly, individual structural connectivity analysis using sMRI may play a crucial role in analyzing brain disease [211] and brainprinting [214]. However, due to inter-individual variability, there is a demand for encoding network into unique representation for each individual. This encoding scheme may also enable efficient storing and transfer of individual networks. Weininger et al. [215] proposed a chemical notation, specifically designed for chemical compound, where its structure is a network, atoms are nodes and bonds are edges. Later, Lukovits I.I. [216] obtained a compact form of a chemical compound, where presence of a bond between carbon atoms was represented as 1 in adjacency matrix. However, this method failed for compounds with more than 15 carbon atoms; thereby demanding more investigations in it. This work was studied in detail and shaped further to construct UBNINs for brain networks, where brain regions are nodes and the inter-regional connections are edges.

Studies have demonstrated PD with GM atrophy, but none have examined how Parkinson's brain networks change with aging. Age is the largest risk factor influencing the clinical progression of PD [217], [218], [219]. However, there are still undiscovered brain processes that underlie how ageing affects PD neurodegeneration. Also, a faster development of motor

disability, severe gait and postural impairment, often leads to dementia in PD patients with age [220]. Also, PD patients have vastly been observed with inability to understand, process information and other cognitive disabilities. Change in human brain structure affects not only memory and learning but also other cognitive skills as people age. It is envisaged that such improper and inefficient information dissemination in PD patients may be due to aberrant brain connectivity.

In this study, the main objective is to create a unique identification number to represent individual brain network. Hence, a model was proposed for automatically creating a unique brain network identification number for each individual. This was inspired by an encoding technique where a chemical structure was represented as adjacency matrix with carbon atoms as nodes and bonds are edges [216]. Additionally, it was hypothesized that disease progression with age may cause disruption of brain network. Therefore, brain network topology was analyzed over a range of sparsities and evaluated connectivity metrics for different age-cohorts of PD. This may address how age affects PD brain network. These analyses were also performed on healthy control individuals.

3.2. MATERIALS AND METHODOLOGY

3.2.1. Participants

Subjects were recruited at Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), India. Every individual had provided written informed consent in compliance with NIMHANS Ethics Committee. The data considered for the current study included 180 PD patients (135 males and 45 females; age: 54.84 ± 9.78 years, age range: 22-72 years) and 70 healthy controls, HC (52 males and 18 females; age: 49.24 ± 10.99 years, age range: 20-73 years). The PD subject characteristics obtained were mean Unified Parkinson's Disease Rating Scale at off state (UPDRS Off, 33.26 ± 8.92), at on

state (UPDRS On, 17.85 ± 6.27), Hoehn and Yahr scale (H&Y, Median: 2), and Age at onset (48.27 ± 9.14). For age-based analysis, the participants were categorized into five age groups (age group A: ≤ 32 years (n1=4, n2=5), B: 33-42 years (n1=18, n2=14), C: 43-52 years (n1=42, n2=23), D: 53-62 years (n1=69, n2=22) and E: ≥ 63 years (n1=46, n2=6), where n1 and n2 are the number of subjects in PD and HC respectively. Brain parcellation for one of the PD subjects in age group D was unsuccessful, due to which it was excluded from the study. The age-based participants' demographics are listed in **Table 8**. The study uses a retrospective dataset which has been previously employed on PD [82], [221].

3.2.2. MRI Scanning and Image Processing

A 3 Tesla Philips Achieva scanner with a 32-channel head coil was used to scan the participants. A high resolution T1-weighted structural MRI scan was obtained with a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (repetition time (TR) = 8.2 ms, echo time (TE) = 3.7 ms, flip angle = 8, field of view (FOV) = 256 * 256 * 165 mm3, 165 sagittal slices, voxel size = 1 * 1 * 1 mm3). The original scans (in

Table 8. Demographic and clinical information of participants

Age-cohort (Age range)	Count		Gender (Male : Female)		UPDRS Off (Mean ± SD)	UPDRS On (Mean ± SD)	H&Y (Median)	Age at Onset (Mean ± SD)
	PD	HC	PD	HC				
A (≤32)	4	5	3:1	5:0	31.69 ± 2.71	16.75 ± 9.91	2	16.83 ± 10.67
B (33-42)	18	14	11:7	7:7	32.94 ± 7.13	16.67 ± 6.16	2	35.19 ± 4.69
C (43-52)	42	23	28:14	19:4	32.33 ± 11.11	17.78 ± 8.15	2	42.83 ± 4.29
D (53-62)	69	22	56:13	16:6	32.79 ± 8.76	17.70 ± 5.37	2	47.47 ± 6.26
E (≥63)	46	6	36:10	5:1	35.17 ± 6.97	19.30 ± 4.78	2	51.54 ± 9.88
p-value					0.58	0.54		3.08×10^{-22} **

A,B,C,D,E: Age-based groups with age range in brackets; H & Y: Hoehn and Yahr scale; p-value: from ANOVA test; SD: Standard Deviation; UPDRS: Unified Parkinson's Disease Rating Scale; ** Significant at 99% confidence interval; H&Y (categorical value) is expressed as median.

DICOM format), were converted into NIfTI by using MRIcron's dcm2nii tool (http://www.nitrc.org/projects/mricron). The pre-processing steps are explained elsewhere [82]. This involved MATLAB-based Computational Anatomy Toolbox (CAT12, http://www.neuro.uni-jena.de/cat/) within Statistical Parametric Mapping (SPM12). Recent study [222] has inferred better age prediction using this pipeline and hence it was used for brain network analysis. The scans were segmented and smoothed before taking the GM map for further study.

3.2.3. Individual Brain Network Construction

The GM map of every individual was parcellated into 56 regions using LPBA40, LONI Probabilistic Brain Atlas [67] (https://resource.loni.usc.edu/resources/atlases-downloads/). The regions are illustrated in the appendix of [82] and **Supplementary Table S**. As PD subjects from both genders were taken and this study intended to observe the effect of age on PD, the effect of gender was disregarded voxel-wise using Linear regression. This adjusts or corrects for discrepancies between measurements obtained at different age.

Each brain region indicated a node in the network for which region-specific grey matter volumes (rGMV) were obtained using CAT12. Individual brain network was constructed [211] for each subject based on correlation (r_{ij}) between rGMV of each pair of regions. It was computed using equation 2 where $rGMV_i$ and $rGMV_j$ are regional GM volume of regions i and j.

$$r_{ij} = \frac{1}{\left(rGMV_i - rGMV_j\right)^2 + 1} \qquad (2)$$

The correlation or association matrix, hence obtained, was visualized as a network in BrainNet Viewer [100], with coefficients representing edges and regions indicating nodes.

3.2.4. Generating Unique Brain Network Identification Number (UBNIN)

The algorithm uses binary adjacency matrix (AM) of a network (N). Every individual association matrix was binarized using consistency-based thresholding to retain 30% of connections [223], [224]. Given that AM is symmetric, it suffices to take into account either upper or lower triangle of AM. Let's consider a sample network (**Figure 9**) with 10 nodes whose upper triangle matrix is shown as AM (

Figure 10). The lower triangular matrix (shaded in blue in

Figure 10, step 1) is zeroed due to symmetric matrix and undirected network, hence not included in computing UBNIN. The entry of any column is either 0 or 1, and the entire column is assumed as a single binary number starting from bottom-to-top (

Figure 10, step 2).

So, every column of AM, above principal diagonal, was transformed into a decimal equivalent (

Figure 10, step 3) [216]. Hence, the decimals thus obtained were 1, 3, 6, 13, 20, 60, 67, 1, 321 to be stored in an array DEC (i.e. decimal equivalents). It is apparent that the minimal

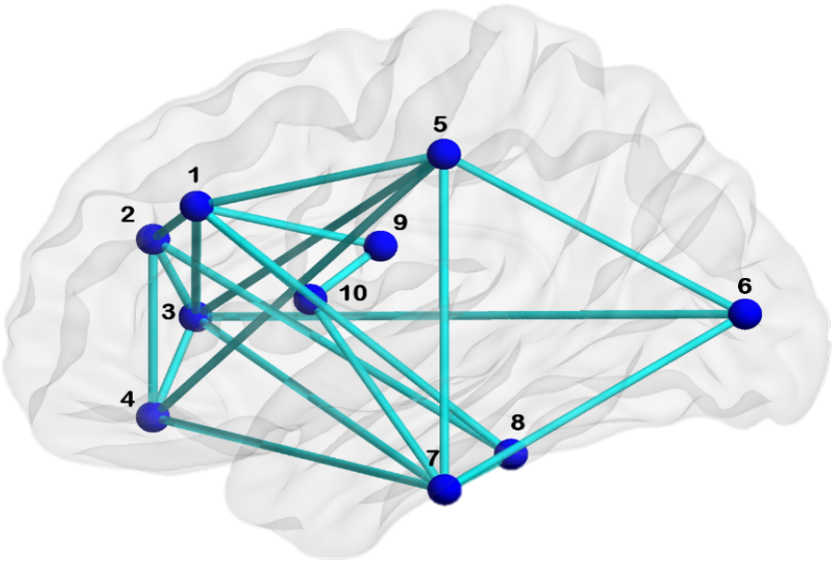


Figure 9. Sample network for deriving UBNIN

decimal value of any column is zero; the maximal value of a column *i* is $2^i - 1$, where *i* = 1 for first column of upper triangle of AM. A node with zero decimal value indicates a disconnected node, while a decimal value of $2^i - 1$ indicates node connecting to every other node. The maximal decimal equivalents of the columns (nodes) are 1, 3, 7, 15, 31…$2^n - 1$ for a fully connected network with n nodes. The values in DEC were further

converted into a single numerical form, named $UBNIN_T$, using equation 3, where T is the total number of nodes in the network or number of columns in the binary matrix.

$$UBNIN_{T=10} = \frac{1}{2^{i=8}} * \left[\left[\frac{1}{2^{i=7}} * \left\{\frac{1}{2^{i=6}} * \left\|\frac{1}{2^{i=5}} * \left(\frac{1}{2^{i=4}} * \left[\frac{1}{2^{i=3}} * \left\langle\frac{1}{2^{i=2}} * \left\lfloor\frac{1}{2^{i=1}} * DEC(2) + DEC(3)\right\rfloor + DEC(4)\right\rangle + DEC(5)\right] + DEC(6)\right) + DEC(7)\right\| + \right. \tag{3}$$

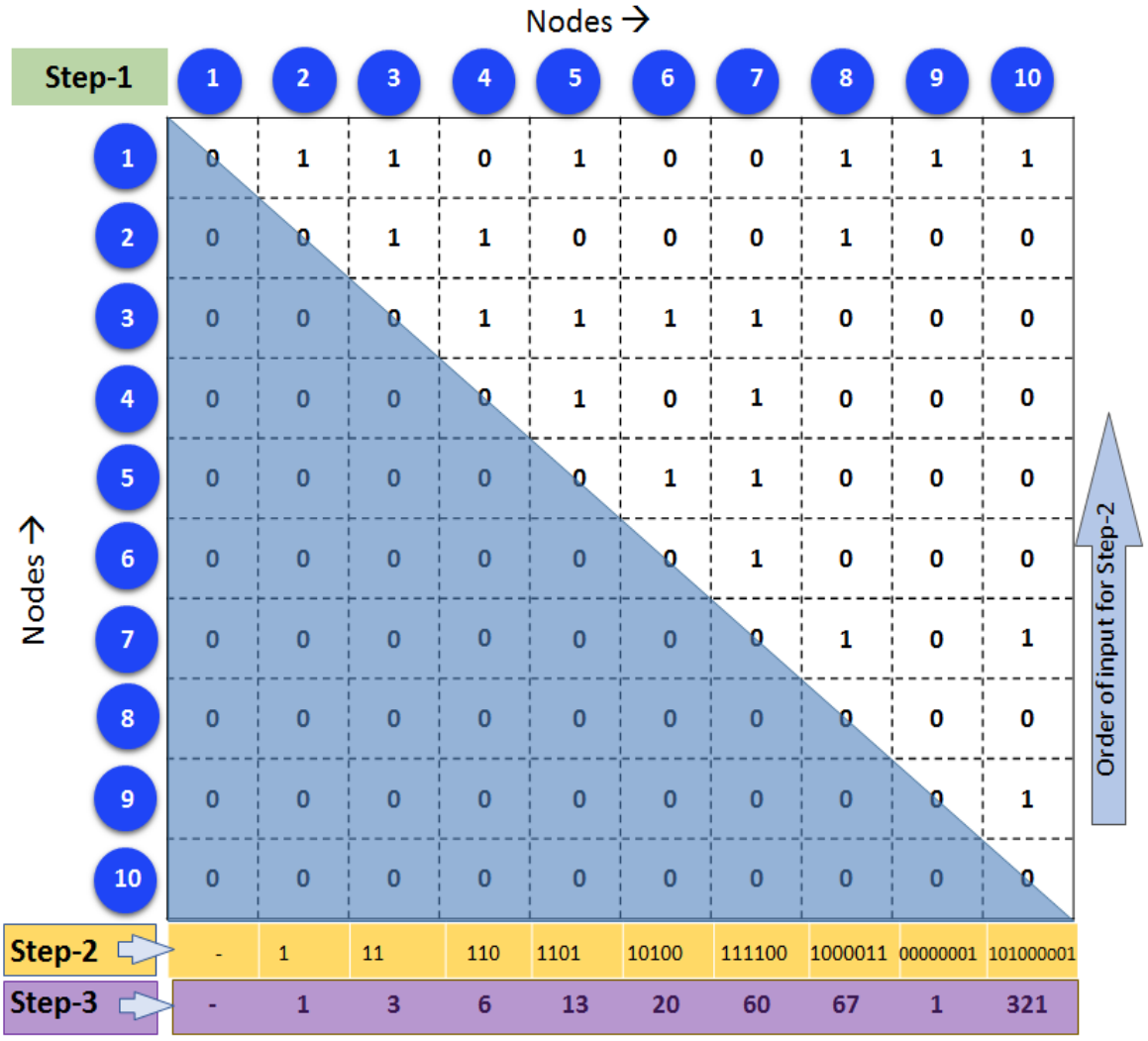


Figure 10. Adjacency matrix for sample network (as in Fig 9) for UBNIN computation. Step 2 and Step 3 are intermediate steps.

Equation 3 upon substituting corresponding decimal and node values from

Figure 10 gives:

$$UBNIN_{\mathrm{T}=10} = \frac{1}{2^8} * \left[\!\left[\frac{1}{2^7} * \left\{\frac{1}{2^6} * \left\|\frac{1}{2^5} * \left(\frac{1}{2^4} * \left[\frac{1}{2^3} * \left\langle\frac{1}{2^2} * \left\lfloor\frac{1}{2^1} * 1 + 3\right\rfloor + 6\right\rangle + 13\right] + 20\right) + 60\right\| + 67\right\} + 1\right]\!\right] + 321$$

This gives a value of $UBNIN_{\mathrm{T}=10} = 321.005979848895$, for the network (**Figure 9**) with 10 nodes.

Algorithm 1: UBNIN for an Adjacency matrix (AM) of a given network

```
N = length(AM); AM = upptriang(AM);
for j = 2:N
    BinNode = AM(j-1: (-1) :1,j);
    for a = 1:length(BinNode)
        Val + = 10^(j-a-1)(BinNode(a));
    end
    DEC_j = decimal(BinNodeVal);
end
temp = DEC_2;
power = 1;
for i = 2:N-1
    UBNIN = temp*(1/2^power);
    UBNIN += DEC_(i+1);
    power = power+1;
end
```

The power of $\frac{1}{2^i}$ starts with i = 1 and the equivalent decimal (*DEC*) starts with the second node of adjacency matrix. The top-to-bottom approach for this methodology was also attempted for multiple adjacency matrices, but the UBNIN hence obtained was larger than that from bottom-to-top. Hence, the binary number (bottom-to-top) of each node was deployed to convert into its corresponding decimal equivalent, and applied in (1) to obtain the UBNIN. The developed algorithm, is illustrated as Algorithm 1.

3.2.5. Group level network metric analysis

A subject-by-regions matrix for each age group was used as input, where subject count varied with age-cohort while number of regions remained constant. A weighted undirected matrix (association matrix) was constructed [82] using Pearson's correlation between regional GM volume of every pair of nodes in MATLAB (MathWorks R2020a). Pearson's correlation gives the measure of association between brain regions (nodes) and is used if the number of subjects is less than the number of regions [47], [201], [225]. Prior to constructing association matrix for age-based cohort, the same was made for the entire PD cohort for comparing the network pattern with that from healthy group.

To facilitate simplified network analysis, association matrix of each age-cohort was binarized across an optimized range of sparsities. Sparsity implies the densities of zeros in a matrix and is used as a measure of threshold, yielding an adjacency matrix by removing spurious edges. A sparsity threshold of 0.8 considers the highest 80% values only and rejects the lowest 20% values. The range is optimized in such a manner that below the lower limit, the network becomes equivalent to a random network (i.e. small-world index close to 1). And above the upper limit, the average degree becomes less than logarithm of the number of nodes and small world index is not estimable [77], [82], [226].

A complex brain network is often measured and analysed in terms of clustering coefficient. Nodal clustering coefficient (C_i) is defined as the fraction of directly linked neighbors which are interconnected among themselves [226], given as equation 4.

$$C_i = \frac{2t_i}{k_i * (k_i - 1)} \tag{4}$$

where i is node; k_i is total number of nodes; t_i is total number of edges between neighbors of i.

The nodal clustering coefficients and local efficiency were estimated using Brain Connectivity Toolbox-derived MATLAB functions [226] and averaged over the network to obtain mean clustering coefficient, separately performed for HC and PD. Structural brain network is the pattern containing edges between two nodes where interregional correlation exists. The networks for each age-cohort was visualized using BrainNet Viewer [100]. The sparsity-based method for binarizing association matrix for PD comparison with healthy was similar to that adopted for age-based groups.

3.2.6. Statistical Analysis

One-way analysis of variance (ANOVA) was performed between PD age-groups (five levels) to find differences in clinical variables (one factor), such as UPDRS Off, On and age at onset. A P-value < 0.05 was concluded to be significant. Inter-group differences in clustering coefficient were examined for both HC and PD using a nonparametric permutation test [82], [202] with 1000 iterations to determine their statistical significance. While preserving the number of subjects in each group, clustering coefficients were randomly assigned to either groups. Randomized group differences determined permutation distribution of the differences. Thus, actual inter-group differences were compared using self-developed MATLAB code.

3.3. RESULTS

3.3.1. UBNIN for Individual Brain Network

In this study, a unique brain network identification number (UBNIN) was proposed. It is a numerical representation scheme which was observed to be distinct for every individual network. The weighted undirected matrix, hence formed, contained 56 nodes and 1540 (= 56 × 55/2) edges. Adjacency matrix for each patient was constructed using 30% consistency-based thresholding. The UBNIN of PD and HC individuals, computed using the given MATLAB-based *.p* code, were all distinct and ranged between 15360 to 17768936615460608 and 12288 to 17733751438064640, respectively. UBNIN values of fully connected networks are illustrated in **Table 9**.

3.3.2. Network metric analysis

A significant difference was found in age at onset ($P = 3.08 \times 10^{-22}$, $F = 36.72$) between the age-based groups of PD at 99% confidence interval (**Table 8**). However, no significant intergroup difference was observed in UPDRS at Off ($P = 0.58$, $F = 0.72$) and On ($P = 0.54$, $F = 0.78$) state (**Table 8** & **Figure 11**). Detailed demographics of various age groups of HC and PD are in **Table 8** and group comparison of clinical features of PD is reported in **Table 8** and **Figure 11**. Pearson correlation was performed to obtain detect relationship between regions. The weighted undirected correlation matrix for PD is shown in top row of **Figure 12**. Strength of the connections are indicated by color-bar. The binary undirected correlation matrix of PD age-groups are shown in middle row of **Figure 12**. The axial view of undirected networks for each PD age group are respectively shown in bottom row of **Figure 12**, where intra- and inter-hemispheric connections can be seen. These are shown at the same sparsity threshold of 0.8 for illustration purpose only.

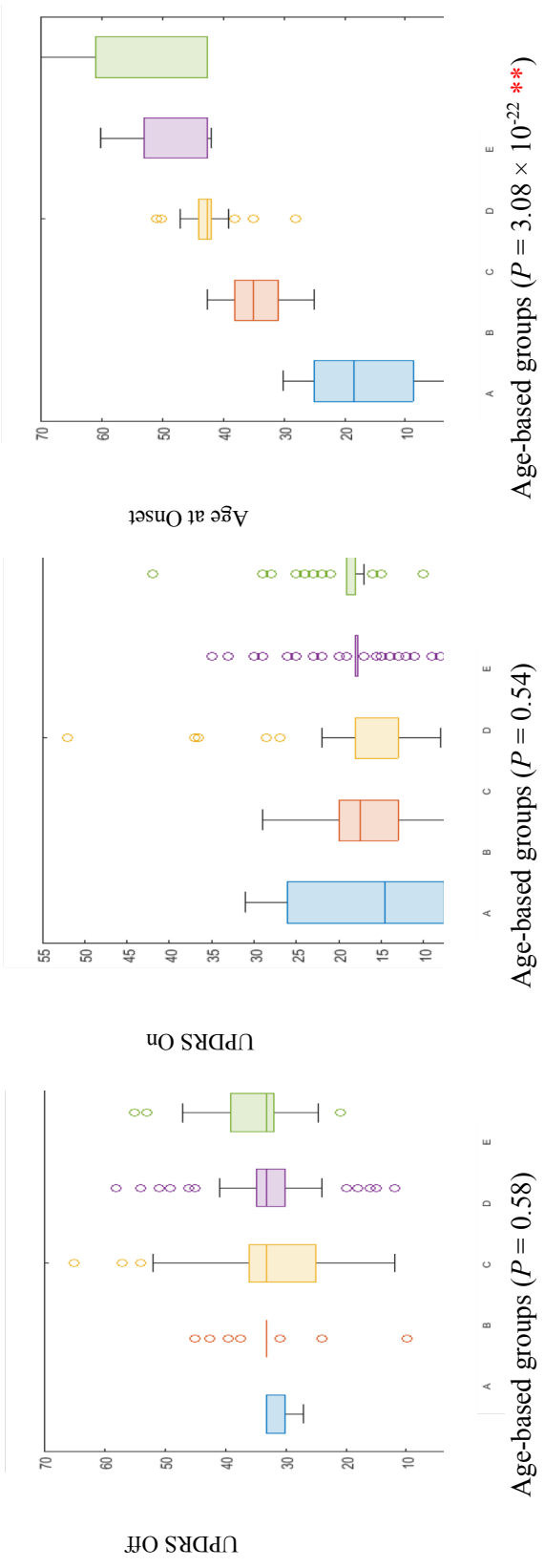


Figure 11. ANOVA of PD clinical features across age-groups. A, B, C, D and E are age-groups; UPDRS: Unified Parkinson's Disease Rating Scale; Notches indicate median; Top, bottom edges of the box indicate 25th, 75th percentiles; ** Significant at 99% confidence interval.

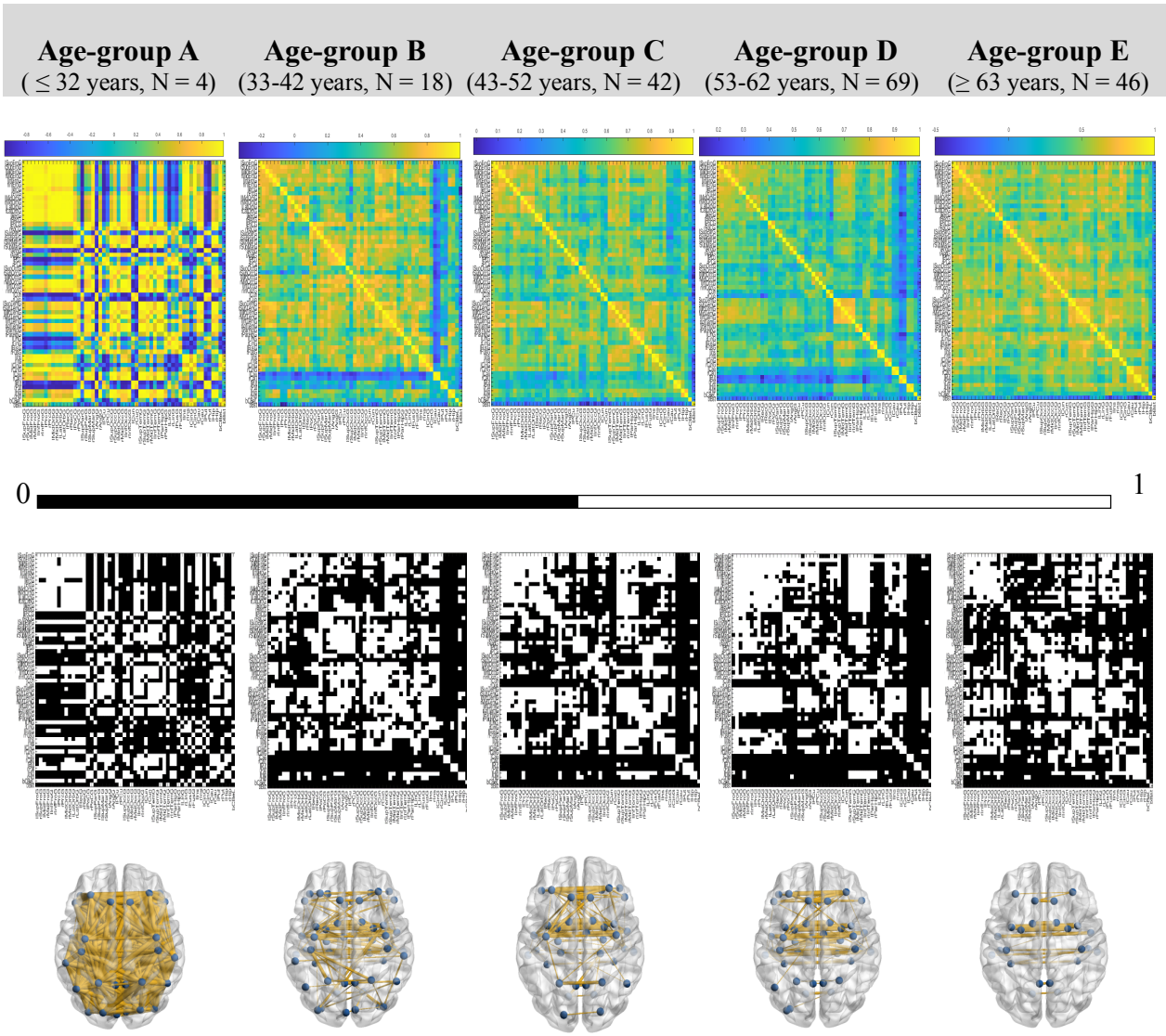


Figure 12. Age-based Weighted Undirected Matrix (top row), Binary Undirected Matrix (middle row) and Undirected Network (bottom row) of PD. Color-bar shows strength of connections. N is number of subjects in specified age range. Networks are at 0.8 threshold for illustration purpose only. The x- & y- axes in weighted & binary undirected matrices are brain regions in order as listed in [13], [22]

Table 9. UBNIN values of fully connected networks

Number of nodes	UBNIN value
10	511.99999999998544808477163314819335937 5
20	524288
30	536870912
40	549755813888
50	562949953421312
100	633825300114114700748351602689
150	$7.136238463529799405291429847247475681 91 \times 10^{44}$
200	$8.034690221294951377709810461705813012 61 \times 10^{59}$
250	$9.046256971665327767466483203803742801 04 \times 10^{74}$
300	$1.018517988167243043134222844204689080 53 \times 10^{90}$
500	$1.636695303948070935006594848413799576 11 \times 10^{150}$
800	$3.334007216439927137039925895360628898 57 \times 10^{240}$
1000	$5.357543035931336604742125245300009052 81 \times 10^{300}$
1020	$5.617791046444737211654078721215702292 56 \times 10^{306}$
1024	$8.988465674311579538646525953945123668 09 \times 10^{307}$
1025	NaN

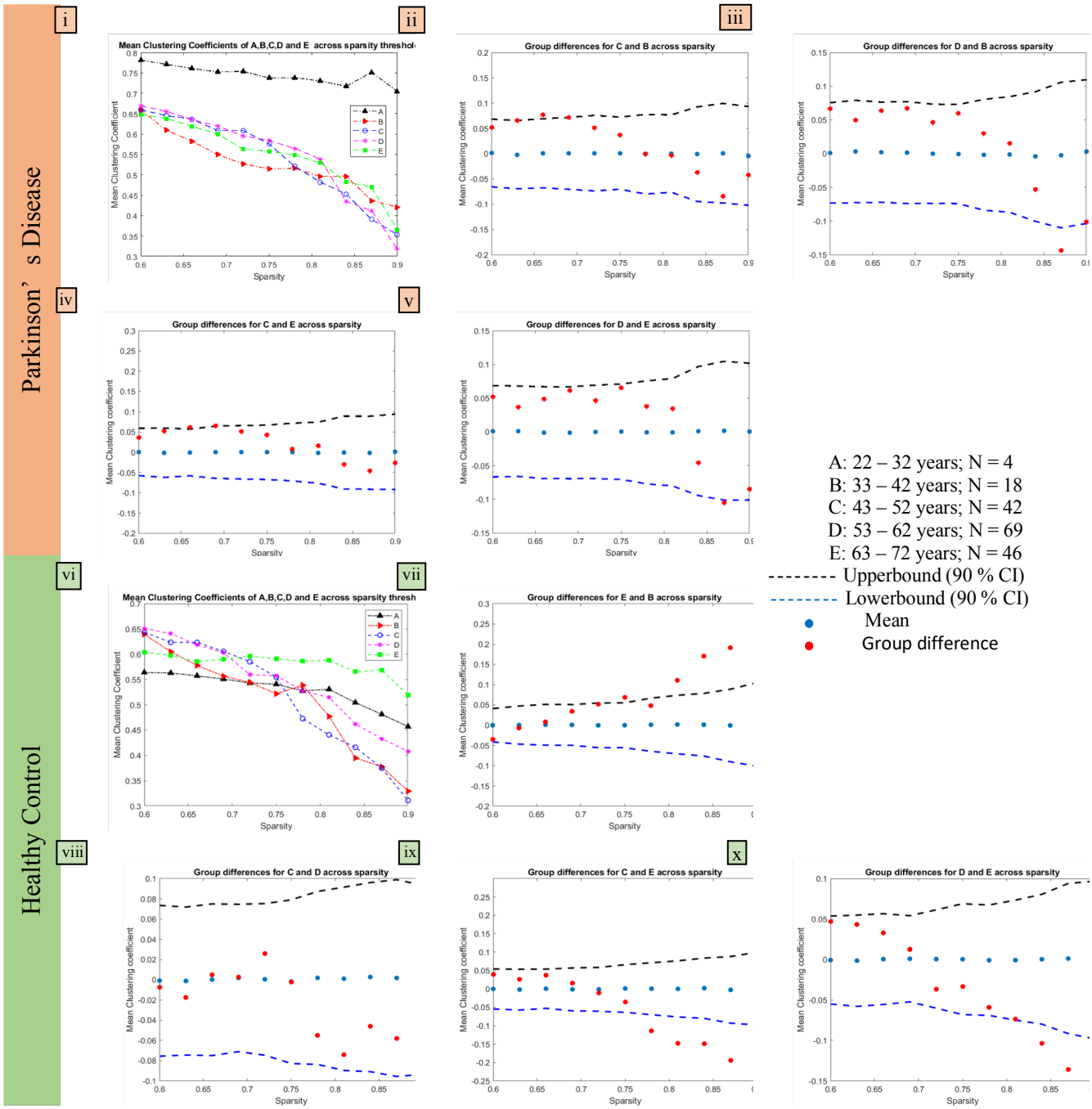


Figure 13. Distribution and intergroup differences in mean clustering coefficient for age-based groups. The subfigures are numbered (i-v for PD and vi-x for HC) and colored as orange for PD and green for HC individuals.

Sparsity threshold was estimated from the data and every age-cohort was analysed in a range of 0.6 to 0.9 with a step size of 0.03 [77], [82]. Small-world index exceeded 1 for every threshold level. Clustering coefficient of networks obtained was higher and at the same time the mean shortest path length was equivalent to that of random network in the

sparsity range. The mean clustering coefficient of PD was found to be lesser than HC at a sparsity of 0.9, which signifies reduced local integration and disrupted communication in the disease (**Supplementary Figure S1**, A.2). Further, reduced local efficiency in PD at sparsities above 0.84 (**Supplementary Figure S1**, B.2) reveals restricted information exchangeability [227]. Nodal local efficiency assesses how efficient communication is among the node's immediate neighbors [201], [226], [228], when a node is removed.

Figure 13 illustrates distribution of mean clustering coefficient and the significant comparison of different age groups as obtained from adjacency matrix. The mean clustering coefficient in PD declined over sparsity except for group A at 0.87 sparsity (**Figure 13**-i), suggestive of cognitive decline with aging in PD patients. However, more investigation is encouraged on large sample of young PD patients to confirm the findings on group A. Interestingly, **Figure 13**-vi illustrates reduction in mean clustering coefficient in HC with age, except for some variability in group B (at 0.78 sparsity) and E (at 0.72 sparsity). This decline in network metrics indicate change in network topology which may be due to natural aging in HC. The blue circles (**Figure 13**-ii to **Figure 13**-v) are mean differences in network metrics obtained from averaging 1000 permutation. The mean clustering coefficient of group A was observed to be significantly higher than that of B, C, D and E age groups. The obvious fact that there are fewer patients at younger ages, especially less than 32 years, calls for further research on a bigger dataset.

In PD, clustering coefficient of group C (between 43 and 52 years) is significantly different from that of PD group B (33-42 years) at 0.66 sparsity level (**Figure 13**-ii). The clustering coefficient of PD group C (between 43 and 52 years) is significantly different from that of PD group E (≥63 years) at 0.66 and 0.69 sparsity (**Figure 13**-iv). It is overall higher in PD group C than PD group E at lower sparsities up to 0.75, but lower at higher sparsities from 0.84 onwards. Also, the clustering coefficient of PD group D (53 – 62 years) is significantly

different from that of PD group B (33-42 years) at higher sparsity levels of 0.87 and 0.9 (**Figure 13**-iii). It is higher in PD group D than B till 0.81 sparsity, but afterwards it is lower. The clustering coefficient in PD group D is significantly different from that of PD group E (≥63years) at 0.87 sparsity (**Figure 13**-v). PD group D has a larger clustering coefficient than PD group E up to 0.81 sparsity, but after that it decreases. However, group comparisons showed no difference between PD group B (33-42 years) with PD group E (≥63 years), and PD group C (43-52 years) with PD group D (53-62 years).

A similar permutation analysis in HC, showed clustering coefficient for age-group A was significantly different from every other groups. However, HC group E (≥ 63 years) was significantly different than that of HC group B (33-42 years) at sparsities of 0.75 and above 0.81 (**Figure 13**-vii), HC group C (43-52 years) at sparsities above 0.78 (**Figure 13**-ix), and HC group D (53-62 years) at sparsities above 0.84 (**Figure 13**-x). Clustering coefficient indicates the local integration and communication between nodes. So, when networks of two groups are compared using permutation test, interaction between the nodes is studied at structural level. Clustering coefficient of HC group C also differed significantly from that of HC group D at sparsity of 0.9 only (**Figure 13**-viii). However, the same analysis may be replicated on a bigger HC dataset, especially in age groups A and E.

3.4. DISCUSSIONS

3.4.1. UBNIN for Individual Brain Network

This proposed algorithm (UBNIN) enables numerical representation of individual brain network and hence, holds significance in encoding. The adjacency matrix comprises of binary data in rows and columns. Each of these rows and columns correspond to a specific node. The values in the columns (from bottom to top) are converted into their corresponding decimal numbers (Figure 10). The decimal value of the 2nd node was initially multiplied by

$1/2^{i=1}$ (i= n-1, where n = 2 for 2nd node) and then added the decimal equivalent of 3rd node to obtain the partial sum. This partial sum is again multiplied by $1/2^{i=2}$ (i= n-1, where n = 3 for 3rd node). Thus, after the partial sums are first divided by increasing powers of 2, these numbers are added one after the other in a sequential manner so that, at the end of each iteration, the fractional part of the partial sum comprises the information of the previous nodes. The finally obtained UBNIN value (using equation 3) has the integer part (which is highly weighted) from the last column (node), since when dividing by the increasing powers of 2, the quotient is always less than 1.

An approach (similar to UBNIN) named BIN [216] was applied to a network, where carbon atoms were nodes and the bonds were edges. Although it could compress the adjacency matrix, it increased exponentially with number of nodes, and hence had limited applicability in chemical structures with 15 nodes only. However, this method can compute the encoded numerical representation for fully connected network with upto 1024 nodes, after which it generates not a number (*NaN*) in MATLAB (**Table 9**). The UBNIN of a fully connected network with 100 nodes would yield just a 30-digit value, whilst the method described in [216] generates infinitely large value (*Inf*) even for a small network with 46 nodes, which is uninterpretable. For networks containing more than 1024 nodes, $\frac{DEC(i)}{2}$ is suggested to use, i.e., half of the equivalent decimals, instead of direct decimals in equation 3. Previous studies have confirmed that structural connections in brain are unique to each individual person [213], [214]. Also, the architecture of region-to-region connectivity reflects the complex wiring system or connectome of the individual which has been used for brain fingerprinting [214], [229]. Hence, the numerical representation system proposed in this study has a potential applicability to be used as a neural signature of a person's unique structural connectivity. The future of individualized mental health care may lie in this brain fingerprinting, which has demonstrated promising marker for mental illness

[230]. Researchers may be able to separate various neuropsychiatric illnesses using connectivity patterns and detect network-level abnormalities in specific patients. In the end, such information might enable an important change in clinical neuroscience and enable personalized prognosis for individual treatment [231]. Another aspect which could be explored is the dynamic behaviour of human brain as a resultant of brain plasticity. The degeneration and compensatory mechanism play a vital role in human brain leading to reorganization of brain structure, functions and connections. Structural MRI helps quantifying spatial patterns and is more stable during data acquisition compared to other neuro-imaging modalities and doesn't depend on the physical condition of the subject. The results reported in the current study is upon implementing UBNIN algorithm using sMRI. Each time, the algorithm produces the same UBNIN value on any given network and hence, is representative of the network. However, any change in MRI scans leads to alterations in brain network, which will be ultimately reflected by a different UBNIN value. The same may be tested on a longitudinal data of PD patients to analyse the network change over time, by tracking changes in UBNIN values. Further it was tested on healthy controls and observed every individual having unique UBNIN. However, the individuals can't be directly categorized based on their UBNIN values because of different network topology and connection patterns. A higher UBNIN value doesn't necessarily imply denser network, but it depends on the nodes to which it is connected. Additionally, the algorithm was also implemented on multiple (count =100000) random binary matrices of dimension 10 x 10 and found distinct UBNIN values each time.

Threshold value affects adjacency matrix for computing individual UBNIN from structural MRI, and hence it is crucial to select an appropriate thresholding method. UBNIN is also advantageous in saving storage memory and offering low-space-high-speed transferring of network information. The adjacency matrix of an individual is always a 2-D symmetric

matrix and the corresponding network consume a lot of computer storage space and takes time and space for transferring. The number of rows and columns in the adjacency matrix, however, depend on the brain atlas used during parcellation. Here in this study, it's 56 × 56, which may increase to higher count in other brain atlases. Also, if sparsity thresholding measure is applied (eg. 10 different sparsities) for 1000 subjects, 10000 numbers of 2D matrices were obtained. This imposes storage burden on computer systems and databases. The idea emerged while performing network analysis on multiple adjacency matrices on the age-cohorts. Not only vast computer memory, but also huge computation time was consumed, which motivated us to find an encoding scheme for these networks. Hence, this encoding approach (UBNIN) may help overcome the problem, and applied to any network of large number of nodes.

3.4.2. Group-level network metric analysis

Reduced mean clustering coefficient in PD indicates reduced local integration and disrupted communication in the disease compared to HC. This corroborates with earlier findings in early drug-naïve PD patients in magnetoencephalography and DTI studies respectively [88], [232]. Previous study on PD with hemiparkinsonism and HC showed no difference in clustering coefficient [201], while another investigation showed higher clustering coefficient in PD compared to HC [202]. As current study finds reduced local efficiency in PD, impaired local efficiency was also reported in PD from a longitudinal magnetoencephalographic study [232] and PD with tremor from a functional network study [227]. Additionally, PD patients with MCI were characterized with reduced local efficiency compared to PD with normal cognition and healthy controls, both [233]. On the contrary, PD patients were observed to have higher local efficiency from a morphological network analysis [227] and no difference from healthy control subjects in another study [201].

Although standard parcellation method has been employed in the current study, use of connectivity based parcellation is encouraged in PD [234]. The results (from 4th and 5th column of **Figure 12**) accentuates progressive decline in structural connectivity with increase in age which indirectly affects the functional characteristics, which corroborates with normal aging in the earlier findings [235]. Brain network states the pattern of connections between neuronal elements and explains information processing inside the brain [47], [58], [82]. Connectivity studies using sMRI have analyzed brain network in PD compared to healthy control [82], [201], [202], compared across gender [236], and with different associated dysfunctions like tremor [227] and mild cognitive impairment [233]. Hence, there is a lack of structural brain connectivity studies in PD spanning the lifetime.

A significantly higher mean clustering coefficient in group A (youngest age group of the current study) was obtained from every other age-cohorts. This suggests higher information integration in young adults i.e., up to 32 years old. This is in agreement with a previous study [237] on healthy individuals from electroencephalogram (EEG) signals. The functional connectivity network from the beta- band working memory task evoked a significant increase in clustering coefficient for younger adults (19-29 years) than that of older participants (58-70 years) [237]. Another study, analysing longitudinal functional connectivity using magnetoencephalography, found early untreated patients displaying lower clustering compared to controls [232]. However, due to limited subjects in group A, additional investigation is suggested with larger sample size. It's intriguing to note that inter-group differences were observed in almost all PD age-groups, B (33-42 years) vs C (43-52 years), B vs D (53-62 years), C vs E (≥63), D vs E. This may be suggestive of the fact that clustering coefficient in PD varies with age, affecting the network organization and its ability to aggregate information from different nodes. Hence, the disease affects differently at different age-groups. Network degeneration is also consistently found in

healthy control due to natural aging. The differences in clustering coefficient is evident in almost all pairs of HC age-groups (except for B vs C and B vs D). This research may be further replicated on other available large datasets. Since, clustering coefficient measures the functional segregation [238], functional connectivity studies have shown it decreasing with age [239], which corroborates with result of this current study. Network metrics have shown highly correlating with each other [223], which may be investigated in age-groups of PD patients from multiple sites and scanners. Due to limited availability of clinical features in the current study, more research may confirm the clinical changes over age. However, previous studies [219] have confirmed that age strongly affects clinical manifestation of PD. Additionally, functional analysis may provide complementary understanding of functional changes taking place along with change in network topology.

The primary risk factor associated with PD, which also adds to its onset and progression, is aging. Results of the present study indicate chronological age is proportional to the age at disease onset. However, age of a patient, rather than age at onset is a factor that greatly affects PD progression, including the emergence of dementia and hallucinations [240]. The clinical manifestation of PD may change as people age, and these changes may include changing pharmacological side effects, increased risk of dementia, hallucinations and increased likelihood of nursing care admission. Hence, age-related brain changes in PD is reflected by motor and non-motor dysfunctions, as phenotypic presentation of the disease [240]. The neuronal processes underlying the consequences of aging on neurodegeneration of Parkinson's disease, however, remain unanswered. In the present research, it was discovered that aging significantly affects grey matter networks of Parkinson's patients. This is evident from the constant fall in clustering coefficient with sparsity in age-groups, with a slight variability which may be due to compensatory mechanism [227]. As grey matter covers the neuronal cell, it is assumed that neuronal decay over time affects the grey

matter volume and thereby lead to network aberration. This may offer potential cause to explain how aging leads to neurodegeneration in PD due to which information is perceived differently. Results from diffusion tensor imaging study showed Parkinson's patients at older age exhibiting disrupted white matter network topology compared to young age groups [241]. The clustering coefficient of PD patients in 69 - 82 years was found lesser than that at 34-69 years. However, the same study did not find any difference in grey matter covariance network using similarity-based method on structural MRI data or functional network over age [241]. Age-associated structural brain networks were analyzed in order to better comprehend the underlying connections across age. It may assist researchers and physicians in offering age-specific treatment. Anti-aging medicines might open new doors for the creation of disease-modifying therapeutics for PD [242], [243] to interfere the pathophysiology of the disorder [244], [245]. It is assumed that investigation of the present study may play a new role in assisting researchers to understand variation in connections across age.

3.4.3. Limitations and future scope

The entirety of UBNIN value can be precisely calculated manually (using paper and pen) as a unique identifier for a binary matrix, so also reconstruction of the original binary matrix. The current computer hardware technology limits the number of significant digits, which slightly hampers accurate reconstruction of larger binary matrix.

For a larger 20 × 20 fully connected binary matrix, 190 digits would be required to represent the collection of decimal values. The UBNIN value hence obtained are rounded up (as shown in appendix) which is a computational constraint. This limitation is illustrated in https://github.com/NeuralLabIITGuwahati/UBNIN/blob/main/UBNIN_reverse_calculation.xlsx where the multiplication factor ($\frac{1}{2^i}$) rounds up at large i (i.e. node) values.

The network pattern changes over time [246], with varying degree of change, both in normal healthy control [237], [247] and different disease conditions [214]. Although previous studies have explored this phenomenon using diffusion MRI [213], [214] or electroencephalography [229], yet structural MRI data, due to less temporal resolution, may offer a more stable result. Besides individual brain networks from sMRI, UBNIN may be applied to other imaging modality as well. It may also be implemented to encode any binary network viz. internet, citation network, economic network and even social networks. Earlier researches [214] have observed that brain's unique pattern changes over time, and also is affected by disease, environment and other factors. This UBNIN, verified using structural MRI, may pave the way towards understanding brain plasticity. The individual subject UBNIN may further be investigated to find association with IQ as well as other clinical features of PD.

Since prodromal stage is a crucial time in PD and is not explored yet, more research is required in this direction. Also, longitudinal study may help understand the underlying change in brain network over time due to the disease. The neurodegeneration in the disease may thus be understood clearly in PD. Although PD is uncommon at younger age and the immensely dense connectivity at age-group (≤ 32 years) is an obvious truth, the connectivity could be investigated in detail with large sample size of younger patients to better understand the effect of Parkinson's disease. The HC individuals were also limited and hence the analysis may be replicated on a larger dataset. Employing connectivity information from other imaging modalities may help draw more firm conclusions about aging in Parkinson's disease as evident from loss of connections and differences in network metric.

3.5. CONCLUSION

A novel encoding system is proposed, namely, Unique Brain Network Identification Number (UBNIN) for individual brain network. To the best of knowledge, this is first of a kind study on in Parkinson's disease employing sMRI information. Moreover, the numerical representation of brain network could render efficient transmission and storage of network information. A change in the encoded number might reflect a change in the network pattern. Since every individual has a unique pattern of brain network, this approach may be employed for creating individual-brain ID to store in a database and also for further analysis. It also has a potential scope in brainprinting as it stores the unique identity of the person and could be taken up as a research prospect for understanding brain plasticity. The age-based network topology was also investigated and connectivity analysis was performed. This investigation on sMRI- based structural network demonstrated varying clustering coefficient at different age in PD. As clustering coefficient reveals statistically significant group difference, it could be treated as a biomarker to understand the information communication at different age. This study could be taken as a base work to analyze other connectivity metrics so as to perform a comprehensive connectivity analysis.

Data and Code Availability: The imaging and clinical data were collected from National Institute of Mental Health and Neurosciences and may be provided on approval as per Institute norms. The code developed for this work is available at below GitHub link:

https://github.com/NeuralLabIITGuwahati/UBNIN

CHAPTER -4

Brain Connectivity of Grey Matter Based Subtypes using Structural MRI for Parkinson's disease

Chapter 4. BRAIN CONNECTIVITY OF GREY MATTER BASED SUBTYPES USING STRUCTURAL MRI FOR PARKINSON'S DISEASE

4.1. INTRODUCTION

Clinical symptoms-based classification of Parkinson's disease (PD) patients has shown overlapping disease subtypes. Several efforts have been made in the past towards predicting and classifying PD patients based on non-motor and motor characteristics [248], as well as age of onset, proving patients could belong to multiple subtypes [47]. As a result, paucity of coherent clinical diagnostic scale and unavailability of standard biomarker encourages further investigation to probe stable PD subtypes [111], [112], [249], [250]. This is crucial not just for data-driven differential diagnosis of the disease, but also for subtype-specific network biomarker, thereby, be used to identify inter-subtype differences. Hence, more research is needed to categorize Parkinson's patients into stable and definitive classes using neuroimaging-based data-driven algorithms. Most of the state-of-the-art methods classify PD patients from healthy control, but limited attempt has been made towards finding subtypes within the disease. Source–based morphometry (SBM) is a neuroimaging-based data-driven approach to decompose brain data into spatially independent components. This multivariate approach employs independent component analysis (ICA) for doing so, prior to determining the significant sources that differ between disease and health [251], [252]. SBM has been applied in wide variety of application such as in healthy aging [253], PD [254] and based on the sMRI information such as grey matter density [204], [251], [255] and cortical thickness [256].

Additionally, it's equally important to analyse brain network differences between the data-driven subtypes. Since network analysis is difficult at a single threshold [227], network metrics have been estimated over a range of threshold values [82]. Sparsity, as a measure of threshold, is derived from the data and plays a crucial role in estimating the metrics, performing brain network analysis and differentiating the patients network from that of the healthy subjects [82]. Hence, based on the fact that grey matter (GM) differences exist in atrophy regions, hypothesize presence of subtypes within the disease is hypothesized. To accomplish this, the current study exploited GM information from sMRI to find PD subtypes. In addition, presence of network topology differences is postulated, which is supported by comparative brain network analysis. The current study employs an international multicentre dataset to discover subtypes within Parkinson's disease and perform a comparative connectivity analysis of deciphered PD subtypes.

4.2. MATERIALS AND METHODOLOGY

The proposed approach for subtyping of Parkinson's disease based on Magnetic Resonance Imaging (MRI) scans basically consisted of three main stages: (1) identification of brain components with significant grey matter difference between Parkinson's patients and healthy control, (2) selection of two components of interest, (3) detection of subtypes within PD only patients based on these components satisfying the selection criteria. This was followed by structural brain network analysis of subtype-specific subjects. The entire study of this chapter is graphically illustrated in **Figure 14**.

4.2.1. Participants Demographics and Clinical Features for subtyping using grey matter volume

For conducting this study, imaging and clinicopathological data of the participants, at their first visit, were obtained from PPMI (Parkinson's Progression Markers Initiative

www.ppmi-info.org/access-data-specimens/download-data), a multicentre database. This is a publicly-shared large multicentre dataset which extends to the full breadth of every individual. The single scanner data (Siemes Trio Tim) of 131 PD patients and 78 healthy controls (HC) from 10 different sites were considered for this study. Demographic data of these subjects such as age, gender, handedness, education and other clinical features are presented in **Table 10**. The missing records in any of the clinical variables of the data set were imputed using mean of the respective available data.

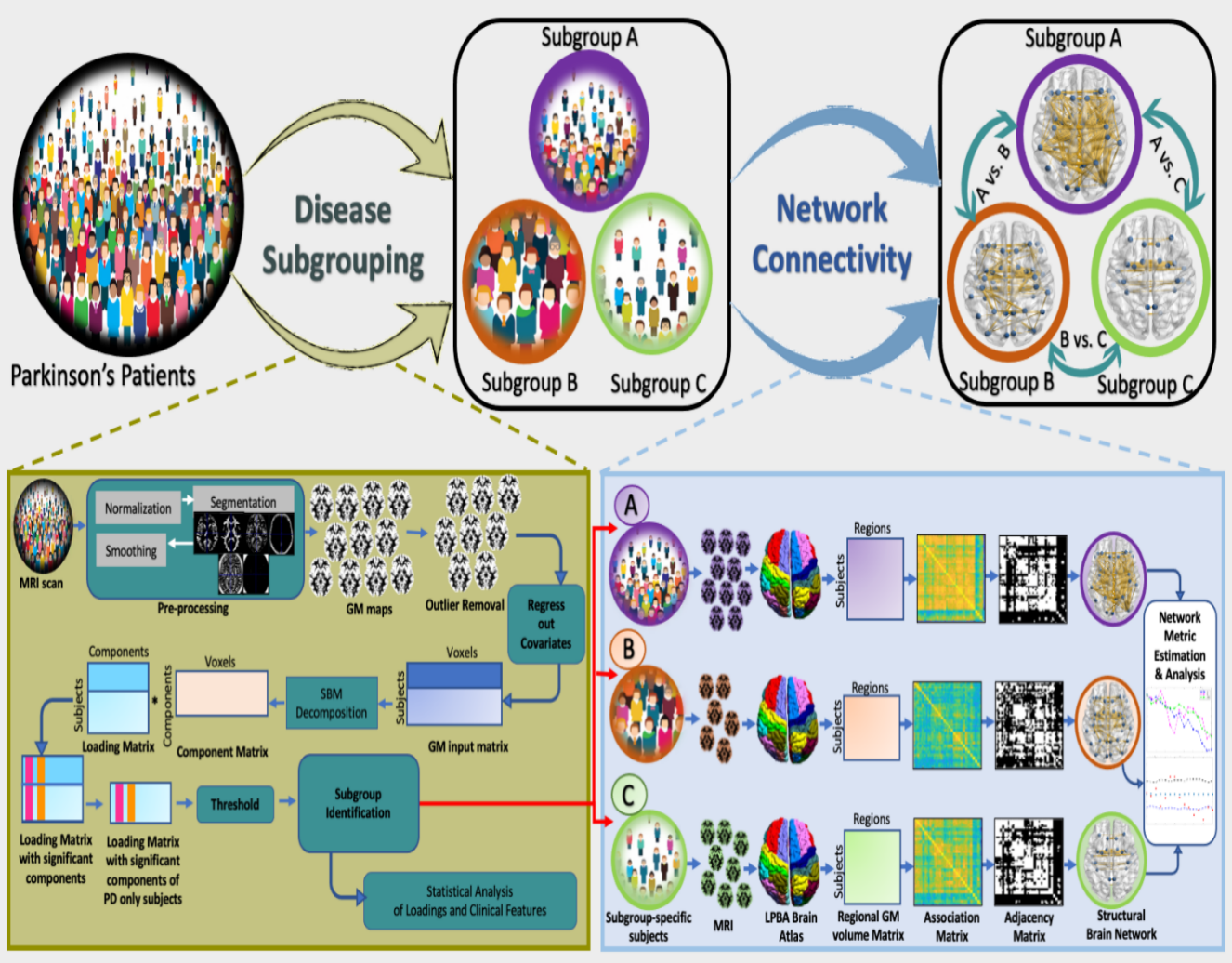


Figure 14. Graphical abstract illustration methodology of this study

Table 10. Demographic characteristics of the participants from PPMI database for subtype identification using single feature

Sl.	Features	Healthy Control	Parkinson's Patients	χ2 or t-value	p-value
1	Count	78	131	-	-
2	Age in years	60.20 ± 10.10	58.16 ± 7.67	1.65	0.1
3	Age Range	32-80	38-70	-	-
4	Gender (Male : Female)	53:25	82:49	0.61	0.43
5	Handedness (R:L:B)	65:5:8	117:10:4	4.72	0.09
6	Education (in years)	15.35 ± 4.57	15.54 ± 2.67	-0.39	0.696
7	Site	10 sites	10 sites	-	-
8	Montreal Cognitive Assessment (MoCA)	28.28 ± 1.15	27.64 ± 2.09	2.48	0.01**
9	Geriatric Depression Scale (GDS)	1	2	-	-
10	State-Trait Anxiety Inventory for Adults Test (STAIT)	57.65 ± 14.08	67.16 ± 22.39	-3.37	8.92×10^{-4}**
11	University of Pennsylvania Smell Identification Test (UPSIT)	-	22.77 ± 8.21	-	-
12	Random Eye Movement Sleep Behavior Disorder (RBD)	2.68 ± 2.13	4.69 ± 3.33	-4.79	3.17×10^{-6}**
13	Scales for Outcomes in Parkinson's Disease- Autonomic (SCOPA-AUT)	6.08 ± 3.83	10.28 ± 7.19	-4.77	3.4×10^{-6}**
14	Letter-Number Sequencing score (LNS)	-	10.75 ± 2.69	-	-
15	Symbol-Digit Modalities score (SDM)	46.42 ± 9.96	41.09 ± 10.11	3.71	2.68×10^{-4} **
16	Unified Parkinson's Disease Rating Scale III (UPDRS-III)	0.71 ± 1.63	23.13 ± 12.44	-15.83	0**
17	Unified Parkinson's Disease Rating Scale IV (UPDRS-IV)	-	1.69 ± 2.96	-	-
18	Questionnaire for Impulsive-Compulsive Disorders (QUIP)	-	0	-	-

SD: Standard Deviation; χ2 test for Gender; Student's t-test for Age, R: Righthanded, L: Lefthanded, B: Ambidextrous; Subject count, gender and handedness are given as count; GDS and QUIP are given as median. All other features are given as Mean ± SD; **significant at 99% confidence interval using t-test

4.2.2. Pre-processing of sMRI

The single scanner T1-weighted structural MRI scans of brain were selected from PPMI dataset as generated by employing Siemens TrioTim 3 Tesla scanner with a 32-channel head coil. Due to heterogeneity of the brain scans from PPMI dataset, it was opted to combine data obtained using identical acquisition parameters—most notably, the manufacturer of the scanner and the magnetic field—in order to remove any potential sources of bias. The single scanner T1-weighted structural MRI scans, generated by Siemens TrioTim 3 Tesla scanner, were acquired in the sagittal plane with Repetition Time (TR) = 2300 ms, Echo time (TE) = 3.0 ms, Inversion time (TI) = 900 ms, Slice thickness = 1.0 mm, Pulse sequence = GR/IR and flip angle = 9°. The original scans were obtained in NIfTI (Neuroimaging Informatics Technology Initiative) format and were preprocessed using Computational Anatomy Toolbox, CAT12 (http://www.neuro.uni-jena.de/cat/) in MATLAB platform (MathWorks, MATLAB 2020a) within Statistical Parametric Mapping, SPM12 (https://www.fil.ion.ucl.ac.uk/spm/).

MRI of every individual was divided into different tissue types, viz. grey matter (GM), white matter and cerebrospinal fluid. These were normalized to the standard Montreal Neurological Institute (MNI-152) template, and modulated using Jacobian determinant. Further, a linear regression analysis was performed (MATLAB 2021a) to voxel-wise remove the effect of nuisance variables such as age, gender and site [257]. The scans were then smoothed using 8 mm (full-width at half maximum) isotropic Gaussian kernel in SPM.

4.2.3. Subtyping of PD patients only - using grey matter volume

Parkinson's patients were categorized into different subtypes by employing sMRI scans. The stepwise procedure followed to obtain subtypes are: (1) construction of input GM

volume matrix, (2) decomposing the input GM matrix into a loading and a component matrix using source-based morphometry (SBM), (3) identifying components with significant grey matter differences between Parkinson's patients and healthy controls, (4) selecting only the significant components, (5) PD subtype identification by thresholding the component loading.

The smoothed GM tissue probability maps obtained from SPM were correlated with the mean scan of each group, to remove the outliers, if any, present in the data. For this, the GM tissue probability maps were averaged across the number of subjects to obtain mean GM scan of HC. Each HC individual was then correlated with the hence obtained mean GM scan of HC. Any correlation coefficient less than 0.8 [258], [259] was considered an outlying scan, and further excluded from the study. This procedure was also followed for Parkinson's patients separately. The smoothed 3D (113 × 137 × 113 = 1749353 voxels) grey matter scan of each subject (78 HC + 131 PD = 209 subjects) was row-wise concatenated to obtain a 2D subject-by-voxel matrix of dimension 209 × 1749353. This 2D matrix containing the subject-wise GM information was used as the input GM matrix. This input matrix was subjected to SBM that involves ICA using Group ICA for fMRI Toolbox (GIFT, https://trendscenter.org/software/gift/) on MATLAB 2021a platform (MathWorks). ICA decomposes data, yielding a loading (mixing or weight) matrix and a component (or spatial) matrix [251], [254], [260]. The number of components was set to 30 [260], [261], [262], and ICASSO (http://research.ics.aalto.fi/ica/icasso/) was used to determine the stability in 20 runs. ICA was performed using Infomax to maximize the information transfer from input to the output matrix. It extracts the linearly combined hidden components which covary in GM concentration among the subjects. The loading matrix expresses how much each component contributes to each subject's data, i.e., how much GM do individual subjects have in corresponding components. The loading matrix obtained

was thus 209 subjects × 30 components in dimension. And component matrix gives the spatial map and expresses the relationship between components and voxels within the brain.

Further, the loading matrix was utilized to identify the components with significant grey matter difference between the groups. Subsequently, components showing significant group differences were identified by student's t-test and the true positives were found by using *fdr_bh* function [263] in MATLAB at $P < 0.05$. FDR-controlling procedures are used to limit the predicted proportion of the rejected null hypotheses that are incorrect, i.e., it helps in reducing the incorrect rejections of the null hypothesis during multiple testing. Two components were selected based on effect size, significance in PD according to literature [145], [264], [265], [266], [267], [268], and an expert radiologist's advice. These component loadings were thresholded using the best-proven method adopted from [205], which outperformed other methods attempted on a simulated dataset. This thresholding method selects the PD patients with loadings > mean, if mean was positive, else selects the PD patients with loadings < mean, if mean was negative. This implies that PD patients with high weight in the components were grouped as subtypes, based on the premise that the feature is highly expressed in those subjects. Finally, patients which were common in both groups were categorized as third subtype. Thus, the loading matrix was used for subtyping the subjects, while component matrix was used to obtain spatial maps of the significant components.

4.2.4. Structural Brain Network Construction using grey matter volume

The pre-processed inlier GM tissue probability map, same as considered for subtyping, were further examined for subtype-specific structural brain network construction and connectivity analysis. The steps involved in brain network construction are (1) brain parcellation, (2) estimation of regional grey matter volume, (3) constructing the association

matrix by measuring the inter-regional correlation, (4) constructing the adjacency matrix by binarizing the association matrix using a threshold, and finally (5) network construction and visualization. Parcellation of grey matter maps of all PD subjects was done using LONI Probabilistic Brain Atlas, LPBA40 [67] (https://resource.loni.usc.edu/resources/atlases-downloads/). The maps were divided into 56 regions of interest (ROI), where each ROI represented a node in the structural network. The CAT12 software was used to obtain regional grey matter volume (rGMV) for each of these 56 ROIs. Further, for each subtype, an n × 56 dimension rGMV matrix was obtained, where n indicates the number of subjects in a subtype. Consequently, for every subtype, a 56 × 56 interregional correlation matrix was constructed using Pearson correlation [201], [225], often called as association matrix Here, any unit, A_{ij} represents correlation of a region i with another region j, serving as an edge in the network. The association and adjacency matrices of subtypes were constructed in MATLAB (MathWorks R2021a, https://in.mathworks.com/). The structural brain networks [100] were visualized using BrainNet Viewer.

4.2.5. Network Metric Analysis of subtypes using grey matter

The association matrices of each subtype were taken across an optimized range of sparsities to facilitate network analysis [77], [82]. In this study, a sparsity threshold of 0.65 implied taking into account the highest 65% values only, while rejecting the lowest 35% values. Below a certain threshold of 0.65 sparsity, the network was virtually identical to a random network (i.e. small-world index close to 1). This was the criteria for selecting the lower limit for network analysis [77], [82], [233]. The number of nodes and edges in a random network remained same as that of the network under study. Hence, the random network in the current research included 56 nodes and the number of edges varied with sparsity. The maximum possible number of edges were 1540 (i.e. 56 × 55/2) for the entire analysis. To

describe the network topological features, a few network metrics viz. clustering coefficient, local efficiency, betweenness centrality, participation coefficient were computed [47]. The aforementioned network metrics were calculated using functions from Brain Connectivity Toolbox at https://sites.google.com/site/bctnet/ [226]. Regions having degree and betweenness centrality higher than one standard deviation above than the respective means were considered as hubs. These were obtained in MATLAB and were visualized using customized settings in BrainNet Viewer.

4.2.6. Statistical analysis

All the statistical analyses were conducted using MATLAB 2021a platform. Components' effect size was calculated using *Hedge's g* test [269], [270]. It is preferable to use *Hedge's g* over *Cohen's d* test in case of sample size larger than 20. The effect size of the differences between HC and PD were computed using bias-corrected *Hedge's g*, according to the equation 5, given as below:

$$Hedge's\ g = \frac{m1 - m2}{s} * \left(1 - \frac{3}{4(n1 + n2) - 9}\right) \quad (5)$$

where $m1$ and $m2$ were the mean scores of HC and PD; $n1$ and $n2$ were the number of subjects in HC and PD; and s was the standard deviation across the two samples.

Chi Square ($\chi 2$) test at $P < 0.05$ was performed for categorical variable (gender, handedness) and *Student's t-test* was performed to compare other clinical features of HC from those of PD. One-way ANOVA (one factor, 3 levels) was performed on subtype-specific demographic and clinical variables such as age (years), education (years), Montreal Cognitive Assessment (MoCA), Geriatric Depression Scale (GDS), Random Eye

Movement Sleep Behavior Disorder (RBD), State-Trait Anxiety Inventory for Adults Test (STAIT), Scales for Outcomes in Parkinson's Disease- Autonomic (SCOPA), University of Pennsylvania Smell Identification Test (UPSIT), Letter-Number Sequencing score (LNS), Symbol-Digit Modalities score (SDM) and Unified Parkinson's Disease Rating Scale (UPDRS III, UPDRS IV). One-way ANOVA (one factor, 3 levels) was separately performed on subtype-specific loading coefficients of selective components to evaluate differences in features amongst PD subtypes. For continuous variables, the Kruskal-Wallis test was used when normality (Shapiro-Wilk < 0.05) was violated [271] and statistical significance was indicated by a P value < 0.05. A group comparison was performed on total GM volume using *Student's t-test*. Chi Square test ($\chi 2$) for gender at $P < 0.05$ was tested between the subtypes. Furthermore, association between these clinical variables and subtype- specific loadings were investigated using Pearson correlation test. Inter-subtype (pairwise) comparison of the network metrics was done using permutation test to find if the metric differed significantly between subtypes. In each permutation, the network metrics were randomly assigned to one of the two subtypes in such a way that each randomized subtype contained the same number of subjects as that of the original subtype. The differences in network metrics between the randomized subtypes were then computed for each permutation and the mean of the differences were plotted resulting to permutation distribution of difference. The actual between-group difference in network measures was then plotted to find if the actual difference lies outside 95% confidence interval. This nonparametric permutation test [82], [202] was performed 1000 times. This procedure was followed at each sparsity threshold to find significant difference in network metrics at that sparsity threshold. Subtype identification and significance of inter-subtype differences in network metrics was examined using a self-written MATLAB code.

4.3. RESULTS

4.3.1. Participants demographics for subtyping using grey matter volume

A count of 78 HC and 131 PD patients were examined in age range of 32-80 and 38-70 years, respectively (**Table 10**). A significant difference was observed in MoCA ($P = 0.01$), STAIT ($P = 8.92 \times 10^{-4}$), RBD ($P = 3.17 \times 10^{-6}$), SCOPA-AUT ($P = 3.4 \times 10^{-6}$), SDM ($P = 2.68 \times 10^{-4}$) and UPDRS-III ($P = 0$) between the PD patients and the healthy control subjects from two-sample *Student's t-test*.

The dataset comprised of 53 males and 25 females in HC, while 82 males and 49 females in PD. The chi-squared test ($\chi 2$ test) on gender showed no significant difference ($\chi 2 = 0.61$, $P = 0.43$) between HC and PD.

4.3.2. SBM analysis and Subtyping using grey matter

The input GM volume matrix was decomposed to a loading matrix of 209 subjects by 30 components and a component matrix of 30 components by 1749353 voxels. Initially, 25 components were found with significant differences (two- sample t-test, $P<0.05$) between PD and HC, of which, all 25 components were true positives after FDR. Finally, two components, i.e., inferior, middle and superior frontal gyri (IFG-MFG-SFG) and inferior, middle and superior temporal gyri (ITG-MTG-STG) were considered for subtyping.

These two components (**Table 11**) were selected due to their high effect size (**Supplementary Figure S2**), role in PD according to literatures, and over consultation with an expert radiologist. The Hedge's g effect size of IFG-MFG-SFG and ITG-MTG-STG were 1.355 and 0.492, respectively. These components' spatial maps, extracted from

GIFT toolbox, are depicted in **Figure 15**.A and the regions in these components are listed in **Table 11**.

Table 11. Components showing group differences between HC and PD

	Positive Talairachs of components using SBM and subtyping					
Components	Regions (Brodmann Area)	Loadings Directionality	Volume (cm^3) L/R	Max Z value (x,y,z) L/R	MNI (x,y,z) L/R	*Hedge's g* effect size
1	Inferior Frontal Gyrus (9)	HC > PD	3.8/ 2.4	8.6(-42,34,11)/ 6.7(42, 35, 11)	(-42,34.5,13.5)/ (42, 36, 13.5)	1.355
	Middle Frontal Gyrus (6,8,9,10,46)		13.8/14.1	9.9(-37, 28, 32)/ 8.1(39, 29, 33)	(-37.5,27,36)/ (39,28.5,37.5)	
	Superior Frontal Gyrus (8,9,10)		2.0/ 3.8	6.1(-37,35, 31)/ 6.8(37, 37, 31)	(-37.5,34.5,36)/ (37.5, 36, 36)	
2	Inferior Temporal Gyrus (19,20,21,37)	HC > PD	1.9/ 0.9	9.4(-18,10,62)/ 10.0(21,12,62)	(-18,7.5,67.5)/ (21,9,67.5)	0.492
	Middle Temporal Gyrus (19,20, 21,22,37,39)		9.9/ 9.0	8(-58,-53,-2)/ 10.6(43,-56,25)	(-58.5,-54,-6)/ (43.5,-58.5,24)	
	Superior Temporal Gyrus (22,39)		5.0/ 2.1	7.7(-45,-54,22)/ 11.6(45,-58,28)	(-45,-57,21)/ (45,-61.5,27)	

Thresholding the two components' loadings, as per N-BiC algorithm, resulted into three subtypes: 1) subtype A (21 PD patients) highly weighted in IFG-MFG-SFG exclusively, 2) subtype B (31 PD patients) highly weighted in ITG-MTG-STG exclusively, and 3) subtype AB (44 PD patients) highly weighted in both components covering inferior, middle and superior frontal and temporal gyri.

The total grey matter volume of the entire brain was significantly different between subtypes A and AB ($P = 4.46 \times 10^{-4}$, t-stat = 3.71), and between B and AB ($P = 0.01$, t-stat = 2.57) (**Table 12**). The inter-subtype regional GM volumes for either components were also compared separately and for both components together. The rGMV of ITG-MTG-STG of subtype A was significantly different than that of subtype AB ($P = 0.003$, t-stat = 3.07). A significant difference was observed in rGMV of IFG-MFG-SFG between subtypes A and AB ($P = 0.003$, t-stat = 3.09) and between subtypes B and AB ($P = 1.51 \times 10^{-4}$, t-stat = 3.99). Similarly, combined GMV of IFG-MFG-SFG and ITG-MTG-STG was different between subtypes A and AB ($P = 0.0019$, t-stat = 3.25) and between B and AB ($P = 0.0025$, t-stat = 3.14).

Table 12. Subtype-specific subjects demographic information

Subtypes	Total count	Female count	Male count	Age Range	Age (Mean ± SD)	Handedness: Count	Grey Matter Volume (Mean ± SD
A exclusively	21	6	15	45-70	59.14 ± 7.63	R:20, L:1, B:0	609.86 ± 48.69
B exclusively	31	14	17	38-69	57.74 ± 9.04	R:28, L:1, B:2	593.92 ± 49.26
AB	44	16	28	49-69	57.64 ± 5.54	R:35, L:7, B:2	567.01 ± 41.00

R: Righthanded; L: Lefthanded; B: Ambidextrous; SD: Standard Deviation; Dark purple color shows the t-test between subtype B and AB; Blue color shows the t-test between subtype A and AB; **significant at 99% confidence interval using t-test

A significant difference was found between the subtype-specific loadings using one-way ANOVA at $P < 0.01$ for IFG-MFG-SFG ($P = 1.12 \times 10^{-17}$, $F = 61.15$) and ITG-MTG-STG ($P = 1.48 \times 10^{-17}$, $F = 60.49$) (**Figure 15**.B). Data are expressed as median (interquartile range 25–75 percentile) in the ANOVA plots. Distribution for loadings of healthy subjects

and various PD subtypes in one component (IFG-MFG-SFG) are plotted against those of the other component (ITG- MTG-STG), as depicted in **Figure 16**.

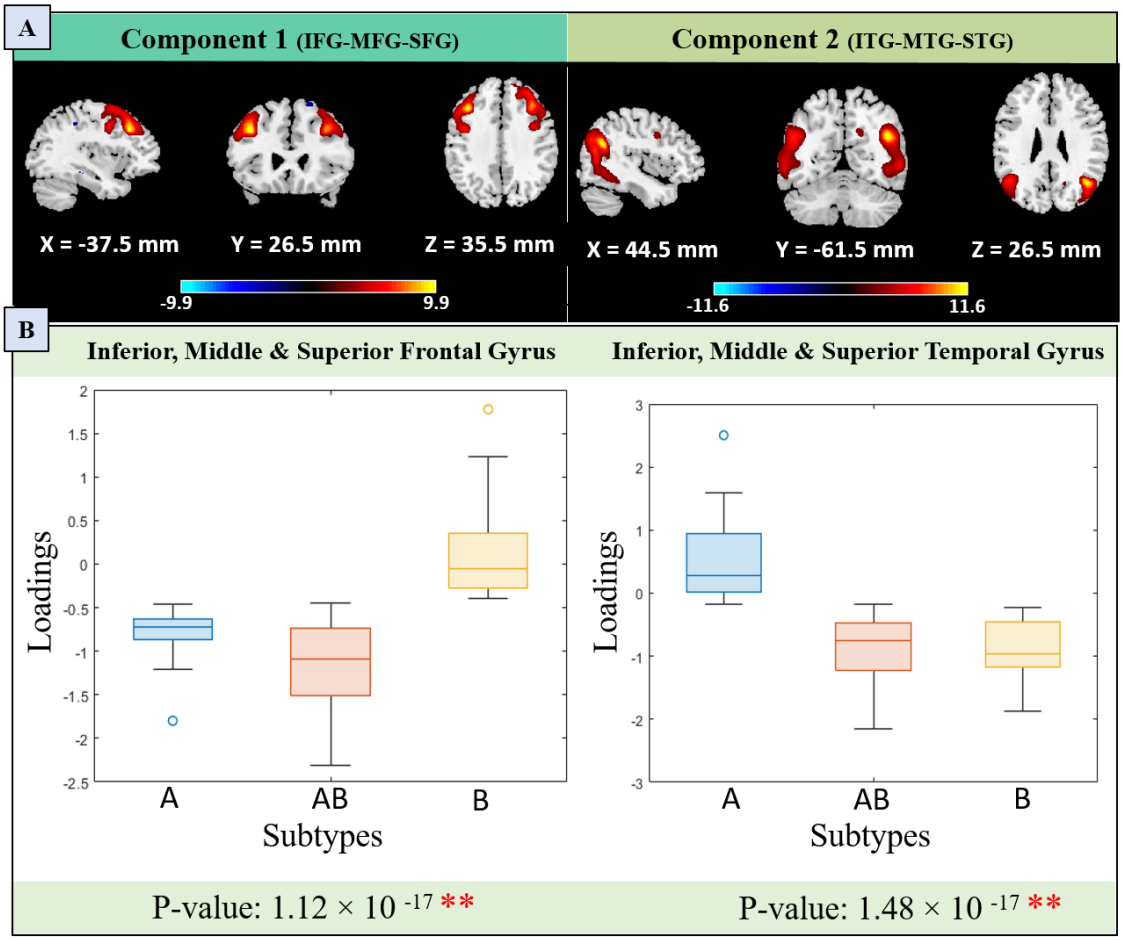


Figure 15. **(A)** Spatial maps of components showing group effect from GIFT toolbox. The voxels above threshold of |Z| > 3.0 are shown. **(B)** Box Plot of Subtype-specific loadings from ANOVA test. A: Subtype A exclusively; B: Subtype B exclusively; AB: Common to subtypes A and B; P-value: Probability value from ANOVA, **significant at 99% confidence interval.

The demographic and clinical data of PD subtypes are summarized in **Table 12** where age and total grey matter volume are expressed as mean ± standard deviation. The subtype-specific subjects (**Supplementary Figure S3**) showed significant difference in UPDRS-III (P = 0.1, H = 3.57) and SDM (P = 0.1, H = 4.3) at 90% confidence interval (CI) using Kruskal-Wallis Test. However, inter-subtype comparison for the other clinical features

such as age ($P = 0.64$, $H = 0.89$), education ($P = 0.45$, $H = 1.6$), MoCA ($P = 0.52$, $H = 1.3$), RBD ($P = 0.68$, $H = 0.77$), GDS (P = 0.71, $H = 0.67$), STAIT ($P = 0.58$, $H = 1.09$), UPSIT ($P = 0.11$, $H = 4.47$), SCOPA ($P = 0.17$, $H = 3.52$), LNS ($P = 0.99$, $H = 0$) and UPDRS IV ($P = 0.71$, $H = 0.69$) showed no significant difference using the same test. The correlation between the loading coefficients and the clinical features were also analyzed. The loading coefficients of subtype A was found to be correlated with Age ($P = 0.0038$, r = 0.6), and UPDRS-IV ($P = 0.04$, r = -0.45) in IFG-MFG-SFG, and with GDS ($P = 0.03$, r = 0.47), SDM ($P = 0.07$, r = -0.39), UPDRS-III ($P = 0.07$, r = 0.39), STAIT ($P = 0.03$, r = 0.46) in ITG-MTG-STG. Similarly, subtype B loadings had correlations with age ($P = 0.59 \times 10^{-4}$, r = -0.58), LNS ($P = 0.03$, r = 0.38) in IFG-MFG- SFG, and with MoCA ($P = 0.06$, r = 0.34), UPDRS-IV ($P = 0.04$, r = 0.36) in ITG-MTG-STG. The subtype AB loading coefficients were observed to be correlated with LNS ($P = 0.001$, r = 0.47), MoCA ($P = 0.02$, r = 0.34), RBD ($P = 0.03$, r = -0.32), SDM ($P = 1.07 \times 10^{-4}$, r = 0.55), UPDRS-III ($P = 0.02$, r = -0.33) in ITG-MTG-STG.

4.3.3. Subtype-specific structural brain network

The weighted-undirected association matrix for each of the subtypes was formed with 56 nodes (corresponding to each ROI, listed in **Supplementary Table S**) and 56*55/2 = 1540 edges. **Figure 17** shows weighted undirected association matrix (1st row), binary undirected adjacency matrix (2nd row) and undirected networks (3rd row) for the deciphered subtypes. For each subtype, the sparsity varied from 0.55 to 0.9 with a step size of 0.025. The clustering coefficient of the network under study was higher than that of random network, small-world index greater than 1, and average shortest path length was equivalent to that of random network for all subtypes.

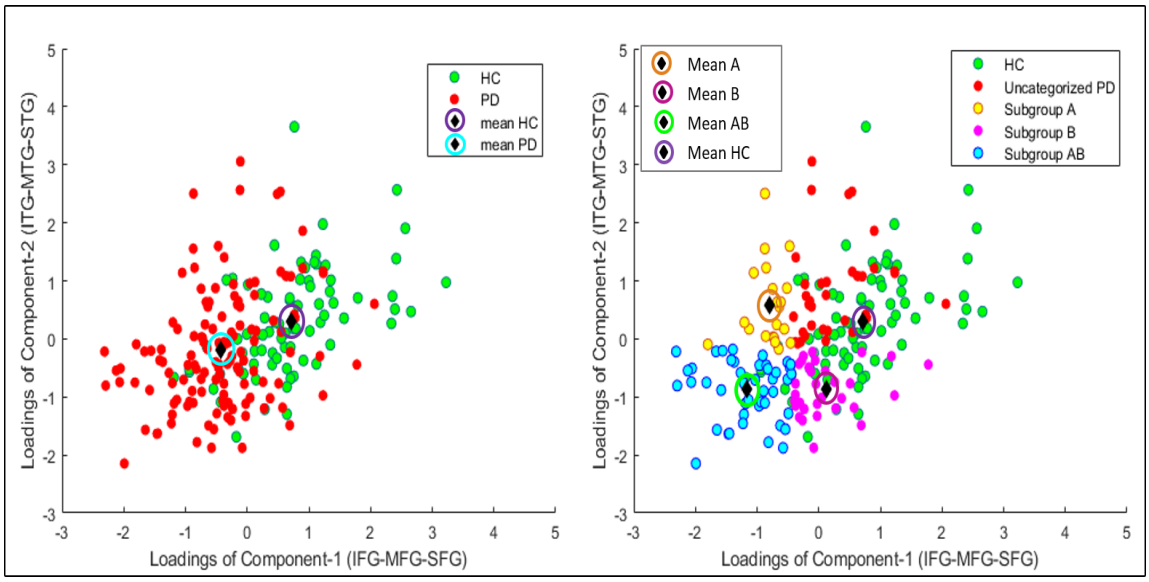


Figure 16. Plot of component loadings of healthy subjects and various PD subtypes in both components before (left panel) and after (right panel) subtyping

4.3.4. **Network Metrics Analysis of subtypes using grey matter**

The distribution and inter-subtype differences of network metrics are shown in **Figure 18**, as obtained from regional GM volume of all the three subtypes. The results suggest that mean clustering coefficient (**Figure 18**.i) and mean local efficiency (**Figure 18**.v) of all the three subtypes overall decays with an increase in sparsity. As suggested from pairwise comparison through permutation test at 95% CI, significant difference was observed in clustering coefficient between subtypes A and B (**Figure 18**.ii, sparsity: 0.55 - 0.675), between subtypes A and AB (**Figure 18**.iii, sparsity: 0.575) and between subtypes B and AB (**Figure 18**.iv, sparsity: 0.55 - 0.6, 0.725). Local efficiency significantly differed between subtypes A and B (**Figure 18**.vi) between sparsity of 0.55 to 625.

The betweenness centrality (**Figure 18**.ix) of PD subtypes exhibited overall increasing trend till sparsity of 0.7 for subtype A and 0.85 for subtype B. It was observed to be significantly varying between subtypes A and B at sparsity of 0.85 (**Figure 18**.x).

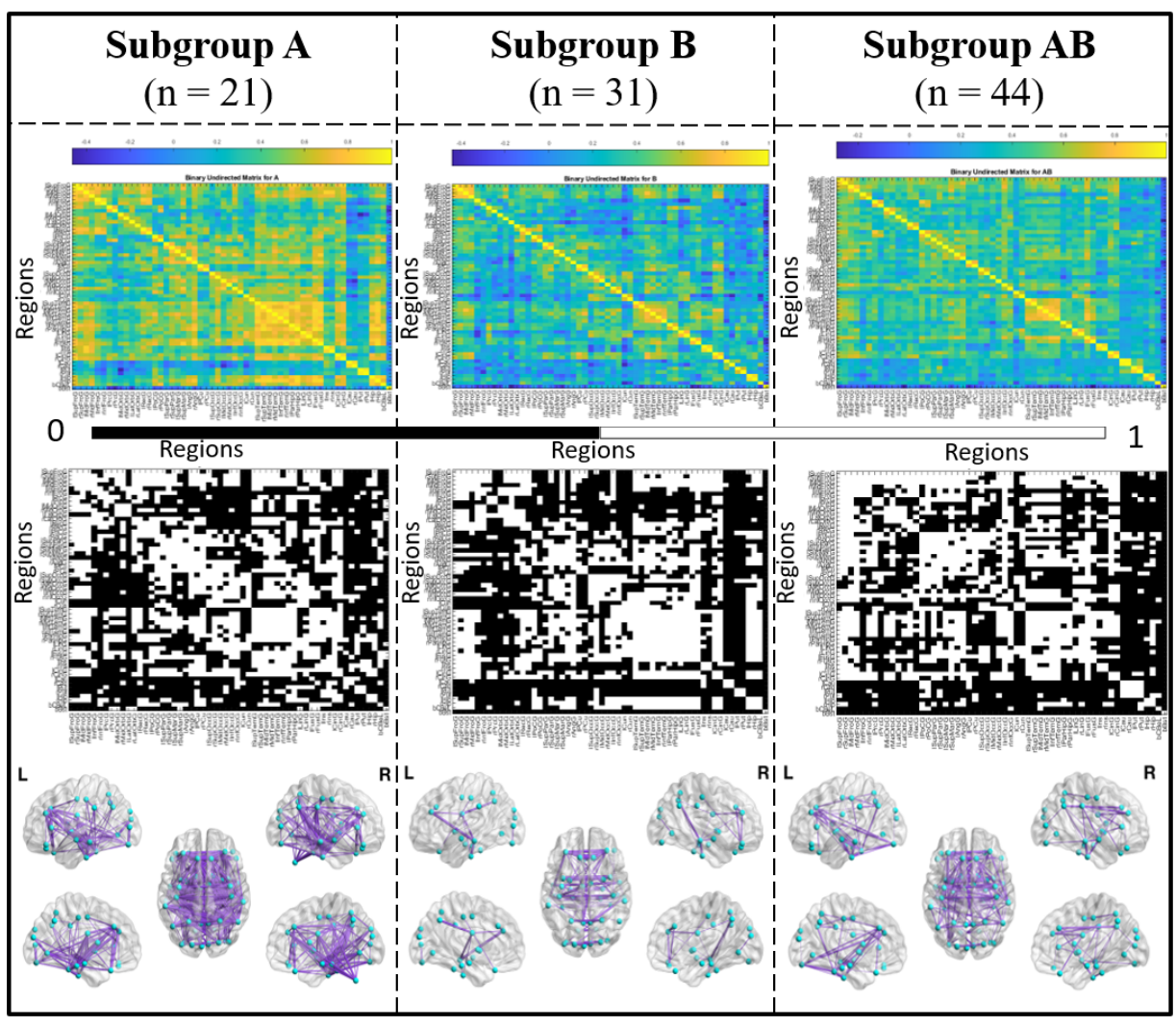


Figure 17. Association and Adjacency Matrices and Networks of Parkinson's subtypes. 1st row shows the association matrix, 2nd row is the adjacency matrix, and 3rd row is the network of the subtypes. Colorbar in the association matrices and width of the edge in the network indicates strength of the connections. In adjacency matrices, white is 1, i.e., presence of edge and black is zero, i.e. no edge.

In addition, participation coefficient didn't show any specific pattern (**Figure 18**.xiii) with sparsity, though it differed significantly between subtypes A and B (**Figure 18**.xiv, sparsity:0.55 - 0.65, 0.75 - 0.85), between subtypes A and AB (**Figure 18**.xv, sparsity: 0.55, 0.675, 0.75, 0.8, 0.875) and between subtypes B and AB (**Figure 18**.xi, at sparsity of 0.575 - 0.625, 0.675, 0.75, 0.775).

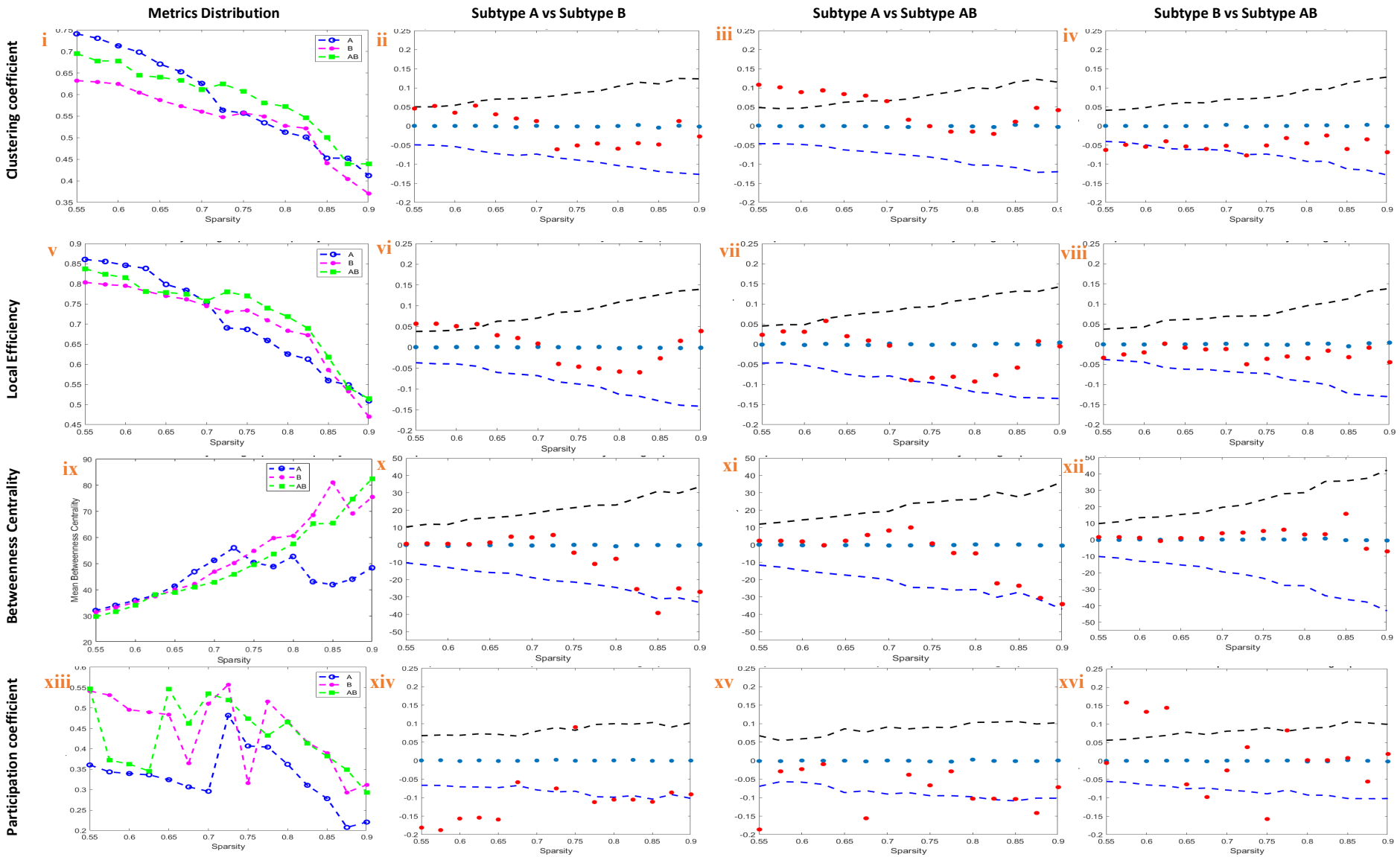


Figure 18. Distribution and intergroup differences of network metrics within/ between PD subtypes. Block i-iv: Clustering Coefficient; Block v-viii: Local Efficiency; Block ix-xii: Betweenness Centrality; Block xiii-xvi: Participation coefficient. Block i, v, ix, xiii: Distribution of network metrics; Block ii, vi, x, xiv: subtype A vs. B; Block iii, vii, xi, xv: subtype A vs. AB; Block iv, viii, xii, xvi: subtype B vs. AB. The upper and lower 95% confidence intervals are shown in dashed lines in blocks ii, iii, iv, vi, vii, viii, x, xi, xii, xiv, xv and xvi; the red asterisk signs lying outside the envelops indicate significant group difference at specific sparsity. Positive values in permutation test indicate former subtype having greater values than the latter.

4.3.5. Hubs detected in networks of various subtypes

Hubs are nodes with both degree and betweenness centrality one standard deviation higher than the respective means (**Figure 19**). The hubs exclusively identified in subtype A (royal blue color) are right angular gyrus, left precuneus, right cingulate gyrus; in subtype B (pink color) are left middle frontal gyrus, left middle temporal gyrus, bilateral parahippocampal gyrus; and in subtype AB (green color) are bilateral superior frontal gyrus, left postcentral gyurs, left inferior occipital gyrus, left fusiform gyrus. However, right fusiform gyrus was conserved between subtypes A and B (mustard yellow color) and left cingulate gyrus (cyan color) between subtypes B and AB. These hubs represent the important nodes for information exchange between the subnetworks.

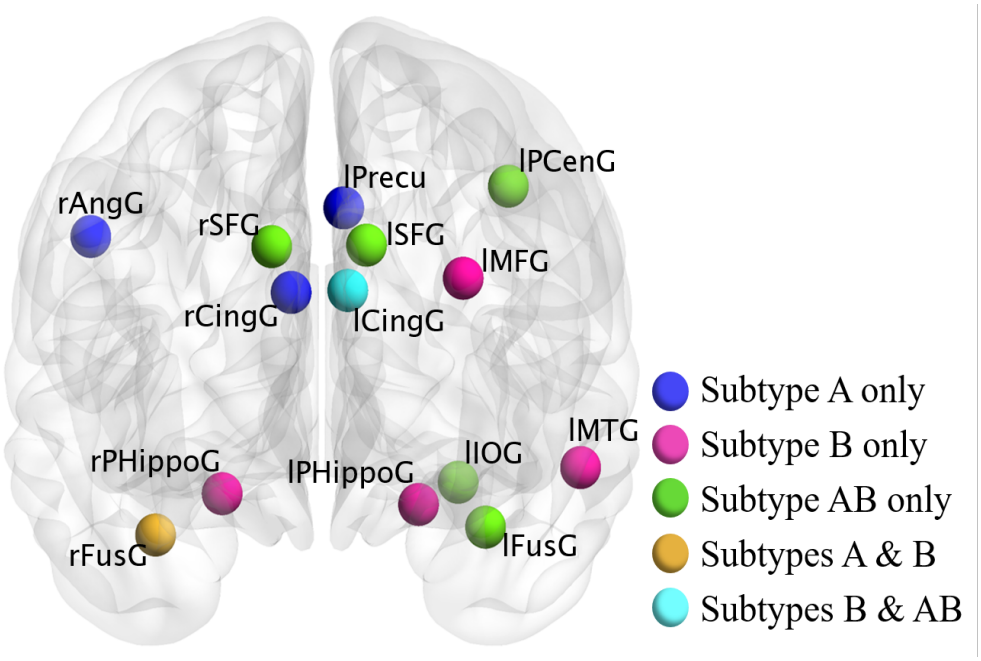


Figure 19. Distribution of hubs in PD subtypes. Hubs exclusively in each subtype and common to any two subtypes are shown in different colors. Color codes are shown on the right panel. Fading colors indicate distance from the surface.

4.4. DISCUSSIONS

4.4.1. Source-based morphometry and subtyping of PD patients using grey matter

SBM is a model-free method to find independent GM volume patterns across groups. In this study, since PD patients were majorly observed with reduction of grey matter across entire brain, two components were utilized to unravel subtypes within the disease. Since the mean of loading coefficients of all PD patients in either components was less than that of HC and the component's spatial maps were predominantly positive, it may be understood that the grey matter volume in PD patients is lesser than that of HC participants [260]. This study could successfully identify two independent patterns of GM volume changes using SBM. The first component involved GM volume change in IFG-MFG-SFG [145], [264], [265], [272], [273], while the second component involved GM differences in ITG-MTG-STG. GM volume atrophy in IFG, MFG and SFG is found in PD patients with cognitive impairment as well as dementia. Dysfunctions in SFG is also associated with involuntary movement [274] in PD patients. Additionally, previous study [266] has identified right MTG-STG as atrophy regions in PD compared to HC. Previous studies using SBM [254], [275] could differentiate PD patients from atypical parkinsonisms. Interestingly, the gender distribution in the current research for each of the subtypes (**Table 12**) confirms that men are more prone to PD than women. A potential reason could be release of female predominant neurotransmitters and hormones like estrogen and oxytocin, which affect dopaminergic neurons and protect female nervous system [276]. A previous study employed multimodal MRI measures (amplitude of low-frequency fluctuation from functional MRI and GMV from structural MRI) and hierarchical clustering analysis to identify two PD subtypes [277]. Subtype 1 (diffuse malignant subtype) had a neurodegenerative pattern characterised by widespread loss in GMV and diminished

amplitude of low-frequency fluctuation in the visual-related brain. Subtype 2 (mild subtype) showed slight reduction in GMV and elevated amplitude of low-frequency fluctuation in the frontal, temporal, and sensorimotor cortices. However, results of the current study show reduced total grey matter volume in subtype AB compared to both subtypes A as well as B.

Usually, machine learning approaches use labelled data for training a classifier and then the unlabelled test data is labelled using the built classifier. The aim of the present study was to subtype Parkinson's patients having no labels. Since machine learning approaches cannot be applied in principle, use data-mining approach named biclustering was opted, aiming to find submatrices by reorganizing the information available in a bigger matrix.

The current study revealed significant differences in total grey matter volume of entire brain between subtypes A and AB, as well as between B and AB. Moreover, the inter-subtype analysis of regional grey matter volumes indicated notable differences in both components. These findings underscore the importance of considering regional grey matter variations when assessing subtype differences, providing light on potential neuroanatomical markers associated with these subtypes. The findings contribute valuable insights into the nuanced neurobiological characteristics of the subtypes, emphasizing the need for further investigation to enhance the understanding of these distinctions and their implications.

4.4.2. Network analysis of subtypes identified using grey matter volume

Past studies have investigated the topological organization of PD patients using sMRI, but either between PD patients and control, or between the PD patients with/ without cognitive issues and control. However, in the current study, network topology has been analysed

among PD subtypes. Network metrics were quatified for three subtypes using graph-theory based connectivity analysis. Mean clustering coefficient and local efficiency of every subtype reduced as sparsity incremented. Increase in sparsity implies reduction in the number of nodes and so also the number of edges, thereby resulting to lowering of clustering coefficient and local efficiency at higher sparsities. Thus, these findings corroborate with the fact that with increase in sparsity, efficiency as well as cluster formation within network reduces due to reduction in the number of paths for information transfer. Clustering coefficient and mean local efficiency of subtype A were significantly higher than those of subtype B at lower sparsity. Also at lower sparsity, clustering coefficient of subtype AB was lower than that of A, while higher than that of B. Previous study revealed aberrant network topology with differences in efficiency and clustering coefficient in PD patients with mild cognitive impairment from HC [233]. Two key factors drive betweenness centrality as a network property. On one hand, it measures the role of each node for information transmission across the network. On the other hand, it was aimed to measure the interplay between various brain regions. Because of this, betweenness centrality was used as a metric to determine information flow within as well as between subnetworks [278]. Hence, it was employed to find the subtype-specific hub nodes. In this direction, betweenness centrality of subtype A is less than that of B at higher sparsity. Participation coefficient of a node indicates if it is a provincial hub or a connector hub, implying if the hub distributes information to distant nodes or to its immediate node. Participation coefficient of subtype A was found to be less than both B and AB. However, subtype AB had lower participation coefficient than that of B at lower sparsity, while the reverse at higher sparsity. In addition to clustering coefficient, participation coefficient was also found to be significantly different between each subtype, and hence they could be used as network biomarkers to distinguish the connectivity pattern of PD subtypes. Whilst

clustering coefficient and local efficiency were found to vary between healthy controls and PD at higher sparsity [82], it's intriguing to find the same network metrics differing between the PD subtypes at lower sparsity levels in this study. Hence, it is speculated that lower sparsities may be ideal for analysing brain network in disease subtypes.

Network analysis using cortical thickness found right fusiform gyrus as hub regions in male. These nodes was also observed as hubs in the current study, in subtypes B and AB. The hubs which corroborated with results of the current study are right fusiform gyrus in subtypes A and B, and right parahippocampal gyrus in subtype B [236]. In PD, fusiform gyrus is associated with high-level visual processing, usually with recognizing words [279]. In addition, right fusiform gyrus has also shown greater grey matter loss in PD patients with psychosis compared to those without psychosis [280]. Cingulate gyrus, another common hub in PD subtypes B and AB, is associated with cognitive functions and non-motor symptoms in PD [281], [282]. Since right fusiform gyrus is detected as a conserved hub in subtypes A and B and left cingulate gyrus in subtypes B and AB, these nodes may be suggested to play crucial role in keeping the communication active in PD subtypes, and hence, may further be investigated to understand the disease pathophysiology. A previous functional connectivity study [283] discovered extensive disrupted areas in predetermined PD subtypes (Tremor dominant and Postural instability/gait difficulty). These primarily affected the lingual gyrus, insula, IFG, cerebellum, STG, pre- and postcentral gyri, and parahippocampal gyrus. Another cross-sectional resting-state functional MRI study [284] found varying visual and sensorimotor networks between tremor dominant and postural instability/gait difficulty. But their results lost significance when dopaminergic drug use was taken into account. Also, the GMV did not differ between the subtypes.

Overall, the current study has revealed frontal and temporal gyrus playing key role in finding subtypes within Parkinson's disease. Moreover, the traditional concept - frontal region to be the primary cortical substrate for working memory and executive function-related tasks - supports the present findings. Previous study [285] on whole brain of early Parkinson's patients observed verbal memory positively correlating with functional connectivity of SFG. In addition, Hou *et al.* [286] demonstrated that functional connectivity between the inferior parietal and bilateral superior frontal gyrus are associated with improved memory function in drug-naïve PD patients. Dopaminergic circuitry loss was originally believed to be responsible for neurotransmitter changes in frontal regions of PD. However, other non-dopaminergic processes are being implicated in frontal dysfunction, which may be the cause of decline [266], [287] in cognitive function in the early stages of Parkinson's disease. Hence, it may be suggested that subtype A patients are in early stage of PD with both verbal memory and immediate recall impairment. Similarly, subtype B patients are related to atrophy in temporal gyrus and are correlated with cognitive, neuropsychological and motor scores as evident from MoCA, LNS and UPDRS scores. Despite the fact that temporal lobe is conventionally regarded as being essential to the memory process, damage to regions connected to the temporal lobe can also impair memory. The limbic system and ventral temporal lobe, as well as other association areas, provide information to the frontal cortex, which is crucial for memory encoding and retrieval [266]. Hence, analysing whole-brain network would reveal better features. Chen et al. analysed white matter brain networks of 74 de-novo PD patients with probable RBD, 97 PD patients without RBD and 73 healthy controls. They suggested altered nodal efficiency in left middle frontal and temporal gyrus, which may have involvement in disease pathology of PD patients with RBD [288]. Hence, the results propose subtype AB have sleep disorder in addition to neuropsychological and motor disorders. Results also

suggest these regions as one of the hubs in subtype B and AB, which might imply different profiles of sleep disorder in the patients. RBD as one of the features helped in revealing varying disease courses and survival rates among the PD subtype AB, which were linked to different rates of Lewy pathology and severity [289]. Letter-Number Sequencing Test, correlated with the loading coefficients of subtypes B and AB, is an assessment of verbal working memory. It has the capacity to temporarily hold and manipulate information, and is widely considered as a fundamental aspect of executive function [290]. The motor assessment test, UPDRS-III differed between the PD subtypes, and also showed its correlation with loadings from IFG-MFG-SFG of subtype A and both components of subtype AB. This suggests motor dysfunctions in subtypes A and AB, besides memory impairment. To conclude, subtypes A showed association with verbal memory and immediate recall impairment, motor impairment, anxiety, depression; subtype B with cognitive, working memory and motor dysfunction; and subtype AB presented cognitive, motor and memory impairment with sleep disorder.

4.4.3. Limitations and Future Scope

There are a few limitations of the work in this chapter that need to be acknowledged and further explored. As the number of SBM component increases, the number of subtypes obtained increases, thereby affecting the total number of patients which are subtyped. This current study did not take into account every SBM component, which may be why some of the participants could not be categorised into any one subtype. This restricts the current study which is intend to look into in future. Also combining multi-modal brain imaging data (say grey and white matter) may provide more information about Parkinson's disease as in other disorders [291], [292]. Deciphering subtypes and correlating with specific

clinical characteristics has potential implications in stratification of PD patients for suitable clinical trials to comprehend disease pathophysiology.

Age, gender, genes and environmental causes are some of the factors affecting the incidence of PD [3], [4], [5], [7], [8]. There is a future scope of study for subgrouping PD taking these factors into consideration in addition to GM information. Although this current study regresses out the effect of age, gender and sites from the structural MRI data and then performs subtyping analysis, other methods such as z-standardisation [293] and federated learning [294] may also be explored. Z-standardization [293] may be used to normalize the correlation coefficients in the association matrices of each subtypes before analysing network metrics. Also, the federated learning approach [294] may be a potential way to preserve the privacy of the raw medical image data from multiple sites.

Brain's complexity may be manifested by understanding the interplay between connectivity, dynamics and information [295]. Although, brain networks were exploited to understand the information transmission, other imaging modalities explaining the trajectory of information flow may provide additional detail. Due to multiple threshold methods used in diverse connectivity studies, the results varied and so also the interpretations. So, a uniform threshold value is to be used in all the networks under study. In addition, a single threshold value might make the analysis bias, suggesting the selection of a range of cut-offs to prefer [99]. Moreover, it's crucial to adopt a common data-driven threshold strategy such as adaptive thresholding [296], [297] for comparing networks of multiple disorders. Adaptive thresholding is advantageous over conventional thresholding method as they automatically adapt to divergent network settings and hence generalizable.

Despite the widespread use of unweighted (or binary) networks, they do not reflect the continuous nature of correlation values between 0 and 1. Any value lower than the threshold loses its information in binary network and is considered 0. However, in weighted network [298], the exact correlation value is taken into account. And hence, instead of binary correlation matrix for network analysis, use of weighted correlation matrix may be encouraged to avoid the informational loss during binarization. Conditional dependency between two nodes has been used as directional information to obtain directed network [299], which may be ventured in structural networks as well.

Hence, the current study may provide valuable information to healthcare professionals to give precise prognosis to individual patients and categorise into subtypes. In addition, standard therapeutic strategies may be laid, separately for each subtype. Growth-mixture models [300] may help understand the disease progression and severity. These, followed by examining the medication response on the individuals, may be used to study the survival of the PD patients. Parallel use of multimodal approaches may complement the findings and draw confirmative conclusion.

4.5. CONCLUSION

This study aimed at finding subtypes within Parkinson's disease and envisaged as a step towards analysing the differences in their connectivity pattern. Inferior, middle and superior frontal and temporal gyri, were investigated based on the variability across subjects to obtain subtypes which vary in GM volume. Clustering coefficient, local efficiency, betweenness centrality, participation coefficient varied between the subtypes. This study on multi-site sMRI - aided morphological network revealed the fact that network metrics may act as a biomarker of Parkinson's subtypes. In addition, use of multiple tissue

information and multimodal imaging, which may provide more information on inter-tissue dependencies as well as interactions, is the way forward.

Data availability statement

The imaging and clinical data were collected from Parkinson's Progressive Marker's Initiative, a multicentre database by *Michael J. Fox Foundation*. All participants provided written informed consent and the data was obtained in compliance with PPMI Data Use Agreement. The MRI scans and demographic information of subjects involved in this study are made taken from PPMI following an application process. The subjects' identity number used in this study from PPMI database and the code developed for this work is available at below GitHub link:

https://github.com/NeuralLabIITGuwahati/SubgroupingConnectivity

CHAPTER -5

Mutual K-Nearest Neighbor-based brain network analysis of Parkinson's subtypes using multi-features

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 5. MUTUAL K-NEAREST NEIGHBOR-BASED BRAIN NETWORK ANALYSIS OF PARKINSON'S SUBTYPES USING MULTI-FEATURES

5.1. INTRODUCTION

Literatures suggest various studies using grey matter (GM) and white matter (WM) information from T1-w MRI for diagnosis and classification of PD from healthy individuals [301], [302], overlooking the inter-tissue relationship. On the other hand, studies have shown morphometric changes in one tissue affects the other, suggesting simultaneous consideration of GM and WM information [161], [303]. Moreover, PD is associated with both grey and white matter differences [47], [161], which are also used in multimodal fusion techniques such as joint-SBM [176] and joint ICA [304].

Joint ICA (JICA) is a multivariate multimodal technique which was applied in previous studies to find joint sources showing group differences [176], [304]. It uses source-based morphometry (SBM) [305] which employs independent component analysis (ICA) technique that enables identification of spatially covarying patterns. The input information matrix (eg. subject by voxel matrix) is decomposed into independent components (brain regions) containing each subject's tissue information in each region. Joint ICA was used on both GM and WM on sMRI data to find differences in Alzheimer's patients and controls [304]. They found three joint sources where patients had reductions in both GM and WM compared to that of HC. Another study on joint-SBM [176] aimed to incorporate multiple tissue types from sMRI and decompose two different sources (grey matter and white matter images) to understand group level differences in the obtained sources. They observed two joint sources where Schizophrenic patients had lower GM and WM than controls, while

two other joint sources with higher GM and WM than controls. The concept aimed to establish that GM regions exhibited similar intersubject covariation to that observed in WM regions. In a recent study, the clinical factors influencing impulse control behaviour (ICB) in PD patients were examined. They evaluated white matter integrity from diffusion MRI, graph characteristics from fMRI, and GM volume from sMRI. Additionally, they established associations between structural changes and anxiety and depression in ICB and identified various brain areas with GM loss, increased axial diffusivity, and aberrant functional connectivity [183]. Other multimodal studies are mentioned in **section 2.2**.

Brain connectivity in PD have found correlation of specific measure (eg. GM volume or cortical thickness) between different brain regions using sMRI. The association matrix was binarized across a range of thresholds (eg. density or sparsity) for computation and analysis of network metrics, since determining a single threshold was a challenge. Also, multimodal brain network analysis [306] uses a unified brain network construction algorithm that jointly correlates the structural data with the functional data for classification of epilepsy patients.

In this direction, mutual K-nearest neighbor (MKNN) is proposed for use in binarization using multiple features. MKNN is a varied form of KNN that has been widely used in classification studies. Jarvis and Patrick [307] introduced the concept of clustering using shared similarity between two near neighbors based on Euclidean distance. Another study [308] intended to find clusters of data points based on inter-point affinities identified by density of the data points. It was based on the idea of two-way affinity of a data point with other points in its vicinity. Graph clustering was also attempted [309] using MKNN approach based on network density as well as edge weights in the graph, thereby finding both cliques and non-clique subnetworks. One of the main advantages of this technique is

that it can find densely connected regions irrespective of it being a clique or not, cliques being the dense clusters in a network. Hence, it is speculated that MKNN concept may be used as a thresholding approach to find connectedness between nodes (brain regions) of a network based on mutual strength of connection.

The present work is motivated from interrelationship between GM and WM to investigate subtypes within Parkinson's disease using sMRI data. Another primary objective of this study is to analyze brain connectivity of the derived subtypes using fused information from both tissue types. Further, the network analysis is performed on adjacency matrix binarized using mutual K-nearest neighbor based thresholding. To the best of knowledge, this is the first study employing MKNN for thresholding brain network.

5.2. MATERIALS AND METHODOLOGY

5.2.1. Participants Demographics

To undertake this research endeavor, the imaging and clinicopathological information of the 70 healthy controls (HC) and 180 PD patients were sourced from Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India. Essential demographic details of these participants, including clinical information such as age, gender, Unified Parkinson's Disease Rating Scale at on (UPDRS-On) and off (UPDRS-Off) state, Hoehn and Yahr scale (H&Y), age at onset are presented in **Table 13**. In situations where clinical variables within the dataset exhibited missing records, imputation was carried out by using winsorized mean based on the available data points pertaining to the respective variables. A 90% winsorization method was adopted in which 10% data was modified (i.e. 5% from each tail). This study uses a retrospective dataset [82], [221], obtained in compliance with NIMHANS Institutional Ethics Committee.

Table 13. Demographics of participants from NIMHANS for JICA-based subtyping

	Healthy Control	Parkinson's Disease	χ^2 or t-value	p-value
Count	70	180	-	-
Age in years (Mean ± SD)	49.24 ± 10.99	54.84 ± 9.78	-3.92	1.15×10^{-4} **
Age Range (years)	20-73	22-72	-	-
Gender (Male : Female)	52 : 18	135 : 45	0.0136	0.9070
UPDRS Off (Mean ± SD)	-	33.26 ± 8.92	-	-
UPDRS On (Mean ± SD)	-	17.85 ± 6.27	-	-
H&Y (Mean ± SD)	-	2.07 ± 1.46	-	-
Age at Onset (Mean ± SD)	-	48.27 ± 9.14	-	-

**Significant at 99% confidence interval

5.2.2. Pre-processing of sMRI

The T1-weighted structural MRI scans were generated using a 3 Tesla Philips scanner equipped with a 32-channel head coil. These brain MRI scans were obtained using a magnetization prepared rapid acquisition gradient echo sequence (repetition time (TR) of 8.2 ms, echo time (TE) of 3.7 ms, flip angle of 8 degrees, 165 sagittal slices, and a voxel size of $1\times1\times1mm^3$. The preprocessing has been explained elsewhere [82]. Thus, the initial images, acquired in DICOM (Digital Imaging and Communications in Medicine) format, were transformed into NIfTI (Neuroimaging Informatics Technology Initiative) format

using dcm2nii tool (https://www.nitrc.org/projects/dcm2nii/). These were subsequently subjected to preprocessing using Computational Anatomy Toolbox, CAT12 (http://www.neuro.uni-jena.de/cat/), in MATLAB platform (MathWorks, MATLAB 2020a), within the Statistical Parametric Mapping, SPM12 (https://www.fil.ion.ucl.ac.uk/spm/).

Each individual's MRI scans were segmented into distinct tissue types, namely grey matter (GM), white matter (WM), and cerebrospinal fluid. These scans were then normalized to conform to the standard Montreal Neurological Institute (MNI-152) template and modulated through the utilization of the Jacobian determinant. Furthermore, voxel-wise regression was conducted in MATLAB 2021a, using a linear regression analysis to perform, aiming to eliminate the influence of extraneous variables such as age, gender, age and gender interaction. Subsequently, smoothing was done using an 8 mm isotropic Gaussian kernel (Full-width at half maximum) within the SPM environment, separately for GM and WM images.

5.2.3. Subtyping of PD patients only - using Joint ICA

Patients diagnosed with Parkinson's disease underwent a process of subtyping through the utilization of structural MRI (sMRI) scans. The sequential methodology employed to achieve these subtypes encompassed the following steps:

- The initial stage involved creation of input matrix based on GM and WM volumes.
- This input GM and WM matrix was further decomposed into two separate matrices: a joint loading matrix and a component matrix.
- Components exhibiting significant differences in grey and white matter between Parkinson's patients and healthy controls were then identified.

- From the previously identified components, only those deemed statistically significant were retained.
- The subsequent phase encompassed the identification of Parkinson's disease subtypes by applying a threshold to loading values of the significant components.

The smoothed GM and WM scans were subjected to correlation analysis with the mean scan of each group. This step aimed to identify and eliminate any potential outliers present within the data. The smoothed 3D GM and WM data matrices were separated reshaped to 1D arrays for every individual. Further, these 1D arrays of GM and WM were concatenated horizontally to form a single row vector. This row vector of each individual was concatenated one below the other to form a 2D input matrix. This concatenation process was conducted for all subjects, encompassing 70 healthy controls and 180 Parkinson's disease patients. Joint ICA was implemented using Fusion ICA toolbox (https://trendscenter.org/software/fit/) on the input matrix (250 subject by 1185790 voxel) within MATLAB 2021a platform (MathWorks). Employing an infomax algorithm within the toolbox, spatial independent component analysis (spatial ICA) was applied to decompose the input data matrix [176] and number of independent components was fixed to 30 [262]. Thus, it yielded both a joint loading (*i.e.* mixing or weight) matrix and a component (*i.e.* spatial or source) matrix [175]. The loading matrix contains columns that signify the extent to which each subject expresses the joint source differences between the two groups. Within the source matrix, there exist separate matrices for GM sources and WM sources. Every row in the source matrix corresponds to a shared source incorporating both GM and WM maps, displaying identical covariance patterns. These shared sources represent spatial components that are maximally independent.

Furthermore, the loading matrix was employed to identify components displaying significant group differences in grey matter and white matter. Components exhibiting statistically significant intergroup differences were identified using *Student's t-test*, and the verification of true positives was carried out using the *fdr_bh* function [263] within MATLAB, adhering to a significance level of $P < 0.05$. This utilization of false discovery rate (FDR) controlling procedures assists in curtailing the anticipated proportion of incorrectly rejected null hypotheses, thus mitigating the occurrence of erroneous null hypothesis rejections during multiple testing.

From the identified components, two components with highest effect size were selected [204]. The component loadings underwent thresholding using a method adopted from [205], which demonstrated superior performance compared to other techniques. This thresholding approach involves selecting PD patients with loadings greater than the mean if mean is positive, else selecting PD patients with loadings less than the mean if mean is negative. Thus, PD patients exhibiting higher weight in each of the two components were categorized as two different subtypes, according to the implication that the feature is strongly displayed in those patients [204].

Consequently, PD patients common to both groups were classified as a third subtype. This delineation was aimed at capturing a subtype characterized by shared attributes across both patient and control groups. Thus, the loading matrix played a pivotal role in categorizing subjects into subtypes. On the other hand, the component matrix facilitated the creation of spatial maps corresponding to the components of significance.

5.2.4. Structural Brain Network Construction- using MKNN

The pre-processed GM and WM scans, used for subtyping, were also considered for construction of subtype-specific structural brain networks and subsequent connectivity assessment. The network construction involved following steps:

Step 1. **Brain Parcellation:** The first step encompassed the parcellation of GM and WM scans of each PD patient into distinct regions using LPBA40 (LONI Probabilistic Brain Atlas, https://resource.loni.usc.edu/resources/atlases-downloads/) [67]. This resulted in the division of the maps into 56 GM regions of interest (ROIs) and 56 WM ROIs for each PD patient, where each ROI represented a node within the structural network.

Step 2. **Regional Grey Matter Volume Estimation:** Following brain parcellation, the estimation of regional grey matter volume (rGMV) and regional white matter volume (rWMV) for each of these regions was carried out. The CAT12 software facilitated in extraction of rGMV and rWMV values for each of these 56 ROIs. Subsequently, for each subtype, an n by 56 dimension rGMV matrix and the same dimension rWMV matrix were generated, where n represents number of subjects within a given subtype.

Step 3. **Association Matrix Construction:** An association matrix was constructed by measuring the inter-regional tissue correlations between different brain regions. For each subtype, a 56 by 56 interregional correlation matrix was established using Pearson correlation between rGMV and rWMV. In this matrix, each unit (r_{ij}) signifies the correlation [201], [225] between GMV of region i and WMV of region j, effectively representing an edge within the network.

Step 4. **Adjacency Matrix Construction:** Subsequently, an adjacency matrix was generated by binarizing the association matrix using mutual K-nearest neighbor (MKNN)

approach. This approach is proposed to be used for thresholding the association matrix based on highest correlation. Here K-nearest neighbor indicates high correlation coefficient. Hence in MKNN-based thresholding, an edge exists between two nodes (*i* and *j*), if *i* is one among the K-nearest neighbor of *j* and vice versa. To obtain this, primarily, the correlation coefficient of rGMV of every node to rWMV of every other node is sorted in descending order. Further, only the top K number of correlation coefficients are considered between the nodes. Although there are multiple ways to find the K parameter, $K = \sqrt{N}$ was used, where N is the total number of nodes [310]. The K parameter was opted in a way that it is not greater than $\sqrt{N}$, and hence considered K = 7. Hence, an edge exists between nodes *i* and *j*, if *i* is amongst the top K correlation coefficients with *j*, and *j* is also amongst the top K correlation coefficients with *i*. This is explained using an example in **Figure 20**. The construction of association and adjacency matrices for the subtypes was accomplished using self-written MATLAB (MathWorks R2021a, https://in.mathworks.com/).

Step 5. **Network Construction and Visualization:** The final stage involved construction of brain network and its subsequent visualization. To visually present the structural brain networks [100], the self-developed MATLAB (R2021a) code was used.

5.2.5. Network Metrics and Analysis

The adjacency matrix, containing mutual connections, was evaluated to enable comprehensive network analysis. To capture the network's topological characteristics, specific network metrics including average degree, density, betweenness centrality, participation coefficient, clustering coefficient, global efficiency, path length, eigen vector centrality, assortativity, transitivity, gamma and lambda were computed. These network metrics were computed using functions from the Brain Connectivity Toolbox [226]

(https://sites.google.com/site/bctnet/), and explained in Samantaray et al., 2022 [47]. Regions demonstrating degrees and betweenness centrality values exceeding one standard deviation above their respective means were identified as hubs. These findings were generated using MATLAB and subsequently visualized with customized settings in the BrainNet Viewer tool, available at https://www.nitrc.org/projects/bnv.

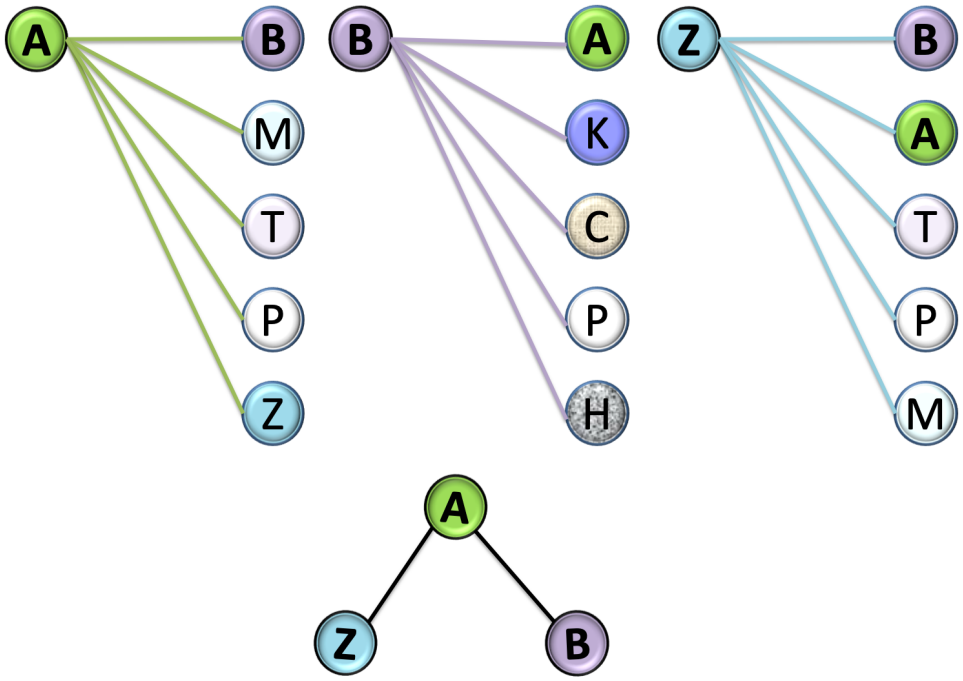


Figure 20. Explanation of MKNN-based connectivity. In the association matrix, there exists no zero value, indicating presence of a correlation value between rGMV and rWMV of any pair of ROIs. This implies the network, in such case, is fully connected with edges between every pairs of nodes. To binarize the matrix, the first step is to sort the correlation coefficients in descending order. Considering the nodes indicated in color-coded circles and assuming K = 5 for this figure, node B is among the five nearest neighbors of node A (Top left), and node A is also among the five nearest neighbors of node B (Top middle), indicating presence of connection between nodes A and B (Bottom figure). Similarly, observe node A is among the five nearest neighbors of node Z (Top right), and vice-versa (Top left), indicating A and Z are mutually connected (Bottom figure). On the contrary, although node B is also among the five nearest neighbors of node Z (Top right), node Z is not among the five nearest neighbors of node B (Top middle), indicating absence of connection between nodes B and Z (Bottom figure). Bottom figure shows network with only mutual (bidirectional) connections.

5.2.6. Statistical Analysis

A *Student's t-test* was conducted to compare differences in age between HC and PD, and Chi Square ($\chi 2$) test for gender, with a significance level of $P < 0.05$. Additionally, *t-test* was also used to assess the differences in joint loading coefficients between HC and PD for every component. The effect size of the components was determined using *Hedge's g test* [269], [270]. Notably, *Hedge's g* is recommended when dealing with sample sizes greater than 20, as it offers advantages over the *Cohen's d test*. A bias-corrected version of *Hedge's g* was used to determine the effect size of the differences between PD and HC groups, computed using equation 5.

These statistical analyses were performed using MATLAB 2021a platform. For comparing the derived subtypes, clinical measures were analyzed. One-way ANOVA (one factor, three levels) was carried out on subtype-specific demographic and clinical variables, including age (in years), Hoehn and Yahr scale (H&Y), Unified Parkinson's Disease Rating Scale at on (UPDRS On) and Off state (UPDRS Off). Additionally, one-way ANOVA (1 factor, 3 levels) was applied separately to subtype-specific loading coefficients of significant components.

Exploring the associations between these clinical variables and subtype-specific loading coefficients involved using *Pearson* correlation tests. For these correlations to be deemed significant, they needed to exhibit a *P* value below 0.05. The inter-subtype differences in connectivity metrics obtained from the brain network were computed using *Student's t-test*. The identification of subtypes and the assessment of inter-subtype differences in network metrics were conducted through the utilization of a self-developed MATLAB code.

5.3. RESULTS

5.3.1. Participants Demographics

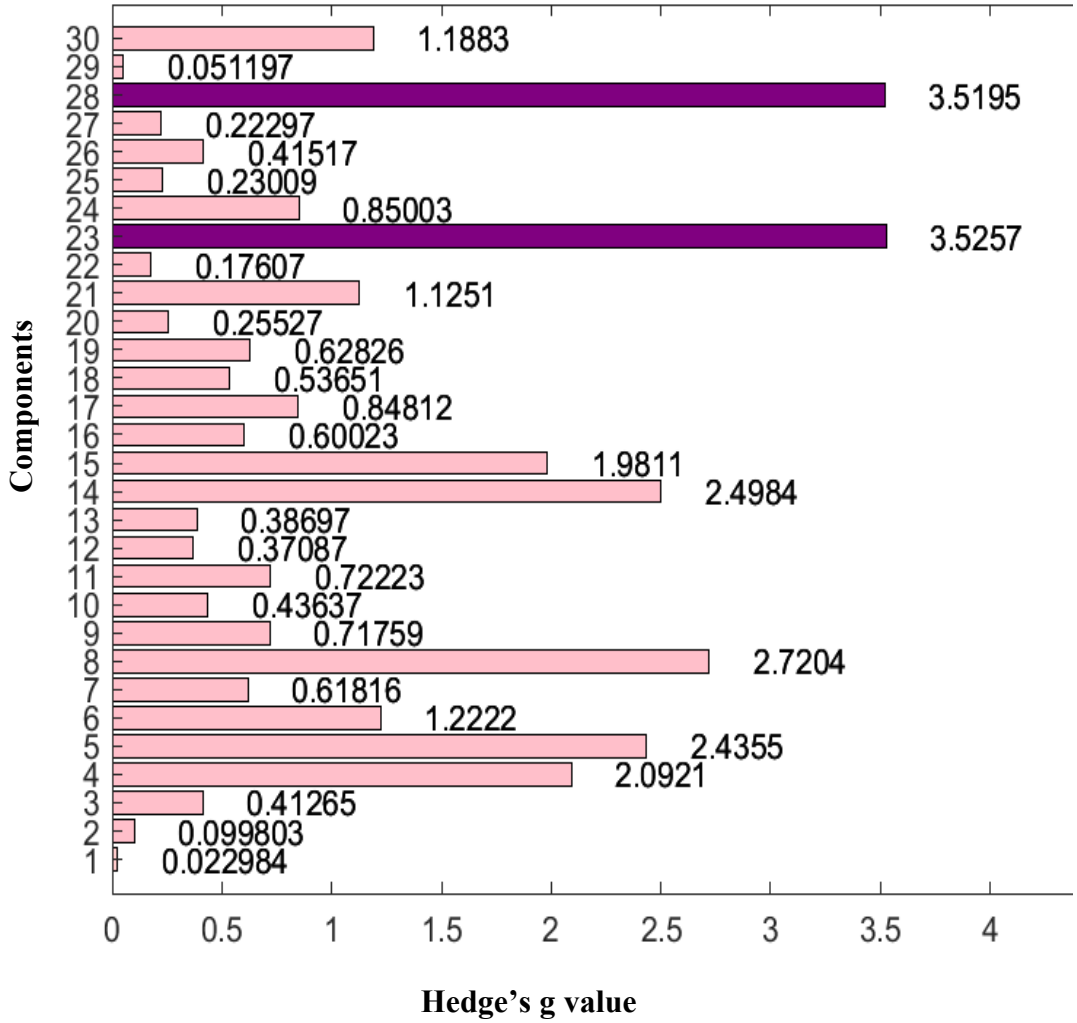
Seventy healthy control subjects and 180 Parkinson's patients, ranging in age from 20-73 and 22-72 years, respectively, were examined (**Table 13**). From a two-sample Student's t-test, it was determined that the age difference between the PD patients and the healthy control participants was statistically significant (t= -3.92, $P = 1.15 \times 10^{-4}$). In the HC dataset, there were 52 men and 18 women, while the PD dataset had 135 men and 45 women. The gender chi-squared test revealed no discernible difference between HC and PD ($\chi 2$ =0.014, $P = 0.91$).

5.3.2. Joint ICA-based subtyping

The joint loading matrix, obtained upon decomposition, was 250 subjects by 30 components. The loading parameters were analyzed using two-sample t-test for finding differences between PD and HC. Significant difference in loading parameters between HC and PD was identified in twenty-three components. Of these, *Hedge's g test* suggested component 23 (*Hedge's g* = 3.5257) had highest effect size, followed by component 28 (*Hedge's g* = 3.5195), **Table 14** and **Figure 21**.

The positive GM map in the first selected component (component 23) covered lentiform nucleus (LN) and thalamus (Th), while the negative GM map in the same component included inferior frontal gyrus (IFG) and rectal gyrus (RG). On the other hand, the positive WM map in component 23 comprised culmen (Cul), while negative WM map of the same component involved extranuclear (EN) and lentiform nucleus (LN). Similarly, the positive GM map in the second selected component (component 28) covered superior and middle temporal gyri (MTH-STG) and inferior parietal lobule (IPL), while the negative GM map

in the same component included subgyral (SubG) and supramarginal (SupMG)regions. Both positive and negative WM maps in component 28 had the complementary regions as reported in GM maps of the same component, i.e. positive WM map covered subgyral and supramarginal regions, while negative WM map covered superior and middle temporal gyri and inferior parietal lobule. These two component maps are shown in **Figure 22** and the component regions are listed in **Table 14**.



Violet colored components have highest effect size and hence taken for further study

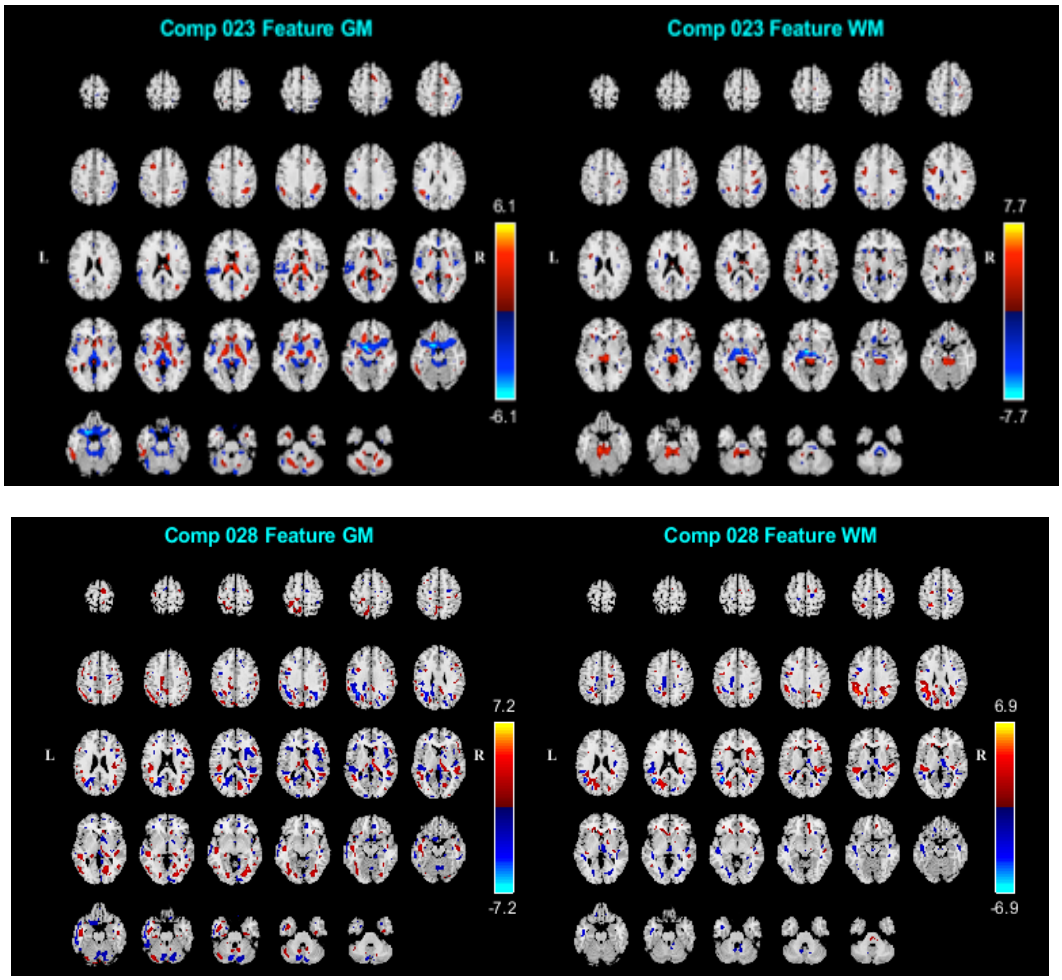


Figure 22. Spatial maps of components showing group effect from Fusion ICA toolbox. Voxels above threshold (|Z| > 3.5 are shown. Left panel shows GM regions and right panel shows WM regions. Top and bottom panel shows maps of the selected top two components based on *Hedge's g* value and previous studies

The joint loadings of significant components were thresholded as per N-way biclustering algorithm [205] to derive subtypes within PD patients. Three subtypes were obtained:

1) Subtype A (51 PD patients, with joint loading coefficients highly expressed in component 23, i.e. LN-Th-IFG-RG-EN-Cul),
2) Subtype B (57 PD patients, with joint loading coefficients highly expressed in component 28, i.e. MTG-STG-IPL-SubG-SupMG), and
3) Subtype AB (36 PD patients, with joint loading coefficients highly expressed in both component 23 and 28).

Table 14. Talairach labels of components showing group differences

GM Talairachs of component 23 using JICA					
Talairach	Region (Brodmann Area)	Volume (cm^3) L/R	Max value (x,y,z) L/R	MNI (x,y,z) L/R	*Hedge's g* effect size
Positive	Lentiform Nucleus	0.1/ 0.3	3.7 (-20,-8, -6)/ 4.4 (20, -10, -6)	(-20,-8,-8)/ (20, -10, -8)	3.5257
	Thalamus	0.2/ 0.4	3.9(-12,-19,14)/ 4.3(12,-23,14)	(-12, -20, 14)/ (12, -24, 14)	
Negative	Inferior Frontal Gyrus (47)	0.3/ 0.5	5.2(-22, 12,-22)/ 4.3 (24, 15, -19)	(-22, 14, -26)/ (24, 16, -22)	
	L Rectal Gyrus (11)	0.4/ 0.0	5.1 (-10, 12, -22)	(-10, 14, -26)	
WM Talairachs of component 23 using JICA					
Positive	Culmen	0.3/ 0.3	5.4(-12,-32,-20)/ 4.6 (14, -34, -20)	(-12, -32, -26)/ (14, -34, -26)	
Negative	Extra-Nuclear	1.0/ 0.5	6.1 (-12, -4, -8)/ 5.6 (12, -4, -8)	(-12, -4, -10)/ (12, -4, -10)	
	R Lentiform Nucleus	0.0/ 0.3	5.2 (18, -8, -6)	(18, -8, -8)	
GM Talairachs of component 28 using JICA					
Positive	Superior Temporal Gyrus (22,39)	1.2/0.4	6.7 (-40,-54,17)/ 4.1 (42, -55, 21)	(-40,-56,16)/ (42, -58, 20)	3.5195
	Middle Temporal Gyrus (19,22,39)	0.5/0.3	5.6(-38,-57,18)/ 5.2 (42, -59, 21)	(-38, -60, 16) /(42, -62, 20)	
	L Inferior Parietal Lobule (7, 40)	0.4/0.0	4.4(-42,-56,47)	(-42,-60,48)	
Negative	Sub-Gyral	0.6/ 0.4	5.7(-34,-53,32)/ 5.1 (36, -55, 32)	(-34,-56,32)/ (36, -58, 32)	
	L Supramarginal	0.5/ 0.0	4.4(-42,-43,30)	(-42,-46,30)	
WM Talairachs of component 28 using JICA					
Positive	Sub-Gyral (22,39)	0.6/0.4	5.8(-34,-53,32)/ 5.3 (36, -55, 32)	(-34,-56,32)/ (36, -58, 32)	
	Supramarginal	0.6/0.1	4.9(-44,-43,30)/ 4.0 (40, -53, 34)	(-44,-46,30)/ (40, -56, 34)	
Negative	Superior Temporal Gyrus (22,39)	1.0/0.1	6.1(-40,-54,17)/ 4.0 (42, -55, 21)	(-40,-56,16)/ (42, -58, 20)	
	Middle Temporal Gyrus (19,39)	0.6/0.1	5.6(-38,-57,18)/ 4.7 (42, -59, 20)	(-38,-60,16)/ (42, -62, 18)	
	L Inferior Parietal Lobule	0.2/0.0	4.7 (-48,-34,22)	(-48, -36, 22)	

The joint loadings between three subtypes were significantly different in both component 23 ($P = 3.22 \times 10^{-33}$, $F = 133.24$) and component 28 ($P = 1.32 \times 10^{-30}$, $F = 116.57$) at $P < 0.01$ (**Figure 23**). **Table 15** summarizes demographic and clinical data of every PD subtype. A significant difference was observed in UPDRS Off ($P = 0.0076$, $F = 5.05$), UPDRS On ($P = 0.0182$, $F = 4.12$), H&Y ($P = 0.0384$, $F = 3.34$), age ($P = 3.3 \times 10^{-5}$, $F = 11.11$) and age at onset ($P = 0.0004$, $F = 8.41$) between the subtypes at 95% confidence interval (**Figure 23**). No difference was observed in total GM volume or total WM volume in any pairs of subtypes. The inter-subtype difference in total GM volume was found to be non-significant using *Student's t-test*, i.e. between subtypes A and B ($t = 1.12$, $P = 0.27$), A and AB ($t = -0.06$, $P = 0.95$), B and AB ($t = -1.06$, $P = 0.29$). Similarly, total WM volume difference was also insignificant between subtypes A and B ($t = 0.59$, $P = 0.55$), A and AB ($t = -0.98$, $P = 0.33$), B and AB ($t = -1.37$, $P = 0.17$). The loading coefficients of component 23 was observed to be correlated with UPDRS Off ($r = 0.3$, $P = 0.03$), age at onset ($r = 0.27$, $P = 0.06$), age ($r = 0.36$, $P = 0.009$) of subtype A-specific subjects. Also, UPDRS Off ($r = 0.26$, $P = 0.05$), H&Y ($r = 0.24$, $P = 0.07$) and age ($r = 0.29$, $P = 0.03$) of subtype B-specific subjects were found to have association with loading coefficients of component LN-Th-IFG-RG-EN-Cul, and age ($r = 0.22$, $P = 0.10$) with loading coefficients of component MTG-STG-IPL-SubG-SupMG. However, none of the clinical variables of subtype AB-specific subjects had significant association with loading coefficients of either component LN-Th-IFG-RG-EN-Cul or MTG-STG-IPL-SubG-SupMG.

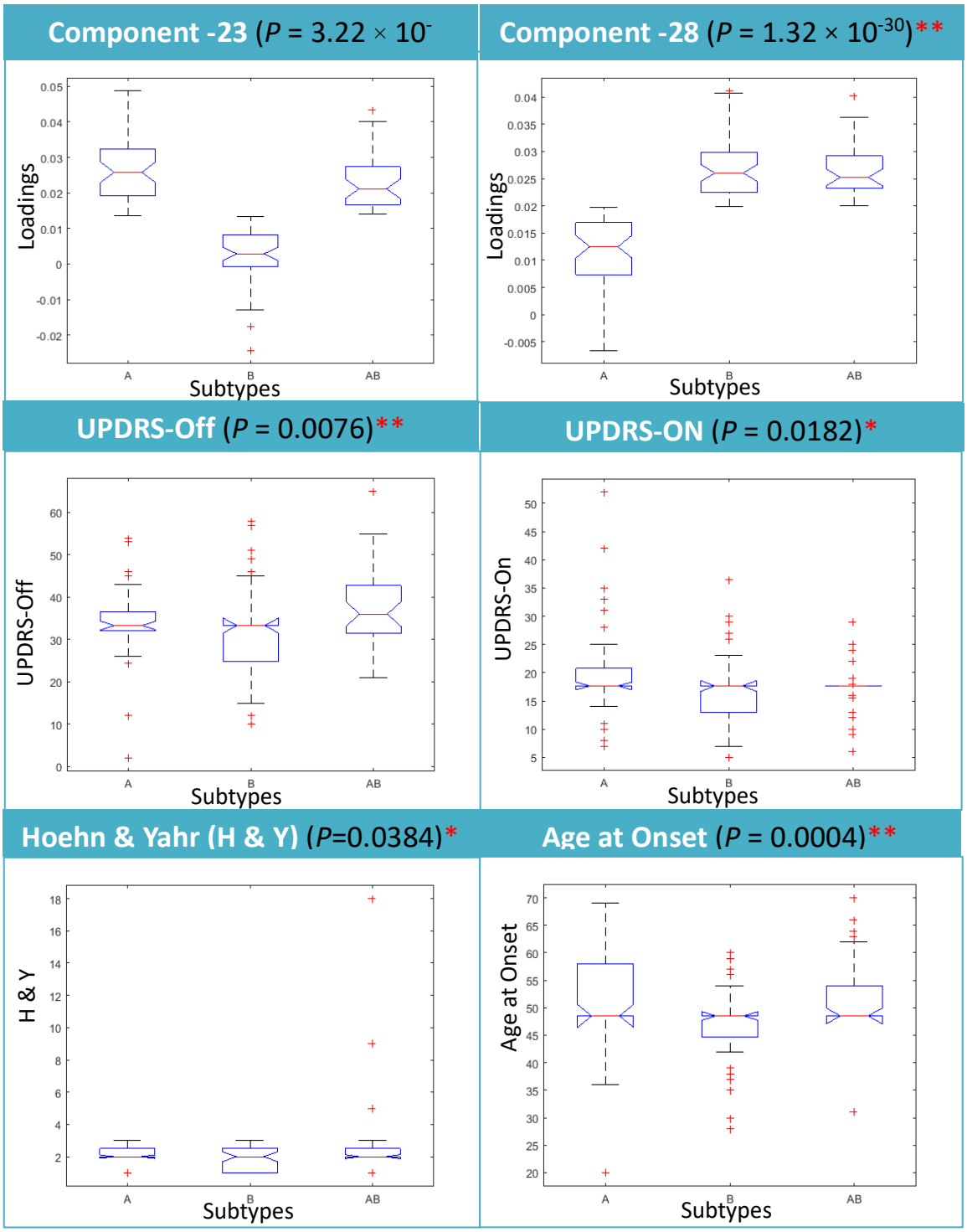


Figure 23. ANOVA of subtype-specific loading coefficients and clinical variables * Significant at 95% confidence interval; ** Significant at 99% confidence interval

Table 15. Subject demographics for JICA based subtypes

Subtypes	A exclusively	B exclusively	AB	*P*-value
Count	51	57	36	-
Gender (F:M)	18:33	9:48	5:31	-
Age	59.0 ± 8.38	52.3 ± 8.66	59.3 ± 8.35	3.33×10^{-5}**
Age Range	32-72	32-70	37-72	-
UPDRS Off	33.94 ± 8.05	31.01 ± 10.2	37.29 ± 9.53	0.0076**
UPDRS On	19.92 ± 7.59	16.65 ± 5.9	17.11 ± 4.27	0.0182*
H & Y	2.03 ± 0.6	1.9 ± 0.67	2.7 ± 2.9	0.0384*
Age at Onset	52.0 ± 8.35	46.63 ± 6.89	51.65 ± 7.09	0.0004*
Total GMV (cm^3)	511.04 ± 45.82	500.22 ± 53.76	511.59 ± 44.22	0.4145
Total WMV (cm^3)	315.82 ± 33.41	311.32 ± 43.37	323.00 ± 33.76	0.3502

Age, UPDRS, H&Y, Total GMV (grey matter volume) and Total WMV (white matter volume) are expressed in mean ± SD, * significance at 95% Confidence interval, **

5.3.3. Brain network of PD subtypes

The weighted- undirected associated matrix for every subtype was comprised of 56 rows and 56 columns corresponding to 56 brain regions, i.e. nodes. Each cell contained the correlation between regional GM and WM volumes. The adjacency matrix was constructed based on mutual connectivity with upto 7 nearest neighboring nodes. Binarization using mutual K-nearest neighbor-based threshold gives 34 edges, 40 edges and 33 edges in networks for subtype A, B and AB, respectively. The bipartite graph, hence obtained, contained a few disconnected nodes in each of the subtypes. Bipartite graph is a network representation of connections between two different networks comprising of different node measures. Further, the completely disconnected nodes are removed from the respective

adjacency matrices and bipartite graphs for better visualization (**Figure 24**). So, 22 nodes were interconnected in subtype A, 25 nodes in subtype B and 24 nodes in subtype AB.

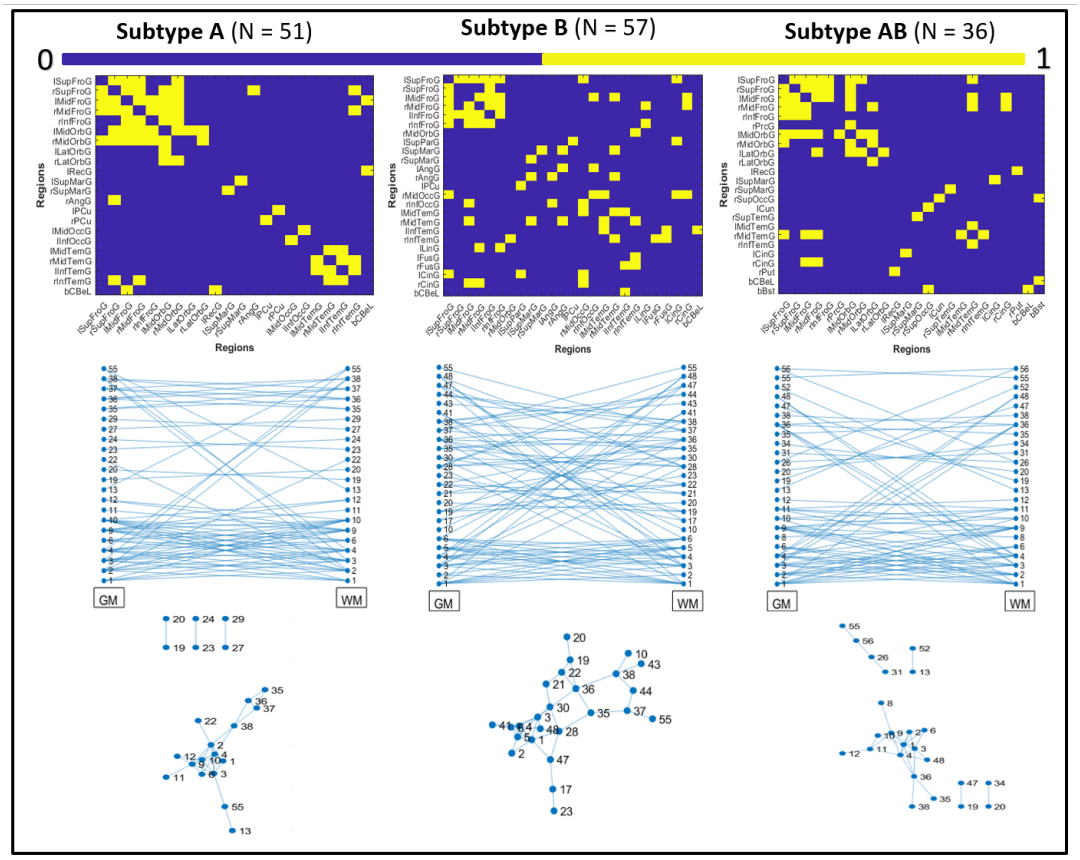


Figure 24. Subtype- specific mutual k-nearest neighbor-based adjacency matrices, bipartite graph and condense network. Adjacency matrices are shown in top row. Middle row shows bipartite graphs indicating mutual connections between GM regions (left nodes indicated by GM) and white matter regions (right nodes indicated by WM). Bottom row shows condensed network, where the same grey matter region (eg. node no. 13 of bipartite) and white matter region (eg. Node no. 13 of bipartite) are indicated by a single region (node no. 13) in condensed network. The node numbers indicate regions, and mentioned in the Supplementary Table S3.

The densities of binary network were found to be 0.0221 in subtype A, 0.0260 in subtype B and 0.0214 in subtype AB (**Figure 25**). The average network degree was observed to be 1.2143 for subtype A, 1.4286 for subtype B and 1.1786 for subtype AB (**Figure 25**). The global efficiency of subtype A was 0.0439, subtype B was 0.0821 and subtype AB was 0.0398 (**Figure 25**). The average assortativity coefficient for subtype A (assortativity = 0.3513) was slightly higher than subtype AB (assortativity = 0.3506) for subtype AB and that for subtype B (assortativity = 0.2352) was observed to be the least (**Figure 25**). A similar trend was notices for transitivity, where the coefficient was found to be 0.5909 for subtype A, 0.5042 for subtype AB and 0.3488 for subtype B (**Figure 25**).

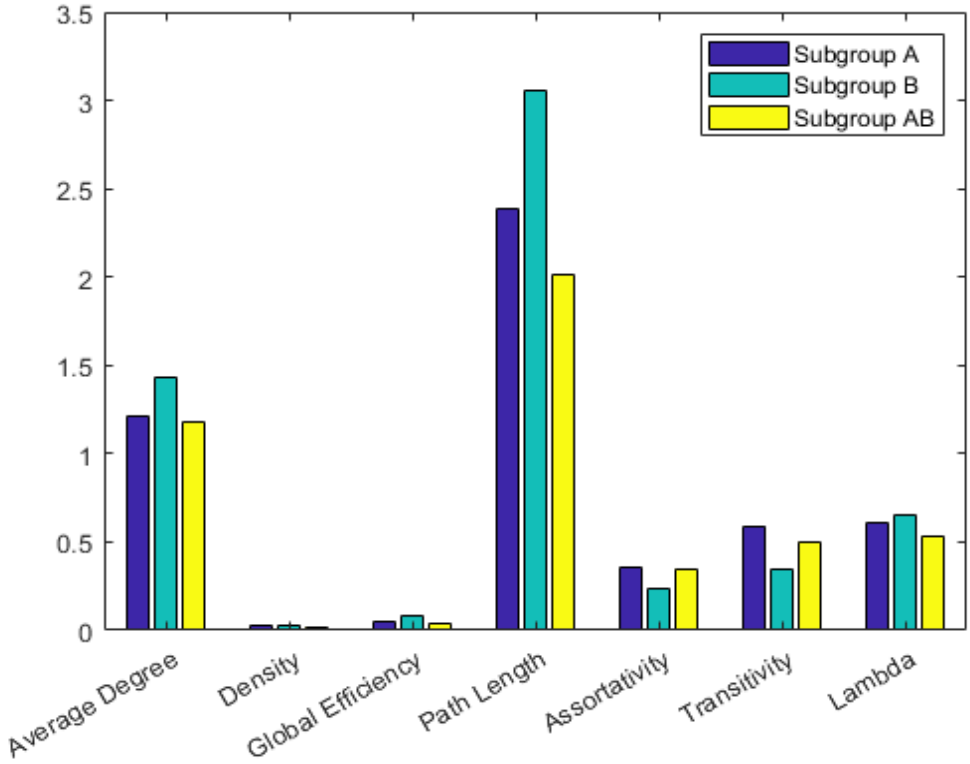


Figure 25. Plot showing network metrics of JICA-based subtypes

Student's t-test suggests significant group difference in betweenness centrality between subtypes A and B (t = -2.34, P = 0.02) and between subtypes B and AB (t = 0.006, P = 2.79). Participation coefficient significantly differed between subtypes A and B (t = -2.16, P = 0.03). However, no significant difference was observed in clustering coefficient, gamma, or eigen vector centrality between the subtype-specific networks.

Hubs were deciphered from betweenness centrality and average degree (**Figure 26**). The hubs identified are right superior frontal gyrus (rSFG, royal blue color) in subtype A, right middle occipital gyrus (rMOG, magenta color) in subtype B and left lateral orbitofrontal gyrus (lLatOrbG, green color) in subtype AB. A few common hubs were also found in different subtypes, such as right inferior temporal gyrus (rITG, cyan color) in Subtypes A and B, left caudate (lCaud, mustard yellow color) in subtypes A and AB, left superior frontal gyrus (lSFG, dark red color) and right middle temporal gyrus (rMTG, dark red color)

in subtypes B and AB. Additionally, left middle frontal gyrus (lMFG, grey color) was found in all three subtypes, which suggests this conserved brain region might be keeping communication active in all Parkinson's patients.

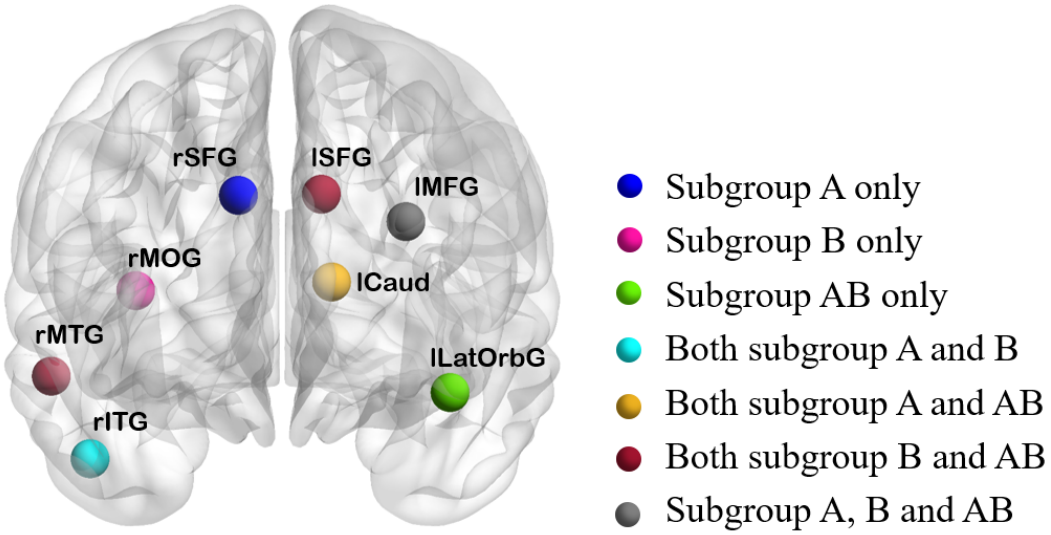


Figure 26. Hub regions in different JICA-based subtypes

5.4. DISCUSSIONS

A joint ICA approach identified joint GM and WM components differing significantly between healthy controls and PD patients. Considering components with top two effect sizes, PD patients were categorized based on loading weights in the respective components.

- The first PD category, subtype A, weighed higher in the component with highest effect size, *i.e.,* component 23 or LN-Th-IFG-RG-EN-Cul.
- The second PD category, subtype B, weighed higher in the component with second highest effect size, *i.e.,* component 28 or MTG-STG-IPL-SubG-SupMG.
- Patients belonging to the third PD category, subtype AB, weighed higher in both the components, i.e., component 23 and 28.

The above subtypes were obtained from fused information of grey and white matter tissues. It's intriguing to notice that while assessing these subtypes with the clinical features, it was observed that patients belonging to:

- subtype B had significantly least chronological age, UPDRS-Off value, UPDRS-On value, H&Y scale and age at onset than the other subtypes B and AB.
- subtype A had chronological age, UPDRS-Off value, UPDRS-On value and H&Y scale higher than those of subtype B, but lower than those of subtype AB.
- subtype AB had significantly highest chronological age, UPDRS-Off value, UPDRS-On value and H&Y scale than the other subtypes A and B.

UPDRS-Off assesses the severity of motor symptoms at no medication state to remove the effect of dopaminergic medications [311]. UPDRS-On measures the motor symptoms of the patients after drug was given. H&Y scale, the staging system indicating PD severity [122], is a standard for measuring disability and impairment [312]. Hence, this study results suggest that subtype B may be the mild-PD type with mild motor symptoms, subtype A is possibly the intermediate-PD type, subtype AB could be the severe-PD with predominance in motor impairment. Age at onset shows the age at which motor symptoms appeared.

The joint loading coefficients were correlated with clinical features of the subtypes to understand associations between them. The investigation suggested loading coefficient of joint component 23 in subtype A were positively correlated with UPDRS-Off, age at onset and chronological age. Similarly, subtype B loading coefficients of joint component 23 were positively correlated with UPDRS-Off, H&Y and chronological age. These subjects' loading coefficients in joint component 28 were also positively correlated with age.

The brain network of each of these derived subtypes were constructed based on association between grey matter and white matter volume of every pairs of brain regions (nodes). Each of the weighted undirected matrices, hence obtained, were binarized using mutual K-nearest neighbor (MKNN) based threshold. MKNN is a node affinity measure that defines

a bidirectional relationship between the nodes based on similarity values between grey matter and white matter information. The factor K in MKNN plays a significant role in network. As K increases, the number of neighboring nodes increase, allowing more number of edges in the network. Thus, extremely higher value of K would allow spurious edges, and hence the insignificant edges as well. This might even lead to a random network. On the other hand, extremely lesser K-value may potentially prevent the significant edges, leading to presence of several disconnected nodes in the network. Network efficiency indicates the number of steps needed for information transfer from one brain region to another. Highly efficient networks are thought to take fewer steps or shorter path lengths than inefficient networks, which usually take several steps. The global efficiencies show that subtype B transmits information faster than subtype A, and subtype A faster than AB. Also, path length values of subtype B is greater than that of subtype A, which is again higher than that of subtype AB. So, global efficiency and pathlength results indicate that subtype B might transfer information efficiently through longer path.

A measure of network's resilience, network assortativity, is association between the degrees of all nodes on two opposite sides of an edge [226]. In assortative networks, vertices with higher degrees tend to be connected together in a sub-network with greater mean degree than the overall network [313]. This phenomenon is reflected in network structure of the three subtypes (**Figure 24**) where subtype A with highest assortativity has more specific subnetwork comprising less number of nodes as compared to that of subtypes AB and B. Corroborating with the phenomenon, same nodes on opposite hemispheres are interconnected, eg. in subtype A, Left and right supramarginal gyrus (nodes 19, 20 in **Figure 24**, bottom left), left and right precuneus (nodes 23, 24 in **Figure 24**, bottom left), left and right middle temporal gyrus (node 35, 36 in **Figure 24**, bottom left), left and right inferior temporal gyrus (nodes 37, 38 in **Figure 24**, bottom left), left and right superior

frontal gyrus (nodes 1, 2 in **Figure 24**, bottom left), left and right middle frontal gyrus (nodes 3, 4 in **Figure 24**, bottom left) and left and right middle orbitofrontal gyrus (nodes 9, 10 in **Figure 24**, bottom left). High-degree and low-degree nodes in assortative networks prefer connecting to other high- and low-degree nodes, respectively [313]. This order tends to erode as the assortativity coefficient falls, and some vertices start forming new connections with vertices that have degrees that are less similar to their own degrees. In PD subtype B, the lower assortativity coefficient might indicate that few regions start establishing or increasing connections with other regions. The decreased assortativity in PD subtype B might suggest that few brain regions begin to connect with other regions, possibly the reason for a bigger subnetwork than those of subtypes A and AB.

Betweenness centrality quantifies involvement of each node in information flow within the network. It also helps examining how different nodes (brain regions) interact with each other. Hence, in order to determine information flow both within and between subnetworks, betweenness centrality was employed as a metric for analysis. The results indicate subtype B differing in betweenness centrality from the other two subtypes (subtypes A and AB). This measure was also used to find hub regions in the networks. Participation coefficient is another metric which was compared between the subtypes. Participation coefficient distinguishes a provincial hub from a connector hub, and therefore indicates whether a node transmits information to distant nodes or to its neighboring nodes. This study suggests a significant difference in participation coefficient between subtypes A and B. Frontal and temporal gyrus were majorly found as hubs in different subtypes, showing these regions with significant role in information transmission through the network.

5.4.1. Challenges and opportunities

There are a few limitations associated with this chapter. Only two joint sources with highest effect size between healthy and PD were utilized for subtyping, possibly why all PD patients could not be categorized into any of the subtypes. Hence, considering all components might help to subtype the entire PD cohort. Beside, multiple features (GM and WM) used from a single modality, multimodal neuroimaging data might be advantageous in PD subtyping. In this direction, imaging transcriptomics, *i.e.,* combined evaluation of neuroimaging and gene expression data to understand molecular changes in PD, may offer deep insights into the disease pathology and unravel disease heterogeneity. Brain network using mutual neighborhood is dependent on K, which limits the average network degree to K. However, the strength of MKNN is implementation of a single threshold value instead of a range of thresholds based on mutual information between GM (neuronal head) and WM (axon) tissue.

5.5. CONCLUSION

Based on GM and WM information, the present study successfully deciphered PD subtypes associated with specific clinical symptoms. Inter-subtype differences in network metrics were also found based on mutuality between the brain regions. Thus, this approach of identifying subtypes within the disease may be useful in precision treatment in contrast to traditional one-size-fits-all approach, where a single treatment is uniformly administered to all PD patients. This study encourages tailoring of subtype-specific treatments by considering respective spectrum of clinical symptoms along with neuroimaging features.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

CHAPTER -6

Optimizing Permutations In Biclustering Algorithm for Simulated and Structural MRI Data

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

Chapter 6. OPTIMIZING PERMUTATIONS IN BICLUSTERING ALGORITHM

6.1. INTRODUCTION

Clustering is a popular unsupervised machine-learning technique that has been used to identify clusters within a dataset. Some of the applications of clustering include image segmentation and recognition, information retrieval, data mining, and so on [314], [315]. Even though it has many uses, clustering doesn't identify data points that may belong to several clusters over varying conditions (local patterns) [316]. For example, in a gene expression matrix, identifying the local patterns would lead to a greater understanding of genes that up-regulate or down-regulate together and the particular sample/condition in which such expression patterns occur. Biclustering is one such technique that can overcome this limitation of clustering. It is a data mining technique that clusters rows and columns together. A bicluster refers to a two-dimensional sub-matrix within a larger matrix. They can be classified based on the type of values (constant, scaled variables, etc.,), structure (Row or column exclusivity and exhaustivity, etc.) or based on numerous evaluation measures like Mean Square Residue (MSR) or Scaled Mean Square Residue, etc. [317].

Biclustering also has applications in disease subtyping [47], [317], [318], [319], [320], [321]. Other than biological data analysis, the technique has also been used in text mining [322], [323]. Some of the popular biclustering algorithms have been integrated into the toolbox BIDEAL, developed by [324]. However, it is a comparatively recent approach as applied to complex medical image data [325]. Some of the advancements concerning biclustering with neuroimaging datasets are Biclustered Independent Component Analysis (BICA) [204], N-component biclustering (N-BiC) [205], and Fuzzy non-negative matrix factorization (FNMF) [203]. In BICA, biclustering was performed on two components

selected based on previous studies, whereas in N-BiC, multi-component biclustering was performed, both studies employing sMRI data [204], [205]. In N-BiC, permutation refers to the order in which components are sent to the modified DFS algorithm. For example, a permutation using three components (say 1, 2, and 3) relates to common subjects between components 1,2, and 3. Similarly, yet another permutation 2,3 and 1 relates to the same set of subjects between components 1, 2 and 3 in different order, which is therefore redundant. A similar FNMF was applied to diffusion tensor imaging (DTI) data resulting in its decomposition to two different matrices with only positive entries. Overall the number of studies in this domain is relatively less, and this warrants research for better and faster algorithms.

This study attempted to improve upon N-BiC, which has been used for identifying the subtypes in a cohort of patients suffering from Schizophrenia [205]. Although the application of this algorithm onto neuroimaging datasets led to a better understanding of disease, the method requires high resources due to the n! permutation-based merging step. The proposed algorithm, Modified N-BiC, which performs biclustering, includes usage of an iterative search method and performing comprehensive merging, thus removing the need for permutations. This optimization of the permutation count brings down the runtime and memory requirements significantly. To illustrate this, two simulated datasets were generated using custom MATLAB code and a publicly available multicentre sMRI dataset was also considered from Parkinson's Progression Marker's Initiative (PPMI) [326].

6.2. MATERIALS AND METHODOLOGY

6.2.1. Datasets

Two artificial datasets with embedded biclusters to mimic the subject by component matrix were created to compare the performance of N-BiC and Modified N-BiC. These were

generated with a normal distribution in size (400×10) with the help of the MATLAB function randn. The matrix was normalized by using min_max_scaling to the range [0-1]. Once the data is generated, the biclusters were embedded in the data matrix (well separated to enable ease of interpretation), by artificially increasing the value of loading coefficients at particular locations. Dataset 1 has 10 columns, while Dataset 2 (**Figure 27**) has 11 columns. Both datasets have three biclusters embedded with the following specifications (**Table 16**). The datasets have 10 and 11 columns respectively, to demonstrate the effect of column count on performance.

Figure 27. Sorted eleven component dataset visualized using Imagesc (MATLAB)

Table 16. Details of simulated dataset

Dataset number	Bicluster number	Columns	Rows
1	1	1,2,3	1 to 133
	2	4,5,6	134 to 266
	3	7,8,9,10	267 to 400
2	1	1,2,3	1 to 133
	2	4,5,6,7	134 to 266
	3	8,9,10,11	267 to 400

Patient and Demographic details. PPMI is a publicly-shared large multicentre dataset [326] from which 165 healthy controls (HC) and 216 Parkinson's disease (PD) patients were considered (**Table 17**). The studies were conducted following the Good Clinical Practice (GCP) principles and Declaration of Helsinki after approval was received from the regional ethics committees of the participating sites.

Table 17. Demographics details for PPMI participants

Measures	HC	PD
Count (Total (Male))	165 (107)	216 (130)
Age in Yrs (Mean ± SD)	60.50 ± 11.03	58.82 ± 7.39
Age range	31-83	38-70

6.2.2. Device Specifications and Software

Following are the device specifications under which the experiments were run. CPU: I5-3470S, GPU: GT 710, System memory: 16 GB, Ram Specifications: DDR3 with 1600 MHz. All program codes were written using MATLAB 2022a software.

6.2.3. Pre-processing of sMRI

The sMRI images were acquired using multiple scanner models of 1.5 or 3T with scan acquisition parameters. The sMRI scans were preprocessed according to the methodology followed in [82]. Computational anatomy toolbox, CAT12, was used within Statistical Parametric Mapping (SPM 12) for preprocessing in MATLAB R2021a. The following steps were involved in the preprocessing:

1. Normalization and registration to standard 152 average T1 Montreal Neurological Institute (MNI) template.
2. Segmentation into multiple tissue types using unified segmentation algorithm [327].

3. Smoothing of grey matter (GM) images using full-width half maximum (FWHM) Gaussian kernel of 8mm.

The preprocessed images were regressed with age, gender, site, scanner, and field strength. These scans in the form of a 4D matrix (3D images over subjects) were flattened to a 2D matrix (subjects by voxel). Source-based morphometry (SBM) [251] was used on the subjects by voxel matrix, to obtain the loading coefficient matrix (subjects by components) and component matrix (components by voxel). The Group ICA of the fMRI toolbox (GIFT, https://github.com/trendscenter/gift) was used to perform decomposition.

After SBM, a loading matrix of 381×30 was used. The number of components used in the current study was followed from previous subtyping studies [204], [205]. Of these 30 components, 10 and 11 number of components were considered separately to provide a balanced comparison between the two algorithms, N-BiC and Modified N-BiC. The N-BiC algorithm requires more than 42GB to store 12! permutations of 12 components. In MATLAB this takes up 12! × 12 × 8 bytes, where 8 is the number of bytes allocated by MATLAB to a single number of double data type. This gives around 42.8 GB to be exact.

6.2.4. Modified N-BiC Algorithm

Overview: A few steps of Modified N-BiC algorithm are similar to the N-BiC algorithm [205] where the loading matrix is sorted according to the best methods explained therein (Step 1 of **Figure 28**). The loading matrix, sorted using these methods, is further used to identify biclusters, by evaluating every combination of components at the given parameters. The loading coefficients of common subjects between pairs or groups of components are stored in a separate data structure (Bicluster Search) (Step 2 of **Figure 28**). At the end of the search, the biclusters are sent to a local validator, where every bicluster is compared with (N-1) biclusters in the set. N is the number of biclusters enlisted and biclusters already compared are excluded. The biclusters are merged based on overlap criteria, and thus final

list contains unique biclusters for further analysis. The total number of such comparisons comes out to be N×(N-1)/2. This validation process is shown in Step 3 of **Figure 28** (Merging). After Step 3, required subgroups are identified from the bicluster list (Step 4 of **Figure 28**).

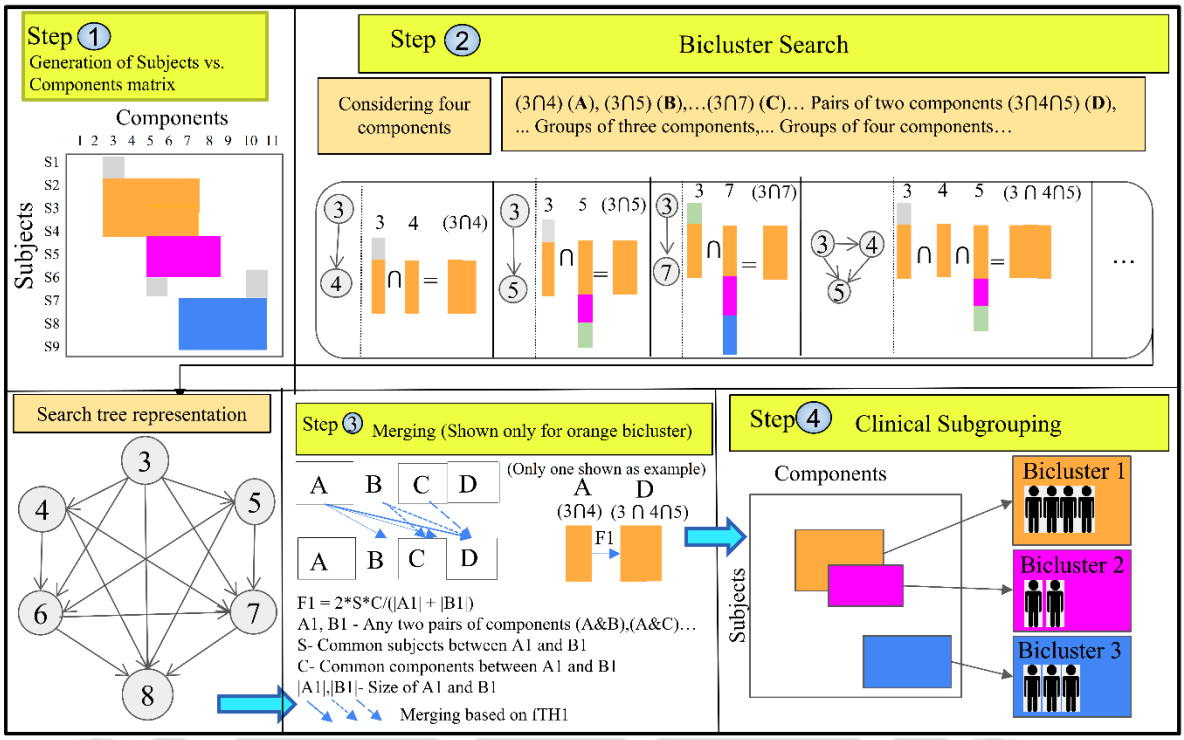


Figure 28. Graphical abstract depicting methodology behind Modified N-BiC depicting bicluster search, merging, and subgrouping

Sorting: The main idea of sorting is to select only those subjects which show high expression for a particular component. For each component, the individual loading coefficients are compared with the mean loading coefficient, and then retained or eliminated based on the threshold. More details on sorting methods is explained in [205].

Parameters and Search Bic function: The algorithm is a type of Depth-First Search (DFS), called iterative deepening DFS [328], which explores the intersection of all possible subsets of components. It combines the approaches of DFS and Breadth-First Search DFS [328] as it looks for the intersection between components at a particular depth, before

moving to a higher depth. For example, intersection for common subjects is performed between pairs of two components such as (1, 2), (1, 3), (1, 4), and so on, after which intersection between groups of three components (1, 2, 3), (1, 2, 4) happens and so on. This process continues till maximum depth is reached (total number of components) as shown in step 2 of **Figure 28**. The total number of such component pairs is given in equation 6.

$$TotalPair(TP) = \sum_{k=2} \binom{n}{k} \tag{6}$$

where k goes from 2 to n and n is the number of components in the matrix.

This study used MINTERSECT, a MATLAB helper file, for performing sequential intersection. The parameters (*NoS, K, O*) referred to in [205] as (*S*, *K*, *O*) are used for present approach except for *fTH2*, which is the overlap criteria for merging over all permutations. This is because, in Modified N-BiC, merging takes place with a single validator, requiring a single overlap parameter (*fTH1*). The overlap between two biclusters is determined by the F1 index or Dice index, which calculates the similarity between two biclusters [205]. The total number of operations (*TO*) in the worst-case scenario for Modified N-BiC, is given by equation 7, while that for N-BiC is given by equation 8.

Let *Pair* represent an array of length *TP* (equation 6) containing all pairs of components, *m* be number of non-zero elements in a particular column, *N* be the total number of enlisted biclusters, *M, P* be number of subjects, components, respectively, in bicluster 1, and *Z, Q* be number of subjects, components, respectively, in bicluster 2 for pairwise comparison.

The total number of operations in the proposed approach, Modified N-BiC are:

$$TO_{Modified-N-BiC} = \sum_{i=1}^{TP} \prod_{k=1}^{length(Pair(i))} m_k + \sum_{j=1}^{\frac{N*N-1}{2}} \left(M_j * Z_j + \left(P_j * Q_j\right)\right) \tag{7}$$

whereas, in N-BiC, the total number of operations is:

$$TO_{N-BiC} = \sum_{a=1}^{NP}\left[\sum_{i=1}^{TP}\prod_{k=1}^{length(Pair(i))} m_k + \sum_{j=1}^{VN}\left(M_j * Z_j + \left(P_j * Q_j\right)\right)\right] \quad (8)$$

where *VN* is $\frac{N*N-1}{2}$ when a bicluster is compared to N-1 remaining biclusters.

Here *NP* is the number of permutations considered, *VN* is the variable number of biclusters, and since the bicluster merging is order dependent, merged bicluster pairs are removed. Here the first term in equation 7 and 8 denotes the number of comparison operations involved in a component pair, which is sum of the product of *m* over all the pairs (*TP*) in a permutation. The second term in equation 7 relates to the intersection between two biclusters, which is the computed intersection of the common subjects/ indices (*M* and *Z*) and common components (*P* and *Q*) summed over *N*(N-1)/2*. In equation 8, this relates to the number of operations in internal validation (internal summation across a variable number of biclusters) and the summation over all the permutation steps). To perform an exhaustive merging across all permutations, $NP = n!$

6.2.5. Evaluation Measures

Mean Square Residue (MSR): Mean square residue [317] is a measure reflecting the homogeneity of a bicluster. With MSR, it is possible to identify biclusters with similar values of loading coefficients, allowing one to better understand the dataset.

$$\text{MSR(B)} = \frac{1}{|I| * |J|} \sum_{i=1}^{i=|I|}\sum_{j=1}^{j=|J|}(b_{ij} - b_{iJ} - b_{Ij} + b_{IJ})^2 \quad (9)$$

Here B is a bicluster, |I|, |J|, b_ij, biJ,bIj,bIJ stand for bicluster row size, bicluster column size, element in the ith row and jth column, row means, column means, and total means respectively.

F1 index and Consensus Score (CS): The F1 index is calculated for each pair of extracted biclusters and original biclusters. Consensus score [329] denotes the consensus between final list of biclusters (identified/ extracted) and the original list of biclusters (ground truth). Let's suppose there are three ground truth biclusters A, B, and C, and three extracted biclusters A1, B1, and C1. Suppose A has more subjects in common with A1, B in common with B1, and C in common with C1, according to F1 indexes. In this case, the F1 index is given by equation 10, and the consensus score is given by equation 11.

$$Numberofpairs = 3\{(A, A1), (B, B1), (C, C1)\}$$

$$\sum F1indexes = F1(A, A1) + F1(B, B1) + F1(C, C1) \quad (10)$$

$$CS = \frac{\sum F1indexes}{Numberofpairs} \quad (11)$$

6.2.6. Comparison with Bimax

The performance of Modified N-BiC was compared with an existing biclustering algorithm called Bimax [330], using the toolbox BIDEAL [324]. Bimax is a divide and conquer algorithm, which divides the whole binarized (1's and 0's) matrix into several partitions and searches for biclusters with only 1's. The algorithm was chosen for its ease of use and interpretation.

6.3. RESULTS AND DISCUSSIONS

Three algorithms namely N-BiC, Modified N-BiC and Bimax were compared using simulated data and PPMI dataset. Positive and negative sorting method, as included within the N-BiC toolbox was the default sorting method for all scenarios.

6.3.1. Simulated Dataset.

Comparison of runtime and average MSR on simulated datasets are shown in **Table 18**. **Table 19** lists the biclusters identified by both algorithms on dataset 1. Comparison of consensus scores of each algorithm over the simulated dataset is shown in **Table 20**. **Table 21** represents biclusters enlisted by both algorithms on dataset 2. The results at 100 permutations are shown in **Table 19** and **Table 21**. Parameter settings used for following experiments: NoS = 125, K = 2, O = 0.8, where NoS – minimum number of subjects in a bicluster, K – minimum number of components in a bicluster, O – Overlap between biclusters. These parameters were set based on the rationale of N-BiC algorithm.

Table 18. Performance of both algorithms on simulated datasets

Number of Components	Number of Permutations	N-BiC (RT)	Modified N-BiC (RT)	N-BiC (AvMSR)	Modified N-BiC (AvMSR)
10	1	0.66	0.82	1.04	0.94
10	2	0.74	-	1.07	-
10	5	0.82	-	1.06	-
10	100	3.00	-	1.06	-
11	1	6.61	0.98	0.86	1.01
11	2	6.50	-	0.89	-
11	5	6.68	-	0.90	-
11	100	9.03	-	0.91	-

RT- Runtime (seconds), *AvMSR*- Average MSR

Biclusters discovered by Bimax on Dataset 2: Parameters settings: min_rows (Minimum rows): 125, min_cols (Minimum columns): 2, NumBiClust (No. Of Biclusters): 20. Biclusters found by Bimax (in the format of {Components, Number of Subjects}): {(2,3), 128}, {(1,3), 127}, {(1,2), 129}, {(1,2,3), 126}, {(8,11), 127}, {(10,11), 128}, {(8,10), 127}, {(8,9), 128}.

Table 19. Biclusters enlisted by both algorithms on Dataset 1

Sl. No.	N-BiC (Number of Components)	Modified N-BiC (Number of Components)	N-BiC (Number of Subjects)	Modified N-BiC (Number of Subjects)
1	[1,2]	[1,2]	125	129
2	[1,3]	[1,3]	126	127
3	[2,3]	[2,3]	126	128
4	[7,8,9]	[7,8,9]	125	127
5	[7,8,10]	[7,8]	125	129
6	[7,9,10]	[7,10]	126	131
7	[8,9,10]	[8,9]	125	128
8	-	[8,10]	-	127
9	-	[9,10]	-	129
10	-	[1,2,3]	-	126
11	-	[7,9]	-	131
12		[8,9]		128

Table 20. Consensus score comparison on simulated datasets

Number of Components	Number of permutations	N-BiC	Modified N-BiC
10	1	0.806	0.7396
10	2	0.793	-
10	5	0.798	-
10	100	0.802	-
11	1	0.706	0.743
11	2	0.691	-
11	5	0.692	-
11	100	0.682	-

Table 21. Biclusters enlisted by both algorithms on Dataset 2

Sl. No	N-BiC (Number of Components)	Modified N-BiC (Number of Components)	N-BiC (Number of Subjects)	Modified N-BiC (Number of Subjects)
1	[1,2]	[1,2]	126	129
2	[1,3]	[1,3]	126	127
3	[2,3]	[2,3]	126	128
4	[8,9]	[8,9]	125	128
5	[8,10]	[8,10]	125	127
6	[8,11]	[8,11]	125	127
7	[9,10]	[9,10]	125	129
8	[9,11]	[9,11]	126	129
9	[10,11]	[10,11]	126	128
10	-	[1,2,3]	-	126
11	-	[9,10,11]	-	126
12	-	[8,9,10]	-	125
13	-	[8,9,11]	-	125

6.3.2. PPMI Dataset.

Parameter settings used for performance comparison using the PPMI dataset are *NoS* = 40, *K* = 2, *and O* = 0.8. The PD-only loadings were considered for biclustering on components: (1:10) and (1:11), where *NoS* – minimum number of subjects in a bicluster, *K* – minimum number of components in a bicluster, *O* – Overlap between the biclusters. A comparison of runtime and average MSR on the loading matrix is shown in **Table 22**. These parameters were set based on the rationale of the N-BiC algorithm, although other configurations may be explored in future studies.

Table 22. Performance on PPMI Dataset

Number of Components	Number of Permutations	N-BiC (RT)	Modified N-BiC (RT)	N-BiC (AvMSR) (×10^8)	Modified N-BiC (AvMSR) (×10^8)
10	1	0.61	0.15	3.58	3.15
	2	0.64		3.71	
	5	0.67		3.33	
	100	2.11		3.15	
11	1	6.51	0.19	4.09	3.38
	2	6.44		3.62	
	5	6.58		3.63	
	100	8.46		3.38	

RT- Runtime (seconds), *AvMSR*- Average MSR

6.3.3. Performance of Modified N-BiC on PPMI dataset

Parameter settings used for bicluster search using the PPMI dataset: *NoS* = 55, *K* = 2, *O* = 0.8, where *NoS* – minimum number of subjects in a bicluster, *K* – minimum number of

components in a bicluster, *O* – Overlap between the biclusters. The PD-only loadings were considered for biclustering on all 30 components. Biclusters discovered: Components (3 and 16) with 56 subjects and components (5 and 16), (5 and 17), (25 and 30) with 55 subjects each. Modified N-BiC is capable of finding biclusters from a loading matrix with 30 components, under the same system configuration. Although the algorithm was executed under various parameters of *NoS*, K, and *O*, the results have only been presented for one parameter setting as proof of execution. Biclusters common to 4 or more components were not generated under various parameters, so the algorithm was executed up to the intersection between three components. Modified N-BiC used 5MB for dataset 1 and 7.2 MB for dataset 2 (both approximate). The memory usage was calculated using *whos* command in MATLAB. N-BiC on the other hand used 276.8 MB for dataset 1 and 3.27 GB for dataset 2 (just for allocation of the space for permutation possibilities) and it requires 42GB of system memory for the same (12 components and more). At the same system configuration, N-BiC failed to run, but Modified N-BiC successfully discovered the biclusters efficiently. This indicates the feasibility of execution for Modified N-BiC and a higher number of valid biclusters returned by exhaustive search approaches in a reasonable time. It can be seen that Bimax performed on par with N-BiC as well as Modified N-BiC on Dataset 1 and Dataset 2. Thus, it is suggest that Modified N-BiC can work comparably better to existing algorithms, although more research on this aspect is required.

It is observed from **Table 18** and **Table 22** that the run time increases for both N-BiC and Modified N-BiC with an increase in the number of components, though the increase in time for N- BiC is more than that for Modified N-BiC. This is because N-BiC is a backtracking algorithm that uses recursive DFS to enumerate search possibilities, and recursion can be less efficient in terms of time and space complexity over iterative approaches, especially with an increase in the number of operations [331].

For N-BiC, the algorithm has to run for a higher number of permutations (dependent on the dataset) to get stable results. This is because the arrangement/permutation plays a role in the enlistment of biclusters. Increasing the number of permutations increases the running time, and has a varying influence on the Average MSR due to a varying number of biclusters enlisted, every time the algorithm is executed. In the end, it can be seen from **Table 18**, **Table 19**, **Table 20**, **Table 21** and **Table 22**, that the Modified N-BiC has an on-par or similar performance with N-BiC without requirement of permutation.

6.3.4. Research Limitations/ Implications

The biclusters enlisted by the algorithm are dependent on the *NoS, K, and O* and hence need to be tuned very properly. These values also affect the runtime of the algorithm (same as in N-BiC [205]). For example, if the overlap (*O*) is set high, highly similar, and too many biclusters are listed. This makes the analysis tedious. In contrast, if the *O* value is set to be too low, an aggressive merging takes place, and too few biclusters are left for analysis.

6.4. ORIGINALITY AND VALUE

This study presents an improvement over N-BiC [205] in terms of reduction in memory requirements and run time while providing on-par performance with the same (bicluster enlistment). The improvement comes in the form of a switch to an iterative deepening DFS over the recursive DFS with backtracking in N-BiC. The merging step was optimized by comparison of a bicluster with other biclusters in a single permutation, which reduces the necessity for n! permutations in the worst case down to one permutation. Hence, the proposed approach seems to provide similar results at a lesser computational cost. The value of Modified N-BiC lies in its potential as an alternative biclustering algorithm that can be used for disease subtyping not only from sMRI data but for other varied applications like cancer and gene analysis.

6.5. CONCLUSION & FUTURE RESEARCH WORK

This study suggests improvements for N-BiC algorithm by using an iterative DFS approach and a comprehensive merging step, thereby, removing the need for permutations. The Modified N-BiC approach works on par with the N-BiC approach and uses lesser resources than the latter. The provided source code and datasets for download will enable further testing and validation of this work. This study also shows the feasibility of execution of exhaustive biclustering algorithms on devices with low specifications, especially for subtyping. This algorithm can still be improved upon, in the future by use of optimizations like dynamic programming approach. It can also be used on other datasets to obtain biologically relevant biclusters.

Code Availability. The simulated data and codes used in the current study are available at the below GitHub link:

https://github.com/NeuralLabIITGuwahati/Permutations_Biclustering

LIMITATIONS & FUTURE SCOPE

LIMITATIONS & FUTURE SCOPE

There are a few limitations of this research that need to be acknowledged and further explored. An algorithm was proposed, namely Unique Brain Network Identification Number (UBNIN), for encoding any given network in Chapter 3. The entire UBNIN value can be precisely calculated manually (using paper and pen). However, the current computer hardware technology limits the number of significant digits, which slightly hampers accurate reconstruction of larger binary matrix. Furthermore, for a 20 × 20 fully connected binary matrix, 190 digits would be required to represent the collection of decimal values. The UBNIN value hence obtained are rounded up which is a computational constraint. This limitation is due to the multiplication factor ($\frac{1}{2^i}$) that rounds up at large i (i.e. node) values.

Age-based network analysis of PD patients was also performed in Chapter 3. Since prodromal stage is a crucial time in PD and is not explored yet, more research is required in this direction. Also, longitudinal study may help understand the underlying change in brain network over time due to the disease. The neurodegeneration in the disease may thus be understood clearly in PD. Although PD is uncommon at younger age and the immensely dense connectivity at younger age (≤ 32 years) is an obvious truth, the connectivity could be investigated in detail with large sample size of younger patients (in this study only 4 PD patients below 32 years old were available) to better understand the effect of Parkinson's disease. The number of HC individuals were also limited and hence the analysis may be replicated on a larger dataset. Although the current study didn't analyse the effect of medication/therapy, the same may be tested by comparing longitudinal changes in PD patients without therapy with those with therapy.

Subtyping of Parkinson's disease (in Chapter 4) was performed using only grey matter information obtained from structural MRI. However, inclusion of other tissue features

might provide additional information of the subjects and help analysing simultaneous changes in these tissues. To address this challenge, white matter information was considered (in Chapter 5) along with grey matter data to identify the brain regions where the tissues covary. Furthermore, connectivity analysis of the derived subtypes was also conducted (in Chapter 5) on adjacency matrix. In this approach, binarization was done using Mutual K-nearest neighbor based thresholding method, where K is the limiting parameter due to which the maximum degree of any node would be restricted to K.

Longitudinal investigations based on multimodal imaging may aid in the discovery of neurophysiological sequences of events in specific brain regions of the deciphered subgroups. This may aid with a thorough understanding of disease pathology.

Deciphering subtypes and correlating with specific clinical characteristics has potential implications in stratification of PD patients for suitable clinical trials to comprehend disease pathophysiology. However, combining functional and genomics data with the imaging data may help find association of subtype-specific brain regions with its specific functional impairment, and to find if a common gene lies behind a specific subtype. Neither of the studies (Chapter 4 & 5) take every SBM component into account, but only the two components with highest effect size. This could have resulted to information of other components being missed, and possibly why some of the participants could not be categorised into any of the subtypes.

Brain networks were exploited to understand the information transmission, however, other imaging modalities explaining the trajectory of information flow may provide additional detail. Also, to have an unbiased analysis, inter-subtype comparisons of brain networks were made at a range of thresholds (sparsity) (Chapter 4 & 5). A common data-driven threshold strategy such as adaptive thresholding [296], [297] may be attempted for

comparison of networks of multiple disorders. Adaptive thresholding is advantageous over conventional thresholding method as they automatically adapt to divergent network settings and hence generalizable.

Despite the widespread use of unweighted (or binary) networks, they do not reflect the continuous nature of correlation values between 0 and 1. Any value lower than the threshold loses its information in binary network and is considered 0. However, in weighted network [298], the exact correlation value is taken into account. And hence, instead of binary correlation matrix for network analysis, use of weighted correlation matrix may be encouraged to avoid the informational loss during binarization. Conditional dependency between two nodes has been used as directional information to obtain directed network [299], which may be ventured in structural networks as well. These are the few factors that restricts the current study, and may be taken up in future.

PD patients with longer disease duration, early age of onset, higher dose and longer duration of levodopa treatment, and certain genetic factors are associated with higher incidence of Levodopa Induced Dyskinesias (LIDs) [332]. Also, PD patients with LID have greater levels of iron deposition in their substantia nigra compared to those without LID. This suggests that a greater quantitative substantia nigra susceptibility mapping (QSM) score could be a potential neuroimaging biomarker for LID early diagnosis [333]. Future studies may be taken up towards identifying early diagnostic biomarkers of LIDs and subgrouping studies may be extended in LIDs.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] J. Parkinson, “An Essay on the Shaking Palsy,” *JNP*, vol. 14, no. 2, pp. 223–236, May 2002, doi: 10.1176/jnp.14.2.223.

[2] O. Hornykiewicz, “Basic research on dopamine in Parkinson’s disease and the discovery of the nigrostriatal dopamine pathway: the view of an eyewitness,” *Neurodegener Dis*, vol. 5, no. 3–4, pp. 114–117, 2008, doi: 10.1159/000113678.

[3] S. K. Van Den Eeden, “Incidence of Parkinson’s Disease: Variation by Age, Gender, and Race/Ethnicity,” *American Journal of Epidemiology*, vol. 157, no. >11, pp. 1015–1022, Jun. 2003, doi: 10.1093/aje/kwg068.

[4] S. Cerri, L. Mus, and F. Blandini, “Parkinson’s Disease in Women and Men: What’s the Difference?,” *J Parkinsons Dis*, vol. 9, no. 3, pp. 501–515, 2019, doi: 10.3233/JPD-191683.

[5] T. H. Reekes, C. I. Higginson, C. R. Ledbetter, N. Sathivadivel, R. M. Zweig, and E. A. Disbrow, “Sex specific cognitive differences in Parkinson disease,” *npj Parkinsons Dis.*, vol. 6, no. 1, p. 7, Dec. 2020, doi: 10.1038/s41531-020-0109-1.

[6] World Health Organization, “Facts on Parkinson disease by WHO.” Accessed: Aug. 19, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

[7] E. R. Dorsey *et al.*, “Global, regional, and national burden of Parkinson’s disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016,” *The Lancet Neurology*, vol. 17, no. 11, pp. 939–953, Nov. 2018, doi: 10.1016/S1474-4422(18)30295-3.

[8] G. Singh *et al.*, “The burden of neurological disorders across the states of India: the Global Burden of Disease Study 1990–2019,” *The Lancet Global Health*, vol. 9, no. 8, pp. e1129–e1144, Aug. 2021, doi: 10.1016/S2214-109X(21)00164-9.

[9] V. W. Sung and A. P. Nicholas, “Nonmotor Symptoms in Parkinson’s Disease: Expanding the View of Parkinson’s Disease Beyond a Pure Motor, Pure Dopaminergic Problem,” *Neurologic Clinics*, vol. 31, no. 3, Supplement, pp. S1–S16, Aug. 2013, doi: 10.1016/j.ncl.2013.04.013.

[10] D. K. Simon, C. M. Tanner, and P. Brundin, “Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology,” *Clinics in Geriatric Medicine*, vol. 36, no. 1, pp. 1–12, Feb. 2020, doi: 10.1016/j.cger.2019.08.002.

[11] M. L. Kramer and W. J. Schulz-Schaeffer, “Presynaptic α-Synuclein Aggregates, Not Lewy Bodies, Cause Neurodegeneration in Dementia with Lewy Bodies,” *The Journal of Neuroscience*, vol. 27, no. 6, pp. 1405–1410, Feb. 2007, doi: https://doi.org/10.1523/JNEUROSCI.4564-06.2007.

[12] J. Levin, A. Kurz, T. Arzberger, A. Giese, and G. U. Höglinger, “The Differential Diagnosis and Treatment of Atypical Parkinsonism,” *Deutsches Aerzteblatt Online*, Feb. 2016, doi: 10.3238/arztebl.2016.0061.

[13] T. Ishizawa, P. Mattila, P. Davies, D. Wang, and D. W. Dickson, "Colocalization of tau and alpha-synuclein epitopes in Lewy bodies," *J Neuropathol Exp Neurol*, vol. 62, no. 4, pp. 389–397, Apr. 2003, doi: 10.1093/jnen/62.4.389.

[14] J. M. Fearnley and A. J. Lees, "Ageing and Parkinson's disease: Substantia nigra regional selectivity," *Brain*, vol. 114, no. 5, pp. 2283–2301, Oct. 1991, doi: 10.1093/brain/114.5.2283.

[15] P. Mahlknecht, K. Seppi, and W. Poewe, "The Concept of Prodromal Parkinson's Disease," *J Parkinsons Dis*, vol. 5(4), pp. 681–697, 2015, doi: 10.3233/JPD-150685.

[16] J. Massano and K. P. Bhatia, "Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management," *Cold Spring Harb Perspect Med*, vol. 2, no. 6, p. a008870, Jun. 2012, doi: 10.1101/cshperspect.a008870.

[17] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, Apr. 2008, doi: 10.1136/jnnp.2007.131045.

[18] L. V. Kalia and A. E. Lang, "Parkinson's disease," *Lancet*, vol. 386, no. 9996, pp. 896–912, Aug. 2015, doi: 10.1016/S0140-6736(14)61393-3.

[19] B. S. Connolly and A. E. Lang, "Pharmacological Treatment of Parkinson Disease: A Review," *JAMA*, vol. 311, no. 16, pp. 1670–1683, Apr. 2014, doi: 10.1001/jama.2014.3654.

[20] A. A. Moustafa *et al.*, "Motor symptoms in Parkinson's disease: A unified framework," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 727–740, Sep. 2016, doi: 10.1016/j.neubiorev.2016.07.010.

[21] E. Qian and Y. Huang, "Subtyping of Parkinson's Disease - Where Are We Up To?," *Aging Dis*, vol. 10, no. 5, pp. 1130–1139, Oct. 2019, doi: 10.14336/AD.2019.0112.

[22] M. Goedert, M. G. Spillantini, K. Del Tredici, and H. Braak, "100 years of Lewy pathology," *Nat Rev Neurol*, vol. 9, no. 1, pp. 13–24, Jan. 2013, doi: 10.1038/nrneurol.2012.242.

[23] H. Braak, K. D. Tredici, U. Rüb, R. A. I. de Vos, E. N. H. Jansen Steur, and E. Braak, "Staging of brain pathology related to sporadic Parkinson's disease," *Neurobiology of Aging*, vol. 24, no. 2, pp. 197–211, Mar. 2003, doi: 10.1016/S0197-4580(02)00065-9.

[24] A. Gauthier-Kemper *et al.*, "The frontotemporal dementia mutation R406W blocks tau's interaction with the membrane in an annexin A2-dependent manner," *J Cell Biol*, vol. 192, no. 4, pp. 647–661, Feb. 2011, doi: 10.1083/jcb.201007161.

[25] R. Brandt, J. Léger, and G. Lee, "Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain," *J Cell Biol*, vol. 131, no. 5, pp. 1327–1340, Dec. 1995, doi: 10.1083/jcb.131.5.1327.

[26] J. Chen, Y. Kanai, N. J. Cowan, and N. Hirokawa, "Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons," *Nature*, vol. 360, no. 6405, Art. no. 6405, Dec. 1992, doi: 10.1038/360674a0.

[27] X. Zhang *et al.*, "Tau Pathology in Parkinson's Disease," *Frontiers in Neurology*, vol. 9, 2018, doi: 10.3389/fneur.2018.00809.

[28] A. Alonso, T. Zaidi, M. Novak, I. Grundke-Iqbal, and K. Iqbal, "Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments," *Proc Natl Acad Sci U S A*, vol. 98, no. 12, pp. 6923–6928, Jun. 2001, doi: 10.1073/pnas.121119298.

[29] K. C. Luk *et al.*, "Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice," *Science*, vol. 338, no. 6109, pp. 949–953, Nov. 2012, doi: 10.1126/science.1227157.

[30] L. Stefanis, "α-Synuclein in Parkinson's Disease," *Cold Spring Harb Perspect Med*, vol. 2, no. 2, Feb. 2012, doi: 10.1101/cshperspect.a009399.

[31] J. Burré, M. Sharma, and T. C. Südhof, "Cell Biology and Pathophysiology of α-Synuclein," *Cold Spring Harb Perspect Med*, vol. 8, no. 3, Mar. 2018, doi: 10.1101/cshperspect.a024091.

[32] J. Burré, M. Sharma, T. Tsetsenis, V. Buchman, M. Etherton, and T. C. Südhof, "α-Synuclein Promotes SNARE-Complex Assembly in vivo and in vitro," *Science*, vol. 329, no. 5999, pp. 1663–1667, Sep. 2010, doi: 10.1126/science.1195227.

[33] T. Arai *et al.*, "Distinct isoforms of tau aggregated in neurons and glial cells in brains of patients with Pick's disease, corticobasal degeneration and progressive supranuclear palsy," *Acta Neuropathol*, vol. 101, no. 2, pp. 167–173, Feb. 2001, doi: 10.1007/s004010000283.

[34] R. A. Armstrong and N. J. Cairns, "Spatial patterns of the tau pathology in progressive supranuclear palsy," *Neurol Sci*, vol. 34, no. 3, pp. 337–344, Mar. 2013, doi: 10.1007/s10072-012-1006-0.

[35] M. G. Spillantini, M. L. Schmidt, V. M.-Y. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert, "α-Synuclein in Lewy bodies," *Nature*, vol. 388, no. 6645, Art. no. 6645, Aug. 1997, doi: 10.1038/42166.

[36] M. Poulopoulos, O. A. Levy, and R. N. Alcalay, "The neuropathology of genetic Parkinson's disease," *Mov Disord*, vol. 27, no. 7, pp. 831–842, Jun. 2012, doi: 10.1002/mds.24962.

[37] E. L. Thacker and A. Ascherio, "Familial aggregation of Parkinson's disease: a meta-analysis," *Mov Disord*, vol. 23, no. 8, pp. 1174–1183, Jun. 2008, doi: 10.1002/mds.22067.

[38] M. Funayama, K. Nishioka, Y. Li, and N. Hattori, "Molecular genetics of Parkinson's disease: Contributions and global trends," *J Hum Genet*, vol. 68, no. 3, Art. no. 3, Mar. 2023, doi: 10.1038/s10038-022-01058-5.

[39] S. J. Frucht and P. Termsarasab, "Phenomenology of Atypical Parkinsonism," in *Movement Disorders Phenomenology: An Office-Based Approach*, S. J. Frucht and P. Termsarasab, Eds., Cham: Springer International Publishing, 2020, pp. 37–61. doi: 10.1007/978-3-030-36975-0_4.

[40] G. K. Wenning, F. Krismer, and W. Poewe, "New insights into atypical parkinsonism," *Curr Opin Neurol*, vol. 24, no. 4, pp. 331–338, Aug. 2011, doi: 10.1097/WCO.0b013e3283480569.

[41] J. G. Graham and D. R. Oppenheimer, "Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy.," *J Neurol Neurosurg Psychiatry*, vol. 32, no. 1, p. 28, Feb. 1969, doi: 10.1136/jnnp.32.1.28.

[42] H. Okazaki, L. E. Lipkin, and S. M. Aronson, "Diffuse Intracytoplasmic Ganglionic Inclusions (Lewy Type) Associated with Progressive Dementia and Quadriparesis in Flexion," *J Neuropathol Exp Neurol*, vol. 20, no. 2, pp. 237–244, Apr. 1961, doi: 10.1097/00005072-196104000-00007.

[43] J. C. Steele, J. C. Richardson, and J. Olszewski, "Progressive Supranuclear Palsy: A Heterogeneous Degeneration Involving the Brain Stem, Basal Ganglia and Cerebellum With Vertical Gaze and Pseudobulbar Palsy, Nuchal Dystonia and Dementia," *Archives of Neurology*, vol. 10, no. 4, pp. 333–359, Apr. 1964, doi: 10.1001/archneur.1964.00460160003001.

[44] J. J. Rebeiz, E. H. Kolodny, and E. P. Richardson Jr, "Corticodentatonigral degeneration with neuronal achromasia: a progressive disorder of late adult life.," *Trans Am Neurol Assoc*, vol. 92, pp. 23–26, Jan. 1967, doi: PMID: 5634049.

[45] W. R. G. Gibb, P. J. Luthert, and C. D. Marsden, "Corticobasal Degeneration," *Brain*, vol. 112, no. 5, pp. 1171–1192, Oct. 1989, doi: 10.1093/brain/112.5.1171.

[46] E. A. Coon, "Autonomic Dysfunction in the Synucleinopathies," *Semin Neurol*, vol. 40, no. 5, pp. 492–501, Oct. 2020, doi: 10.1055/s-0040-1713844.

[47] T. Samantaray, J. Saini, and C. N. Gupta, "Subgrouping and structural brain connectivity of Parkinson's disease – past studies and future directions," *Neuroscience Informatics*, vol. 2, no. 4, p. 100100, Dec. 2022, doi: https://doi.org/10.1016/j.neuri.2022.100100.

[48] J. Ballinger, D. Bell, and T. Foster, "MRI physics," *Radeiopaedia*, Apr. 07, 2013. Accessed: Aug. 23, 2023. [Online]. Available: https://doi.org/10.53347/rID-22558

[49] S. Currie, N. Hoggard, I. J. Craven, M. Hadjivassiliou, and I. D. Wilkinson, "Understanding MRI: basic MR physics for physicians," *Postgrad Med J*, vol. 89, no. 1050, pp. 209–223, Apr. 2013, doi: 10.1136/postgradmedj-2012-131342.

[50] H. H. Schild, *MRI made easy (... well almost)*. Berlin/Bergkamen: Schering AG, 1990. [Online]. Available: https://rads.web.unc.edu/wp-content/uploads/sites/12234/2018/05/Phy-MRI-Made-Easy.pdf

[51] C. R. Jack *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *J Magn Reson Imaging*, vol. 27, no. 4, pp. 685–691, Apr. 2008, doi: 10.1002/jmri.21049.

[52] M. Symms, H. R. Jäger, K. Schmierer, and T. A. Yousry, "A review of structural magnetic resonance neuroimaging," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 9, pp. 1235–1244, Sep. 2004, doi: 10.1136/jnnp.2003.032714.

[53] S. Duchesne, Y. Rolland, and M. Vérin, "Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI," *Academic Radiology*, vol. 16, no. 1, pp. 61–70, Jan. 2009, doi: 10.1016/j.acra.2008.05.024.

[54] H.-J. Huppertz *et al.*, "Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification," *Movement Disorders*, vol. 31, no. 10, pp. 1506–1517, 2016, doi: 10.1002/mds.26715.

[55] M. Hutchinson and U. Raff, "Parkinson's disease: a novel MRI method for determining structural changes in the substantia nigra," *J Neurol Neurosurg Psychiatry*, vol. 67, no. 6, pp. 815–818, Dec. 1999, doi: 10.1136/jnnp.67.6.815.

[56] J. S. Meyer, J. Huang, and M. H. Chowdhury, "MRI confirms mild cognitive impairments prodromal for Alzheimer's, vascular and Parkinson-Lewy body dementias," *Journal of the Neurological Sciences*, vol. 257, no. 1, pp. 97–104, Jun. 2007, doi: 10.1016/j.jns.2007.01.016.

[57] "Center for functional MRI, Department of Radiology, UCSD." [Online]. Available: http://fmri.ucsd.edu/Howto/3T/structure.html

[58] O. Sporns, G. Tononi, and R. Kötter, "The Human Connectome: A Structural Description of the Human Brain," *PLoS Comp Biol*, vol. 1, no. 4, p. e42, 2005, doi: 10.1371/journal.pcbi.0010042.

[59] K. J. Friston, "Functional and Effective Connectivity: A Review," vol. 1, no. 1, pp. 13–36, Jun. 2011, doi: 10.1089/brain.2011.0008.

[60] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg, "Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia," *Journal of Neuroscience*, vol. 28, no. 37, pp. 9239–9248, Sep. 2008, doi: 10.1523/JNEUROSCI.1929-08.2008.

[61] K. Wu *et al.*, "The Overlapping Community Structure of Structural Brain Network in Young Healthy Individuals," *PLoS ONE*, vol. 6, no. 5, p. e19608, May 2011, doi: 10.1371/journal.pone.0019608.

[62] J. Liu *et al.*, "Complex Brain Network Analysis and Its Applications to Brain Disorders: A Survey," *Complexity*, vol. 2017, pp. 1–27, 2017, doi: 10.1155/2017/8362741.

[63] N. Tzourio-Mazoyer *et al.*, "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject

Brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002, doi: 10.1006/nimg.2001.0978.

[64] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *Neuroimage*, vol. 53, no. 1, pp. 1–15, Oct. 2010, doi: 10.1016/j.neuroimage.2010.06.010.

[65] L. Fan *et al.*, "The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture," *Cerebral cortex (New York, N.Y. : 1991)*, vol. 26, no. 8, pp. 3508–3526, Aug. 2016, doi: 10.1093/cercor/bhw157.

[66] K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig : Barth, 1909. [Online]. Available: http://archive.org/details/b28062449

[67] D. W. Shattuck *et al.*, "Construction of a 3D Probabilistic Atlas of Human Cortical Structures," *Neuroimage*, vol. 39, no. 3, pp. 1064–1080, Feb. 2008, doi: 10.1016/j.neuroimage.2007.09.031.

[68] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, Jul. 2006, doi: 10.1016/j.neuroimage.2006.01.021.

[69] N. Makris *et al.*, "Decreased volume of left and total anterior insular lobule in schizophrenia," *Schizophrenia Research*, vol. 83, no. 2, pp. 155–171, Apr. 2006, doi: 10.1016/j.schres.2005.11.020.

[70] J. A. Frazier *et al.*, "Structural Brain Magnetic Resonance Imaging of Limbic and Thalamic Volumes in Pediatric Bipolar Disorder," *AJP*, vol. 162, no. 7, pp. 1256–1265, Jul. 2005, doi: 10.1176/appi.ajp.162.7.1256.

[71] J. M. Goldstein *et al.*, "Hypothalamic Abnormalities in Schizophrenia: Sex Effects and Genetic Vulnerability," *Biological Psychiatry*, vol. 61, no. 8, pp. 935–945, Apr. 2007, doi: 10.1016/j.biopsych.2006.06.027.

[72] M. F. Glasser *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, Aug. 2016, doi: 10.1038/nature18933.

[73] G. Auzias, O. Coulon, and A. Brovelli, "MarsAtlas: A cortical parcellation atlas for functional mapping," *Hum Brain Mapp*, vol. 37, no. 4, pp. 1573–1592, Jan. 2016, doi: 10.1002/hbm.23121.

[74] A. Klein and J. Tourville, "101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol," *Front. Neurosci.*, vol. 6, 2012, doi: 10.3389/fnins.2012.00171.

[75] B. Holla *et al.*, "A series of five population-specific Indian brain templates and atlases spanning ages 6–60 years," *Human Brain Mapping*, vol. 41, no. 18, pp. 5164–5175, 2020, doi: https://doi.org/10.1002/hbm.25182.

[76] J. Wang *et al.*, "Parcellation-dependent small-world brain functional networks: a resting-state fMRI study," *Hum Brain Mapp*, vol. 30, no. 5, pp. 1511–1523, May 2009, doi: 10.1002/hbm.20623.

[77] G. Sanabria-Diaz *et al.*, "Surface area and cortical thickness descriptors reveal different attributes of the structural human brain networks," *NeuroImage*, vol. 50, no. 4, pp. 1497–1510, May 2010, doi: 10.1016/j.neuroimage.2010.01.028.

[78] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nat Rev Neurosci*, vol. 10, no. 3, pp. 186–198, Mar. 2009, doi: 10.1038/nrn2575.

[79] A. Alexander-Bloch, J. N. Giedd, and E. Bullmore, "Imaging structural co-variance between human brain regions," *Nat Rev Neurosci*, vol. 14, no. 5, pp. 322–336, May 2013, doi: 10.1038/nrn3465.

[80] Y. Zhang *et al.*, "Abnormal topological organization of structural brain networks in schizophrenia," *Schizophrenia Research*, vol. 141, no. 2–3, pp. 109–118, Nov. 2012, doi: 10.1016/j.schres.2012.08.021.

[81] Z. Dai and Y. He, "Disrupted structural and functional brain connectomes in mild cognitive impairment and Alzheimer's disease," *Neurosci. Bull.*, vol. 30, no. 2, pp. 217–232, Apr. 2014, doi: 10.1007/s12264-013-1421-0.

[82] T. Samantaray, J. Saini, and C. N. Gupta, "Sparsity Dependent Metrics Depict Alteration of Brain Network Connectivity in Parkinson's Disease," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2022, pp. 698–701. doi: 10.1109/EMBC48229.2022.9871258.

[83] N. Amoroso, M. La Rocca, A. Monaco, R. Bellotti, and S. Tangaro, "Complex networks reveal early MRI markers of Parkinson's disease," *Medical Image Analysis*, vol. 48, pp. 12–24, Aug. 2018, doi: 10.1016/j.media.2018.05.004.

[84] J. B. Pereira *et al.*, "Aberrant cerebral network topology and mild cognitive impairment in early Parkinson's disease: Aberrant Brain Network Topology in Early PD," *Hum. Brain Mapp.*, vol. 36, no. 8, pp. 2980–2995, Aug. 2015, doi: 10.1002/hbm.22822.

[85] S. K. Yadav *et al.*, "Gender-based analysis of cortical thickness and structural connectivity in Parkinson's disease," *J Neurol*, vol. 263, no. 11, pp. 2308–2318, Nov. 2016, doi: 10.1007/s00415-016-8265-2.

[86] D. Zhang, J. Wang, X. Liu, J. Chen, and B. Liu, "Aberrant Brain Network Efficiency in Parkinson's Disease Patients with Tremor: A Multi-Modality Study," *Front. Aging Neurosci.*, vol. 7, Aug. 2015, doi: 10.3389/fnagi.2015.00169.

[87] C. Li *et al.*, "Impaired topological architecture of brain structural networks in idiopathic Parkinson's disease: a DTI study," *Brain Imaging and Behavior*, vol. 11, no. 1, pp. 113–128, Feb. 2017, doi: 10.1007/s11682-015-9501-6.

[88] S. Nigro *et al.*, "Characterizing structural neural networks in de novo Parkinson disease patients using diffusion tensor imaging," *Human Brain Mapping*, vol. 37, no. 12, pp. 4500–4510, 2016, doi: 10.1002/hbm.23324.

[89] M. Ingalhalikar *et al.*, "Sex differences in the structural connectome of the human brain," *Proceedings of the National Academy of Sciences*, vol. 111, no. 2, pp. 823–828, Jan. 2014, doi: 10.1073/pnas.1316909110.

[90] "Parkinson's Progressive Markers Initiative (PPMI)." Accessed: Jun. 18, 2022. [Online]. Available: www.ppmi-info.org/access-data-specimens/download-data

[91] "Parkinson's Disease Biomarkers Program (PDBP)." Accessed: Aug. 04, 2022. [Online]. Available: https://pdbp.ninds.nih.gov

[92] "Dementias Platform UK." Accessed: Aug. 04, 2022. [Online]. Available: https://portal.dementiasplatform.uk

[93] L. Badea, M. Onu, T. Wu, A. Roceanu, and O. Bajenaru, "Exploring the reproducibility of functional connectivity alterations in Parkinson's disease," *PLOS ONE*, vol. 12, no. 11, p. e0188196, Nov. 2017, doi: 10.1371/journal.pone.0188196.

[94] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010, doi: 10.1016/j.neuroimage.2009.10.003.

[95] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, and Y. He, "GRETNA: a graph theoretical network analysis toolbox for imaging connectomics," *Front. Hum. Neurosci.*, vol. 9, 2015, doi: 10.3389/fnhum.2015.00386.

[96] S. M. H. Hosseini, F. Hoeft, and S. R. Kesler, "GAT: A Graph-Theoretical Analysis Toolbox for Analyzing Between-Group Differences in Large-Scale Structural and Functional Brain Networks," *PLoS ONE*, vol. 7, no. 7, p. e40709, Jul. 2012, doi: 10.1371/journal.pone.0040709.

[97] A. S. Ribeiro, L. M. Lacerda, and H. A. Ferreira, "Multimodal Imaging Brain Connectivity Analysis (MIBCA) toolbox," *PeerJ*, vol. 3, Jul. 2015, doi: 10.7717/peerj.1078.

[98] M. Mijalkov, E. Kakaei, J. B. Pereira, E. Westman, and G. Volpe, "BRAPH: A graph theory software for the analysis of brain connectivity," *PLoS One*, vol. 12, no. 8, Aug. 2017, doi: 10.1371/journal.pone.0178798.

[99] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," *NeuroImage*, vol. 53, no. 4, pp. 1197–1207, Dec. 2010, doi: 10.1016/j.neuroimage.2010.06.041.

[100] M. Xia, J. Wang, and Y. He, "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLoS ONE*, vol. 8, no. 7, Jul. 2013, doi: https://doi.org/10.1371/journal.pone.0068910.

[101] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūtė, C. Suo, and A. Fornito, "Consistency and differences between centrality measures across distinct classes of

networks," *PLOS ONE*, vol. 14, no. 7, p. e0220061, Jul. 2019, doi: 10.1371/journal.pone.0220061.

[102] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, Art. no. 6684, Jun. 1998, doi: 10.1038/30918.

[103] M. M. Wickremaratchi *et al.*, "The motor phenotype of Parkinson's disease in relation to age at onset," *Movement Disorders*, vol. 26, no. 3, pp. 457–463, 2011, doi: 10.1002/mds.23469.

[104] A. K. Srivastava, K. Jeberson, and W. Jeberson, "A systematic review on Data Mining Application in Parkinson's disease," *Neuroscience Informatics*, vol. 2, no. 4, p. 100064, Dec. 2022, doi: https://doi.org/10.1016/j.neuri.2022.100064.

[105] R. M. Hendricks and M. T. Khasawneh, "A Systematic Review of Parkinson's Disease Cluster Analysis Research," *Aging Dis*, vol. 12, no. 7, pp. 1567–1586, Oct. 2021, doi: 10.14336/AD.2021.0519.

[106] S. Lewis, T. Foltynie, A. Blackwell, T. Robbins, A. Owen, and R. Barker, "Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach," *J Neurol Neurosurg Psychiatry*, vol. 76, no. 3, pp. 343–348, Mar. 2005, doi: 10.1136/jnnp.2003.033530.

[107] B. Post, J. D. Speelman, R. J. de Haan, and on behalf of the CARPA-study group, "Clinical heterogeneity in newly diagnosed Parkinson's disease," *J Neurol*, vol. 255, no. 5, pp. 716–722, May 2008, doi: 10.1007/s00415-008-0782-1.

[108] J. S. A. M. Reijnders, U. Ehrt, R. Lousberg, D. Aarsland, and A. F. G. Leentjens, "The association between motor subtypes and psychopathology in Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 15, no. 5, pp. 379–382, Jun. 2009, doi: 10.1016/j.parkreldis.2008.09.003.

[109] S.-M. Fereshtehnejad, Y. Zeighami, A. Dagher, and R. B. Postuma, "Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression," *Brain*, vol. 140, no. 7, pp. 1959–1976, Jul. 2017, doi: 10.1093/brain/awx118.

[110] J. Mu, K. R. Chaudhuri, C. Bielza, J. de Pedro-Cuesta, P. Larrañaga, and P. Martinez-Martin, "Parkinson's Disease Subtypes Identified from Cluster Analysis of Motor and Non-motor Symptoms," *Front Aging Neurosci*, vol. 9, p. 301, Sep. 2017, doi: 10.3389/fnagi.2017.00301.

[111] T. Samantaray, J. Saini, and C. N. Gupta, "Meta-Analysis of Clinical Symptoms and Data Driven Subtyping Approaches in Parkinson's Disease," The Brain Conference, 2021, Mar. 04, 2021. Accessed: Oct. 19, 2022. [Online]. Available: https://thebrainconference.co.uk/wpcontent/uploads/2021/02/R2_51_Samantaray_Tanmayee_MovementDisorders_51.png

[112] D. Berg *et al.*, "Time to redefine PD? Introductory statement of the MDS Task Force on the definition of Parkinson's disease," *Mov Disord*, vol. 29, no. 4, pp. 454–462, Apr. 2014, doi: 10.1002/mds.25844.

[113] B.-A. Sieber *et al.*, "Prioritized research recommendations from the National Institute of Neurological Disorders and Stroke Parkinson's Disease 2014 conference," *Annals of Neurology*, vol. 76, no. 4, pp. 469–472, 2014, doi: https://doi.org/10.1002/ana.24261.

[114] J. S. Perlmutter, "Assessment of Parkinson disease manifestations," *Curr Protoc Neurosci*, vol. Chapter 10, p. Unit10.1-Unit10.1, Oct. 2009, doi: 10.1002/0471142301.ns1001s49.

[115] J. Pitton Rissardo and A. Fornari Caprara, "Parkinson's disease rating scales: a literature review," *Ann Mov Disord*, vol. 3, no. 1, p. 3, 2020, doi: 10.4103/AOMD.AOMD_33_19.

[116] C. Marras and K. R. Chaudhuri, "Nonmotor features of Parkinson's disease subtypes," *Movement Disorders*, vol. 31, no. 8, pp. 1095–1102, 2016, doi: https://doi.org/10.1002/mds.26510.

[117] A. Sauerbier, P. Jenner, A. Todorova, and K. R. Chaudhuri, "Non motor subtypes and Parkinson's disease," *Parkinsonism & Related Disorders*, vol. 22, pp. S41–S46, Jan. 2016, doi: 10.1016/j.parkreldis.2015.09.027.

[118] A. H. V. Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," *Nature Reviews Neuroscience*, vol. 18, no. 7, Art. no. 7, Jul. 2017, doi: 10.1038/nrn.2017.62.

[119] J. C. Greenland, C. H. Williams-Gray, and R. A. Barker, "The clinical heterogeneity of Parkinson's disease and its therapeutic implications," *European Journal of Neuroscience*, vol. 49, no. 3, pp. 328–338, 2019, doi: 10.1111/ejn.14094.

[120] S. M. van Rooden, W. J. Heiser, J. N. Kok, D. Verbaan, J. J. van Hilten, and J. Marinus, "The identification of Parkinson's disease subtypes using cluster analysis: A systematic review," *Movement Disorders*, vol. 25, no. 8, pp. 969–978, Jun. 2010, doi: 10.1002/mds.23116.

[121] G. Pagano, N. Ferrara, D. J. Brooks, and N. Pavese, "Age at onset and Parkinson disease phenotype," *Neurology*, vol. 86, no. 15, pp. 1400–1407, Apr. 2016, doi: 10.1212/WNL.0000000000002461.

[122] M. M. Hoehn and M. D. Yahr, "Parkinsonism: onset, progression and mortality," *Neurology*, vol. 17, no. 5, pp. 427–442, May 1967, doi: 10.1212/wnl.17.5.427.

[123] S. M. van Rooden *et al.*, "Clinical subtypes of Parkinson's disease," *Movement Disorders*, vol. 26, no. 1, pp. 51–58, 2011, doi: https://doi.org/10.1002/mds.23346.

[124] E. Gasparoli *et al.*, "Clinical predictors in Parkinson's disease," *Neurol Sci*, vol. 23, no. 2, pp. s77–s78, Sep. 2002, doi: 10.1007/s100720200078.

[125] R. Erro *et al.*, "The Heterogeneity of Early Parkinson's Disease: A Cluster Analysis on Newly Diagnosed Untreated Patients," *PLoS One*, vol. 8, no. 8, p. e70244, Aug. 2013, doi: 10.1371/journal.pone.0070244.

[126] M. Lawton *et al.*, "Parkinson's Disease Subtypes in the Oxford Parkinson Disease Centre (OPDC) Discovery Cohort," *J Parkinsons Dis*, vol. 5, no. 2, pp. 269–279, 2015, doi: 10.3233/JPD-140523.

[127] M. Lawton *et al.*, "Developing and validating Parkinson's disease subtypes and their motor and cognitive progression," *J Neurol Neurosurg Psychiatry*, vol. 89, no. 12, pp. 1279–1287, Dec. 2018, doi: 10.1136/jnnp-2018-318337.

[128] L.-Y. Ma, P. Chan, Z.-Q. Gu, F.-F. Li, and T. Feng, "Heterogeneity among patients with Parkinson's disease: cluster analysis and genetic association," *J Neurol Sci*, vol. 351, no. 1–2, pp. 41–45, Apr. 2015, doi: 10.1016/j.jns.2015.02.029.

[129] T. D. van Balkom *et al.*, "Profiling cognitive and neuropsychiatric heterogeneity in Parkinson's disease," *Parkinsonism Relat Disord*, vol. 28, pp. 130–136, Jul. 2016, doi: 10.1016/j.parkreldis.2016.05.014.

[130] D. Belvisi *et al.*, "The Pathophysiological Correlates of Parkinson's Disease Clinical Subtypes," *Mov Disord*, vol. 36, no. 2, pp. 370–379, Feb. 2021, doi: 10.1002/mds.28321.

[131] S.-M. Fereshtehnejad, S. R. Romenets, J. B. M. Anang, V. Latreille, J.-F. Gagnon, and R. B. Postuma, "New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes," *JAMA Neurol*, vol. 72, no. 8, pp. 863–873, Aug. 2015, doi: 10.1001/jamaneurol.2015.0703.

[132] K.-C. Wong, "A Short Survey on Data Clustering Algorithms," in *2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI)*, Nov. 2015, pp. 64–68. doi: 10.1109/ISCMI.2015.10.

[133] C. Pont-Sunyer *et al.*, "The Onset of Nonmotor Symptoms in Parkinson's disease (The ONSET PD Study)," *Movement Disorders*, vol. 30, no. 2, pp. 229–237, 2015, doi: https://doi.org/10.1002/mds.26077.

[134] P. Liu, T. Feng, Y. Wang, X. Zhang, and B. Chen, "Clinical heterogeneity in patients with early-stage Parkinson's disease: a cluster analysis," *J Zhejiang Univ Sci B*, vol. 12, no. 9, pp. 694–703, Sep. 2011, doi: 10.1631/jzus.B1100069.

[135] R. Erro *et al.*, "Clinical clusters and dopaminergic dysfunction in de-novo Parkinson disease," *Parkinsonism & Related Disorders*, vol. 28, pp. 137–140, Jul. 2016, doi: 10.1016/j.parkreldis.2016.04.026.

[136] J. M. Graham and H. J. Sagar, "A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes," *Mov Disord*, vol. 14, no. 1, pp. 10–20, Jan. 1999, doi: 10.1002/1531-8257(199901)14:1<10::aid-mds1005>3.0.co;2-4.

[137] H. Xiong, J. Wu, and J. Chen, "K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 318–331, Apr. 2009, doi: 10.1109/TSMCB.2008.2004559.

[138] F. Rodriguez-Sanchez *et al.*, "Identifying Parkinson's disease subtypes with motor and non-motor symptoms via model-based multi-partition clustering," *Sci Rep*, vol. 11, no. 1, Art. no. 1, Dec. 2021, doi: 10.1038/s41598-021-03118-w.

[139] G. M. L. Eglit *et al.*, "Delineation of Apathy Subgroups in Parkinson's Disease: Differences in Clinical Presentation, Functional Ability, Health-related Quality of Life, and Caregiver Burden," *Movement Disorders Clinical Practice*, vol. 8, no. 1, pp. 92–99, 2021, doi: 10.1002/mdc3.13127.

[140] M. Dąbrowska *et al.*, "The role of neuroimaging in the diagnosis of the atypical parkinsonian syndromes in clinical practice," *Neurologia i Neurochirurgia Polska*, vol. 49, no. 6, Art. no. 6, 2015, doi: 10.1016/j.pjnns.2015.10.002.

[141] A. Hotter, R. Esterhammer, M. F. H. Schocke, and K. Seppi, "Potential of advanced MR imaging techniques in the differential diagnosis of parkinsonism," *Movement Disorders*, vol. 24, no. S2, pp. S711–S720, 2009, doi: 10.1002/mds.22648.

[142] L. A. Massey *et al.*, "Conventional magnetic resonance imaging in confirmed progressive supranuclear palsy and multiple system atrophy," *Movement Disorders*, vol. 27, no. 14, pp. 1754–1762, 2012, doi: 10.1002/mds.24968.

[143] Y. Ya *et al.*, "Machine Learning Models for Diagnosis of Parkinson's Disease Using Multiple Structural Magnetic Resonance Imaging Features," *Frontiers in Aging Neuroscience*, vol. 14, 2022, doi: https://doi.org/10.3389/fnagi.2022.808520.

[144] J. Ashburner and K. J. Friston, "Why Voxel-Based Morphometry Should Be Used," *NeuroImage*, vol. 14, no. 6, pp. 1238–1243, Dec. 2001, doi: 10.1006/nimg.2001.0961.

[145] X. Xu, Q. Han, J. Lin, L. Wang, F. Wu, and H. Shang, "Grey matter abnormalities in Parkinson's disease: a voxel-wise meta-analysis," *European Journal of Neurology*, vol. 27, no. 4, pp. 653–659, 2020, doi: 10.1111/ene.14132.

[146] M. Minnerop *et al.*, "Voxel-based morphometry and voxel-based relaxometry in multiple system atrophy—A comparison between clinical subtypes and correlations with clinical parameters," *J Neurol Neurosurg Psychiatry*, vol. 36, no. 4, pp. 1086–1095, Jul. 2007, doi: https://doi.org/10.1016/j.neuroimage.2007.04.028.

[147] A. Padovani *et al.*, "Diffusion tensor imaging and voxel based morphometry study in early progressive supranuclear palsy," *J. Neurol. Neurosurg. Psychiatry*, vol. 77, no. 4, pp. 457–463, Apr. 2006, doi: 10.1136/jnnp.2005.075713.

[148] S. E. Lee *et al.*, "Clinicopathological correlations in corticobasal degeneration," *Ann. Neurol.*, vol. 70, no. 2, pp. 327–340, Aug. 2011, doi: 10.1002/ana.22424.

[149] E. J. Burton *et al.*, "Patterns of cerebral atrophy in dementia with Lewy bodies using voxel-based morphometry," *Neuroimage*, vol. 17, no. 2, pp. 618–630, Oct. 2002, doi: PMID: 12377138.

[150] F. Lauretani, Y. Longobucco, G. Ravazzoni, E. Gallini, M. Salvi, and M. Maggio, "Imaging the Functional Neuroanatomy of Parkinson's Disease: Clinical Applications and Future Directions," *International Journal of Environmental*

Research and Public Health, vol. 18, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/ijerph18052356.

[151] S. Nigro *et al.*, "Magnetic Resonance Parkinsonism Index: diagnostic accuracy of a fully automated algorithm in comparison with the manual measurement in a large Italian multicentre study in patients with progressive supranuclear palsy," *Eur Radiol*, vol. 27, no. 6, pp. 2665–2675, Jun. 2017, doi: 10.1007/s00330-016-4622-x.

[152] H. Braak, E. Ghebremedhin, U. Rüb, H. Bratzke, and K. Del Tredici, "Stages in the development of Parkinson's disease-related pathology," *Cell Tissue Res.*, vol. 318, no. 1, pp. 121–134, Oct. 2004, doi: 10.1007/s00441-004-0956-9.

[153] P.-L. Chiang *et al.*, "White matter damage and systemic inflammation in Parkinson's disease," *BMC Neuroscience*, vol. 18, no. 1, p. 48, Jun. 2017, doi: 10.1186/s12868-017-0367-y.

[154] K. Kamagata *et al.*, "Relationship between cognitive impairment and white-matter alteration in Parkinson's disease with dementia: tract-based spatial statistics and tract-specific analysis," *Eur Radiol*, vol. 23, no. 7, pp. 1946–1955, Jul. 2013, doi: 10.1007/s00330-013-2775-4.

[155] A. Radziunas *et al.*, "Brain MRI morphometric analysis in Parkinson's disease patients with sleep disturbances," *BMC Neurol*, vol. 18, no. 1, p. 88, Dec. 2018, doi: 10.1186/s12883-018-1092-6.

[156] H. Watanabe *et al.*, "Cortical and subcortical brain atrophy in Parkinson's disease with visual hallucination," *Movement Disorders*, vol. 28, no. 12, pp. 1732–1736, 2013, doi: 10.1002/mds.25641.

[157] A. Sharma, S. Kaur, N. Memon, A. Jainul Fathima, S. Ray, and M. W. Bhatt, "Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis," *Neuroscience Informatics*, vol. 1, no. 3, p. 100012, Nov. 2021, doi: 10.1016/j.neuri.2021.100012.

[158] P. Chagué *et al.*, "Radiological classification of dementia from anatomical MRI assisted by machine learning-derived maps," *Journal of Neuroradiology*, vol. 48, no. 6, pp. 412–418, Nov. 2021, doi: 10.1016/j.neurad.2020.04.004.

[159] X. Li *et al.*, "Presurgical differentiation between malignant haemangiopericytoma and angiomatous meningioma by a radiomics approach based on texture analysis," *Journal of Neuroradiology*, vol. 46, no. 5, pp. 281–287, Sep. 2019, doi: 10.1016/j.neurad.2019.05.013.

[160] C. Park, P. H. Lee, S.-K. Lee, S. J. Chung, and N.-Y. Shin, "The diagnostic potential of multimodal neuroimaging measures in Parkinson's disease and atypical parkinsonism," *Brain and Behavior*, p. e01808, 2020, doi: https://doi.org/10.1002/brb3.1808.

[161] B. Rana *et al.*, "Relevant 3D local binary pattern based features from fused feature descriptor for differential diagnosis of Parkinson's disease using structural MRI,"

Biomedical Signal Processing and Control, vol. 34, pp. 134–143, Apr. 2017, doi: 10.1016/j.bspc.2017.01.007.

[162] C. Salvatore *et al.*, "Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy," *Journal of Neuroscience Methods*, vol. 222, pp. 230–237, Jan. 2014, doi: 10.1016/j.jneumeth.2013.11.016.

[163] N. K. Focke *et al.*, "Individual voxel-based subtype prediction can differentiate progressive supranuclear palsy from idiopathic Parkinson syndrome and healthy controls," *Hum Brain Mapp*, vol. 32, no. 11, pp. 1905–1915, Nov. 2011, doi: 10.1002/hbm.21161.

[164] S. Duchesne, Y. Rolland, and M. Vérin, "Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI," *Academic Radiology*, vol. 16, no. 1, pp. 61–70, Jan. 2009, doi: 10.1016/j.acra.2008.05.024.

[165] A. Bhan, S. Kapoor, M. Gulati, and A. Goyal, "Early Diagnosis of Parkinson's Disease in brain MRI using Deep Learning Algorithm," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Feb. 2021, pp. 1467–1470. doi: 10.1109/ICICV50876.2021.9388571.

[166] J. Mei, C. Tremblay, N. Stikov, D. Desrosiers, and J. Frasnelli, "Differentiation of Parkinson's disease and non-Parkinsonian olfactory dysfunction with structural MRI data," *Medical Imaging 2021: Computer-Aided Diagnosis*, vol. 115971E, no. 15 February 2021, Feb. 2021, doi: https://doi.org/10.1117/12.2581233.

[167] S. Chakraborty, S. Aich, and H.-C. Kim, "Detection of Parkinson's Disease from 3T T1 Weighted MRI Scans Using 3D Convolutional Neural Network," *Diagnostics*, vol. 10, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/diagnostics10060402.

[168] B. Gong *et al.*, "Neuroimaging-based diagnosis of Parkinson's disease with deep neural mapping large margin distribution machine," *Neurocomputing*, vol. 320, pp. 141–149, Dec. 2018, doi: 10.1016/j.neucom.2018.09.025.

[169] S. Shinde *et al.*, "Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI," *NeuroImage: Clinical*, vol. 22, p. 101748, Jan. 2019, doi: 10.1016/j.nicl.2019.101748.

[170] G. Solana-Lavalle and R. Rosas-Romero, "Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 198, p. 105793, Jan. 2021, doi: 10.1016/j.cmpb.2020.105793.

[171] N. M. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun, "Canonical Correlation Analysis for Data Fusion and Group Inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010, doi: 10.1109/MSP.2010.936725.

[172] C. L. Grady, M. V. Springer, D. Hongwanishkul, A. R. McIntosh, and G. Winocur, "Age-related changes in brain activity across the adult lifespan," *Journal of*

Cognitive Neuroscience, vol. 18, no. 2, pp. 227–241, 2006, doi: 10.1162/jocn.2006.18.2.227.

[173] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway Partial Least Squares," *NeuroImage*, vol. 22, no. 3, pp. 1023–1034, Jul. 2004, doi: 10.1016/j.neuroimage.2004.03.038.

[174] V. D. Calhoun, T. Adali, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson, "A method for multitask fMRI data fusion applied to schizophrenia," vol. 27, pp. 598–610, 2006, doi: 10.1002/hbm.20204.

[175] X. Guo, Y. Han, K. Chen, Y. Wang, and L. Yao, "Mapping joint grey and white matter reductions in Alzheimer's disease using joint independent component analysis," *Neurosci Lett*, vol. 531, no. 2, p. 10.1016/j.neulet.2012.10.038, Dec. 2012, doi: 10.1016/j.neulet.2012.10.038.

[176] L. Xu, G. Pearlson, and V. D. Calhoun, "Joint source based morphometry identifies linked gray and white matter group differences," *NeuroImage*, vol. 44, no. 3, pp. 777–789, Feb. 2009, doi: 10.1016/j.neuroimage.2008.09.051.

[177] A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich, "Linked independent component analysis for multimodal data fusion," *NeuroImage*, vol. 54, no. 3, pp. 2198–2217, Feb. 2011, doi: 10.1016/j.neuroimage.2010.09.073.

[178] J. Liu, G. Pearlson, A. Windemuth, G. Ruano, N. I. Perrone-Bizzozero, and V. Calhoun, "Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA," *Human Brain Mapping*, vol. 30, no. 1, pp. 241–255, 2009, doi: 10.1002/hbm.20508.

[179] J. Sui, T. Adali, G. D. Pearlson, and V. D. Calhoun, "An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques," *NeuroImage*, vol. 46, no. 1, pp. 73–86, May 2009, doi: 10.1016/j.neuroimage.2009.01.026.

[180] J. Sui *et al.*, "Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model," *NeuroImage*, vol. 57, no. 3, pp. 839–855, Aug. 2011, doi: 10.1016/j.neuroimage.2011.05.055.

[181] E. Damaraju, R. F. Silva, T. Adali, and V. D. Calhoun, "A multimodal IVA fusion approach to identify linked neuroimaging markers," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico: IEEE, Nov. 2021, pp. 3928–3932. doi: 10.1109/EMBC46164.2021.9631027.

[182] R. F. Silva, S. M. Plis, T. Adali, M. S. Pattichis, and V. D. Calhoun, "Multidataset Independent Subspace Analysis With Application to Multimodal Fusion," *IEEE Trans. on Image Process.*, vol. 30, pp. 588–602, 2021, doi: 10.1109/TIP.2020.3028452.

[183] H. Baagil *et al.*, "Neural correlates of impulse control behaviors in Parkinson's disease: Analysis of multimodal imaging data," *NeuroImage: Clinical*, vol. 37, p. 103315, 2023, doi: 10.1016/j.nicl.2023.103315.

[184] A. Shoeibi *et al.*, "Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: A review," *Information Fusion*, vol. 93, pp. 85–117, May 2023, doi: 10.1016/j.inffus.2022.12.010.

[185] T. Zhou, S. Ruan, Y. Guo, and S. Canu, "A Multi-Modality Fusion Network Based on Attention Mechanism for Brain Tumor Segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA: IEEE, Apr. 2020, pp. 377–380. doi: 10.1109/ISBI45749.2020.9098392.

[186] L. Liu, S. Chen, F. Zhang, F.-X. Wu, Y. Pan, and J. Wang, "Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI," *Neural Comput & Applic*, vol. 32, no. 11, pp. 6545–6558, Jun. 2020, doi: 10.1007/s00521-019-04096-x.

[187] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 11384, A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. Van Walsum, Eds., in Lecture Notes in Computer Science, vol. 11384. , Cham: Springer International Publishing, 2019, pp. 37–48. doi: 10.1007/978-3-030-11726-9_4.

[188] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, "Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network," *Journal of Healthcare Engineering*, vol. 2018, pp. 1–14, 2018, doi: 10.1155/2018/4940593.

[189] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015, doi: 10.1016/j.neuroimage.2014.12.061.

[190] B. Chen *et al.*, "Automatic classification of MSA subtypes using Whole-brain gray matter function and Structure-Based radiomics approach," *European Journal of Radiology*, vol. 161, p. 110735, Apr. 2023, doi: 10.1016/j.ejrad.2023.110735.

[191] X. Zhuang, Z. Yang, and D. Cordes, "A technical review of canonical correlation analysis for neuroscience applications," *Hum Brain Mapp*, vol. 41, no. 13, pp. 3807–3833, Sep. 2020, doi: 10.1002/hbm.25090.

[192] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical Correlation Analysis: An Overview with Application to Learning Methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004, doi: 10.1162/0899766042321814.

[193] X. Zhuang, Z. Yang, T. Curran, R. Byrd, R. Nandy, and D. Cordes, "A family of locally constrained CCA models for detecting activation patterns in fMRI," *NeuroImage*, vol. 149, pp. 63–84, Apr. 2017, doi: 10.1016/j.neuroimage.2016.12.081.

[194] N. M. Correa, T. Eichele, T. Adalı, Y.-O. Li, and V. D. Calhoun, "Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI," *NeuroImage*, vol. 50, no. 4, pp. 1438–1445, May 2010, doi: 10.1016/j.neuroimage.2010.01.062.

[195] L. Liang *et al.*, "Fusion analysis of gray matter and white matter in subjective cognitive decline and mild cognitive impairment by multimodal CCA-joint ICA," *NeuroImage: Clinical*, vol. 32, p. 102874, Jan. 2021, doi: 10.1016/j.nicl.2021.102874.

[196] R. C. Wolf *et al.*, "A Neural Signature of Parkinsonism in Patients With Schizophrenia Spectrum Disorders: A Multimodal MRI Study Using Parallel ICA," *Schizophrenia Bulletin*, vol. 46, no. 4, pp. 999–1008, Jul. 2020, doi: 10.1093/schbul/sbaa007.

[197] Y. He, Z. J. Chen, and A. C. Evans, "Small-World Anatomical Networks in the Human Brain Revealed by Cortical Thickness from MRI," *Cerebral Cortex*, vol. 17, no. 10, pp. 2407–2419, Oct. 2007, doi: 10.1093/cercor/bhl149.

[198] A. Cronin-Golomb, "Parkinson's Disease as a Disconnection Syndrome," *Neuropsychol Rev*, vol. 20, no. 2, pp. 191–208, Jun. 2010, doi: 10.1007/s11065-010-9128-8.

[199] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Greicius, "Neurodegenerative diseases target large-scale human brain networks," *Neuron*, vol. 62, no. 1, pp. 42–52, Apr. 2009, doi: 10.1016/j.neuron.2009.03.024.

[200] T. Samantaray, J. Saini, and C. N. Gupta, "Sparsity dependent metrics depict alteration of brain network connectivity in Parkinson's disease," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Accepted.

[201] X. Xu *et al.*, "Brain Atrophy and Reorganization of Structural Network in Parkinson's Disease With Hemiparkinsonism," *Frontiers in Human Neuroscience*, vol. 12, p. 117, 2018, doi: 10.3389/fnhum.2018.00117.

[202] Q. Wu, Y. Gao, A.-S. Liu, L.-Z. Xie, L. Qian, and X.-G. Yang, "Large-scale cortical volume correlation networks reveal disrupted small world patterns in Parkinson's disease," *Neuroscience Letters*, vol. 662, pp. 374–380, Jan. 2018, doi: 10.1016/j.neulet.2017.10.032.

[203] J. Arnedo *et al.*, "Decomposition of brain diffusion imaging data uncovers latent schizophrenias with distinct patterns of white matter anisotropy," *NeuroImage*, vol. 120, pp. 43–54, Oct. 2015, doi: 10.1016/j.neuroimage.2015.06.083.

[204] C. N. Gupta *et al.*, "Biclustered Independent Component Analysis for Complex Biomarker and Subtype Identification from Structural Magnetic Resonance Images in Schizophrenia," *Front Psychiatry*, vol. 8, Sep. 2017, doi: 10.3389/fpsyt.2017.00179.

[205] M. A. Rahaman *et al.*, "N-BiC: A Method for Multi-Component and Symptom Biclustering of Structural MRI Data: Application to Schizophrenia," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 110–121, Jan. 2020, doi: 10.1109/TBME.2019.2908815.

[206] V. D. Calhoun, T. Adali, N. R. Giuliani, J. J. Pekar, K. A. Kiehl, and G. D. Pearlson, "Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data," *Hum Brain Mapp*, vol. 27, no. 1, pp. 47–62, Jan. 2006, doi: 10.1002/hbm.20166.

[207] V. M. Vergara, A. Ulloa, V. D. Calhoun, D. Boutte, J. Chen, and J. Liu, "A three-way parallel ICA approach to analyze links among genetics, brain structure and brain function," *NeuroImage*, vol. 98, pp. 386–394, Sep. 2014, doi: 10.1016/j.neuroimage.2014.04.060.

[208] X. Ouyang *et al.*, "Simultaneous changes in gray matter volume and white matter fractional anisotropy in Alzheimer's disease revealed by multimodal CCA and joint ICA," *Neuroscience*, vol. 301, pp. 553–562, Aug. 2015, doi: 10.1016/j.neuroscience.2015.06.031.

[209] L. Amorim *et al.*, "Poor Sleep Quality Associates With Decreased Functional and Structural Brain Connectivity in Normative Aging: A MRI Multimodal Approach," *Frontiers in Aging Neuroscience*, vol. 10, p. 375, 2018, doi: 10.3389/fnagi.2018.00375.

[210] M. G. Puxeddu, J. Faskowitz, R. F. Betzel, M. Petti, L. Astolfi, and O. Sporns, "The modular organization of brain cortical connectivity across the human lifespan," *NeuroImage*, vol. 218, p. 116974, Sep. 2020, doi: 10.1016/j.neuroimage.2020.116974.

[211] Y. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang, and J. Liu, "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier," *Neurocomputing*, vol. 324, pp. 63–68, Jan. 2019, doi: 10.1016/j.neucom.2018.04.080.

[212] C. He *et al.*, "Individual-based morphological brain network organization and its association with autistic symptoms in young children with autism spectrum disorder," *Hum Brain Mapp*, vol. 42, no. 10, pp. 3282–3294, May 2021, doi: 10.1002/hbm.25434.

[213] J. M. Vettel, N. Cooper, J. O. Garcia, F. Yeh, and T. D. Verstynen, "White Matter Tractography and Diffusion-Weighted Imaging," in *eLS*, 1st ed., John Wiley & Sons, Ltd, Ed., Wiley, 2017, pp. 1–9. doi: 10.1002/9780470015902.a0027162.

[214] F.-C. Yeh *et al.*, "Quantifying Differences and Similarities in Whole-Brain White Matter Architecture Using Local Connectome Fingerprints," *PLOS Computational Biology*, vol. 12, no. 11, p. e1005203, Nov. 2016, doi: 10.1371/journal.pcbi.1005203.

[215] D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for generation of unique SMILES notation," *J. Chem. Inf. Comput. Sci.*, vol. 29, no. 2, pp. 97–101, May 1989, doi: 10.1021/ci00062a008.

[216] I. Lukovits, "A Compact Form of the Adjacency Matrix," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 5, pp. 1147–1150, Sep. 2000, doi: 10.1021/ci990152h.

[217] V. Kaasinen, J. Joutsa, T. Noponen, J. Johansson, and M. Seppänen, "Effects of aging and gender on striatal and extrastriatal [123I]FP-CIT binding in Parkinson's disease," *Neurobiol Aging*, vol. 36, no. 4, pp. 1757–1763, Apr. 2015, doi: 10.1016/j.neurobiolaging.2015.01.016.

[218] T. J. Collier, N. M. Kanaan, and J. H. Kordower, "Ageing as a primary risk factor for Parkinson's disease: evidence from studies of non-human primates," *Nat Rev Neurosci*, vol. 12, no. 6, pp. 359–366, Jun. 2011, doi: 10.1038/nrn3039.

[219] K. Szewczyk-Krolikowski *et al.*, "The influence of age and gender on motor and non-motor features of early Parkinson's disease: Initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort," *Parkinsonism & Related Disorders*, vol. 20, no. 1, pp. 99–105, Jan. 2014, doi: 10.1016/j.parkreldis.2013.09.025.

[220] G. Levy, "The Relationship of Parkinson Disease With Aging," *Archives of Neurology*, vol. 64, no. 9, pp. 1242–1246, Sep. 2007, doi: 10.1001/archneur.64.9.1242.

[221] R. M. Naduthota *et al.*, "Imaging biomarker correlates with oxidative stress in Parkinson's disease," *Neurology India*, vol. 65, no. 2, p. 263, Mar. 2017, doi: 10.4103/neuroindia.NI_981_15.

[222] G. Antonopoulos, S. More, F. Raimondo, S. B. Eickhoff, F. Hoffstaedter, and K. R. Patil, "A systematic comparison of VBM pipelines and their application to age prediction," BioRxiv, preprint, Jan. 2023. doi: 10.1101/2023.01.23.525151.

[223] J. W. Madole *et al.*, "Strong intercorrelations among global graph-theoretic indices of structural connectivity in the human brain," *NeuroImage*, vol. 275, p. 120160, Jul. 2023, doi: 10.1016/j.neuroimage.2023.120160.

[224] C. R. Buchanan *et al.*, "The effect of network thresholding and weighting on structural brain networks in the UK Biobank," *NeuroImage*, vol. 211, p. 116443, May 2020, doi: 10.1016/j.neuroimage.2019.116443.

[225] A. Zalesky, A. Fornito, and E. Bullmore, "On the use of correlation as a measure of network connectivity," *Neuroimage*, vol. 60, no. 4, pp. 2096–2106, May 2012, doi: 10.1016/j.neuroimage.2012.02.001.

[226] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010, doi: 10.1016/j.neuroimage.2009.10.003.

[227] D. Zhang, J. Wang, X. Liu, J. Chen, and B. Liu, "Aberrant Brain Network Efficiency in Parkinson's Disease Patients with Tremor: A Multi-Modality Study," *Front. Aging Neurosci.*, vol. 7, Aug. 2015, doi: 10.3389/fnagi.2015.00169.

[228] X. Liao, A. V. Vasilakos, and Y. He, "Small-world human brain networks: Perspectives and challenges," *Neuroscience & Biobehavioral Reviews*, vol. 77, pp. 286–300, Jun. 2017, doi: 10.1016/j.neubiorev.2017.03.018.

[229] L. A. Farwell, "Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials," *Cognitive Neurodynamics*, vol. 6, no. 2, p. 115, Apr. 2012, doi: 10.1007/s11571-012-9192-2.

[230] D. F. Hermens, C. Russo, Z. Shan, and J. Lagopoulos, "Brain fingerprinting: A promising future application for predicting mental illness," *Futures*, vol. 152, p. 103211, Sep. 2023, doi: https://doi.org/10.1016/j.futures.2023.103211.

[231] T. D. Satterthwaite, C. H. Xia, and D. S. Bassett, "Personalized Neuroscience: Common and Individual-Specific Features in Functional Brain Networks," *Neuron*, vol. 98, no. 2, pp. 243–245, Apr. 2018, doi: 10.1016/j.neuron.2018.04.007.

[232] K. T. E. Olde Dubbelink *et al.*, "Disrupted brain network topology in Parkinson's disease: a longitudinal magnetoencephalography study," *Brain*, vol. 137, no. 1, pp. 197–207, Jan. 2014, doi: 10.1093/brain/awt316.

[233] J. B. Pereira *et al.*, "Aberrant cerebral network topology and mild cognitive impairment in early Parkinson's disease: Aberrant Brain Network Topology in Early PD," *Hum. Brain Mapp.*, vol. 36, no. 8, pp. 2980–2995, Aug. 2015, doi: 10.1002/hbm.22822.

[234] Y. Li *et al.*, "Connectivity-based Brain Parcellation for Parkinson's Disease," *IEEE Trans Biomed Eng*, vol. PP, Nov. 2022, doi: 10.1109/TBME.2022.3222072.

[235] L. Farràs-Permanyer *et al.*, "Age-related changes in resting-state functional connectivity in older adults," *Neural Regen Res*, vol. 14, no. 9, pp. 1544–1555, Sep. 2019, doi: 10.4103/1673-5374.255976.

[236] S. K. Yadav *et al.*, "Gender-based analysis of cortical thickness and structural connectivity in Parkinson's disease," *J Neurol*, vol. 263, no. 11, pp. 2308–2318, Nov. 2016, doi: 10.1007/s00415-016-8265-2.

[237] F. Hou *et al.*, "Age-Related Alterations in Electroencephalography Connectivity and Network Topology During n-Back Working Memory Task," *Frontiers in Human Neuroscience*, vol. 12, 2018, doi: https://doi.org/10.3389/fnhum.2018.00484.

[238] R. Sala-Llonch *et al.*, "Changes in whole-brain functional networks and memory performance in aging," *Neurobiology of Aging*, vol. 35, no. 10, pp. 2193–2202, Oct. 2014, doi: 10.1016/j.neurobiolaging.2014.04.007.

[239] J. Samogin, L. R. Delgado, G. A. Taberna, S. P. Swinnen, and D. Mantini, "Age-Related Differences of Frequency-Dependent Functional Connectivity in Brain Networks and Their Link to Motor Performance," *Brain Connectivity*, Apr. 2022, doi: 10.1089/brain.2021.0135.

[240] J. V. Hindle, "Ageing, neurodegeneration and Parkinson's disease," *Age and Ageing*, vol. 39, no. 2, pp. 156–161, Mar. 2010, doi: 10.1093/ageing/afp223.

[241] Z. Chen, B. Wu, G. Li, L. Zhou, L. Zhang, and J. Liu, "Age and sex differentially shape brain networks in Parkinson's disease," *CNS Neuroscience & Therapeutics*, doi: 10.1111/cns.14149.

[242] T. Baluchnejadmojarad, S.-M. Eftekhari, N. Jamali-Raeufy, S. Haghani, H. Zeinali, and M. Roghani, "The anti-aging protein klotho alleviates injury of nigrostriatal dopaminergic pathway in 6-hydroxydopamine rat model of Parkinson's disease: Involvement of PKA/CaMKII/CREB signaling," *Experimental Gerontology*, vol. 100, pp. 70–76, Dec. 2017, doi: 10.1016/j.exger.2017.10.023.

[243] D.-K. Kim *et al.*, "Anti-aging treatments slow propagation of synucleinopathy by restoring lysosomal function," *Autophagy*, vol. 12, no. 10, pp. 1849–1863, Oct. 2016, doi: 10.1080/15548627.2016.1207014.

[244] D. Su, Y. Su, B. Xu, J. K. Chhetri, and P. Chan, "Age as a risk factor for orthostatic hypotension induced by the levodopa challenge test in patients with Parkinson's disease: Results from a single-center trial," *Medicine (Baltimore)*, vol. 102, no. 9, p. e33161, Mar. 2023, doi: 10.1097/MD.0000000000033161.

[245] J. E. Ahlskog, "Seniors with Parkinson's Disease: Initial Medical Treatment," *J Clin Neurol*, vol. 6, no. 4, pp. 159–166, Dec. 2010, doi: 10.3988/jcn.2010.6.4.159.

[246] M. Filippi *et al.*, "Longitudinal brain connectivity changes and clinical evolution in Parkinson's disease," *Mol Psychiatry*, vol. 26, no. 9, Art. no. 9, Sep. 2021, doi: 10.1038/s41380-020-0770-0.

[247] H. Javaid, E. Kumarnsit, and S. Chatpun, "Age-Related Alterations in EEG Network Connectivity in Healthy Aging," *Brain Sciences*, vol. 12, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/brainsci12020218.

[248] S. Shakya, J. Prevett, X. Hu, and R. Xiao, "Characterization of Parkinson's Disease Subtypes and Related Attributes," *Frontiers in Neurology*, vol. 13, 2022, doi: https://doi.org/10.3389/fneur.2022.810038.

[249] E. Kaplan *et al.*, "Novel nested patch-based feature extraction model for automated Parkinson's Disease symptom classification using MRI images," *Computer Methods and Programs in Biomedicine*, vol. 224, p. 107030, Sep. 2022, doi: 10.1016/j.cmpb.2022.107030.

[250] A. Sarica, A. Quattrone, and A. Quattrone, "Explainable machine learning with pairwise interactions for the classification of Parkinson's disease and SWEDD from clinical and imaging features," *Brain Imaging and Behavior*, May 2022, doi: 10.1007/s11682-022-00688-9.

[251] L. Xu, K. M. Groth, G. Pearlson, D. J. Schretlen, and V. D. Calhoun, "Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia," *Human Brain Mapping*, vol. 30, no. 3, pp. 711–724, 2009, doi: 10.1002/hbm.20540.

[252] C. N. Gupta, J. A. Turner, and V. D. Calhoun, "Source-based morphometry: a decade of covarying structural brain patterns," *Brain Struct Funct*, vol. 224, no. 9, pp. 3031–3044, Dec. 2019, doi: 10.1007/s00429-019-01969-8.

[253] M. Eckert, N. Keren, D. Roberts, V. Calhoun, and K. Harris, "Age-related changes in processing speed: unique contributions of cerebellar and prefrontal cortex," *Frontiers in Human Neuroscience*, vol. 4, 2010, doi: https://doi.org/10.3389/neuro.09.010.2010.

[254] I. Rektorova, R. Biundo, R. Marecek, L. Weis, D. Aarsland, and A. Antonini, "Grey Matter Changes in Cognitively Impaired Parkinson's Disease Patients," *PLOS ONE*, vol. 9, no. 1, p. e85595, Jan. 2014, doi: 10.1371/journal.pone.0085595.

[255] F. Knolle *et al.*, "A multicentre study on grey matter morphometric biomarkers for classifying early schizophrenia and parkinson's disease psychosis," *npj Parkinsons Dis.*, vol. 9, no. 1, Art. no. 1, Jun. 2023, doi: 10.1038/s41531-023-00522-z.

[256] M. D. Steenwijk *et al.*, "Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant," *Brain*, vol. 139, no. 1, pp. 115–126, Jan. 2016, doi: 10.1093/brain/awv337.

[257] C. Wachinger, A. Rieckmann, and S. Pölsterl, "Detect and correct bias in multi-site neuroimaging datasets," *Medical Image Analysis*, vol. 67, p. 101879, Jan. 2021, doi: 10.1016/j.media.2020.101879.

[258] H. Akoglu, "User's guide to correlation coefficients," *Turk J Emerg Med*, vol. 18, no. 3, pp. 91–93, Aug. 2018, doi: 10.1016/j.tjem.2018.08.001.

[259] Y. H. Chan, "Biostatistics 104: Correlational Analysis," *Singapore Med J*, vol. 44, no. 12, pp. 614–619, 2003, doi: https://paulogentil.com/pdf/Biostatistics%20104.%20Correlational%20analysis.pdf .

[260] C. N. Gupta *et al.*, "Patterns of Gray Matter Abnormalities in Schizophrenia Based on an International Mega-analysis," *Schizophrenia Bulletin*, vol. 41, no. 5, pp. 1133–1142, Sep. 2015, doi: 10.1093/schbul/sbu177.

[261] J. A. Turner *et al.*, "Heritability of Multivariate Gray Matter Measures in Schizophrenia," *Twin Research and Human Genetics*, vol. 15, no. 3, pp. 324–335, Jun. 2012, doi: 10.1017/thg.2012.1.

[262] Y. Zeighami *et al.*, "Network structure of brain atrophy in de novo Parkinson's disease," *eLife*, vol. 4, p. e08440, Sep. 2015, doi: 10.7554/eLife.08440.

[263] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995, doi: https://www.jstor.org/stable/2346101.

[264] N. Shao, J. Yang, and H. Shang, "Voxelwise meta-analysis of gray matter anomalies in Parkinson variant of multiple system atrophy and Parkinson's disease using

anatomic likelihood estimation," *Neurosci. Lett.*, vol. 587, pp. 79–86, Feb. 2015, doi: 10.1016/j.neulet.2014.12.007.

[265] P. L. Pan, W. Song, and H. F. Shang, "Voxel-wise meta-analysis of gray matter abnormalities in idiopathic Parkinson's disease," *Eur J Neurol*, vol. 19, no. 2, pp. 199–206, Feb. 2012, doi: 10.1111/j.1468-1331.2011.03474.x.

[266] J. Ceccarini *et al.*, "Regional changes in the type 1 cannabinoid receptor are associated with cognitive dysfunction in Parkinson's disease," *Eur J Nucl Med Mol Imaging*, vol. 46, no. 11, pp. 2348–2357, Oct. 2019, doi: 10.1007/s00259-019-04445-x.

[267] A. R. E. Potgieser, A. van der Hoorn, A. M. Meppelink, L. K. Teune, J. Koerts, and B. M. de Jong, "Anterior Temporal Atrophy and Posterior Progression in Patients with Parkinson's Disease," *Neurodegenerative Diseases*, vol. 14, no. 3, pp. 125–132, Aug. 2014, doi: 10.1159/000363245.

[268] J. Zhang, Y.-T. Zhang, W.-D. Hu, L. Li, G.-Y. Liu, and Y.-P. Bai, "Gray matter atrophy in patients with Parkinson's disease and those with mild cognitive impairment: a voxel-based morphometry study," *Int J Clin Exp Med*, vol. 8, no. 9, pp. 15383–15392, Sep. 2015.

[269] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs," *Front. Psychol.*, vol. 4, 2013, doi: 10.3389/fpsyg.2013.00863.

[270] L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*. Academic Press, 2014.

[271] J.-Y. Lee, N. P. Lao-Kaim, J. Pasquini, G. Deuschl, N. Pavese, and P. Piccini, "Pallidal dopaminergic denervation and rest tremor in early Parkinson's disease: PPMI cohort analysis," *Parkinsonism & Related Disorders*, vol. 51, pp. 101–104, Jun. 2018, doi: 10.1016/j.parkreldis.2018.02.039.

[272] A. M. Meppelink, B. M. de Jong, L. K. Teune, and T. van Laar, "Regional cortical grey matter loss in Parkinson's disease without dementia is independent from visual hallucinations," *Movement Disorders*, vol. 26, no. 1, pp. 142–147, 2011, doi: https://doi.org/10.1002/mds.23375.

[273] V. S. Kostić *et al.*, "Regional patterns of brain tissue loss associated with depression in Parkinson disease," *Neurology*, vol. 75, no. 10, pp. 857–863, Sep. 2010, doi: 10.1212/WNL.0b013e3181f11c1d.

[274] Y.-T. Shen *et al.*, "Dysfunction in superior frontal gyrus associated with diphasic dyskinesia in Parkinson's disease," *npj Parkinsons Dis.*, vol. 6, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41531-020-00133-y.

[275] E. Premi *et al.*, "Source-Based Morphometry Multivariate Approach to Analyze [123I]FP-CIT SPECT Imaging," *Mol Imaging Biol*, vol. 19, no. 5, pp. 772–778, Oct. 2017, doi: 10.1007/s11307-017-1052-3.

[276] I. N. Miller and A. Cronin-Golomb, "Gender differences in Parkinson's disease: Clinical characteristics and cognition," *Movement Disorders*, vol. 25, no. 16, pp. 2695–2703, 2010, doi: 10.1002/mds.23388.

[277] K. Cao *et al.*, "Identifying and validating subtypes of Parkinson's disease based on multimodal MRI data via hierarchical clustering analysis," *Frontiers in Human Neuroscience*, vol. 16, 2022, Accessed: Aug. 31, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2022.919081

[278] V. V. Makarov *et al.*, "Betweenness centrality in multiplex brain network during mental task evaluation," *Phys. Rev. E*, vol. 98, no. 6, p. 062413, Dec. 2018, doi: 10.1103/PhysRevE.98.062413.

[279] A. Mechelli, G. W. Humphreys, K. Mayall, A. Olson, and C. J. Price, "Differential effects of word length and visual contrast in the fusiform and lingual gyri during | Proceedings of the Royal Society of London. Series B: Biological Sciences," vol. 267, no. 1455, Sep. 2000, doi: https://doi.org/10.1098/rspb.2000.1229.

[280] S. Pisani *et al.*, "Grey matter volume loss in Parkinson's disease psychosis and its relationship with serotonergic gene expression: A meta-analysis," *Neuroscience & Biobehavioral Reviews*, vol. 147, p. 105081, Apr. 2023, doi: 10.1016/j.neubiorev.2023.105081.

[281] J. Chen *et al.*, "Gray and white matter alterations in different predominant side and type of motor symptom in Parkinson's disease," *CNS Neuroscience & Therapeutics*, vol. 28, no. 9, pp. 1372–1379, 2022, doi: 10.1111/cns.13877.

[282] B. A. Vogt, "Cingulate cortex in Parkinson's disease," in *Handbook of Clinical Neurology*, vol. 166, B. A. Vogt, Ed., in Cingulate Cortex, vol. 166. , Elsevier, 2019, pp. 253–266. doi: 10.1016/B978-0-444-64196-0.00013-3.

[283] L.-Y. Ma, X.-D. Chen, Y. He, H.-Z. Ma, and T. Feng, "Disrupted Brain Network Hubs in Subtype-Specific Parkinson's Disease," *European Neurology*, vol. 78, no. 3–4, pp. 200–209, Aug. 2017, doi: 10.1159/000477902.

[284] A. F. Wolters *et al.*, "Brain network characteristics and cognitive performance in motor subtypes of Parkinson's disease: A resting state fMRI study," *Parkinsonism & Related Disorders*, vol. 105, pp. 32–38, Dec. 2022, doi: 10.1016/j.parkreldis.2022.10.027.

[285] F. Cao, X. Guan, Y. Ma, Y. Shao, and J. Zhong, "Altered Functional Network Associated With Cognitive Performance in Early Parkinson Disease Measured by Eigenvector Centrality Mapping," *Frontiers in Aging Neuroscience*, vol. 12, 2020, doi: https://doi.org/10.3389/fnagi.2020.554660.

[286] Y. Hou *et al.*, "Resting-state network connectivity in cognitively unimpaired drug-naïve patients with rigidity-dominant Parkinson's disease," *Journal of the Neurological Sciences*, vol. 395, pp. 147–152, Dec. 2018, doi: 10.1016/j.jns.2018.10.003.

[287] S. S. Cho *et al.*, "The Relationship Between Serotonin-2A Receptor and Cognitive Functions in Nondemented Parkinson's Disease Patients with Visual Hallucinations," *Mov Disord Clin Pract*, vol. 4, no. 5, pp. 698–709, 2017, doi: 10.1002/mdc3.12466.

[288] A. Chen *et al.*, "Disrupted Brain Structural Network Connection in de novo Parkinson's Disease With Rapid Eye Movement Sleep Behavior Disorder," *Front Hum Neurosci*, vol. 16, p. 902614, Jul. 2022, doi: 10.3389/fnhum.2022.902614.

[289] E. De Pablo-Fernández, A. J. Lees, J. L. Holton, and T. T. Warner, "Prognosis and Neuropathologic Correlation of Clinical Subtypes of Parkinson Disease," *JAMA Neurol*, vol. 76, no. 4, pp. 470–479, Apr. 2019, doi: 10.1001/jamaneurol.2018.4377.

[290] B. R. Matthews, "Memory Dysfunction," *Continuum (Minneap Minn)*, vol. 21, no. 3 Behavioral Neurology and Neuropsychiatry, pp. 613–626, Jun. 2015, doi: 10.1212/01.CON.0000466656.59413.29.

[291] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of Neuroscience Methods*, vol. 204, no. 1, pp. 68–81, Feb. 2012, doi: 10.1016/j.jneumeth.2011.10.031.

[292] V. D. Calhoun and J. Sui, "Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness," *Biol Psychiatry Cogn Neurosci Neuroimaging*, vol. 1, no. 3, pp. 230–244, May 2016, doi: 10.1016/j.bpsc.2015.12.005.

[293] L.-L. Zeng *et al.*, "Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI," *EBioMedicine*, vol. 30, pp. 74–85, Apr. 2018, doi: 10.1016/j.ebiom.2018.03.017.

[294] L.-L. Zeng *et al.*, "Gradient Matching Federated Domain Adaptation for Brain Image Classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2022, doi: 10.1109/TNNLS.2022.3223144.

[295] O. Sporns, "The complex brain: connectivity, dynamics, information," *Trends in Cognitive Sciences*, Oct. 2022, doi: 10.1016/j.tics.2022.08.002.

[296] F. H. Y. Chan, F. K. Lam, and H. Zhu, "Adaptive thresholding by variational method," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 468–473, Mar. 1998, doi: 10.1109/83.661196.

[297] I. A. Higgins, S. Kundu, K. S. Choi, H. S. Mayberg, and Y. Guo, "A difference degree test for comparing brain networks," *Human Brain Mapping*, vol. 40, no. 15, pp. 4518–4536, 2019, doi: 10.1002/hbm.24718.

[298] F. Skidmore, D. Korenkevych, Y. Liu, G. He, E. Bullmore, and P. M. Pardalos, "Connectivity brain networks based on wavelet correlation analysis in Parkinson fMRI data," *Neuroscience Letters*, vol. 499, no. 1, pp. 47–51, Jul. 2011, doi: 10.1016/j.neulet.2011.05.030.

[299] G. Zhang *et al.*, "Detecting abnormal connectivity in schizophrenia via a joint directed acyclic graph estimation model," *NeuroImage*, vol. 260, p. 119451, Oct. 2022, doi: 10.1016/j.neuroimage.2022.119451.

[300] N. Ram and K. J. Grimm, "Growth Mixture Modeling: A Method for Identifying Differences in Longitudinal Change Among Unobserved Groups," *Int J Behav Dev*, vol. 33, no. 6, pp. 565–576, 2009, doi: 10.1177/0165025409343765.

[301] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, and D. Shen, "Feature Selection Based on Iterative Canonical Correlation Analysis for Automatic Diagnosis of Parkinson's Disease," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 1–8. doi: 10.1007/978-3-319-46723-8_1.

[302] C. Song *et al.*, "Stability Evaluation of Brain Changes in Parkinson's Disease Based on Machine Learning," *Frontiers in Computational Neuroscience*, vol. 15, 2021, Available: https://www.frontiersin.org/articles/10.3389/fncom.2021.735991

[303] A. Pfefferbaum, D. H. Mathalon, E. V. Sullivan, J. M. Rawles, R. B. Zipursky, and K. O. Lim, "A Quantitative Magnetic Resonance Imaging Study of Changes in Brain Morphology From Infancy to Late Adulthood," *Archives of Neurology*, vol. 51, no. 9, pp. 874–887, Sep. 1994, doi: 10.1001/archneur.1994.00540210046012.

[304] X. Guo, Y. Han, K. Chen, Y. Wang, and L. Yao, "Mapping joint grey and white matter reductions in Alzheimer's disease using joint independent component analysis," *Neuroscience Letters*, vol. 531, no. 2, pp. 136–141, Dec. 2012, doi: 10.1016/j.neulet.2012.10.038.

[305] L. Xu, K. M. Groth, G. Pearlson, D. J. Schretlen, and V. D. Calhoun, "Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia," *Hum. Brain Mapp.*, vol. 30, no. 3, pp. 711–724, Mar. 2009, doi: 10.1002/hbm.20540.

[306] Q. Zhu, J. Yang, B. Xu, Z. Hou, L. Sun, and D. Zhang, "Multimodal Brain Network Jointly Construction and Fusion for Diagnosis of Epilepsy," *Frontiers in Neuroscience*, vol. 15, 2021, Accessed: Sep. 04, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2021.734711

[307] R. A. Jarvis and E. A. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors," *IEEE Transactions on Computers*, vol. C–22, no. 11, pp. 1025–1034, Nov. 1973, doi: 10.1109/T-C.1973.223640.

[308] Z. Hu and R. Bhatnagar, "Clustering algorithm based on mutual K-nearest neighbor relationships," *Statistical Analy Data Mining*, vol. 5, no. 2, pp. 100–113, Apr. 2012, doi: 10.1002/sam.10149.

[309] D. Sardana and R. Bhatnagar, "Graph Clustering Using Mutual K-Nearest Neighbors," in *Active Media Technology*, D. Ślęzak, G. Schaefer, S. T. Vuong, and Y.-S. Kim, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 35–48. doi: 10.1007/978-3-319-09912-5_4.

[310] A. Hassanat, "Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach," vol. 12, no. 8, 2014, [Online]. Available: arXiv preprint arXiv:1409.0919

[311] J. C. Greenland, M. Camacho, and C. H. Williams-Gray, "Precision Medicine in Neurodegenerative Disorders, Part I," in *Handbook of Clinical Neurology*, vol. 192, Elsevier, 2023.

[312] C. G. Goetz *et al.*, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations The Movement Disorder Society Task Force on rating scales for Parkinson's disease," *Movement Disorders*, vol. 19, no. 9, pp. 1020–1028, 2004, doi: 10.1002/mds.20213.

[313] M. E. J. Newman, "Assortative Mixing in Networks," *Phys. Rev. Lett.*, vol. 89, no. 20, p. 208701, Oct. 2002, doi: 10.1103/PhysRevLett.89.208701.

[314] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[315] B. Adepu, J. Gyani, and G. Narsimha, "A Novel Multi-objective Differential Evolution Algorithm for Clustering Data Streams: 2nd International Conference on Image Processing and Capsule Networks, ICIPCN 2021," *2nd International Conference on Image Processing and Capsule Networks, ICIPCN 2021*, pp. 67–78, 2022, doi: 10.1007/978-3-030-84760-9_7.

[316] P. Orzechowski, K. Boryczko, and J. H. Moore, "Scalable biclustering — the future of big data exploration?," *GigaScience*, vol. 8, no. 7, p. giz078, Jul. 2019, doi: 10.1093/gigascience/giz078.

[317] B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *Journal of Biomedical Informatics*, vol. 57, pp. 163–180, Oct. 2015, doi: 10.1016/j.jbi.2015.06.028.

[318] M. D. M. Noronha, R. Henriques, S. C. Madeira, and L. E. Zárate, "Impact of metrics on biclustering solution and quality: A review," *Pattern Recognition*, vol. 127, p. 108612, Jul. 2022, doi: 10.1016/j.patcog.2022.108612.

[319] M. Ramkumar *et al.*, "Healthcare Biclustering-Based Prediction on Gene Expression Dataset," *BioMed Research International*, vol. 2022, p. e2263194, Feb. 2022, doi: 10.1155/2022/2263194.

[320] Y. K. Wang, C. G. Print, and E. J. Crampin, "Biclustering reveals breast cancer tumour subgroups with common clinical features and improves prediction of disease recurrence," *BMC Genomics*, vol. 14, no. 1, p. 102, Feb. 2013, doi: 10.1186/1471-2164-14-102.

[321] J. Sun, J. Bi, and H. R. Kranzler, "Multi-view biclustering for genotype-phenotype association studies of complex diseases," in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, Shanghai, China: IEEE, Dec. 2013, pp. 316–321. doi: 10.1109/BIBM.2013.6732509.

[322] P. A. D. De Castro, F. O. De França, H. M. Ferreira, and F. J. Von Zuben, "Applying Biclustering to Text Mining: An Immune-Inspired Approach," in *Artificial Immune Systems*, vol. 4628, L. N. De Castro, F. J. Von Zuben, and H. Knidel, Eds., in Lecture Notes in Computer Science, vol. 4628. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 83–94. doi: 10.1007/978-3-540-73922-7_8.

[323] I. S. Dhillon, S. Mallela, and D. S. Modha, "Information-theoretic co-clustering," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington, D.C.: ACM, Aug. 2003, pp. 89–98. doi: 10.1145/956750.956764.

[324] N. K. Verma, T. Sharma, S. Dixit, P. Agrawal, S. Sengupta, and V. Singh, "BIDEAL: A Toolbox for Bicluster Analysis—Generation, Visualization and Validation," *SN COMPUT. SCI.*, vol. 2, no. 1, p. 24, Jan. 2021, doi: 10.1007/s42979-020-00411-9.

[325] E. N. Castanho, H. Aidos, and S. C. Madeira, "Biclustering fMRI time series: a comparative study," *BMC Bioinformatics*, vol. 23, no. 1, p. 192, Dec. 2022, doi: 10.1186/s12859-022-04733-8.

[326] K. Marek *et al.*, "The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort," *Ann Clin Transl Neurol*, vol. 5, no. 12, pp. 1460–1477, Dec. 2018, doi: 10.1002/acn3.644.

[327] J. Ashburner and K. J. Friston, "Voxel-Based Morphometry—The Methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, Jun. 2000, doi: 10.1006/nimg.2000.0582.

[328] R. E. Korf, "Depth-first iterative-deepening: An optimal admissible tree search," *Artificial Intelligence*, vol. 27, no. 1, pp. 97–109, Sep. 1985, doi: 10.1016/0004-3702(85)90084-0.

[329] S. Hochreiter *et al.*, "FABIA: factor analysis for bicluster acquisition," *Bioinformatics*, vol. 26, no. 12, pp. 1520–1527, Jun. 2010, doi: 10.1093/bioinformatics/btq227.

[330] A. Prelić *et al.*, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Bioinformatics*, vol. 22, no. 9, pp. 1122–1129, May 2006, doi: 10.1093/bioinformatics/btl060.

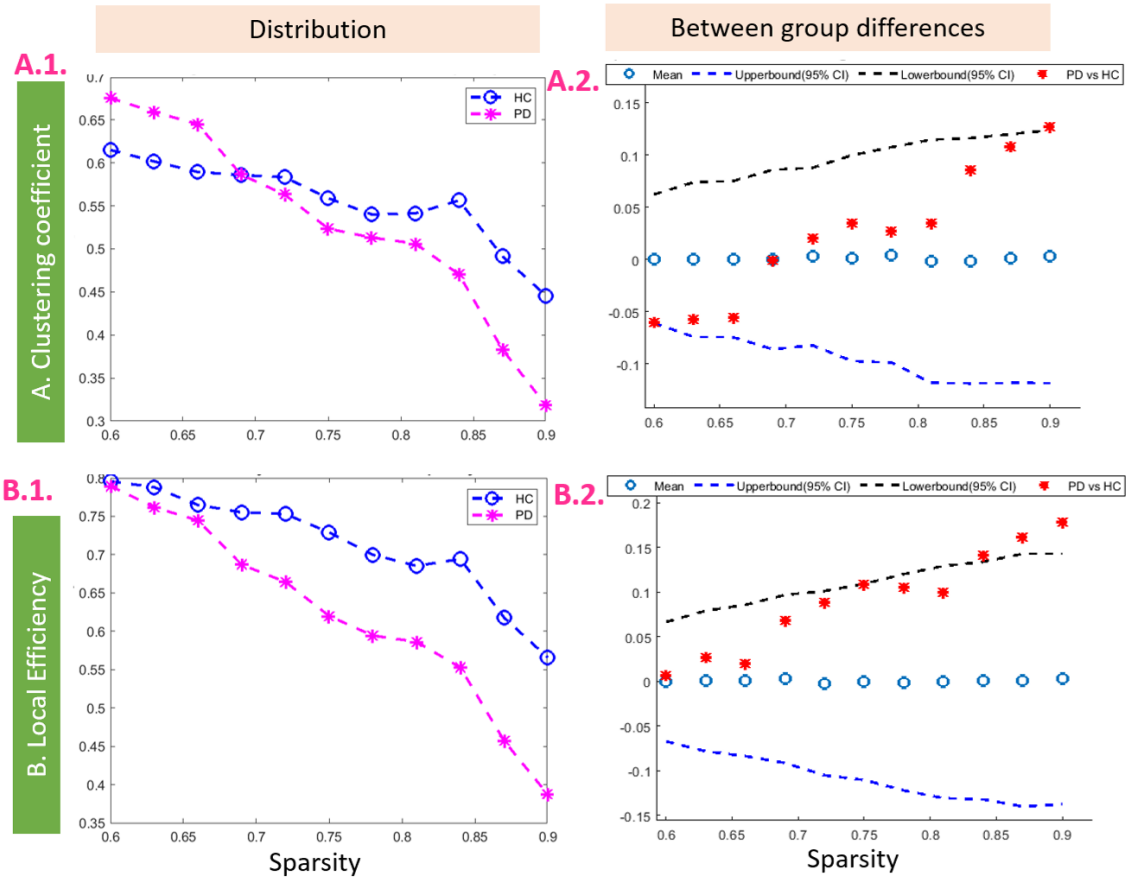
[331] M. Rubio-Sanchez, "Basic Concepts of Recursive Programming," in *Introduction to Recursive Programming*, 1st ed., Taylor & Francis Group, p. 26. Accessed: Jun. 02, 2023. [Online]. Available: https://doi.org/10.1201/9781315120850

[332] B. Thanvi, N. Lo, and T. Robinson, "Levodopa-induced dyskinesia in Parkinson's disease: clinical features, pathogenesis, prevention and treatment," *Postgraduate Medical Journal*, vol. 83, no. 980, pp. 384–388, Jun. 2007, doi: 10.1136/pgmj.2006.054759.

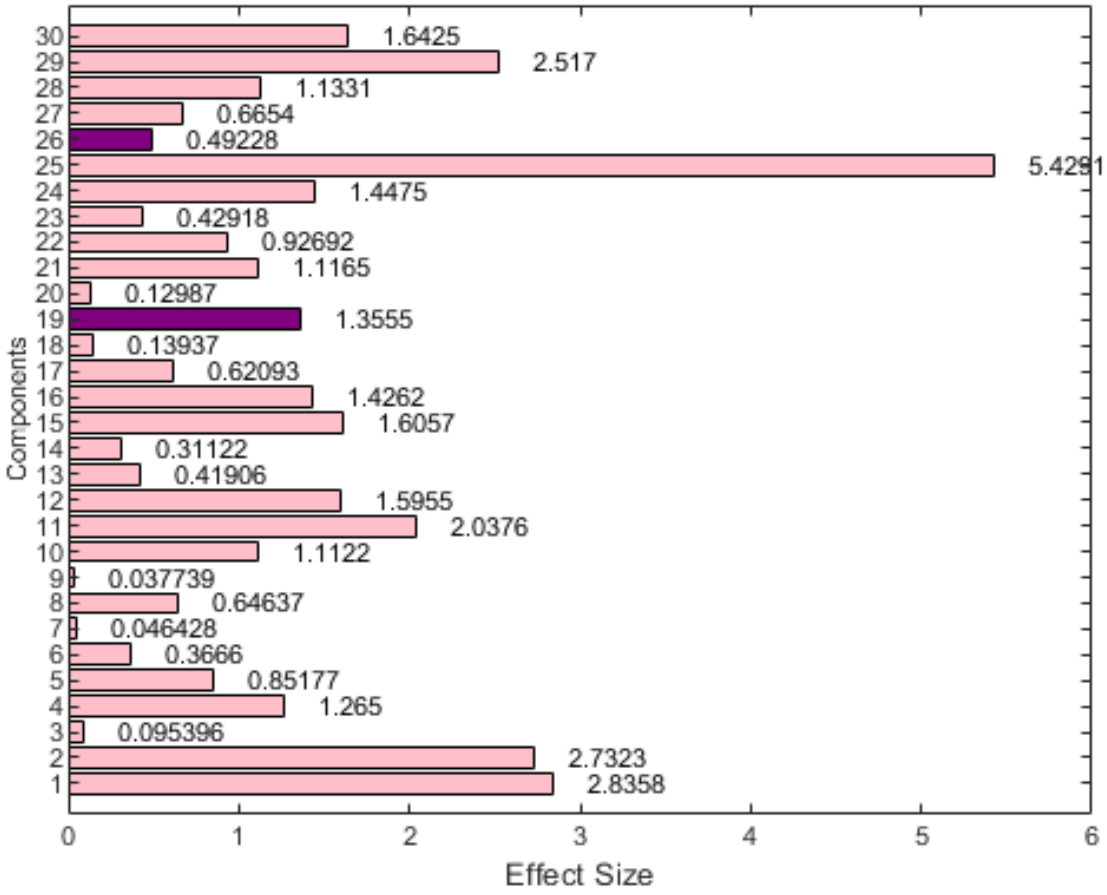
[333] T. Song et al., "Nigral Iron Deposition Is Associated With Levodopa-Induced Dyskinesia in Parkinson's Disease," Frontiers in Neuroscience, vol. 15, 2021, Accessed: Feb. 01, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2021.647168

APPENDIX

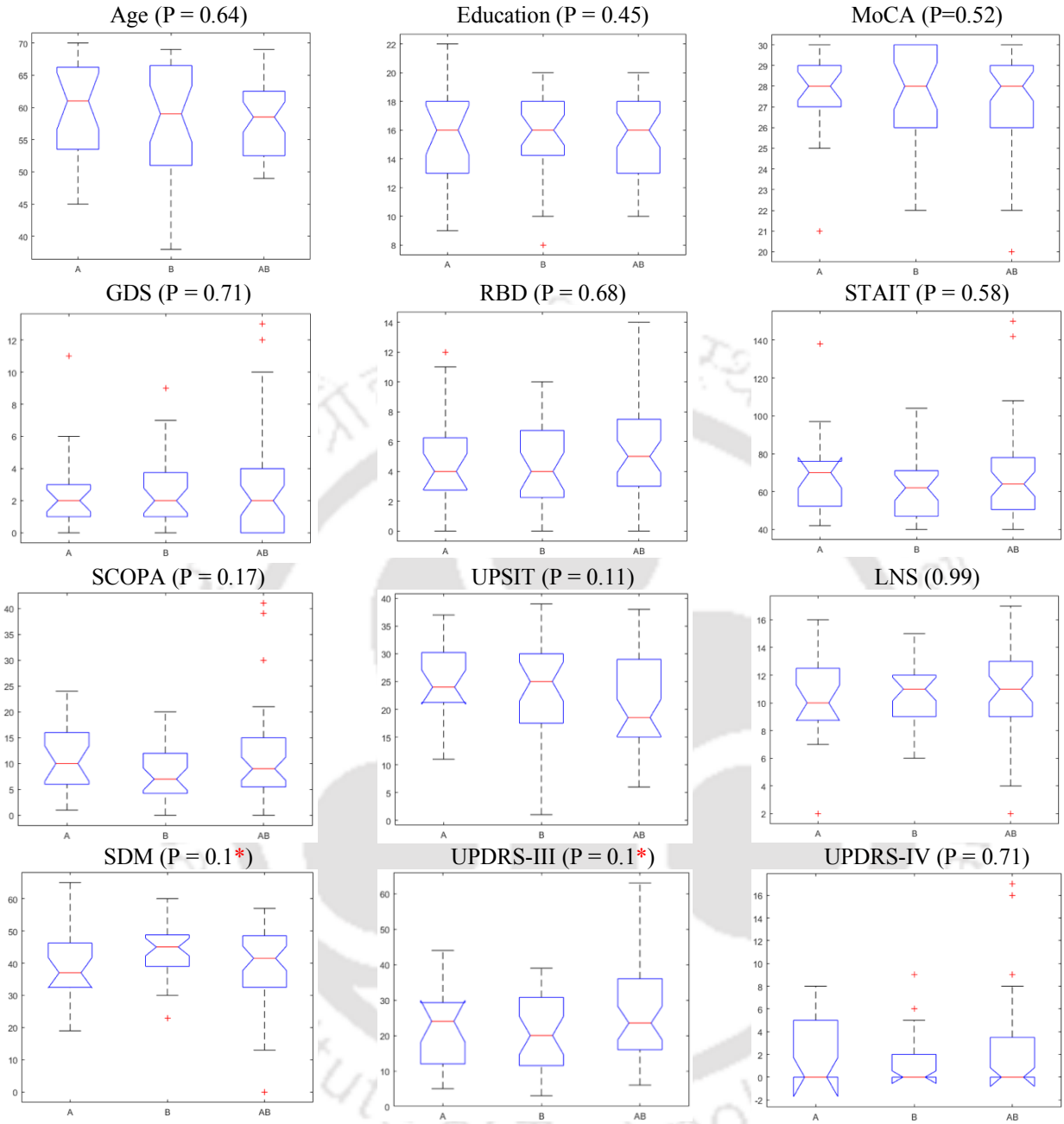
APPENDIX



Supplementary Figure S1. Distribution and group difference between PD and HC in clustering coefficient and local efficiency.



Supplementary Figure S2. Hedge's g effect size of components showing difference between healthy control and Parkinson's patients



Supplementary Figure S3. ANOVA of subtype-specific clinical features from GM information

A: Subtype A exclusively; AB: Common to A and B; B: Subtype B exclusively; GDS: Geriatric Depression Scale; LNS: Letter Number Sequencing; MoCA: Montreal Cognitive Assessment; P: Probability value; QUIP: Questionnaire for Impulsive-Compulsive Disorders; RBD: Random Eye Movement sleep behaviour disorder; SCOPA: Scales for Outcomes in Parkinson's Disease-Autonomic; SDM: Symbol Digit Modality; STAIT: State-Trait Anxiety Inventory for Adults Test; UPSIT: University of Pennsylvania Smell Identification Test; UPDRS: Unified Parkinson's Disease Rating Scale, *: Significant at 90% CI, Line in the middle of the box indicate the median; Top and bottom edges of the box indicate the 25th and 75th percentiles, respectively.

Supplementary Table S1. Regions of LPBA atlas [67]

Sl.	Abbreviations	Region Names
1.	lSupFroG	L.superior.frontal.gyrus
2.	rSupFroG	R.superior.frontal.gyrus
3.	lMidFroG	L.middle.frontal.gyrus
4.	rMidFroG	R.middle.frontal.gyrus
5.	lInfFroG	L.inferior.frontal.gyrus
6.	rInfFroG	R.inferior.frontal.gyrus
7.	lPrcG	L.precentral.gyrus
8.	rPrcG	R.precentral.gyrus
9.	lMidOrbG	L.middle.orbitofrontal.gyrus
10.	rMidOrbG	R.middle.orbitofrontal.gyrus
11.	lLatOrbG	L.lateral.orbitofrontal.gyrus
12.	rLatOrbG	R.lateral.orbitofrontal.gyrus
13.	lRecG	L.gyrus.rectus
14.	rRecG	R.gyrus.rectus
15.	lPoCG	L.postcentral.gyrus
16.	rPoCG	R.postcentral.gyrus
17.	lSupParG	L.superior.parietal.gyrus
18.	rSupParG	R.superior.parietal.gyrus
19.	lSupMarG	L.supramarginal.gyrus
20.	rSupMarG	R.supramarginal.gyrus
21.	lAngG	L.angular.gyrus
22.	rAngG	R.angular.gyrus
23.	lPCu	L.precuneus
24.	rPCu	R.precuneus
25.	lSupOccG	L.superior.occipital.gyrus
26.	rSupOccG	R.superior.occipital.gyrus
27.	lMidOccG	L.middle.occipital.gyrus
28.	rMidOccG	R.middle.occipital.gyrus
29.	lInfOccG	L.inferior.occipital.gyrus

30.	rInfOccG	R.inferior.occipital.gyrus
31.	lCun	L.cuneus
32.	rCun	R.cuneus
33.	lSupTemG	L.superior.temporal.gyrus
34.	rSupTemG	R.superior.temporal.gyrus
35.	lMidTemG	L.middle.temporal.gyrus
36.	rMidTemG	R.middle.temporal.gyrus
37.	lInfTemG	L.inferior.temporal.gyrus
38.	rInfTemG	R.inferior.temporal.gyrus
39.	lParHipG	L.parahippocampal.gyrus
40.	rParHipG	R.parahippocampal.gyrus
41.	lLinG	L.lingual.gyrus
42.	rLinG	R.lingual.gyrus
43.	lFusG	L.fusiform.gyrus
44.	rFusG	R.fusiform.gyrus
45.	lIns	L.insular.cortex
46.	rIns	R.insular.cortex
47.	lCinG	L.cingulate.gyrus
48.	rCinG	R.cingulate.gyrus
49.	lCau	L.caudate
50.	rCau	R.caudate
51.	lPut	L.putamen
52.	rPut	R.putamen
53.	lHip	L.hippocampus
54.	rHip	R.hippocampus
55.	bCBeL	Cerebellum
56.	bBst	Brainstem

ADDENDUMS

During the tenure of PhD, a few additional works were done and recognized.

1. Computational intelligence solutions for healthcare

This was a part of an international competition - 2022 IEEE CIS Student Hackathon on Computational Intelligence in Biomedicine and Healthcare (https://ieeessci2022.org/competition.html, accessed on 08th September 2023). The task was to present computational intelligence solutions for core issues in smart medicine, smart healthcare and smart hospital including but not limited to personalized, predictive, preventive and participatory medicine and healthcare solutions.

An MRI-based healthcare support system was developed which may assist clinicians, radiologists or other healthcare professions to suggest precise treatments. To attain this, a MATLAB-based toolbox was designed and developed for subgrouping patients and analyzing brain connectivity using MRI scans. It is a GUI that handles MRI data and by providing additional clinical variable, the entire task will be done in just few clicks.

Initially, an entire proposal was submitted, which was later shortlisted for the final round of the competition where the software was presented and demonstrated (https://sites.google.com/view/cishackathon-2022/final-contest).

2. Multilingual Alzheimer's Dementia Recognition through Spontaneous Speech

This was a part of ADReSS Signal Processing Grand Challenge for identification of Alzheimer's Dementia using speech data (https://luzs.gitlab.io/madress-2023/, accessed on 08th September 2023), a part of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. It was a challenging automatic prediction problem of societal and medical importance. The challenge comprised of two tasks:

a. Classification task, and

b. Prediction task, where you create a model to infer the subject's Mini Mental Status Examination (MMSE) score based on speech data;

The first task, classification task, was to create a model so as to distinguish healthy control speech from AD/MCI speech. The labelled speech data for training was provided in English and the prediction was to be done on Greek speech data.

The second task, prediction task, was to predict MMSE (Mini-Mental State Examination) score of the test subjects. Feature extraction and selection was done based on the idea that features should be conserved across languages.

The results submitted was recognized in the competition as provided in the below link: https://luzs.gitlab.io/madress-2023/ranking.html

LIST OF PUBLICATIONS

LIST OF PUBLICATIONS

Publications:

- **T. Samantaray**, J. Saini, P. K. Pal and C. N. Gupta, "Brain Connectivity for Subgroups of Parkinson's Disease using Grey Matter Volume",Biomedical Physics and Engineering Express, vol 10, no. 2, p. 025012, Jan 2024. doi: 10.1088/2057-1976/ad1e77
- **T. Samantaray**, J. Saini, and C. N. Gupta, "Unique Brain Network Identification Number of Age-groups in Parkinson's Disease", Brain Sciences, vol. 13, p.1297, Sep 2023, doi: https://doi.org/10.3390/ brainsci13091297
- Aditya S[1], **T. Samantaray[1]** and C. N. Gupta, "Optimizing Permutations in Biclustering Algorithms, 4th International Conference on Image Processing and Capsule Networks 2023, ([1]Joint 1st author). https://icipcn.com/2023/presentation-schedule.html
- **T. Samantaray**, J. Saini, and C. N. Gupta, "Subgrouping and structural brain connectivity of Parkinson's disease – past studies and future directions," Neuroscience Informatics, vol. 2, no. 4, p. 100100, Dec. 2022, doi: https://www.sciencedirect.com/science/article/pii/S2772528622000620
- **T. Samantaray**, J. Saini, and C. N. Gupta, "Sparsity Dependent Metrics Depict Alteration of Brain Network Connectivity in Parkinson's Disease," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2022, pp. 698–701. doi: 10.1109/EMBC48229.2022.9871258.
- **T. Samantaray**, J. Saini, P. K. Pal, S. M. Hazarika, V. V. Saradhi, and C. N. Gupta, Mutual K-NN based brain network analysis of PD subtypes: Integrating Grey and White Matter Information ((Manuscript Under submission)
- Brain Connectivity analysis using Hybrid-model based on Mutual k-Nearest Neighbor. (Manuscript Under preparation)
- Spectrogram-based speech connectivity in Alzheimer's disease (Manuscript Under preparation)

Achievements:

- Twelfth (12th) rank in IEEE - International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023) Signal Processing Grand Challenge: Multilingual Alzheimer's Dementia Recognition through Spontaneous Speech
  Link: https://luzs.gitlab.io/madress-2023/ranking.html
- Final contestant in IEEE - CIS Student Hackathon, Computation Intelligence in Biomedicine and Healthcare, SSCI, 5 December 2023
  Link: https://sites.google.com/view/cishackathon-2022/final-contest?authuser=0

Conferences:

- T. Samantaray, J. Saini, and C. N. Gupta, Poster on "Age-based network analysis', European Molecular Biology Organization, IIT Gandhinagar, 12-17 December 2022.
- T. Samantaray, J. Saini, and C. N. Gupta, Poster presentation on "Subgrouping and Structural Brain Connectivity of Parkinson's disease- A review", North Eastern Research Conclave, IIT Guwahati, 20-22 May 2022.
- Oral and poster presentation on "Structural Brain Connectivity Analysis of Patients with Parkinson's Disease", The Brain Conference 2022, 24-25 Feb 2022.
- T. Samantaray, J. Saini, and C. N. Gupta, Poster presentation on "Meta-Analysis of Clinical Symptoms and Data Driven Subtyping Approaches in Parkinson's Disease," The Brain Conference, 2021, Mar. 04, 2021. Accessed: Oct. 19, 2022. [Online]. Available:https://thebrainconference.co.uk/wp-content/uploads/2021/02/R2_51_Samantaray_Tanmayee_MovementDisorders_51.png\

Workshops

- Worskshop on Application of statistics & Machine Learning in Environmental Research, Indian Statistical Institute North Eastern Centre, Tezpur, 1-3 March 2023.
- SERB-sponsored Karyashala on Fundamentals of Deep Learning, IIT Guwahati, 12-18 September 2022.
- Presented on "Construction and analysis of Brain Network" in SERB-sponsored Karyashala on Brain-Machine Interface, BIT Mesra, Ranchi, 22-28 August 2022.
- Brain Connectivity Workshop 2021 from 25th – 28th May 2021
- Talk on "Glimpse on structural brain connectivity" in Departmental Retreat, Dept. of BSBE, 2021.

Additional Works Done:

- Developed a software for automated subgrouping of disease for a given set of input MRI.
- Developed a manual for brain network construction and visualization.
- Assisted in designing two Matlab tutorials and conducting Viva for BT640 (Neural Imaging and Signal Systems) and BT301 (Biophysics) courses at IIT Guwahati.