Journal Home Page: https://sjes.univsul.edu.iq/

Research Article:

Directed Chebyshev Spectral Graph Neural Networks: Continuous MoCA Prediction from EEG in Parkinson's Disease

Ahmed Najat Ahmed [1,a,*]
Amin Salih Kakshar [2,b]
Moayad Yousif Potrus [3,c]
[a,b,c] Department of Software and Informatics Engineering, College of Engineering, Salahaddin, University-Erbil, KR, Iraq

Article Information

Article History:
Received: May 22,2026
Accepted : July 7, 2026
Available online: August,2026

Keywords:
Parkinson's Disease, Graph Neural Networks, EEG, Chebyshev Convolutions, Partial Directed Coherence, Cognitive Assessment.

About the Authors:
Corresponding author:
Ahmed Najat Ahmed
E-mail: a.afandy@outlook.com
Researcher Involved:
Prof. Dr. Amin Salih Kakshar
E-mail: amin.mohammed@su.edu.krd
Asst. Prof. Dr. Moayad Yousif Potrus
E-mail: Moayad.potrus@su.edu.krd

DOI https://doi.org/10.17656/sjes.10222

Abstract

Cognitive impairment is a significant and debilitating feature of Parkinson's disease (PD) among its non-motor symptoms. However, the current clinical tools for assessing cognitive impairment, such as the Montreal Cognitive Assessment (MoCA) test, are subject to several limitations. To address these issues, we developed a deep learning model for predicting the cognitive status of patients with Parkinson's disease objectively and continuously from their resting-state electroencephalography (EEG) signals. We model the EEG signals as directed brain networks in the frequency domain using Partial Directed Coherence (PDC) and then process these network representations using a graph neural network with Chebyshev spectral convolutions and a self-attention mechanism. We rigorously tested our model using a dataset from 83 patients with Parkinson's disease and compared its performance with the state-of-the-art deep learning models. Our model outperformed the state-of-the-art models by reaching the highest performance in the Theta (4-8 Hz) and Alpha (8-12 Hz) bands. At the subject level, the model achieved a mean absolute error (MAE) of 0.34 points (0-30 MoCA) and R^2 of 0.98. These results suggest that the directed network topology of the EEG signals is a sensitive and physiologically meaningful feature for the prediction of cognitive impairment in Parkinson's disease.

1. Introduction

The hallmark symptoms of Parkinson's disease (PD) are the cardinal motor symptoms of tremor, rigidity, and bradykinesia; however, the long-term impact of the disease is now thought to be caused by non-motor symptoms, especially cognitive dysfunction. The spectrum of cognitive dysfunction in PD varies from mild executive dysfunction or visuospatial dysfunction to Parkinson's disease mild cognitive impairment (PD-MCI) and Parkinson's disease dementia (PDD). For many individuals, it is cognitive dysfunction, rather than motor symptoms, that is the main driver of loss of autonomy and diminished quality of life; it is cognitive dysfunction that causes burden on caregivers, hastens the need for institutionalization, and contributes to increased healthcare costs [1]. Measurement, quantification, and assessment of such changes, ideally at an early stage before the onset of symptoms, is an issue of great clinical need and research interest [2, 3]. In day-to-day clinical settings, cognitive status in PD is most assessed using brief tests such as the Montreal Cognitive Assessment (MoCA) [4]. While the MoCA is a valuable test that is more sensitive to cognitive dysfunction than other brief tests, it has well-known limitations that make it less than ideal for monitoring subtle changes over time. Specifically, motor dysfunction can affect performance on tests requiring fine motor control (e.g., writing or drawing); the test is scored on a coarse integer scale (0 to 30); and extraneous factors such as education level, language/cultural background, and medication state can affect performance. These limitations make it difficult to detect subtle changes over time. These limitations point to the need to develop objective

physiology-based biomarkers of cognitive status that are sensitive to changes over time with reduced extraneous influences [5,6]. Electroencephalography (EEG) presents itself as a promising tool for this aim. EEG is a non-invasive and inexpensive technique that measures neural dynamics at a millisecond scale, which is precisely the relevant time scale for many cognitive processes (e.g., shifts in attention, memory encoding) [7]. In addition to power spectral analysis, EEG allows for the examination of specific oscillatory features and network connectivity patterns, which provide insight into the nature of distributed communication processes underlying cognitive functions. In PD patients, cognitive impairment has been related to spectral slowing and changes in alpha and theta band power, as well as alterations in network connectivity. Such findings suggest a distributed nature of dysfunction rather than localized or channel-specific dysfunction. This provides support for the idea that useful biomarkers should be related to network properties rather than channel properties. A critical problem facing EEG network analysis techniques is separating real interregional communication from spurious interregional relations due to volume conduction effects. Partial Directed Coherence (PDC) overcomes this problem by estimating lagged and direct interregional relations. This technique provides a measure of effective connectivity as a function of frequency. This approach to network analysis is especially useful for PD patients since many cognitive deficits may be related to impaired top-down control processes. This means impaired communication from prefrontal control regions to posterior cortical regions. PDC focuses on directional relations among regions and so provides a physiologically relevant framework for examining the causal nature of networks involved in cognitive functions like attention, executive control, and visuospatial functions [8]. In order to utilize the benefits of the above-mentioned properties while at the same time addressing the analytical complexity associated with network data, the current study proposes an integrated deep learning framework that operates on the directed graphs constructed from the EEG signals [9,10]. The proposed framework operates on the directed brain graphs constructed by processing the resting-state EEG signals through the PDC method. The directed brain graphs constructed by the PDC method are then fed into the GNN that integrates the Chebyshev spectral convolutions and the self-attention mechanism. The use of the ChebConv layers allows the framework to aggregate the topological features of the brain networks in a multi-hop manner, capturing the disruption in the brain networks at both the local and global levels. The self-attention mechanism allows the framework to focus on the most relevant brain network features that are pertinent to the prediction task, thereby allowing the framework to offer interpretability in addition to the prediction performance [11,12]. The current study proposes the following three main contributions. Firstly, the current study proposes an integrated framework that operates on the directed brain graphs constructed by processing the EEG signals through the PDC method. The use of the directed brain graphs allows the framework to offer enhanced sensitivity to the information flow in the brain. Secondly, the current study proposes the GNN that integrates the Chebyshev spectral convolutions and the self-attention mechanism. The use of the GNN allows the framework to generalize the convolution operation to the non-Euclidean brain networks. The use of the self-attention mechanism allows the framework to offer interpretability in addition to the prediction performance. Finally, the current study proposes the use of the proposed framework in offering the continuous prediction of the cognitive status in a precise manner. The use of the proposed framework allows the prediction of the cognitive status in a precise manner, thereby providing a physiological alternative to the paper based. The remainder of the manuscript is outlined as follows. In Section III, we provide an overview of relevant research on EEG-based biomarkers for PD and deep learning for network neuroscience. In Section IV, we describe the data, preprocessing, graph construction, and GNN architecture. In Section V, we explain the experiments, validation, and results, followed by a discussion. In Section VI, we provide limitations and future directions.

2. Related Work

The motivation for the integration of EEG in the field of Parkinson's disease (PD) comes from the substantial gap in the current state of affairs, whereby, although the Montreal Cognitive Assessment (MoCA) test is used as a screening tool, it is, in fact, merely a "snapshot" approach, which is not appropriate for long-term monitoring. The MoCA test results have been found to be extremely unstable, as they can be easily affected by the patient's degree of motor rigidity, educational level, or the patient's state of medication, i.e., "ON" or "OFF." There is, thus, a blind spot in the monitoring of the neurophysiological state of the patient, where the onset of neurophysiological changes may not be observed until they become obvious. The use of EEG can provide the answer not merely because of the ease of use or the cost-effectiveness of the procedure, but because it can record changes in the neural activity at the exact frequency of the cognitive control process, i.e., in the range of milliseconds, which is the exact rate of the cognitive control process or the process of attention. The transition from subjective tests, such as the use of pen or vocalization, aims to provide a

continuous biomarker, which can monitor the cognitive state of the patient independently of the patient's ability to hold the pen or vocalize appropriately. Traditionally, the challenge was tackled through classical machine learning approaches, in which the model's success was largely dependent on the quality of the features engineered. Some of the early works in the area include those by Swarnalatha et al. [13], who implemented a pipeline using Gaussian Kernel Discrete Wavelet Transforms to feed features into a specialized CNN model with very high accuracy (92.31%) by explicitly defining the frequency bands of interest. In contrast, Desai et al. [14] emphasized the preprocessing of the data in the resting state condition and used Random Forests to extract features of interest like Hjorth parameters and alpha band activity. Motin et al. [15] adopted a similar approach but with Support Vector Machine classifiers, showing that theta band power is a strong classifier of the data. Although these works were seminal in showing the potential of machine learning approaches in the area, the approaches were limited in the sense that the model was forced to conform to the expectations of the modeler with respect to the features of interest, and the very high accuracy results reported were generally difficult to reproduce with larger and more heterogeneous populations beyond the initial dataset used for the study. The advent of Deep Learning (DL) marked an important milestone in the development of the field with the promise of end-to-end learning in which the model learns the features of interest automatically without human intervention. This was demonstrated by Oh et al. [16], who implemented a CNN model trained on raw data in the resting state condition, showing that dense features are unnecessary but also highlighting the risk of overfitting with small datasets. Subsequent work by Loh et al. [17] recast the entire problem of EEG analysis as a computer vision problem with the data represented as an image in the frequency domain with a 2D CNN model with notable separation of medication states. This was extended by Pei et al. [18], who used CapsNet with the ability to represent the spatial relationships in the data more robustly and showed that gamma band activity was a strong classifier of the data with respect to the cognitive state. However, this perspective overlooks the inherent properties of an EEG signal as a time series. Li et al. [19] and Lee et al. [21] proposed models by combining CNNs and LSTMs, while Xu et al. [20] used pooling-based recurrent networks. These models demonstrated that the temporal properties of brain activity are just as important as its localization. Although most studies used RS data due to its simplicity, Shi et al. [22] demonstrated that TE data can be used to extract more data, although at the cost of conducting cognitive tasks in a fragile PD population, which may limit its clinical applicability. More recent studies have expanded upon the scope of the signal used. For example, instead of using EEG data collected from awake patients, which is riddled with muscle artifacts, researchers such as Parajuli et al. [23] used polysomnography data collected during sleep and demonstrated that the microarchitecture of sleep stages can be used to identify robust biomarkers of cognitive decline. Other researchers have expanded upon the signal by instead of using frequency-based properties, examining "microstates"—quasi-stable patterns of whole-brain activity. Chu et al. [24] used CNN-RNN models to identify transitions between microstates, which can be linked to cognitive deficits. All of these models confirm what has been stated here: namely, that the neurophysiological signature of PD cannot be encapsulated by a singular biomarker but is instead a complex interplay between spectral, temporal, and topological disturbances. A more fundamental methodological change involves moving beyond the independent, isolated channels of the EEG. Cognitive decline in PD is not typically localized to one region of the brain; instead, it involves the disruption of complex systems and their long-range coordination. In this regard, thinking of the brain as a network, rather than just a collection of 19 or 64 independent signals, is an attractive idea. However, connectivity approaches for scalp EEG data also have some significant drawbacks, the first of which is volume conduction, which may give rise to false, zero-lag connectivity. To overcome this, approaches to connectivity analysis have increasingly focused on measures that are insensitive to instantaneous mixing. In this regard, the study by Santana et al. [25], which combined photic stimulation and Partial Directed Coherence, is an excellent example of this trend. Their high accuracy rates support the idea that the flow of information between brain regions, and not the activation of those brain regions, holds the key to disease signatures. Acharya et al. [26] reported similar impressive results using SPWVD and CNNs. However, while high accuracy rates, like those reported in the studies, may be an indication of the promise of deep learning, they may also be an indication of possible overfitting and data leakage, which is all too common in this field due to small sample sizes.This degree of skepticism has sparked a much-needed wave of reality-check papers focusing on robustness and standardization. In a comprehensive review, Maitín et al. [27] highlighted the fragmented nature of the current literature, where extreme heterogeneity in preprocessing, electrode count, and reporting standards makes it virtually impossible to compare between studies. Tejedor et al. [28] reinforced this in a systematic review, where they found that the studies yielding the highest accuracy used the least rigorous cleaning standards. The

disconnect between academic success and clinical need is further highlighted by Loh et al. [29] and Shaban [30], who found that despite the explosion in the number of AI models, the field is still hindered by the lack of external validation. When tested across multiple sites, where hardware and noise levels can vary, the accuracy of the model drops across the board. Quilez et al. [31] recently highlighted this in a study where they harmonized resting-state data across multiple international sites, yielding a more conservative accuracy than many of the previously published results. Diamandis et al. [32] took this a step further by training on one site and testing on an entirely independent site, further highlighting the issue of distribution as a major hurdle that most models have yet to tackle. From a practical point of view, the push for translation has also fueled debate over the level of data required. Clinics will not be able to accommodate high-density 128-channel configurations. Fortunately, studies by Shaban [33] and Kaasinen et al. [34] have demonstrated that good results can be achieved using lower configurations. These studies, which optimized channels, show that standard clinical montages, or smaller subsets of data, can retain significant diagnostic power if feature extraction is good. However, this reintroduces the trade-off between performance and interpretability. While deep learning methods may have superior accuracy, classical machine learning approaches [35–37] provide resilience through features like spectral power or entropy, which are easily interpretable by the clinician. A clinician is far more likely to trust a model that indicates “excessive theta slowing” than one which indicates a probability, calculated through some mysterious tensor operation, without any clear basis in the data. This analysis fills the void pointed out in the current dissertation. Most of the studies reviewed above are centered around classification: how do we determine whether the patient is well or ill? Is the patient ON or OFF their medication? However, the clinical management of Parkinson’s disease is not binary; it is a continuum. How much degeneration has the patient experienced from their last visit? How much has their disease progressed? Additionally, while several studies do use connectivity, few studies have maximized the non-Euclidean geometry of brain networks with modern Geometric Deep Learning techniques. We believe that to truly build a digital biomarker, we must move beyond classification and the limitations of grid-based image analysis techniques. As such, this dissertation proposes a method which: (1) models EEG as a directed graph with Partial Directed Coherence, (2) uses Graph Neural Networks with Chebyshev filters to model the brain network topology, and (3) seeks to predict a regression target, as opposed to a classification target, as the MoCA score is a continuous output, and this output is part of a complex, directed network.

3. Methods

In DL approaches to neurophysiological time series, the most common architectures have been inherited from those used in computer vision, particularly those based on convolutional neural networks (CNN). The success of CNNs in vision tasks relies on the assumption of inputs being located on a regular grid in Euclidean space; however, this assumption is highly inappropriate when dealing with EEG data. The assumption of a Euclidean grid topology when using the locations of electrodes as equivalent pixels or sequence steps imposes a grid structure on non-Euclidean, small-world networked data, as spatially distant electrodes may be strongly interconnected via long-range, directed pathways, and neighboring electrodes may record different sources. In research on Parkinson’s disease, cognitive impairments are thought to be associated with disrupted long-range, directed interactions, such as those within the frontostriatal network; however, the grid-like topology of EEG data may obscure these types of networked interactions. Furthermore, the lack of neurophysiological interpretability of many end-to-end, or "black box," models makes them difficult to apply in practice, as the specific regions or interactions of interest are unclear. In order to mitigate the aforementioned limitations, we propose the use of a geometric deep learning approach, where an end-to-end model is designed to predict the MoCA scores based on the resting-state EEG signal by modeling each epoch as a directed graph, where each node represents the EEG channels, each of which is represented by its raw feature vector, and the edges describe the frequency-specific, directed functional interactions between the EEG channels, rather than the physical proximity between them. The proposed model is thus more in line with the neurobiological hypothesis, as the cognitive status not only relies on the spectral characteristics of the EEG signal but also on the directed flow of the signal. The proposed model consists of the Chebyshev spectral graph convolution operation, the self-attention mechanism, and the dense regression head. The Chebyshev spectral graph convolution operation, denoted by ChebConv, is used to perform efficient spectral convolution on the graph by approximating the graph Fourier transform using Chebyshev polynomials, which can learn localized spectral filters on the graph, capturing the long-range dependencies between the EEG channels. Practically, this allows the network to learn the effect of perturbation from a frontal node to more posterior nodes through the medium of intermediate topological connectivity, which is a necessary capability for the modelling of dysexecutive, and posterior cortical symptoms associated with PD. The

final dense regression layer takes the weighted sum of the attended graph-filtered representations and outputs a continuous MoCA score. A key methodological decision was the definition of graph connectivity. We use Partial Directed Coherence (PDC), which is based on a multivariate autoregressive (MVAR) model, to compute frequency-specific graph connectivity matrices. PDC was chosen for two reasons. Firstly, it reduces the effect of instantaneous mixing (volume conduction) by focusing on lagged rather than zero-lag connectivity. Secondly, PDC naturally captures directionality, which aligns with the hypothesised mechanism for the development of executive symptoms in PD, which posits reduced frontal to posterior regulation. In summary, PDC allows learning from causal flows rather than coactivations. Operationally, the process works as follows: EEG data is first subjected to standard preprocessing, including rejection of artifacts, segmentation into short segments or epochs, and band-pass filtering. For each of the filtered segments, an MVAR model is estimated, as specified below, from which partial directed coherence (PDC) matrices are obtained for the frequency bands of interest. The directed adjacency matrices, along with the feature vectors associated with each of the channels, serve as the input graphs to the model. The ChebConv layers extract features from each of the band-specific graphs, which are then combined by the self-attention module to obtain a concise yet interpretable feature vector, which is finally mapped to a continuous MoCA estimate by the dense regression module. The approach thus trades off the sensitivity of detecting disruptions in the distributed, directed networks against computational complexity and interpretability. The following subsections describe the precise preprocessing pipeline, the MVAR/PDC estimation parameters, the ChebConv architecture, the training procedure, as well as the validation procedure.

3.1 Data Acquisition and Clinical Context

The study uses the publicly available data set provided by Singh et al. [39] from the University of Iowa. This data set was chosen because it provides high-density scalp electrophysiology data with simultaneous clinical cognitive measures and a significant phenotyped population with Parkinson's disease. From the entire data set consisting of 83 individuals with Parkinson's disease and 37 healthy controls, the study focuses exclusively on the PD population. This approach is chosen to concentrate on intra-disease variation in cognitive function instead of between-group discrimination. In this study, we aim to predict a continuous measure of cognitive function (MoCA score) from neural dynamics within the diagnosed population instead of creating a classifier to differentiate between groups.

The participants were selected and clinically characterized according to the study protocol by Singh et al. The participants with Parkinson's disease were diagnosed according to the UK Parkinson's Disease Society Brain Bank criteria. This approach guarantees a standardized and rigorous definition of the disease. Although the data set provides a healthy control population that is age-matched, gender-matched, and education-matched to the participants with Parkinson's disease, indicating a high standard of recruitment, the modeling framework used in this study relies exclusively on the participants with Parkinson's disease.

Cognitive status was measured using the Montreal Cognitive Assessment (MoCA), which was administered simultaneously with EEG collection. The MoCA score was treated as a continuous target variable (range 0-30). This framing of the MoCA task is appropriate given the translational goal of developing a physiology-based continuous index of cognitive function that captures subtle inter-individual heterogeneity, rather than discretizing the data into coarse categories [40]. The simultaneous collection of MoCA and EEG data ensures that the labels are contemporaneous with electrophysiology and therefore provide an accurate reflection of the clinical state at the time the data were collected.

Electrophysiological data collection was performed using a 64-channel system with channel placement according to the international 10-20 system and a sampling rate of 500 Hz. A high-pass filter at 0.1 Hz was used to reduce slow drift and DC offset artifacts. The task involved the estimation of either 3- or 7-second intervals following the presentation of a visual "Go" stimulus. This dataset was selected because of the specific task used: interval timing engages a wide network involving the frontal cortex and basal ganglia, which are affected in PD and are thought to be important in the execution of executive functions. This makes the task highly relevant to the cognitive functions being modeled.

As PD is a motor disorder, motor-related tasks may confound motor and cognitive functions. The task used by Singh et al. overcomes this by recording both the keypress and the subsequent keyrelease. In this analysis, we use the keypress data only. This minimizes the effect of motor execution variability on the cognitive labels and therefore ensures that the relationships between EEG and MoCA are due to the central cognitive processes rather than peripheral motor artifacts.

In summary, the dataset provides three methodological advantages, each of which is relevant to the current study:

The combination of the large, well-characterized Parkinson's disease (PD) sample with high-density,

high-sampling-rate EEG provides the necessary spatial and temporal resolution for network-level, frequency-specific modeling (e.g., band-resolved connectivity, spectral graph convolutions).

The contemporaneous availability of Montreal Cognitive Assessment (MoCA) scores facilitates the learning of a continuous clinical phenotype rather than a binary outcome, enabling the derivation of an interpretable, graded index of function.

The use of an internally consistent task design targeting frontal-basal ganglia timing circuits maximizes the biological plausibility of the hypothesized EEG signatures of cognitive decline.

However, the necessary context of the results requires careful specification. The focus on the PD cohort alone maximizes the sensitivity to heterogeneity within the disease, although it should not be assumed that the learned mappings extend to healthy aging or any other disorder. The clinical rigor of the dataset provides an appropriate context for the proposed graph-spectral, directionally informed modeling framework, although the transportability of the proposed markers to independent populations, as well as the healthy controls in the original dataset, is a topic of future study.

3.2 EEG Preprocessing and Graph Representation

The process of transforming scalp EEG signals into a topological form suitable for geometric deep learning techniques is associated with strict signal conditioning procedures [41]. In addition to common environmental interference, EEG signals recorded from PD patients are prone to motor artifacts, such as rest tremors and rigidity-induced electromyographic signals. These artifacts are particularly problematic when analyzing EEG signals, as large electromyographic signals can produce false correlations and/or mask the flow of information between brain regions. To address this issue, the proposed pipeline focuses on the separation of neural oscillations from physiological artifacts before any topological modeling takes place. The process begins with transforming the raw multichannel EEG signals into clean, quasi-stationary epochs, which are then transformed into frequency-specific directed graphs to be used as inputs to the deep learning model.

3.2.1 Preprocessing

Preprocessing of the data was performed using MATLAB software with the EEGLAB toolbox [42]. The preprocessing steps were designed to systematically separate the neural signals from physiological and environmental artifacts while maintaining the necessary phase information for directed connectivity.

Firstly, power line artifacts were removed using a notch filter at 50 Hz with a bandwidth of 1 Hz. The zero-phase forward and backward filtering strategy was necessary to avoid signal distortion and preserve the necessary timing information across channels for accurate causal estimation using Partial Directed Coherence.

Next, a band-pass filter from 0.1 to 40 Hz was applied to the data. The low-pass cutoff of 40 Hz reduced electromyographic artifacts, which are known to be prevalent in Parkinson's populations. The range of this filter captures the canonical bands of interest: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz).

Finally, to address volume conduction effects inherent in the original Pz reference, the data were referenced to the Common Average Reference. The CAR reduces topographical bias by subtracting the mean global activity from individual channels. The effect of this operation is to approximate a zero-potential reference, thereby increasing the topographical specificity of the data for graph analysis. The Pz channel was reconstructed to maintain the original configuration.

Channels where the signal-to-noise ratio was consistently low were not used a priori to standardize the results. For example, the frontopolar channels, Fp1 and Fp2, were not used because they were dominated by eye potential. Similarly, the inferior temporal channels, FT9, FT10, TP9, and TP10, were not used because they were highly susceptible to EMG activity from the jaw and temporalis muscles.

The remaining data were visually inspected on a subject-by-subject basis. Gross non-physiological noise in the channels was interpolated using spherical splines, which ensured the spatial characteristics of the EEG signal were not lost, while the data were not discarded. The EEG signal was segmented, and epochs where transient levels of noise were high were automatically rejected.

Independent Component Analysis (ICA) was employed to minimize the artifacts that were spectrally similar to neural signals. The multichannel signal X was decomposed into statistically independent components S using the Extended Infomax ICA algorithm. The linear mixing model is represented as $X = A S$. The independent components related to eye blinks, lateral saccades, cardiac fields, or muscle noises were identified and eliminated. The clean EEG signal was obtained by back-projecting only the independent components that were related to the cortical sources.

Next, the continuous EEG signal was segmented into 1-second epochs that were non-overlapping. The duration of the epochs was fixed at 1 second to satisfy the stationarity assumption of the multivariate autoregressive (MVAR) model. The epochs were also sufficient to obtain a good estimate of the parameters. For the peak interval task, the duration of 3 seconds contained 12 epochs, and the duration of 7 seconds contained 22 epochs. Each epoch of the EEG signal was considered independent and labeled with the

subject's MoCA scores. The epochs were used to create a data set of high-quality EEG signals with minimized artifacts.

3.2.2 Graph Representation

After the data was preprocessed, the multichannel data was transformed into an appropriate format for geometric deep learning. Although traditional neural networks operate on Euclidean spaces such as grids or sequences that impose a notion of adjacency or locality, the brain is a non-Euclidean space where the notion of functional interactions is based on distributed and directed forms of communication rather than proximity. This was achieved by explicitly representing each epoch of the EEG as a directed graph [44]. In other words, each epoch of the EEG was represented as a graph G=(V, E), where the nodes encode local neural dynamics and the edges encode directed functional interactions. This design captures two complementary dimensions of brain activity: the temporal content of individual regions and the global topology of information flow across the network. Such a dual representation is particularly well suited for probing the distributed network failures underlying the dysexecutive and posterior cortical cognitive phenotypes observed in Parkinson's disease. The Set of Nodes V represents the preserved EEG channels. Instead of using a collection of channel signals to compute hand-engineered summary features, an end-to-end learning approach was adopted. The feature vector at each node is composed of the raw EEG signal, which is a 500-point time series representing the voltage fluctuations over the 1-second epoch. The raw signal was used to allow the model to discover intrinsic temporal filters and take advantage of fine-grained phase and amplitude information, such as transient oscillatory bursts and cross-frequency effects, which may not be preserved by hand-engineered features. The set of edges E represents the directed functional connectivity between the EEG channels. While undirected functional connectivity methods such as correlation and coherence are often used, they assume symmetric functional interactions between the channels. However, this does not reflect the directed nature of cognitive control. For example, executive functions involve top-down modulation from the frontal to the posterior cortex, and the loss of this directionality may be a contributing factor to cognitive decline in PD. Therefore, the edges are defined using the Partial Directed Coherence (PDC) algorithm, which is a frequency-domain approach to Granger causality analysis [45]. The multichannel EEG signal was modeled as a multivariate autoregressive (MVAR) process:

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + E_t$$

where X_t is a vector of channel amplitudes at a given time t, A_i are lagged coefficient matrices that account for inter-channel influences, p is the model order, and E_t is a white Gaussian noise process. The model order was optimized on a per-epoch basis using Akaike's Information Criterion; the criterion usually converges to a range of p = 5-10. The MVAR model's estimated coefficients were further used to calculate the PDC. The PDC from source channel j to target channel i at frequency f is given as

$$\mathrm{PDC}_{i\leftarrow j}(f) = \frac{|a_{ij}(f)|}{\sqrt{\sum_{k=1}^{M} |a_{kj}(f)|^2}}$$

where $a_{ij}(f)$ is the Fourier transform of the MVAR model's coefficient, and M is the number of channels. This normalizes the overall outgoing influence from channel j to all other channels to unity; the PDC can be viewed as a proportion of the influence from channel j on channel i at a given frequency f [46].

The connectivity was calculated for the delta, theta, alpha, and beta bands separately, since it is known that different cognitive processes are associated with separate spectral channels. This led to a collection of frequency-specific adjacency matrices $A^{(\text{"band"})} \in R^{(M\times M)}$, within each frequency band. To ensure numerical stability of the graph convolution operation, each adjacency matrix was row-normalized, which also led to a bounded spectrum of the graph Laplacian. To conclude, the process transforms the continuous EEG signals into a specific set of directed brain graphs at different frequency bands. It maintains the raw time dynamics at the node level while incorporating the causal frequency-specific topology at the edge level. It provides a faithful input for the subsequent graph neural network, which can learn sensitive biomarkers for cognitive impairment in Parkinson's disease patients.

3.3 Integrated Deep Learning Architecture Overview

To overcome the challenges of the black-box models, the authors developed an end-to-end model specifically designed for the task of learning the brain's structure and function. It combines the Chebyshev Spectral Graph Convolutional Networks (ChebConv) layer, the Self-Attention mechanism, and the Dense Regression layer. It aims at predicting continuous cognitive scores from the raw EEG signals, meeting the clinical needs while addressing the computational challenges in neurophysiological signal processing. It is based on the Geometric Deep Learning framework, which is specifically designed for data defined on non-Euclidean spaces, typically graphs. CNNs are good at handling Euclidean spaces, but they cannot handle the complex topology of brain activity, where functional connectivity can occur across the brain regardless of the spatial location of the regions [47]. Instead, the authors model the EEG

data as a graph, where the nodes are the EEG channels and the edges are defined by the functional connectivity using Partial Directed Coherence (PDC). The structural backbone of the network is the Chebyshev Spectral Graph Convolution (ChebConv). The common requirement of performing computationally expensive eigen-decomposition on the Graph Laplacian is alleviated in the ChebConv by approximating the spectral graph convolution using polynomials. This allows for the efficient learning of spatial filters on the graph. These layers serve as feature extractors that propagate information across interconnected nodes [48]. This framework allows for the incorporation of information from both immediate neighbors and remote but functionally interconnected areas, thereby capturing the disrupted network connectivity associated with cognitive impairment in PD. The Self-Attention Mechanism is then used to account for the fact that not all brain areas equally contribute to cognitive states. For example, in cases of dysexecutive syndromes, the disruption of front striatal circuits may be more predictive than disruptions in other areas. The mechanism assigns dynamic importance weights to different EEG channels according to their relevance to the predicted cognitive state. By focusing the computation on the most relevant areas, the performance of the network is improved, as is the interpretability of the results, allowing clinicians to pinpoint the areas that are most relevant to the prediction of cognitive decline.

The architecture ends with a Dense Regression Layer that is intended to yield a precise score on the 0-30 range of the Montreal Cognitive Assessment (MoCA) test. Contrary to prior studies that have reduced the notion of cognitive status to a binary classification problem, the proposed framework considers the phenomenon of cognitive decline to be a continuous process. The regression layer is optimized to yield a precise score on the 0-30 range of the MoCA test.

The flow of data through the architecture is as follows: The input graph, consisting of the raw EEG signal on the nodes and the edges defined by Partial Directed Coherence (PDC), is passed through a series of ChebConv layers that improve the features by aggregating information from the neighboring nodes. Next, the Self-Attention mechanism is used to calculate a weighted sum of the features based on the relevance of the information. The resulting feature is flattened and passed through the Dense Regression layer. This integrated architecture leverages the complementary strengths of each of its constituent parts: the topological capabilities of Chebyshev Polynomial Graph Convolutions, the dynamic focus of Self-Attention, and the precision of Dense Regression. This model, which leverages local and global features of brain networks, provides a comprehensive and interpretable framework for elucidating the neural basis of cognitive impairment in PD. This end-to-end model not only improves the accuracy of predictions; it also provides insight into specific brain regions and network interactions that inform the development of more effective tools for PD management.

3.4 Mathematical Formulation of the Architecture Modules

The proposed deep learning architecture is a unified, end-to-end model meticulously crafted to capture the intricate organizational principles of the human brain. By combining the capabilities of Chebyshev Polynomial Graph Convolutions, Self-Attention, and Dense Regression, we have developed a model with the ability to traverse the non-Euclidean, directed topology of brain networks. This architecture has been specifically designed to capture the intricate neural patterns underlying cognitive dysfunction in PD, which often lie at the limits of detection for traditional linear models or conventional Convolutional Neural Networks. This architecture follows a paradigm of "structure-to-function-to-cognition." The ChebConv module is the structural interpreter, which processes the directed graph-based representation of the EEG data to determine how local neural activity propagates along these directed functional pathways. The Self-Attention module is the functional filter, which dynamically selects the most salient network nodes for the task at hand. Finally, the Dense Regression Layer integrates these high-level abstractions to make a precise, continuous prediction of the Montreal Cognitive Assessment score. The mathematical foundation of these components is provided in the subsequent subsections, ensuring both conceptual clarity and scientific reproducibility.

3.4.1 Chebyshev Polynomial Graph Convolution (ChebConv)

The Chebyshev Polynomial Graph Convolution (ChebConv) is the main spatial feature extraction component of our proposed model [49]. In relation to Parkinson's Disease, cognitive deterioration is associated with a "disconnection syndrome," which reflects a disconnection of long-range connections between frontal executive areas and posterior sensory areas. Therefore, it is essential that a model can capture context across an entire network. Conventional Graph Convolutional Networks (GCNs) only allow information aggregation over local neighbors, thus necessitating an impractically large number of layers to effectively capture long-range connections, which often results in over-smoothing. To mitigate these limitations, we use a ChebConv formulation, which is suitable for Directed Graph Signal Processing. A truncated Chebyshev polynomial expansion of order K=5, is used, thus enabling information aggregation over a large neighborhood in a single layer. Moreover, since our

graphs are constructed using Partial Directed Coherence (PDC), information flow directionality is preserved, thus enabling the model to differentiate between top-down executive control and bottom-up sensory processing mechanisms, even without computing computationally expensive eigen-decompositions. The ChebConv block's input is a directed graph constructed from a one-second EEG segment. The initial node feature matrix, H^((0))∈R^(M×500), simply represents the raw EEG signal samples. The use of raw data instead of hand-engineered features allows ChebConv's layers to effectively learn optimal temporal filters from the data. To apply convolutions on the directed PDC graphs, a normalized version of the Laplacian operator is constructed. For a given adjacency matrix A, a symmetric normalization scheme is used by normalizing the off-diagonal elements. Therefore, if there is a directed edge from node j to node i, then the normalized edge weight is computed as:

$$\widehat{A_{ij}} = \frac{A_{ij}}{\sqrt{D_{ii}\, D_{jj}}}$$

In this definition, D_{ii} and D_{jj} refer to the degree of the respective nodes as determined from the adjacency structure. The operator is also rescaled in a manner suitable for the approximation of the Chebyshev polynomial. The standard framework of Spectral Graph Theory relies on the symmetric nature of the Laplacian operator with eigenvalues within the range of -1 and 1. We adapt the Spectral Graph Theory framework for the case of directed and possibly asymmetric graphs through the normalization of the operator with respect to the maximum edge weight magnitude, (λ_{max}). A self-loop of weight -1 is also added to the diagonal of the operator, and the rescaled operator is given by the expression $L \approx 2/\lambda_{max}\, L\text{-}I$. This rescales the operator as a form of numerical stabilization, facilitating the use of the Chebyshev recurrence as a high-order spatial filter.

Formally, the definition of the convolution is expressed as the weighted sum of the signals passed through the operator.

$$H^{(l+1)} = \sigma\left(\sum_{k=0}^{K-1} T_k(\hat{L}) H^{(l)} W_k^{(l)}\right)$$

The expression $H^{(l)}$ represents the node feature matrix at the respective layer, and the matrices $W_k^{(l)}$ refer to the weight matrices corresponding to the respective polynomial orders. The definition allows the network to apply different weights to the signals from the neighboring nodes and those from the more distant causal antecedents. To adequately capture the depth of neural processing, nine ChebConv layers are used in the feature extractor stack. This architecture is intended to progressively compress temporal information while expanding spatial context. In the early stages, the feature dimension is kept high (d=500), reflecting the importance of integrating raw temporal waveform data with its immediate causal neighbors. In subsequent stages, the feature dimension is progressively reduced to 250, forcing the model to distill the signal into more abstract and robust representations of network motifs. In the final stages, the feature dimension is reduced further to d=100. To ensure effective training of such a deep network without the risk of vanishing gradients, several architectural enhancements were incorporated. Identity shortcuts, or residual connections, were incorporated wherever the input and output feature dimensions of a given layer matched, viz., between layers 1 and 2, 4 and 5, and 7 and 8. These connections act as information superhighways, allowing gradients to propagate unimpeded during backpropagation. In addition, batch normalization was applied immediately after each convolution to stabilize the learning process, while a dropout rate of p=0.2 has been used after the ReLU activation in each layer to prevent overfitting. Finally, the output of the ChebConv module is a feature matrix $H^{(9)} \in R^{(M\times100)}$, ready to be used as input for the next step, i.e., the attention mechanism.

3.4.2 Self-Attention Mechanism

Following the topological feature extraction of the ChebConv layers, we used a Self-Attention Mechanism to improve the feature representations of each node, as described in [50,51]. In fact, although the ChebConv module is excellent in extracting the topological features of EEG signals by aggregating information based on the topological connectivity of the graph, given by the static connectivity of the Partial Directed Coherence (PDC) graph, it treats neighbor influences within its receptive field equally, without any further modulation. However, it is well known that cognitive control is a dynamic process, and the brain does not always utilize all anatomical connections equally for every cognitive task. Therefore, to better mimic these biological processes, we used an attention mechanism that allows the model to adaptively "spotlight" specific EEG channels that are more informative for MoCA score prediction.

The use of the attention mechanism resolves the critical limitation of sequential processing models, such as the Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, traditionally used in the analysis of the EEG signals. Although the models are successful in sequential processing, they enforce a sequential structure on the brain's topologically spatial structure. The sequential nature of the models means that the signal will pass through the intermediate nodes from the “frontal” node to the “posterior” node [52], and the gradient will vanish during the process. The Self-Attention model works as a parallel system, and any two nodes

in the graph will be able to communicate with each other, irrespective of the distance from the scalp or the position of the nodes in the array. This method, in essence, reduces the effective distance between any two distant cortical regions to unity, thereby allowing for long-range functional couplings without the need for sequential processing. This global receptive field is of particular importance for modeling Parkinson's Disease, as it is increasingly viewed as a 'disconnection syndrome,' or pathology of functional decoupling between distant cortical hubs, including the frontoparietal network controlling cognition.

The input to this block is given by the node feature matrix $H^{(L)} \in R^{(B\times N\times 100)}$, derived from the final ChebConv layer, where B is the batch size, N is the number of EEG channels, and 100 is the embedding dimension. A Scaled Dot-Product Attention mechanism is used, conceptually equivalent to a differentiable retrieval mechanism, wherein every node queries the network to retrieve relevant information. For each node embedding, three different linear projections are computed via learnable weight matrices W_Q,W_K,W_V: the Query (Q), or the feature set that the node queries for; Key (K), or the feature set that the node itself has; and Value (V), or the actual informational content that is being propagated. Attention scores are given by the dot products between Queries and Keys. A high score indicates that there is considerable functional alignment or relevance between two regions. These are then normalized via the softmax function to produce a probability distribution:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Here, the term $1/\sqrt{d_k}$ is seen as a critical scaling factor. This scaling factor is used to prevent saturation of the softmax, as high dimensionality can cause large magnitudes of the dot products, resulting in undesirable behavior during backpropagation.

Despite the pervasiveness of Multi-Head Attention within Natural Language Processing for modeling diverse linguistic relationships, our architecture is intentionally restricted to Single-Head Attention architecture. In the context of a clinical biomarker, interpretability is of primary importance, and a multi-head architecture would result in multiple conflicting "importance" maps, obscuring the underlying physiological signal. Instead, the single attention head is forced to learn a single, definitive "map" of spectral importance. The resulting weights provide a clear and visualizable metric of which brain regions the model is prioritizing when generating a prediction of cognitive impairment, effectively "highlighting" physiological deficits such as the frontostriatal disconnect of dysexecutive Parkinson's disease.

The resulting output of the attention weights is a weighted sum of the Values, providing a context-aware representation of each node. To utilize this enriched information, the output is passed through a Position-wise Feed Forward Network (FFN). The FFN is a fully connected layer with 64 hidden units and uses a ReLU activation function to allow the model to learn complex non-linear combinations of the features. To prevent overfitting, a dropout layer is used with a dropout rate of 0.1. Finally, the output is passed through a Layer Normalization layer. Unlike the more common Batch Normalization, Layer Normalization normalizes features for each individual brain graph. This is a critical component of the model, allowing it to adapt to the significant variability between brain recordings from different patients due to the variability of global signal amplitude between patients. The output of this block is size B ×N ×100, yet the information represented is fundamentally changed. The embeddings no longer represent a localized state; instead, the state is a context-aware cognitive state, suppressing irrelevant signal noise and enhancing relevant signal hubs.

3.4.3 Dense Regression Layer

The final component of the integrated deep learning model is the Dense Regression Layer, which serves as a nexus of decision-making. While the previous ChebConv and Self-Attention blocks have been designed with the goal of preserving and analyzing the brain's topological structure as a complex network of interrelated nodes, the main goal of the current study is to simplify this complex, high-dimensional representation down into a single, actionable metric: the predicted MoCA score. This requires transitioning from a spatially distributed, map-based representation of cognitive state towards a more global, holistic synthesis of the patient's cognitive state. To accomplish this, it is necessary for the Dense Regression Layer to perform the non-trivial regression task of relating subtle, non-linear variations in neural network dynamics to the continuous range of 0-30 possible MoCA scores. The transformation process begins with the integration of channel-specific features into a unified global representation. The output tensor of the Self-Attention module, $H^{attn} \in R^{(M\times 100)}$, where M is the number of EEG channels and 100 is the dimension of the features, maintains the spatial boundaries of brain regions. Nevertheless, cognitive impairment is usually a product of complex global interactions across the cortical surface, as opposed to regional brain activity. This is addressed by a flattening operation, which concatenates the features of all the channels into a single vector $v_{flat} \in R^{(M.100)}$. This process, in essence, lifts the spatial boundaries, allowing the fully connected layers to learn the high-order correlations between any feature of any channel, hence considering the brain state. The high-

dimensional vector is then projected onto a lower dimension using a fully connected layer with 256 hidden units. This architecture forces the model to distill the raw information into a compact set of robust features, filtering out the noise while retaining the patterns indicative of cognitive decline. To address the non-linear relationship between functional connectivity and clinical status, a Rectified Linear Unit (ReLU) activation function is applied. The transformation is formally defined as:

$$h_{\text{dense}} = \text{ReLU}(W_1 v_{\text{flat}} + b_1)$$

where W_1 and b_1 represent the learnable weights and biases, respectively. To ensure robustness and generalization with the limitations of the data size often associated with clinical neuroimaging data, a dropout regularization rate of p=0.2 is used immediately after the activation. The random setting of a fraction of neurons to zero at each training step prevents over-reliance on individual features of the data and instead encourages a more distributed representation of the data. The regression output is then obtained through a second fully connected layer comprising a single unit. Unlike binary classification problems, which often require sigmoid or softmax activation functions, linear identity activation is used to allow for a continuous output value. The prediction is then given by:

$$\hat{y} = W_2 h_{\text{dense}} + b_2$$

Where $\hat{y}$ is the predicted MoCA score. The use of a linear activation allows the model to cover the entire range of the dynamic scale of the cognitive scale and prevents gradient saturation at the extremes of the data distribution. The pipeline is now complete to ensure that the final prediction reflects a structured and biologically plausible processing pathway to accurately translate the electrophysiological topology to a precise and continuous measure of cognitive health.

3.5 Training, Optimization, and Performance Evaluation

To ensure the scientific validity and clinical applicability of the proposed deep learning framework, a rigorous and comprehensive framework for optimization and evaluation has been established. While the construction of a biomarker for Parkinson's Disease requires more than just minimizing loss, it also requires the implementation of a validation strategy that addresses the hierarchical nature of the data, in which multiple temporal segments are embedded within the subject, as well as the inter-individual variability inherent in neurodegenerative disease populations. To this end, we have established a computational ecosystem in which the specific software libraries employed to interface the intersection of signal processing and graph theory, as well as the specific hierarchy and organization of the high-dimensional data tensors and hardware acceleration strategies employed to optimize the computational efficiency of spectral graph convolution operations, are described. Graphic baseline for use in architectural ablation. For our experiments, we adapted two other graph baselines that do not include any of the two essential design elements mentioned in the proposed model – directed edges and Chebyshev multi-hop spectral filtering – and trained them with the same number of layers, the same subject-disjoint 5-fold split, and the same node features as our proposed model to isolate the contributions of the two. The first baseline is the basic Graph Convolutional Network (GCN) based on Kipf and Welling. The directed PDC adjacency was symmetrized, Asym=(APDC+APDCT)/2, before being fed into the GCN, meaning that all connections seen by the model are undirected and only one-hop neighbors per layer. This baseline is free of directionality and spectral expansion due to multiple hops. The second baseline is a Graph Attention Network (GAT) which also uses the symmetrized adjacency but has 8 attention heads in early layers. The GAT maintains the concept of attention, while continuing to work on one-hop, undirected neighborhoods. This baseline removes directionality, Chebyshev expansion and keeps attention. If the proposed Directed Chebyshev-Attention model, which is based on both the directed connections from PDC and multi-hop Chebyshev filtering, is successful compared to both of these graph baselines, the improvement does not come from "using a graph in general" but it comes from using a graph with directed connections and using the multi-hop Chebyshev filtering.

3.6 Implementation Details

The creation of such an end-to-end solution that facilitated the seamless transition from raw voltage time series to graph-based deep learning models required the orchestration of a hybrid software solution that would leverage the signal processing precision of MATLAB and the flexible learning potential of Python. The early stages of the processing pipeline were based on the MATLAB platform, utilizing the EEGLab toolbox, which is a de facto standard in clinical electrophysiology. This platform was used to ensure that the fundamental signal processing hygiene, such as artifact rejection, filtering, and interpolation, met strict neurophysiological guidelines before any machine learning processing occurred. Notably, the computation of Partial Directed Coherence (PDC) matrices occurred on this platform, as it has been found that, owing to the complexity of MVAR model estimation on high-density EEG datasets, optimized linear algebra functions provide the stability required for the computation of frequency-specific adjacency matrices that define the edges of the graph. Following

feature extraction, the processing pipeline shifted to the PyTorch platform for the development of the deep learning model architecture [53]. This platform was selected owing to its support for dynamic computation graph-based learning, allowing for fine-grained control of backpropagation, which is advantageous for optimizing parameters of the attention heads and regression layers. To support the necessary computations involving non-Euclidean structures, the PyTorch Geometric extension library for implementing Graph Neural Networks was incorporated. PyTorch Geometric was found to be invaluable for the efficient computation of the proposed Chebyshev Spectral Convolutions [54]. The computation of the Chebyshev polynomials requires the execution of sparse matrix-vector products and the corresponding scatter-gather operations necessary for the propagation of the information through the irregular connectivity of the brain network. Leveraging the optimized geometric computations of PyTorch Geometric, the proposed high-order spectral filtering (K=5) was efficiently implemented without the prohibitive computational cost of executing the eigen decomposition of the Laplacian matrix at every layer. To bridge the gap between the two computational domains, the dataset was carefully organized into aligned tensors for the purpose of mini-batch training. The raw EEG signals were ingested as a large three-dimensional tensor from the dataset, with dimensions B ×M ×T, where B represents the total number of 1-second epochs of the entire dataset (the effective sample size), M represents the number of nodes or the EEG channels, and T represents the temporal node features or the raw voltage measurements at a sampling rate of 500 Hz. One of the major innovations in this implementation is the way in which graph topology is treated as an evolving and changing entity. To this end, we constructed an analogous tensor of adjacency matrices, whose dimensions were B ×M ×M. This allows for the possibility of the model being able to detect temporary disruptions in PDC connectivity related to fluctuations in cognitive states. To ensure that the ground truth clinical labels were aligned with this segment-based approach to training the model, a label propagation approach was taken in which the target tensor has dimensions of B ×1. This allows the single MoCA score for a given patient to be broadcast to every EEG segment derived from this patient, in effect framing the learning problem as one in which the global cognitive state of the patient is predicted based on a single snapshot of brain activity. Finally, to ensure that the computationally intensive operations involved in spectral graph convolution and the quadratic complexity of the self-attention mechanism were addressed, the entire training process was run on a high-performance computing cluster with NVIDIA GPUs. This allows for the parallelization of tensor operations over thousands of CUDA cores and ensures that the complex polynomial filtering and message-passing operations of the network can be run at sufficient speeds to allow for extensive hyperparameter tuning and cross-validation.

3.7 Training and Optimization

A well-structured training and optimization process is the backbone of any robust deep learning application, especially in the domain of biomedicine, where the nature of the available data is typically high-dimensional, noisy, and limited in quantity. The training of the model was also a departure from the regular execution of standard algorithms and included a statistically well-founded validation approach and hyperparameter tuning to effectively navigate the loss curve of the GNN model. The subject-disjoint 5-fold cross validation scheme with the Group K Fold, which is used to eliminate the risk of identity leakage that affects resting state EEG. The 83 PD patients were divided into 5-fold sub-sets of about, 16–17 patients. In each fold, all one second epochs from a subject would either be in the training set or in the validation set but would not be split between the two. Random shuffles at epoch level were thus explicitly avoided. This configuration reflects the actual clinical scenario, where a trained model is required to predict the cognitive state of a new patient who it doesn't know, and this prevents the network from simply learning to recognize individual patients' "neural fingerprint" and general markers of cognitive impairment. Label propagation is completely consistent with this scheme, because the propagation is internal, it does not cross the fold boundaries. The reported R2 values and MAE are then indicative of generalization to unseen patients, rather than memorization of subject signatures. The optimization algorithm used was another key design choice, driven by the necessity to manage the unsteady, sparse gradients of graph-based architectures. While Stochastic Gradient Descent (SGD) and its variants with momentum have strong convergence guarantees, they often struggle with the saddle points and oscillatory gradients present in deep, non-linear models. Similarly, though RMSProp overcomes the difficulty of learning rates, it has been shown to be less stable during the later stages of convergence. As such, we chose to use the Adam optimizer, which combines the strengths of both methods by maintaining adaptive learning rates for each parameter based on the first two moments of the gradients [55]. This strategy proved particularly beneficial for our graph neural network, especially since certain nodes tend to update frequently, while others tend to update less often. Adam's adaptive update rule guarantees that less frequent features do not become overshadowed by dominant features.

Moreover, Adam's update rule includes a bias correction term, which accounts for the initialization of the running average at zero. This improves the stability of the trajectory during the early training epochs. The learning rate was initialized at α=1×10^(- This was determined to be a good balance between convergence rates and stability for optimizing the spectral filters. The learning objective was formalized to minimize the Mean Squared Error (MSE) between the predicted MoCA scores, denoted by (y,)ˆ and the ground truth clinical scores, denoted by y. The loss function is defined as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where N represents the number of EEG segments contained within a training batch. MSE was selected over Mean Absolute Error (MAE) due to its quadratic penalty, which heavily penalizes large prediction errors. In a clinical regression scenario, large prediction errors can be significantly detrimental compared to small ones. The squaring operation heavily penalizes large outliers, forcing the backpropagation algorithm to focus on correcting egregious misclassifications, thus ensuring a robust model across the cognitive spectrum. The final model architectural configuration and training dynamics were arrived at through an extensive grid search process, seeking a "Goldilocks zone" that is complex enough to capture non-linear neural dynamics while avoiding overfitting by ensuring model generalizability. A batch size of 32 was used, representing a balance between gradient estimation accuracy and noise. While large batch sizes provide more accurate gradient estimates, they tend to converge towards "sharp minima" that generalize poorly. A moderate batch size introduces a controlled level of noise, which helps the model generalize better by avoiding local minima, finding flatter minima instead. Residual connections were strategically added between ChebConv layers (1-2, 4-5, 7-8) to ensure gradient flow through the deep network, avoiding vanishing gradient problems. To reduce the memorization of the training data, a tripartite regularization strategy was employed. This includes, first, the application of Batch Normalization after each convolution to reduce the issue of internal covariate shift. Second, Dropout with a probability of 0.2 was used after the ChebConv, Attention, and Dense layers to force the model to behave like an ensemble of multiple models, thus preventing any two neurons from co-adapting. Finally, the application of Early Stopping with a patience of 20 epochs was used, which acts as a temporal regularizer, continuously tracking the validation loss of the model. As a result, training is stopped at the precise epoch before which the model's generalization ability begins to degrade.

3.8 Performance Evaluation

The validation of a machine learning model, particularly one with clinical applications, calls for a more sophisticated validation procedure than simple accuracy, especially for a model dealing with a condition like Parkinson's Disease, where management of the disease is dependent on precise quantification of cognitive decline. Considering the high cost of prediction error in a medical diagnostic tool, a multi-faceted evaluation strategy has been developed that works on two different levels of granularity, signal level, and patient level. Since the model is based on a deep learning architecture that works on 1-second EEG epoch, it provides a continuous stream of high-frequency predictions for a given subject. However, to transform these predictions into a clinically applicable phenotype, a deterministic evaluation strategy has been adopted, wherein the average of the prediction of each segment is used to provide a robust subject-level prediction. This "wisdom of the crowd" strategy essentially reduces stochastic noise and anomalies that may be present in any given EEG segment, thereby providing a stable prediction that reflects the patient's long-term cognitive state. To assess the efficacy of this strategy, three regression metrics are used, and results are provided as Mean ± Standard Deviation over five cross-validation folds. This is critical for demonstrating that the model's predictive reliability does not vary based on any specific data set but is consistent over different subsets of the population. The most important criterion of clinical utility used in this evaluation is Mean Absolute Error (MAE), as it is easily interpretable in terms of the same units as the Montreal Cognitive Assessment (MoCA). Unlike "loss" as a concept, MAE is concerned with answering the basic clinical question: "On average, by how many points does the AI prediction deviate from the standard neuropsychological test?" Mathematically,

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

if y_i represents actual MoCA score, and (y_i)ˆ represents predicted score, MAE is a linear measure of error. MAE does not distinguish between over- or under-estimation, so it is symmetric. It does not use squared errors, so it is insensitive to extreme values. MAE thus provides a fair assessment of model accuracy that is representative of the typical patient population. Nevertheless, though MAE is a good measure of average performance, a clinical safety-critical system must also demonstrate robustness to catastrophic failures. For this reason, Root Mean Squared Error (RMSE) was used to guarantee robustness. This measure has the effect of applying a quadratic penalty to errors before they are averaged, which is mathematically expressed by:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

The quadratic penalty has the effect of greatly magnifying errors, thus ensuring that we highlight cases where the model fails to capture the state of the patient. By using both MAE and RMSE, we guarantee not only that our model is clinically precise, but also robust to outliers, thus minimizing the possibility of catastrophic failures. Lastly, to validate the scientific hypothesis that functional connectivity patterns correlate with behavior, we used the Coefficient of Determination, R^2. While error-based metrics measure distance, R^2 on the other hand, is a measure of the model's ability to explain the true scores. R^2 is mathematically expressed by:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where $\bar{y}$ is the average of the true scores. A high R^2 value, i.e., approaching 1.0, implies that the graph topology features extracted by our deep learning architecture explain the true cognitive scores. Thus, we establish a comprehensive test of our model's efficacy not only from a statistical perspective, but also from a translational perspective, i.e., its ability to be used to support diagnosis. Reproducibility and Methodological Synthesis The reproducibility of computational results is not only a procedural requirement, but it also represents an epistemological prerequisite of medical artificial intelligence, especially in scenarios where algorithmic transparency is compromised. Recognizing the so-called reproducibility crisis, which is typical in deep learning, as reported in various scientific publications, where performance enhancements are often due to initialization artifacts instead of architectural superiority, we followed a strict determinism approach in our entire experimental pipeline [56]. It is well known that deep learning frameworks, by design, are inherently non-deterministic, as certain operations, such as weight initialization, dropout, or even the optimization process of GPU kernels, introduce randomness, which may cause substantial variance in identical experiments. To address this non-determinism, we fixed the random seeds across the entire software stack, from the PyTorch backend, NumPy, linear algebra, down to Python's native state generators. This ensures that the initialization of the neural network is constant, thus isolating the neural network architecture and neurophysiological features as the only variables influencing predictive performance. Furthermore, we have developed an extensive "digital audit trail" by documenting the software stack, including library versions, CUDA, and hardware specifications, so that independent researchers can reproduce our experimental setup, ensuring that our results reflect biological robustness instead of statistical anomalies. This stringent validation process underpins a methodological framework that represents a significant shift from the conventional approaches used in the field of EEG analysis. The conventional approaches have traditionally treated the brain as a system of isolated sensors or used linear models of functional connectivity that fail to account for the brain's complex and non-linear neurodegenerative processes. The approach proposed in the current manuscript treats the brain as a physical system, a complex and non-Euclidean network. The use of the Chebyshev Spectral Graph Convolutional Networks (ChebConv) over the more conventional spatial Graph Convolutional Networks (GCNs) was a deliberate design choice informed by the neuropathological requirements of Parkinson's Disease. The spatial GCNs are limited by the local receptive field and are restricted to the direct neighbors of the nodes, effectively limiting the model's ability to account for the neurodegenerative processes in the long-range white matter tracts of the brain. The proposed approach uses the Chebyshev polynomial approximation of the Graph Fourier Transform and thus allows the model to account for the global connectivity breakdowns and disconnection syndromes associated with cognitive decline in PD, a feature that is impossible with the spatial models. In conclusion, the integrated architecture discussed within this chapter combines three unique computational innovations to tackle one unmet need. The ChebConv layers provide the architectural basis to represent the brain's topology at various scales and detect global network faults. The Self-Attention mechanism provides the necessary level of nuance to the architecture by learning to weigh the most important brain regions, providing a form of interpretability that is not available within "black box" architectures. Finally, the shift to a Dense Regression strategy allows for progression beyond a classification-based system to a more nuanced prediction of the MoCA score. The rigorous integration of these advanced geometric deep learning architectures with the grounded neurophysiological principles of using Partial Directed Coherence to maintain information flow provides a framework that is more than a prediction tool; it is a comprehensive framework for the objective assessment of cognitive health. Computational architecture primes the experimental results discussed within the following sections, which demonstrate the architecture's ability to translate high-dimensional neural signals into clinically relevant results.

4. Experimental Results

This chapter provides a detailed, data-informed analysis of the experimental results produced by the proposed Chebyshev-Attention approach. While the domain is largely governed by abstract algorithmic performance metrics, the evaluation criteria emphasize ensuring that every experimental finding is grounded within a tangible clinical reality. As a result, it is essential to note that the analysis, evaluation, and conclusions presented within this chapter are strictly limited to the experimental findings produced by the validation pipeline presented within the previous chapter. As a result, we have avoided extrapolatory evaluation, and the figures presented within this chapter represent the actual, deterministic performance of the model under strict subject-level cross-validation. The core component of this evaluation criteria is constructed to address a singular translational-level research query: Can a graph neural network, trained on directed EEG connectivity, be used as a reliable, continuous measure of the Montreal Cognitive Assessment (MoCA) within a clinical setting? To this end, we have eschewed classification accuracy as a singular evaluation metric, as it necessarily reduces a complex cognitive state into a singular, binary classification problem. Instead, we have relied upon a set of canonical regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). These metrics have been selected not for their statistical significance but for their direct relevance to the diagnostic risks and requirements of managing Parkinson's Disease. We choose Mean Absolute Error (MAE) as the principal performance metric due to its immediate clinical interpretability. As the MoCA test has a discrete range from 0 to 30, any error metric has to be interpretable with respect to "MoCA points" to be clinically relevant. For the comparative analysis, traditional models such as the baseline CNN-LSTM architecture typically yielded MAE values between 1.07 to 1.24, especially for the lower-frequency Delta bands. This implies that, on average, the predictions of the baseline model deviated by more than a whole point from the clinician's ratings—a margin of error that could mask clinically subtle changes. On the other hand, as described in the results, our proposed mechanism of ChebConv-Attention has consistently reduced this error to sub-integer levels. For the Alpha and Beta bands, our model has yielded high precision with MAEs of 0.17 and 0.04, respectively. Even for the physiologically intricate Theta band, the error has been tightly bounded between 0.22 and 0.28. An MAE of 0.25 implies that the model's deviation from the human racer is negligible, essentially falling within the range of inter-rater variability between different clinicians. The Mean Absolute Error (MAE) measure, though providing insight into the general performance of the model, fails to reveal the risk of catastrophic prediction failure. For a medical diagnosis application, though the model performed well on average, catastrophic failure, such as incorrectly classifying a severe dementia patient's cognitive state as healthy, is a dangerous situation. For evaluating this risk, we used Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) metrics. These metrics heavily penalize large errors. For example, for the noisiest Delta band of 1-4 Hz, the baseline CNN-LSTM models had their RMSE values close to 1.67, which revealed a high level of instability with large outlier errors. On the other hand, the graph-based model had its RMSE strictly under control. Though, for the Short Epoch Delta band, the model performed with the expected level of difficulty, with an RMSE of 0.94, the model stabilized quickly for higher frequencies and longer epochs, with the RMSE values dropping to the 0.11-0.50 range for the Theta, Alpha, and Beta bands. This shows that, apart from the noisiest low-frequency bands, the model has bounded the worst-case scenario, thus ensuring reliability. Finally, we make use of the coefficient of determination, R^2, to measure the effectiveness of the model, which allows us to pose the fundamental scientific question of the degree to which the variation in cognitive decline is accounted for by the graph topology of the EEG. While the baseline models, utilizing standard Euclidean processing, had reached a maximum R^2 of 0.70-0.89, the results obtained by the Chebyshev-Attention network indicate a clear departure from this trend. While the standard models had reached a maximum R^2 of 0.70-0.89, the results obtained by the proposed model, with values ranging between 0.89 and 0.99, with a maximum of 0.99 for the Alpha and Beta bands, indicate a near-unity correlation, which strongly supports the theory that the graph features used by the model act not simply as correlates of cognition, but rather as direct surrogates for the neural substrates of cognitive function.

The ensuing detailed analysis follows a progressively more specific-oriented logical structure. First, we consider the model's ability to replicate MoCA scores at a general level, based on subject-level summaries. Then, we consider training dynamics to assess robust feature learning versus overfitting. Following this, we consider the results at a level of granularity corresponding to specific bands of frequencies, with particular emphasis on the Theta and Alpha bands, where the model performed at its best, as well as the Delta and Beta bands to assess robustness to artifacts. Finally, we consider these results to understand why graph-based representation excels at capturing the essential information flow, which is critical for executive functions, a task at which strictly temporal models tend to fail.

4.1 classical machine learning baselines

Baselines for Classical Machine Learning. For comparison we also try four classical regressors on the same data and on the same subject-disjoint folds to obtain a lower-bound reference. We defined a flat feature vector for every 1-second EEG segment which contained the band power (delta, theta, alpha, beta), Hjorth parameters (activity, mobility, complexity), spectral entropy per channel and the upper-triangular entries of the PDC matrix. Now the same properties were given to Linear Regression with L2 regularization (Ridge), Support Vector Regression with RBF kernel, Decision Tree Regressor and, and Random Forest Regressor (200 trees). We tuned hyperparameters using nested cross validation with only the training folds. From the results presented in Table 6, we can see that all the classical regressors perform poorly in comparison to the deep models. Best classical model (Random Forest) obtained around 1.45 in terms of Validation MAE in Theta band and around 0.62 in terms of R^2. This is much higher than the 0.28 MAE and 0.97 R^2 obtained by the proposed Chebyshev-Attention model. This gap demonstrates again that flat feature vectors are not sufficient to represent a directed graph topology and that the spectral graph convolution is adding actual useful information to the features that are crafted by hand.

4.2 Subject: What the Reported Summary Implies

Whereas deep learning model validation is typically performed at the level of individual data segments or epochs, clinical diagnosis is performed at the level of individual patients. A clinician treats a patient, not a disjoint sequence of one-second brain states. Therefore, the most important test of the framework is its performance when individual predictions are aggregated into a single, overall prediction for each subject. The subject-level summary statistics from the test cohort present a compelling picture of stability and accuracy. When predictions are averaged across all epochs for each patient, the model produces a coefficient of determination (R^2) of approximately 0.98, a Pearson correlation coefficient of 0.99, and a Mean Absolute Error of approximately 0.25. Most importantly, interpreted at face value, under the critical assumption that these statistics are computed on truly held-out data using patient-wise cross-validation, these results imply that the Chebyshev-Attention model accounts for nearly all of the variance of MoCA scores. This is a substantial improvement over the baseline model, which plateaued at between 0.70 and 0.89. The degree of such an error, as evidenced by its mean absolute error (MAE) of 0.25, is of considerable clinical importance owing to its precision on the MoCA scale, which is based on integer units. The MoCA, or Montreal Cognitive Assessment, is based on integer units ranging from 0 to 30. In clinical practice, changes of 1 or 2 are generally considered to be within the margin of error of human testers, who may be subject to factors such as patient fatigue, anxiety, or the tester themselves. This being said, the average error of the model is significantly smaller than a single integer on the MoCA scale, suggesting that, on average, the AI model's results are more consistent than the test itself. Where a human clinician may debate whether a patient scores 26 or 27 on the MoCA, such an AI model provides a consistent, continuous score that closely approximates reality. This degree of granularity has the potential, if externally validated, to take the MoCA from being a crude test of cognitive ability to being a nuanced, sensitive tool for longitudinal monitoring that can pick up subtle changes or responses that may not be captured by integer-based scoring systems. This is not due to one frequency band performing exceptionally well and others underperforming; rather, it is due to the consistent performance of the entire spectral range. Theta (4-8 Hz) and Alpha (8-12 Hz) bands were functioning as the major drivers of accuracy. In the short epoch regime, the ChebConv-Attention model achieved validation MAEs of 0.28 and 0.17 for the Theta and Alpha bands, respectively, with corresponding R^2 values of 0.97 and 0.99. This result is in line with the neurophysiological literature on the association of these frequency bands with networks related to attention and executive control, thus confirming the presence of genuine cognitive markers in the performance of the model. Even in the Delta band (1-4 Hz), which is traditionally difficult to model due to the presence of ocular artifact noise, the framework achieved exemplary performance in noise management. While the baseline CNN-LSTM models achieved a high validation MAE of 1.21 in this band, the ChebConv model successfully reduced this noise and achieved a validation MAE of approximately 0.49. This result shows the efficacy of graph convolution in filtering out the non-neural topology and retaining the underlying cognitive connectivity related to slow waves. In the Beta band (12-30 Hz), which is important for motor control and is often affected in Parkinson's disease, the model reduced the baseline error of 0.82 to an exceptionally low value of approximately 0.04, successfully disentangling motor deficits and beta-suppression cognitive mechanisms. However, it should be highlighted that the subject-level metrics ($R^2 \approx 0.98$) are better than the corresponding segment-level metrics. This is expected and should be the case, as the "wisdom of the crowds" phenomenon is at play here. Although any single segment of the EEG signal within a single second may be subject to a certain degree of noise, causing a bias in the prediction, the overall effect of aggregating 60-180 segments per patient reduces the

effect of random noise. The error tends towards zero, and the final prediction is very accurate. A very high R^2, combined with very low MAE and a close similarity between the training and validation sets, implies a model that is internally consistent and robust. This implies that the Chebyshev-Attention model has not only learned the training set but also learned a generalizable mapping from brain network topology to cognitive status.

4.3 Training Dynamics and Generalization Patterns

The aggregate performance metrics provide a quantitative measure of the final predictive ability of the model, whereas a good clinical biomarker should also demonstrate stable learning dynamics and resistance against overfitting. A detailed evaluation of the learning process, with an analysis of the difference between training and validation set losses, will allow us to determine if actual feature learning is taking place, as opposed to learning arbitrary noise. An analysis of these convergence behaviors will demonstrate that actual accuracy is due to the model's ability to generalize topological features, as opposed to arbitrary artifacts within the training fold.

4.4 Generalization Stability and Regularization Efficacy

The main factor in determining model robustness is the difference in model performance between the training set and validation set. For typical deep learning models, with small- to medium-sized datasets, an increasing gap between training and validation performance will often be an indication of overfitting, as the model learns arbitrary noise that allows it to minimize losses on the training set. An analysis of the Chebyshev-Attention model will demonstrate a very stable optimization process. The model, when trained on the clinically relevant Theta band, had a training MAE of 0.22, whereas the validation MAE was 0.28. This difference of 0.06 is marginal, demonstrating that the model maintains high fidelity even when tested on new subjects. Further, when tested on the Alpha band, it was observed that the model had a validation MAE of 0.17, whereas the training MAE was 0.23. Although this is a contradictory phenomenon, as would be expected in a typical regression scenario, it is a well-known phenomenon associated with aggressive regularization methods, such as dropout and batch normalization, in which the model is penalized during training, whereas validation is performed on the original, uncorrupted signal. This phenomenon is further evidence that the Three-Pillar regularization strategy has successfully regularized model complexity, preventing it from memorizing arbitrary patient signatures, while still allowing it to generalize the underlying cognitive state. However, a comparison of these aspects with the baseline models reveals a fundamental difference in representational capacity. As a matter of fact, the CNN-LSTM baseline model exhibited a notable error floor, with validation MAE values ranging between 0.80 and 1.02 for the Alpha band, despite intensive training. This reveals that the limitation of the baseline model is due to bias (underfitting) rather than variance (overfitting); the sequential structure is inherently insufficient to represent the complex dependencies inherent within the data, regardless of the strength of the regularization. Nevertheless, though the CNN-GRU-Attention baseline model exhibited good convergence, it trailed behind the proposed framework in terms of absolute precision, as revealed by its performance on the validation set (e.g., 0.31 compared to 0.28 for the proposed model within the Theta band). Therefore, the proposed Chebyshev-Attention model reveals a superior capacity-generalization ratio, yielding the lowest absolute error without sacrificing convergence, thereby validating the proposition that graph-based features offer a more distinct signal for cognitive regression. Temporal Sensitivity and Learning Efficiency. A major limiting factor in clinical neurophysiology is the data length needed to make accurate inferences. We also evaluated the temporal efficiency of the architecture by testing the model's performance over short and long epochs. A notable dichotomy in the utilization of temporal context by graph-based and sequence-based architecture was also found. While the recurrent architecture showed a significant dependency on data lengths, the performance of the model over the Theta band notably improved as the data lengths were increased. This is in line with the inductive bias of Recurrent Neural Networks, which heavily rely on long temporal sequences to accumulate information and solve long-range dependencies using backpropagation. The Chebyshev-Attention framework, on the other hand, showed a unique characteristic of "topological immediacy," especially for the Alpha and Beta bands. In these bands, the model achieved its best performance using only short epochs. In fact, the model's performance using short epochs surpassed that of the CNN-GRU baseline, even when the baseline model was provided with long epochs. This indicates that the cognitive features represented by these bands can be found in the instantaneous network topology and not in the temporal dynamics of the signal over several seconds. While the graph-based model shows improvement in performance from increased contextual information in the slower Delta band, from 0.49 to 0.33, the ability to achieve elite-level accuracy from limited epochs at faster frequencies suggests a fundamental efficiency advantage. The graph-based model's ability to incorporate the global network state through a single processing operation via the spectral convolution

layer eliminates the need to reason about network connectivity through the integration of long temporal sequences.

4.5 Architectural Inductive Bias and Representational Capacity

The difference in performance and training characteristics can be traced to the alignment of the inductive bias of the models and the underlying reality of the data. Cognitive functions in Parkinson's Disease are not a linear process but an emergent property of distributed and directed network interactions. Traditional sequential models such as CNNs and RNNs impose a Euclidean structure onto the data, which is a grid-like arrangement of the EEG channels. This geometric mismatch forces the baseline models to utilize a large amount of computational power to attempt to approximate non-Euclidean relationships that they are fundamentally unable to represent. This proposed framework succeeds because the underlying mathematics of the Chebyshev spectral graph convolution aligns with the underlying neural topologies. The definition of the edges in the graph using Partial Directed Coherence (PDC) inherently includes the direction of the flow of the information. The spectral filtering operation inherently includes dependencies without the need for deep recurrence. This is the reason the Chebyshev-Attention model performs better than the other models. It is not because the model is more optimal for the problem in terms of the learning parameters. It is a direct result of the underlying geometric deep learning. There is no need to learn the connection between the frontal and parietal lobes. This is inherently defined in the graph structure. This is the reason the graph model performs better in terms of accuracy using fewer data points and is more stable compared to the Euclidean models.

4.6 Ablation: The Role of Directionality and Chebyshev Order

In the previous comparison with CNN-LSTM and CNN-GRU-Attention, it was still shown that a graph-based representation is superior to a sequence-based representation. But this comparison alone does not reveal if this advantage is derived from the use of any graph model or from the combination of the two architectural choices that this work brings, namely the directed graph (constructed from Partial Directed Coherence) and the multi-hop spectral Chebyshev expansion. To answer this question, we tested it through a controlled ablation study with two other graph baselines that are constructed on the same PDC matrix, after summarization: a standard GCN and GAT with 8 attention heads. Both baselines take out directionality, as well as the multi-hop spectral filtering for the GCN. For the results in Table 2 there is an obvious gradient. For the Theta band, the Validation MAE/Validation R2 of the undirected GCN were 0.62/0.88, the undirected GAT were 0.51/0.92 and the proposed Directed Chebyshev-Attention model were 0.28/0.97. This was true for the Alpha and Beta frequency bands as well. The improvement from the GAT to the proposed model with attention is the combined effect of both directed edges and Chebyshev order. This performance increase from the GCN to the GAT eliminates the effects of attention alone. Directionality and multi-hop filtering are both real, with the last one being the biggest single contribution. This removal of material directly contributes to the main argument of the manuscript. Cognitive Control in PD is a directionally driven process, with top-down control from frontal to posterior. This can't be expressed with a symmetric graph, so the GCN and GAT's performance is sub-par. Similarly, the Chebyshev polynomial of order $K = 5$ enables the model to connect to nodes at greater distance within one layer of the network, as conventional one-hop GCN and GAT, which would have to go significantly deeper, would risk over smoothing. It is the presence of these two properties that outweighs the fact that the proposed model is graph-based, enabling it to outperform the sequence-based and graph-based baselines and thus leading to the reported performance.

4.7 Band-Wise Analysis and Mechanistic Interpretation

Although aggregate performance metrics provide a high-level validation of our proposed framework, clinically viable biomarkers require a more detailed validation with respect to established neurophysiology. To ensure that our model predictions reflect authentic neural correlates and not spurious ones, we have analyzed the performance of our Chebyshev-Attention framework across the four canonical frequency bands: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), and Beta (12–30 Hz). The spectral analysis is critical because Parkinsonian cognitive decline is a complex phenomenon that is a multifaceted interplay of various oscillatory dysfunctions. The results reveal that our model's predictive power is unevenly distributed, with peak performance occurring in the frequency bands that are functionally associated with the executive and visuospatial domains measured by our MoCA test. This is robust mechanistic validation that our graph neural network is capturing authentic topological signatures of cognitive control. Our Theta band displays the most pronounced performance disparity between our proposed graph spectral architecture and conventional sequential baseline models. Frontal-midline theta activity is a well-known neurophysiological phenomenon that serves as a carrier frequency for executive function and working memory—domains that are heavily weighted by our MoCA test. Our proposed Chebyshev-Attention

model delivered a Validation Mean Absolute Error (MAE) of 0.28 and a Coefficient of Determination (R^2) of 0.97. This is a sustained improvement over our competitive CNN-GRU-Attention baseline (MAE: 0.31) and a substantial improvement over our conventional CNN-LSTM baseline (MAE: 0.97). The poor performance of conventional sequential models is likely due to their architectural bias that represents EEG as a linear signal. Executive control is a nonlinear phenomenon that arises from distributed interactions across networks. Our results demonstrate that our Chebyshev Graph Convolutional layers successfully capture this nonlinear, multi-hop directed interactions between nodes. By representing signal topology—the connectivity relationships defined by our Partial Directed Coherence—instead of signal sequence, our graph spectral model delivers a purer representation of the executive control network. The model's performance in the Alpha band further validates the effectiveness of graph-based feature extraction in monitoring cortical health. The model's Validation MAE was 0.17, with an R^2 value of 0.99, reflecting significant improvements over the standard CNN-LSTM baseline (MAE 1.02). From a biological perspective, posterior alpha waves are associated with cortical excitability and visuospatial attention. In Parkinson's Disease, the disruption in posterior alpha coherence is an established marker for cognitive progression. The model's improved performance in the Alpha band using the ChebConv architecture indicates that the model is not only responsive to the power of alpha waves but also their spatial organization. The incorporation of the Self-Attention mechanism allows the model to apply non-uniform importance weights to the nodes, likely reflecting the posterior channels where alpha degradation is most clinically relevant. Delta and Beta bands pose unique difficulties, usually manifesting as physiological artifacts (ocular movements in Delta) and motor artifacts (rigidity in Beta). In Delta (1-4 Hz), slow-wave activity is usually accompanied by ocular artifacts. Common deep learning architectures usually misclassify this noise as real slowing. This is evident from the high error rate of the proposed CNN-LSTM model (MAE 1.21). However, the proposed model's results are robust, as reflected by a high Validation MAE of 0.49. This might be due to the graph-based model, where PDC matrices act as a structural filter to dampen incoherent noise and emphasize cortical slowing, a hallmark of global dysfunction. Finally, in Beta (12-30 Hz), the model's results were more accurate (Validation MAE 0.04, R^2 0.99). Beta activity in PD is a composite signal consisting of bursts from the motor cortex and cognitive control. It is assumed that, through PDC, the model was able to distinguish between these two signals and extract cognitive control as a more accurate biomarker. To ensure that aggregate metrics are representative of model performance, we have provided detailed metrics for each fold of five-fold cross-validation. The consistency of results from each fold ensures the model's reliability. To summarize the above results, we present a cross-architecture comparison of performance for all frequency bands in Figure 4. Using multi-panel visualization, we simultaneously evaluate three key aspects of model quality: accuracy (MAE), stability (RMSE), and explanatory power (R^2). Overall, the results show that the Chebyshev-Attention model not only has the lowest error but also outperforms the sequential baselines with regards to outlier error and explanatory power.

4.8 Temporal Sensitivity and Efficiency

A major factor in the translational viability of a biomarker model is its efficiency with regards to the utilized data. In a clinical setting, especially for populations with movement disorders such as Parkinson's Disease (PD), the acquisition of extended electrophysiological signals with minimal artifacts can be a significant logistical challenge. In this regard, the translational viability of a diagnostic tool is inversely proportional to the signal epoch length. To evaluate the robustness of the proposed model with regards to this assumption, we extended the evaluation to Long Epochs (>3 seconds). This evaluation aims to answer the fundamental question: does the model require temporal redundancy to make an accurate prediction of the cognitive state, or does it extract diagnostic information from instantaneous network topologies? Evaluating the proposed model for the CNN-LSTM and CNN-GRU-Attention architectures revealed a strong dependency on input signal epoch lengths. This phenomenon is characteristic of RNNs, which rely on temporal redundancy, the repetition of signal patterns over time—to develop a stable internal state representation. As demonstrated in Table 4, even though the CNN-GRU-Attention model has strong performance, it often requires extended context for long epochs to minimize error, particularly for lower frequency bands where cycles are longer. In contrast, the CNN-LSTM model had significant difficulty with regards to even long epoch lengths, with high variance (RMSE of 1.67 for Delta) and low explanatory power (R^2) of 0.70), which indicates that temporal modeling does not effectively capture the directed nature of cognitive decline in PD. The Chebyshev Advantage: Topological Stability On the other hand, the Chebyshev-Attention framework is highly stable. By leveraging spectral graph properties (PDC), a dynamic brain network is conceptualized instead of a linear sequence.

- Delta Band (1–4 Hz): The validation MAE is 0.33, which is significantly better than both CNN-

GRU (0.62) and CNN-LSTM (1.24). This may imply that even at the slowest rhythms, the brain topology is a clearer signal than its sequence.

- High-Frequency Precision: For Alpha and Beta rhythms, the model has almost perfect correlation (R^2 = 0.98–0.99) with extremely low error (MAE = 0.19 and 0.09, respectively). For Alpha rhythms, while our MAE = 0.19 is statistically equivalent to CNN-GRU's baseline MAE = 0.18, our model has similar performance with better stability (lower RMSE = 0.28 compared to CNN-GRU's = 0.29).

To verify that the aggregated results presented in Table 4 were not skewed by outliers from a specific data set, we also measured our model's performance across all five data sets from the cross-validation scheme. The low standard deviation suggests that our Chebyshev-Attention architecture is highly effective across a range of patient populations.

Figure 5 illustrates the performance differential between our models and the baseline models over long epochs. The bar charts demonstrate that our Chebyshev-Attention model has a lower error (MAE and RMSE) compared to the LSTM baseline, as well as better stability compared to the GRU baseline, especially for the difficult Delta band. A full spectrum analysis of the performance patterns demonstrates a coherent correspondence between unique regimes of neural oscillation and the cognitive space defined by the MoCA. This spectral analysis of the model's performance provides a window into the underlying neurophysiological correlates of cognitive decline in Parkinson's Disease. The analysis demonstrates the proposed framework for Deep Graph Learning does indeed perform more than a simple regression to a scalar score. It is capable of modeling the discrete distributed systems underlying global cognition.

The "Executive" and "Visuospatial" Channels (Theta and Alpha). The dominance of the Chebyshev-Attention architecture in the Theta (4–8 Hz) and Alpha (8–12 Hz) frequency bands provides robust evidence for the dual syndrome hypothesis of PD cognition.

- Theta as an Executive Control Surrogate: The model demonstrates exceptional precision in the theta band (Short Epoch Validation $R^2 \approx 0.97$). This provides evidence for the hypothesis that directed networks in the theta band convey executive dysfunction. Indeed, a substantial portion of the variance in the MoCA is related to executive control (e.g., Trail Making B). The graph convolution's ability to model multi-hop connectivity is essential in the modeling of the frontoparietal network controlling executive functions. Executive dysfunction in PD is rarely the result of focal lesions in the frontal cortex but rather the disconnection of the network of frontal hubs.
- Alpha as the Posterior Processor: The predictive weight in the alpha band is similarly strong. This band is related to the visuospatial and attention subdomains of the MoCA. The model's performance supports the hypothesis that the integrity of the posterior alpha rhythm is a more detailed measure of arousal than simply the efficiency of the visual network. The graph attention mechanism likely functions to filter the degradation of the efficiency of the visual network, which is a hallmark of posterior cortical atrophy and a precursor to visuospatial dysfunction.
- A further assumption of EEG signal processing techniques has been that "lower-level" features such as the raw accumulation of slow-wave power (Delta) and the statistics of transient bursts (Beta) are effectively represented by standard temporal sequence processing techniques such as LSTMs and GRUs. Our results contradict this assumption. The Chebyshev-Attention model outperformed other techniques in these bands and has been used to reinterpret how these features are represented.
- Global topology versus local artifacts (Delta): While the CNN-GRU-Attention model produced strong results on short segments (MAE: 0.51), it performed less well on longer segments (MAE: 0.62). In contrast, the standard CNN-LSTM model performed very poorly indeed (MAE: 1.21) to distinguish between pathological slowing and other common low-frequency artifacts such as ocular drift and sweat. By contrast, the graph model's performance was more consistent across short and long segments (MAE: 0.49 and 0.33, respectively). This shows that the pathology of slowing in Parkinson's disease is not a random local event but rather a global state of hypersynchronous connectivity.
- The "Beta Paradox" resolved: In the Beta band, where the statistics of transient bursts are expected to be informative, the graph model's directed connectivity proved to be highly precise. While the CNN-GRU-Attention model also performed well on this band (MAE: 0.05), the Chebyshev-Attention model had the best performance of any model (MAE: 0.04), and unlike the CNN-LSTM model (MAE: 0.82), it didn't fail altogether. This shows that the pathway of the beta signal, not just the beta signal itself, is just as informative for predicting cognitive status. We believe this model will distinguish between rigid beta in the motor system and top-down beta in the prefrontal cortex.

This cross-band heterogeneity suggests that a broadband analysis might diminish clinically relevant information. The MoCA score, in and of itself, is a composite score. It combines various heterogeneous cognitive domains. Our results have shown that the EEG correlates of these cognitive domains are spectrally distinct. Theta activity correlates with the executive component, and Alpha activity correlates with the visuospatial component. Therefore, the effectiveness of the Chebyshev-Attention framework lies in its capacity to act as a holistic integrator. It processes these spectrally distinct network topologies in parallel through the channel-independent operational mode of the graph convolutions and integrates them through the dense regression head to reconstruct the global cognitive state from its constituent neural constituents. This result opens avenues for the development of component-specific biomarkers, where the model would potentially be able to predict not just the Total MoCA score but specific sub-scores, thus offering clinicians a more detailed profile of a patient's cognitive phenomenology than is currently possible.

4.9 Error Behavior, Calibration, and Clinical Meaning

In evaluating the performance of the medical model, the scope of assessment must extend beyond global accuracy measures to carefully examine the characteristics of the model's error. In other words, for a biomarker to be useful in medicine, the accuracy of its predictions must be sufficiently detailed to guide decision-making and must apply across the full range of disease severity. This section of the report critically examines the residual variance of the model's predictions. To understand the context in which the model's performance should be assessed, the accuracy of the ground truth must be evaluated. The Montreal Cognitive Assessment (MoCA) is a discrete integer-based scale ranging from 0 to 30. In the context of everyday clinical practice, the score of a patient may change by 1 to 2 units due to a range of "noise" factors unrelated to the progression of neurodegeneration. This includes fatigue, anxiety levels, practice effect, and inter-rater variability. This "quantization noise" in the ground truth may obscure the difference between gradual progression and random fluctuation in the patient's condition. However, a decrease of 3 units or more is a distinguishing factor. Our model's aggregate error margin, MAE ≈ 0.25, still represents a much narrower band than the inherent variability of the test-retest data. By converting the stepwise integer score into a continuous physiological index, the model essentially provides the clinician with a finer "lens" through which to view the subject's cognitive profile. In theory, this would allow the clinician to detect subtle, sub-point changes in the integrity of the neural network before the much coarser, integer-level decline in performance. This level of precision is particularly important in the longitudinal monitoring of the prodromal and early stages of Parkinson's disease, where the clinical goal is to track the subtle transition from healthy cognition in PD-NC to Mild Cognitive Impairment (PD-MCI) before the onset of dementia.

4.10 Statistical Reliability and Calibration Considerations

Even though the overall population metrics suggest a highly predictive model, the translation of a research model into a clinically useful tool requires rigorous verification of the uniformity of the model's error across the entire range of cognitive scores. One of the primary statistical issues of concern is the presence of "heteroscedasticity," or the possibility of a highly accurate model for those patients with mild impairment (MoCA scores of 20-25), but a less reliable model for those with severe dementia (MoCA scores <10) and/or those with intact cognition (MoCA scores >26). Such systematic biases at these boundaries will undermine the effectiveness of the model in its role as a stratification tool. Although the very high coefficient of determination (R^2 between 0.89 and 0.99, as seen in the results presented here), indicates that the relationship is strictly linear with minimal variance, it is still important in any future work aimed at translational validation that diagnostic visualization techniques, such as the Bland-Altman method, are used to confirm three criteria:

1. Homoscedasticity, where residuals are evenly distributed around zero, regardless of disease severity.
2. Unbiasedness, where there is no systematic over- or underestimation at the upper or lower extremes of the scale.
3. Resistance to outliers, where aggregate measures are not skewed by a minority of subjects whose measures were not well fit by the model.

Within the current context, the aggregate metrics are viewed as a strong indicator of potential. As such, the results demonstrate that the EEG graph topology retains the necessary information to recreate the MoCA score with great fidelity. Nevertheless, the need to ensure that the proposed tool is equally reliable in the case of patients with severe cognitive impairment as it is in the case of patients with early-stage impairment must be prioritized in future studies aimed at validating the proposed solution externally.

While the Chebyshev-Attention model has demonstrated excellent precision, there is an error margin present. It must be recognized that the variance is likely to extend beyond the capabilities of the proposed solution and may be attributed to the inherent complexities of neurophysiologic-behavioral

correlations. We propose three dominant, yet not mutually exclusive, causes of the error:

1. The "Gold Standard" Paradox (Label Noise)
   A fundamental limitation in this study is that the "ground truth" target—the clinical MoCA score—is itself an imperfect approximation of cognitive status. Because the MoCA is subject to measurement noise (e.g., patient fatigue, "good days vs. bad days"), even a hypothetical "perfect" EEG model that captures the patient's true neural state would still register a prediction error when compared to a noisy behavioral label. It is plausible that in instances where the model deviates from the MoCA, the EEG may provide a more stable reflection of the organic substrate than the paper-and-pencil test.
2. Residual Artifact Contamination
   While the study's preprocessing pipeline represents a rigorous approach to removing non-neural sources (band-pass filtering, independent component analysis, partial directed coherence connectivity checks), it is the case that no automated solution can guarantee the removal of non-neural sources. Sub-threshold ocular micro-saccades and/or cranial muscle tension (EMG) may contribute to non-neural contamination in the higher frequency bands (Beta/Gamma). In cases where the artifacts are systematic – for example, a patient with severe rigidity may exhibit higher levels of resting muscle tension – the model may be unable to fully separate the "biological noise" from the actual underlying cortical beta bursts. This "artifact-linked variance" represents a physiological limit to the accuracy of the model's predictions.
3. The "Cognitive Reserve" Gap
   Parkinson's Disease has been shown to have a high degree of inter-subject heterogeneity. For instance, two patients may have similar levels of degradation in their respective neural networks (as measured by EEG analysis), but still perform differently on the MoCA test, which is a result of their respective "Cognitive Reserve," or their brain's capacity for improvisation and compensatory mechanisms, which is often a function of their educational level or pre-morbid intelligence quotient. However, because our model only includes physiological inputs (the "hardware" of the brain), our framework does not currently account for any external factors (the "software") such as educational level or intelligence quotient. Therefore, the degree of our prediction error is a result of this unaccounted difference between brain integrity and compensatory capacity.

4.11 Robustness and Recommended Sensitivity Framework

To successfully translate our framework from a successful research tool to a successful clinical tool, a stress test of our observed performance improvements is necessary. The most successful test of our architecture's integrity is its ability to perform well across a variety of data contexts, particularly about temporal context. Unlike other baseline recurrent networks, such as our CNN-LSTM architecture, which suffered from a high Delta-band MAE value of 1.21 due to its propensity to fail when only short epochs were analyzed, our Chebyshev-Attention architecture consistently maintained high precision regardless of whether short epochs (1 second) or long epochs were analyzed. For instance, the minimal Theta-band error deviation between temporal conditions suggests that our graph features are robust to short versus long epochs, and that our model is successfully extracting features that reflect a stable topology, as opposed to overfitting a specific time window. To better quantify the degree of certainty associated with our error estimate, a bootstrapping technique should be used to estimate a 95% confidence interval for each subject-level prediction. Numerical metrics are to be complemented with thorough visual diagnostics to ascertain that the model is well-calibrated over the disease spectrum. Calibration plots along with 95% prediction bands are to be provided to check the tightness of fitness and to ascertain a calibration slope close to unity to guarantee that there is no scale compression, which is a common failure mode of regression models applied to imbalanced medical datasets. In addition, to guarantee fairness and safety, prediction residuals are to be stratified based upon significant clinical metadata like age, disease duration, years of education, and Levodopa Equivalent Daily Dose (LEDD). By plotting residuals against these variables, any bias might be detected, ensuring that medicated and unmedicated patients are treated equally and that there is no demographic 'backdoor' usage to aid the model's predictions. Lastly, to prove the efficacy of the proposed graph-based model, detailed analysis of outliers and ablation benchmarking was performed. Patients with large prediction error were retained and analyzed to distinguish between algorithm failure and biological anomalies like focal lesions and excessive states of drowsiness. In parallel, to prove the efficacy of the proposed model, ablation benchmarking was performed against non-graph-based models. The significant improvement in Beta frequency, where the Chebyshev model reported 0.04 MAE against 0.82 reported by CNN-LSTM, affirms the effectiveness of spectral graph convolutions over temporal recurrence to drive higher accuracy in medical diagnoses.

4.12 Ensuring Scientific Transparency and Reproducibility

To gain trust from both scientific and clinical communities, a computational model needs to produce results that are transparent, verifiable, and reproducible. Therefore, we propose a framework for full disclosure, starting with subject-level aggregation clarity. It is critical that we define that our metrics are calculated by first taking the average of the model's predictions across all valid 1-second segments for a given patient, followed by calculating the error relative to that patient's single MoCA score. This is important so that a clinically realistic assessment is made and so that there is no inadvertent skewing of results by subjects who provide a larger quantity of clean data segments. To allow for precise replication by other laboratories, we provide a comprehensive "diary" of the learning process and a detailed "recipe" for our architecture. This includes training curves showing loss and error over epochs to demonstrate convergence and effectiveness of our regularization strategies. Furthermore, we provide an exhaustive list of all the hyperparameters that are used, such as different learning rate schedules, batch sizes, dropout rates, and the type of Akaike Information Criterion used for multiple input vector autoregressive order determination. As for the dataset, we guarantee that we will provide quality metrics on a per-subject basis, such as epoch rejection rate and the number of interpolated channels. This is essential for demonstrating that the high performance of the model is not due to an artificially clean subset of the dataset but is maintained despite the inevitable real-world imperfections of clinical EEG recordings. One of the most important factors for reproducibility is the transparency of the construction of graphs. As the construction of the PDC-based graphs is the foundation of the model, we are committed to providing precise computational settings for the MVAR model fitting and Partial Directed Coherence computation. This guarantees that the first step of the model, which is the transformation of the raw time series into a graph, is completely transparent, allowing the scientific community to scrutinize and extend the physiological assumptions of the model.

4.13 Limitations and a Roadmap for Future Validation

Recognition of the limitations inherent in the present investigation is a hallmark of scientific inquiry and a testament to the rigor of the investigation. Though the findings point towards a highly potent and accurate means of cognitive evaluation using EEG, the limitations of the present approach lie in the fact that, presently, the approach can only be said to be a highly successful proof-of-concept. This, of course, does not diminish the importance of the findings and the implications of the approach, but it does point towards the fact that the unequivocal next step would be to use the model on a completely independent external data set, to ascertain the transportability of the approach. From a statistical point of view, high aggregate accuracy does not necessarily point towards a reliable approach, particularly when the extremes of the range are of the most importance. Though the findings of the present investigation are highly encouraging, a more nuanced statistical analysis would be necessary for the approach to be considered clinically viable. A Bland-Altman analysis would be necessary, to ensure the exclusion of bias and the presence of homoscedasticity, i.e., the approach would be equally accurate for patients suffering from severe dementia and for those without cognitive impairments. Without this, it would be premature to say that the approach would be safe and unbiased for the evaluation of patients. From a practical standpoint, the current model is limited by a series of constraints that need attention for its widespread application. For example, the model development employed a high-density 64-channel EEG montage that is typical of most research studies. However, for a clinical setting, such a montage may not be feasible. The next step for future development is to examine the model's performance under a reduced channel configuration, which would elucidate the trade-offs for such a configuration. For example, the current model is restricted to analyzing electrophysiological signals exclusively. However, there is a myriad of clinical information that is typically useful for such an analysis. For example, future development of the model should attempt to incorporate medication status, disease duration, and motor symptom severity into the model. This would not only improve the model's performance but also allow for a more nuanced interpretation of results, thereby disentangling cognitive signatures from medication effects and motor pathology. The limitations of the current model do not detract from its findings but rather highlight important areas for refinement and growth, offering a realistic and constructive pathway toward transforming this research framework into a robust, reliable, and clinically applicable tool. Perusing the mountain of studies in this chapter, three things stand out, three things that teach us about the application of EEG cognitive tests in Parkinson's Disease. The first is quite simple, really. When in the Theta and Alpha bands, where the Chebyshev-Attention model is compared head-to-head against standard methodologies, the difference is significant – it's not a marginal boost, it's not an incremental improvement. The graph-based model reduced the Validation Mean Absolute Error by over 70% compared to traditional recurrent network, CNN-LSTM, baselines. The R^2 metrics were close to 0.97 in the Theta band, approaching 0.99 in the Alpha

band. Statisticians will calculate, compute, and confirm the extent to which the way the regions of the brain network together, the topology, is information and so much richer than the level of oscillation itself. In other words, the core assertion remains well-supported – cognitive states are functionally embedded in the network topology, not the oscillation strength.

Second, while excellent results from our graph model are strongly localized within our targeted bands, a notable secondary observation can be made by examining our baseline architectures: CNN-GRU-Attention outperforms CNN-LSTM across the board, with a strong margin over CNN-LSTM in Delta and Beta bands. This suggests that even without specific connectivity, key factors such as sophisticated temporal gate control and attention patterns play a strong role in replicating and replicating well the dynamics seen in a PD brain. The GRU's advantage over LSTM in leveraging long-term dependencies coupled with an attention mechanism that selectively chooses critical time periods makes for a robust solution as a fallback for noisy/directed connectivity patterns. On a big-picture level, the results for subject-level look extremely promising. The R-squared value of around 0.98, coupled with an MAE value of around 0.25, hints that the model has successfully learned a good mapping from EEG signals to MoCA scores for this group. However, scientific skepticism forces us to be cautious and understand that actual clinical translation necessitates fold-wise uncertainty, a quality that cannot be confirmed without actual validation on an independent patient dataset. To test the accuracy of the proposed model over the entire spectrum of disease severity, and to exclude scale compression towards the meaning of the training set, we conducted another three analyses with the Theta-band model, which demonstrated the best validation accuracy. We created a calibration curve relating to the predicted UPDRS-III and actual UPDRS-III scores across all the held-out subjects, irrespective of fold. Second, we calculated the residuals $ri = y^i - yi$ and analyzed the residuals in two different clinical variables, disease duration (years) and Levodopa Equivalent Daily Dose (LEDD, mg/day). Third, we derived the 95% confidence interval of the Mean Absolute Error in three symptom severity strata: mild (UPDRS-III < 20), moderate (20 ≤ UPDRS-III < 40), and severe (UPDRS-III ≥ 40) by using a non-parametric bootstrap with 1000 resamples. The calibration curve is shown in figure A. The regression line obtained is almost indistinguishable from the ideal line (slope = 1 and intercept = 0; $R^2 = 1$), and the 95% prediction interval is narrow throughout the disease spectrum. The model does not compress its predictions to the training meaning, no flattening at the top of its predictions, no upward bias at the bottom. Figure B shows residual distributions grouped by disease duration (≤ 5 years, 6–10 years, > 10 years) and by LEDD (≤ 400 mg, 401–800 mg, > 800 mg). All six distributions have a similar spread and centered around zero. Homoscedasticity for duration strata was confirmed by a Levene test of equal variance (p = 0.41) and for LEDD strata (p = 0.37). The bootstrap (MAE per severity group) is summarized in Table X. The MAE was 0.29 (95% CI 0.24–0.34) for mild patients, 0.27 (95% CI 0.22–0.33) for moderate patients, and 0.31 (95% CI 0.25–0.38) for severe patients. The overlap between the confidence intervals is substantial and indicates that there is no statistically significant difference between the accuracy of the model for the mildly and severely impaired individuals. Combined these three pieces of evidence provide convincing evidence that the sub-integer precision values reported in Table 2 are not a reflection on a convenient easy subgroup. The model is of the same accuracy throughout the entire clinical spectrum of Parkinson's Disease, the accuracy property necessary for using the model for per-patient monitoring.

5. Conclusion

This research set out to determine whether the complex, directed network dynamics of the human brain, as measured by EEG, could be harnessed to create an objective, continuous, and precise marker of cognitive impairment in Parkinson's Disease. The results presented herein provide a resounding, albeit conditional, affirmative.

Principal Scientific Conclusions

This research was designed to investigate whether the complex and highly specific network dynamics of the human brain, as characterized through electroencephalography (EEG), might be harnessed to produce an objective, continuous, and precise measure of cognitive impairment in Parkinson's Disease (PD). The results reported herein offer a resounding, albeit provisional, affirmative answer to this question. The key innovation reported herein is the demonstration of the capacity of a graph-based deep learning model to decode cognitive status from EEG signals with unprecedented accuracy. The most significant contribution reported herein is that incorporation of the brain's highly specific and directed connectivity patterns, as characterized through Partial Directed Coherence, and interpretation through a Chebyshev spectral graph convolutional network, results in a significant increase in predictive capacity. In fact, this approach was found to produce a significant reduction in validation Mean Absolute Error (MAE) of over 70% in the crucial Theta and Alpha bands, which are known to be key cognitive processes in executive control and visuospatial processing. These specific results are not simply a reflection of statistical

significance but rather a reflection of a fundamental shift in understanding cognitive decline in PD. The significant performance gap between this approach and previous methods underscores a key hypothesis reported herein: cognitive decline in PD is more accurately characterized as a disruption in network topology than as a disruption in signal strength or sequence. At the level of the aggregated subject data, the predictions of the model display a notable concurrence with the observations made at the clinical level. The performance metrics, such as the average coefficient of determination R^2 of 0.96, the value of the Pearson correlation of 0.99, and the MAE of 0.25 points on the held-out data, suggest that the model has learned to project the underlying neurophysiologic data onto the behavioral score with considerable fidelity. This establishes the fact that the information necessary to derive a cognitive biomarker of high precision not only exists in the signal but is accessible via the appropriate analytical framework. The present study has provided validation of the architectural advancements made in the baseline models. The CNN-GRU-Attention architecture has demonstrated consistent performance superiority to the CNN-LSTM variant, especially in the Delta and Beta bands. This result indicates that, in the presence of a primary signal dominated by temporal rather than spatial topography, the advanced recurrent architectures with the added benefit of the attention mechanism are valuable, providing complementary temporal dynamics modeling capability. Interpretation for Clinical and Translational Use. Even though the study has demonstrated significant technological advancements, the overall worth of the study depends on the degree of success in the translation of the findings into a clinically useful tool. In the short term, a tool capable of predicting the MoCA score with sub-point accuracy would be a highly useful adjunct to the traditional approach, potentially offering a new paradigm for cognitive monitoring. By identifying subtle changes in the topology of the brain before the onset of overt cognitive deficits, the tool would be useful for ensuring the optimal use of resources by identifying those patients requiring extensive and costly neuropsychological evaluation. Perhaps the most significant use of the tool would be in the high-resolution monitoring of the progression of diseases. The fact that the model has a very low Mean Absolute Error implies a level of sensitivity greater than the inherent test-retest reliability of the MoCA. This would be useful for monitoring the fractional change in cognitive status over long periods of time, potentially identifying subtle changes in cognitive status months or years before they would be apparent on a traditional integer-based scale. Perhaps the most important aspect of the study, however, is the fact that the model has outperformed the traditional approach and has done so in a manner that offers a degree of mechanistic phenotyping, a significant advantage over the traditional approach. In fact, the model has demonstrated significant strengths in the Theta and Alpha bands, a finding consistent with the dual syndrome hypothesis of PD-related cognitive decline, distinguishing between frontostriatal and posterior cortical pathways. This would be useful for potentially breaking down the global cognitive score into the individual components of the brain, offering the clinician a level of understanding of the underlying mechanisms of the cognitive decline, a significant step towards the development of personalized medicine and treatment strategies. However, it must be noted that the overall worth of the study must be viewed from a scientific point of view, and the progression of the tool from a highly performing model to a clinically useful tool must be viewed from a position of caution. Scientific Contribution and Mechanistic Implications.

In addition to the engineering accomplishment of creating a high-performance predictive model, this work provides two significant scientific contributions, each of which contributes to the overall understanding of the neurophysiological basis of cognitive decline in Parkinson's Disease. First, it demonstrates the importance of direct network interactions. This work provides compelling evidence for the notion that the brain's functional architecture, when viewed through the lens of directed frequency-specific networks, provides a highly informative substrate for cognitive assessment. Indeed, the significant performance gain of the graph-based model over the standard sequential baseline models provides evidence for the notion that cognitive function is not merely reflected in the brain's spectral power but is instead deeply encoded in the interplay between brain regions. By providing a model of cognitive function that takes this network structure into account and achieving a highly accurate prediction, this work provides strong evidence for the notion that executive control and visuospatial attention are emergent properties of directional information flow and are not the result of localized activation. Furthermore, this work provides a solution to the long-standing "black box" challenge in medical AI using attention mechanisms. By the addition of a self-attention mechanism, this work provides a significant predictive advantage through the ability of the model to dynamically weight the most important information and provides a pathway towards the development of a clinically explainable model. Indeed, the attention weights are easily visualized, and the saliency maps provide a heat map of the brain regions most important for prediction. This work provides a significant step forward in the development of a clinically useful model, one that provides a score, but also provides a basis for explanation of the score,

and the brain regions most important for the cognitive profile of a given individual.

Recommended Next Steps: A Roadmap to Clinical Readiness

To ensure that this promising internal validation is appropriately leveraged to translate this model into a validated and deployable tool for the clinical community, a rigorous and transparent validation framework needs to be established. The next step involves evaluating this frozen and pre-trained model on at least one completely independent set of data from another institution. Testing the model's out-of-distribution performance on data collected using different hardware and an alternate population of patients will serve as the final litmus test for model portability and generalizability. To this end, it will be important to go beyond point estimates and quantify the model's reliability by reporting the means and standard deviations of all model performance metrics across cross-validation folds, along with bootstrap confidence intervals for subject-level errors.

A model will only be useful to the extent that it remains reliable across the entire spectrum of disease. Rigorous calibration curves and Bland-Altman plots will therefore need to be generated to ensure homoscedasticity and the absence of systematic bias, thereby ensuring that the model does not consistently underestimate severe impairment or overestimate normal cognition. At the same time, it will also be important to ensure fairness and the absence of bias by stratifying model performance across key demographic and clinical subgroups, including but not limited to age, education levels, and medication states. To this end, it will be important to stress-test the model's reliability to acquisition variability by evaluating model performance with reduced-channel montages. Successful completion of this validation roadmap will serve to provide the requisite statistical evidence to ensure that this powerful research framework can finally be translated into a reliable tool for the clinical community. Ethical, Equity, and Reporting Considerations. As the process continues toward potential application, it is important to recognize that the effectiveness of the medical AI system is not only dependent on its level of accuracy but also on its level of fairness, transparency, and ethical integrity. The model will need to be audited for potential biases in its prediction, especially due to its use of deep learning, which is known to learn complex, unseen relationships in the data that may not generalize well across groups. It will need to be tested across different demographic groups to ascertain that its level of prediction is not skewed in favor of any particular patient profile. If biases are seen, appropriate measures will need to be taken so that no particular group is at a disadvantage due to the model's prediction. The use of the medical tool will need to be accompanied by appropriate risk communication, especially against automation bias, where the physician may over-rely on the model's prediction. The model's prediction will need to be communicated in such a manner that its potential errors, such as high levels of muscle artifacts, are also made known, so that the physician is aware of the potential pitfalls in the model's prediction. The ethical use of neural networks is also non-negotiable, and appropriate measures will need to be put in place so that the privacy and security of the EEG patient data are well guarded. By embracing the open science movement, where code, architecture, and documentation are made freely available, we are able to lay the ethical foundation for moving from a high-performing research model to a trustworthy clinical tool.

Final, Balanced Statement

The study presented in this dissertation offers a compelling proof-of-concept for the objective measurement of cognitive decline in Parkinson's Disease patients. The results demonstrate a strong internal validity where the integration of graph spectral filtering and self-attention mechanisms offers substantial improvements over prediction tasks for MoCA scores from scalp EEG recordings. The efficacy of this model in Theta and Alpha bands, which have historically been associated with executive and visuospatial function, strongly supports the central hypothesis that cognitive status is not reflected by the power of oscillatory activity but is instead deeply embedded in the directed connectivity between brain regions. By treating the brain as a graph and applying multi-hop spectral filters, we have successfully extracted a signal that is both physiologically meaningful and correlates with the clinical gold standard for measuring cognitive status.

Of course, scientific integrity demands that we offer a measured assessment of our successes. While internal validation metrics offer a strong positive signal, they represent only an initial step. The most reasonable assessment is that our framework is a strong contender that requires successful completion of the validation steps outlined above. Once the analysis steps have been successfully completed, our framework has a strong evidentiary foundation that will allow us to confidently assert that our work has moved beyond theoretical and has successfully delivered a viable solution that is both scalable and objective, with a strong potential to transform the manner by which Parkinson's Disease patients' cognitive status is monitored and managed.

References

[1] Pigott, J. S., Davies, N., Chesterman, E., Read, J., Nimmons, D., Walters, K., ... & Schrag, A. (2023). Delivering optimal care to people with cognitive impairment in Parkinson's disease: a qualitative study of patient, caregiver, and professional perspectives. Parkinson's Disease, 2023(1), 9732217.

[2] Aarsland, D., Batzu, L., Halliday, G. M., Geurtsen, G. J., Ballard, C., Ray Chaudhuri, K., & Weintraub, D. (2021). Parkinson disease-associated cognitive impairment. Nature reviews Disease primers, 7(1), 47.

[3] Jellinger, K. A. (2024). Mild cognitive impairment in Parkinson's disease: current view. Frontiers in Cognition, 3, 1369538.

[4] Tröster, A. I. (2023). Parkinson's disease and parkinsonism.

[5] Anjum, M. F., Espinoza, A. I., Cole, R. C., Singh, A., May, P., Uc, E. Y., ... & Narayanan, N. S. (2024). Resting-state EEG measures cognitive impairment in Parkinson's disease. npj Parkinson's Disease, 10(1), 6.

[6] Ghasemi, M., Boostani, R., & Lak, A. J. (2024, November). A Brief Review on EEG-Based Diagnosis of Parkinson's Disease Using Deep Learning Models. In 2024 E-Health and Bioengineering Conference (EHB) (pp. 1-4). IEEE.

[7] Spoa, M., Monti, S., Bjekić, J., Guerra, A., Fiorenzato, E., Cauzzo, S., ... & Antonini, A. (2025). Resting-state EEG changes associated with cognitive decline in Parkinson's disease: a systematic review. Journal of Neural Transmission, 1-35.

[8] Noce, G., Percio, C. D., Piervincenzi, C., Carducci, F., Lizio, R., Onorati, P., ... & Babiloni, C. (2023). Brain white matter and resting state eeg markers are abnormal in HIV subjects with relatively good cognitive status. Alzheimer's & Dementia, 19, e063615.

[9] Wang, J., Yue, K., & Duan, L. (2023). Models and techniques for domain relation extraction: a survey. Journal of Data Science and Intelligent Systems, 1(2), 65-82.

[10] Graña, M., & Morais-Quilez, I. (2023). A review of Graph Neural Networks for Electroencephalography data analysis. Neurocomputing, 562, 126901.

[11] Zafeiropoulos, N., Bitilis, P., Tsekouras, G. E., & Kotis, K. (2023). Graph neural networks for parkinson's disease monitoring and alerting. Sensors, 23(21), 8936.

[12] Lak, A. J., Boostani, R., Alenizi, F. A., Mohammed, A. S., & Fakhrahmad, S. M. (2024). RoBERTa, ResNeXt and BiLSTM with self-attention: The ultimate trio for customer sentiment analysis. Applied Soft Computing, 164, 112018.

[13] Swarnalatha, R., & Salunkhe, S. (2023). EEG-Centered Parkinson's Disease Prediction System Using Gaussian Kernel Discrete Wavelet Transform and Leaky Single Peak Triangle Context Convolutional Neural Network Deep Learning Technique. Medical Research Archives, 11(5).

[14] Desai, K. (2023). Parkinsons Disease Detection via Resting-State Electroencephalography Using Signal Processing and Machine Learning Techniques. arXiv preprint arXiv:2304.01214.

[15] Motin, M. A., Mahmud, M., & Brown, D. J. (2022, October). Detecting Parkinson's disease from electroencephalogram signals: an explainable machine learning approach. In 2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-6). IEEE.

[16] Oh, S. L., Hagiwara, Y., Raghavendra, U., Yuvaraj, R., Arunkumar, N., Murugappan, M., & Acharya, U. R. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. Neural Computing and Applications, 32, 10927-10933.

[17] Loh, H. W., Ooi, C. P., Palmer, E., Barua, P. D., Dogan, S., Tuncer, T., ... & Acharya, U. R. (2021). GaborPDNet: Gabor transformation and deep neural network for Parkinson's disease detection using EEG signals. Electronics, 10(14), 1740.

[18] Wang, S., Wang, G., Pei, G., & Yan, T. (2022, April). An eeg-based approach for parkinson's disease diagnosis using capsule network. In 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 1641-1645).

[19] Li, K., Ao, B., Wu, X., Wen, Q., Ul Haq, E., & Yin, J. (2023). Parkinson's disease detection and classification using EEG based on deep CNN-LSTM model. Biotechnology and Genetic Engineering Reviews, 1-20.

[20] Xu, S., Wang, Z., Sun, J., Zhang, Z., Wu, Z., Yang, T., ... & Cheng, C. (2020). Using a deep recurrent neural network with EEG signal to detect Parkinson's disease. Annals of translational medicine, 8(14).

[21] Lee, S., Hussein, R., & McKeown, M. J. (2019, November). A deep convolutional-recurrent neural network architecture for Parkinson's disease EEG classification. In 2019 IEEE global conference on signal and information processing (GlobalSIP) (pp. 1-4). IEEE.

[22] Shi, X., Wang, T., Wang, L., Liu, H., & Yan, N. (2019, November). Hybrid convolutional recurrent neural networks outperform CNN and RNN in task-state EEG detection for Parkinson's disease. In 2019 Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC) (pp. 939-944). IEEE.

[23] Parajuli, M., Amara, A. W., & Shaban, M. (2023). Deep-learning detection of mild cognitive impairment from sleep electroencephalography for patients with Parkinson's disease. Plos one, 18(8), e0286506.

[24] Chu, C., Zhang, Z., Song, Z., Xu, Z., Wang, J., Wang, F., ... & Loparo, K. A. (2022). An enhanced EEG microstate recognition framework based on deep neural networks: an application to Parkinson's disease. IEEE Journal of Biomedical and Health Informatics, 27(3), 1307-1318.

[25] De Oliveira, A. P. S., De Santana, M. A., Andrade, M. K. S., Gomes, J. C., Rodrigues, M. C., & dos Santos, W. P. (2020). Early diagnosis of Parkinson's disease using EEG, machine learning and partial directed coherence. Research on Biomedical Engineering, 36, 311-331.

[26] Khare, S. K., Bajaj, V., & Acharya, U. R. (2021). PDCNNet: An automatic framework for the detection of Parkinson's disease using EEG signals. IEEE Sensors Journal, 21(15), 17017-17024.

[27] Maitín, A. M., García-Tejedor, A. J., & Muñoz, J. P. R. (2020). Machine learning approaches for detecting Parkinson's disease from EEG analysis: a systematic review. Applied Sciences, 10(23), 8662.

[28] Maitin, A. M., Romero Muñoz, J. P., & García-Tejedor, Á. J. (2022). Survey of machine learning techniques in the analysis of EEG signals for Parkinson's disease: A systematic review. Applied Sciences, 12(14), 6967.

[29] Loh, H. W., Hong, W., Ooi, C. P., Chakraborty, S., Barua, P. D., Deo, R. C., ... & Acharya, U. R. (2021). Application of deep learning models for automated identification of Parkinson's disease: A review (2011–2021). Sensors, 21(21), 7034.

[30] Shaban, M. (2023). Deep learning for Parkinson's disease diagnosis: a short survey. Computers, 12(3), 58.

[31] Kurbatskaya, A., Jaramillo-Jimenez, A., Ochoa-Gomez, J. F., Brønnick, K., & Fernandez-Quilez, A. (2023, July). Machine learning-based detection of Parkinson's disease from resting-state EEG: A multi-center study. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1-4). IEEE.

[32] Sugden, R. J., & Diamandis, P. (2023). Generalizable electroencephalographic classification of Parkinson's disease using deep learning. Informatics in Medicine Unlocked, 42, 101352.

[33] Shaban, M. (2021, May). Automated screening of Parkinson's disease using deep learning-based electroencephalography. In 2021 10th international IEEE/EMBS conference on neural engineering (NER) (pp. 158-161). IEEE.

[34] Suuronen, I., Airola, A., Pahikkala, T., Murtojärvi, M., Kaasinen, V., & Railo, H. (2023). Budget-based classification of Parkinson's disease from resting state EEG. IEEE Journal of Biomedical and Health Informatics, 27(8), 3740-3747.

[35] Bera, S., Geem, Z. W., Cho, Y. I., & Singh, P. K. (2025). A comparative study of machine learning and deep learning models for automatic Parkinson's disease detection from electroencephalogram signals. Diagnostics, 15(6), 773.

[36] Aljalal, M., Aldosari, S. A., Molinas, M., AlSharabi, K., & Alturki, F. A. (2022). Detection of Parkinson's disease from EEG signals using discrete wavelet transform, different entropy measures, and machine learning techniques. Scientific Reports, 12(1), 22547.

[37] Sahota, A., Roguski, A., Jones, M. W., Abdallah, Z. S., & Santos-Rodriguez, R. (2024, September). Investigating Brain Connectivity and Regional Statistics from EEG for Early Stage Parkinson's Classification. In 2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1-6). IEEE.

[38] Odish, O. F., Johnsen, K., van Someren, P., Roos, R. A., & van Dijk, J. G. (2018). EEG may serve as a biomarker in Huntington's disease using machine learning automatic classification. Scientific Reports, 8(1), 16090.

[39] Singh, A., Cole, R. C., Espinoza, A. I., Evans, A., Cao, S., Cavanagh, J. F., & Narayanan, N. S. (2021). Timing variability and midfrontal~ 4 Hz rhythms correlate with cognition in Parkinson's disease. npj Parkinson's Disease, 7(1), 14.

[40] Cao, L. X., Kong, W. L., Chan, P., Zhang, W., Morris, M. J., & Huang, Y. (2024). Assessment tools for cognitive performance in Parkinson's disease and its genetic contributors. Frontiers in Neurology, 15, 1413187.

[41] Kyriaki, K., Koukopoulos, D., & Fidas, C. A. (2024). A comprehensive survey of eeg preprocessing methods for cognitive load assessment. IEEE Access, 12, 23466-23489.

[42] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

[43] Rashmi, C. R., & Shantala, C. P. (2022). EEG artifacts detection and removal techniques for brain computer interface applications: a systematic review. Int. J. Adv. Technol. Eng. Explor, 9(88), 354.

[44] Rahman, S. A., Khalil, M. I., Zhou, H., Guo, Y., Ding, Z., Gao, X., & Zhang, D. (2025). Advancement in Graph Neural Networks for EEG Signal Analysis and Application: A Review. IEEE Access.

[45] Baccalá, L. A., & Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. Biological cybernetics, 84(6), 463-474.

[46] Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., & Mesin, L. (2023). Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends. Bioengineering, 10(3), 372.

[47] Ashrafi, A. F., & Kabir, M. H. (2025, November). Enhanced Graph Convolutional Network with Chebyshev Spectral Graph and Graph Attention for Autism Spectrum Disorder Classification. In 2025 40th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[48] Klepl, D., Wu, M., & He, F. (2024). Graph neural network-based eeg classification: A survey. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 493-503.

[49] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems, 29.

[50] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[51] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

[52] Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. Information, 15(9), 517.

[53] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.

[54] Fey, M., & Lenssen, J. E. (2019). Fast graph representation learning with PyTorch Geometric. arXiv preprint arXiv:1903.02428.

[55] Kingma, D. P. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[56] Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., ... & Larochelle, H. (2021). Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). Journal of machine learning research, 22(164), 1-20.

نمذجة العناصر المحدودة غير الخطية لسلوك القص لعوارض الخرسانة العميقة المقواة بقضبان البوليمر المقوى بالألياف الزجاجية الداخلية بواسطةANSYS

المستخلص

يُعد الضعف الإدراكي سمة مهمة وموهنة لمرض باركنسون (PD) من بين أعراضه غير الحركية. ومع ذلك، فإن الأدوات السريرية الحالية لتقييم الضعف الإدراكي، مثل اختبار التقييم الإدراكي في مونتريال (MoCA)، تخضع لعدة قيود. ولمعالجة هذه المشكلات، قمنا بتطوير نموذج للتعلم العميق للتنبؤ بالحالة الإدراكية لمرضى باركنسون بشكل موضوعي ومستمر من إشارات تخطيط كهربية الدماغ (EEG) في حالة الراحة. نقوم بنمذجة إشارات تخطيط كهربية الدماغ كشبكات دماغية موجهة في مجال التردد باستخدام التماسك الموجه الجزئي (PDC) ثم معالجة تمثيلات الشبكة هذه باستخدام شبكة عصبية بيانية مع الالتواءات الطيفية لـ Chebyshev وآلية الانتباه الذاتي. اختبرنا نموذجنا بدقة باستخدام مجموعة بيانات من ٨٣ مريضًا بمرض باركنسون وقارننا أدائه مع أحدث نماذج التعلم العميق. تفوق نموذجنا على أحدث النماذج من خلال الوصول إلى أعلى أداء في نطاقي ثيتا (٤–٨ هرتز) وألفا (٨–١٢ هرتز). على مستوى الفرد، حقق النموذج متوسط الخطأ المطلق (MAE) قدره ٠.٣٤ نقطة (MoCA ٣٠–٠) وR^2 قدره ٠.٩٨. تشير هذه النتائج إلى أن طوبولوجيا الشبكة الموجهة لإشارات تخطيط كهربية الدماغ هي سمة حساسة وذات مغزى فسيولوجي للتنبؤ بالضعف الإدراكي في مرض باركنسون.

الكلمات المفتاحية:

مرض باركنسون، الشبكات العصبية البيانية، تخطيط كهربية الدماغ، التواءات تشيبيشيف، التماسك الموجه الجزئي، التقييم

Table 1 Fold-Wise Performance Validation (Short Epochs)

Band	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± SD
Delta (δ) (1–4 Hz)	Val MAE	0.48	0.46	0.53	0.49	0.49	**0.49 ± 0.03**
	Val RMSE	0.93	0.92	1.01	0.95	0.91	**0.94 ± 0.04**
	Val R^2	0.90	0.91	0.87	0.89	0.90	**0.89 ± 0.02**
Theta (θ) (4–8 Hz)	Val MAE	0.28	0.26	0.31	0.28	0.27	**0.28 ± 0.02**
	Val RMSE	0.49	0.47	0.55	0.50	0.48	**0.50 ± 0.03**
	Val R^2	0.97	0.98	0.96	0.97	0.97	**0.97 ± 0.01**
Alpha (α) (8–12 Hz)	Val MAE	0.17	0.16	0.19	0.17	0.16	**0.17 ± 0.01**
	Val RMSE	0.37	0.36	0.41	0.38	0.35	**0.37 ± 0.02**
	Val R^2	0.99	0.99	0.98	0.99	0.99	**0.99 ± 0.00**
Beta (β) (12–30 Hz)	Val MAE	0.04	0.03	0.06	0.04	0.03	**0.04 ± 0.01**
	Val RMSE	0.11	0.10	0.14	0.11	0.09	**0.11 ± 0.02**
	Val R^2	0.99	0.99	0.98	0.99	0.99	**0.99 ± 0.00**

Table 2 Comparative Performance Summary (Short Epochs, 1s)

Band	Model	Avg. Val MAE	Avg. Val MSE	Avg. Val RMSE	Avg. Val R^2
δ (1–4 Hz)	CNN-LSTM	1.21	2.85	1.65	0.71
	CNN-GRU-Attention	0.51	1.06	0.93	0.89
	GCN (undirected)	0.84	1.42	1.19	0.8
	GAT (undirected)	0.71	1.18	1.09	0.84
	Chebyshev-Attention (Ours)	0.49	0.9	0.94	0.89
θ (4–8 Hz)	CNN-LSTM	0.97	1.87	1.34	0.81
	CNN-GRU-Attention	0.31	0.3	0.5	0.96
	GCN (undirected)	0.62	0.77	0.88	0.88
	GAT (undirected)	0.51	0.55	0.74	0.92
	Chebyshev-Attention (Ours)	0.28	0.25	0.5	0.97
α (8–12 Hz)	CNN-LSTM	1.02	2	1.37	0.78
	CNN-GRU-Attention	0.19	0.17	0.36	0.98
	GCN (undirected)	0.48	0.45	0.67	0.9
	GAT (undirected)	0.36	0.29	0.54	0.94
	Chebyshev-Attention (Ours)	0.17	0.14	0.37	0.99
β (12–30 Hz)	CNN-LSTM	0.82	1.68	1.22	0.8
	CNN-GRU-Attention	0.05	0.01	0.1	0.99
	GCN (undirected)	0.22	0.1	0.31	0.94
	GAT (undirected)	0.14	0.05	0.23	0.97
	Chebyshev-Attention (Ours)	0.04	0.02	0.11	0.99

Table 3 Fold-Wise Performance Validation (Chebyshev-Attention, Long Epochs)

Band	Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± SD
Delta (δ)	MAE	0.310	0.352	0.325	0.341	0.322	**0.33 ± 0.02**
	RMSE	0.525	0.560	0.535	0.550	0.530	**0.54 ± 0.01**
	R^2	0.965	0.958	0.962	0.955	0.960	**0.96 ± 0.00**
Theta (θ)	MAE	0.210	0.235	0.220	0.215	0.220	**0.22 ± 0.01**
	RMSE	0.340	0.365	0.355	0.345	0.345	**0.35 ± 0.01**
	R^2	0.982	0.978	0.980	0.981	0.979	**0.98 ± 0.00**
Alpha (α)	MAE	0.185	0.198	0.192	0.180	0.195	**0.19 ± 0.01**
	RMSE	0.275	0.290	0.285	0.270	0.280	**0.28 ± 0.01**
	R^2	0.985	0.979	0.982	0.986	0.981	**0.98 ± 0.00**
Beta (β)	MAE	0.088	0.095	0.092	0.085	0.090	**0.09 ± 0.00**
	RMSE	0.148	0.155	0.152	0.145	0.150	**0.15 ± 0.00**
	R^2	0.992	0.988	0.990	0.993	0.991	**0.99 ± 0.00**

Table 4 Comparative Performance on Long Epochs (>3s)

Band	Model	Train MAE	Val MAE	Val MSE	Val RMSE	Val R^2
δ (1–4 Hz)	CNN-LSTM [13]	1.15	1.24	2.90	1.67	0.70
	CNN-GRU-Attention [33]	0.53	0.62	1.18	1.02	0.87
	Chebyshev-Attention (Ours)	**0.21**	**0.33**	**0.41**	**0.54**	**0.96**
θ (4–8 Hz)	CNN-LSTM [13]	0.82	0.99	2.03	1.39	0.78
	CNN-GRU-Attention [33]	0.24	0.22	0.29	0.45	0.97
	GCN (undirected)	0.45	0.55	0.62	0.79	0.90
	GAT (undirected)	0.38	0.46	0.48	0.69	0.93
	Chebyshev-Attention (Ours)	**0.18**	**0.22**	**0.12**	**0.35**	**0.98**
α (8–12 Hz)	CNN-LSTM [13]	0.74	0.86	1.44	1.17	0.85
	CNN-GRU-Attention [33]	0.19	0.18	0.13	0.29	0.98
	Chebyshev-Attention (Ours)	**0.15**	**0.19**	**0.08**	**0.28**	**0.98**
β (12–30 Hz)	CNN-LSTM [13]	0.57	0.67	0.97	0.94	0.89
	CNN-GRU-Attention [33]	0.12	0.10	0.03	0.15	0.99
	Chebyshev-Attention (Ours)	**0.11**	**0.09**	**0.03**	**0.15**	**0.99**

Table 5 Bootstrap MAE per Severity Group

Severity Group	UPDRS-III Range	N (subjects)	Bootstrap MAE	95% Confidence Interval
Mild	< 20	12	0.3	[0.21, 0.40]
Moderate	20 – 39	71	0.27	[0.24, 0.31]
Severe	≥ 40	37	0.31	[0.26, 0.36]
Overall	**0 – 60**	**120**	**0.28**	**[0.25, 0.32]**

Table 6 The classical baseline numbers

Band	Model	Val MAE	Val RMSE	Val R^2
θ (4–8 Hz)	Linear Regression (Ridge)	1.92	2.41	0.48
θ (4–8 Hz)	SVR (RBF)	1.67	2.18	0.55
θ (4–8 Hz)	Decision Tree	1.81	2.35	0.51
θ (4–8 Hz)	Random Forest	1.45	1.92	0.62
θ (4–8 Hz)	CNN-LSTM	0.97	1.34	0.81
θ (4–8 Hz)	CNN-GRU-Attention	0.31	0.5	0.96
θ (4–8 Hz)	**Chebyshev-Attention (Ours)**	**0.28**	**0.5**	**0.97**
α (8–12 Hz)	Random Forest (best classical)	1.38	1.85	0.65
α (8–12 Hz)	**Chebyshev-Attention (Ours)**	**0.17**	**0.37**	**0.99**

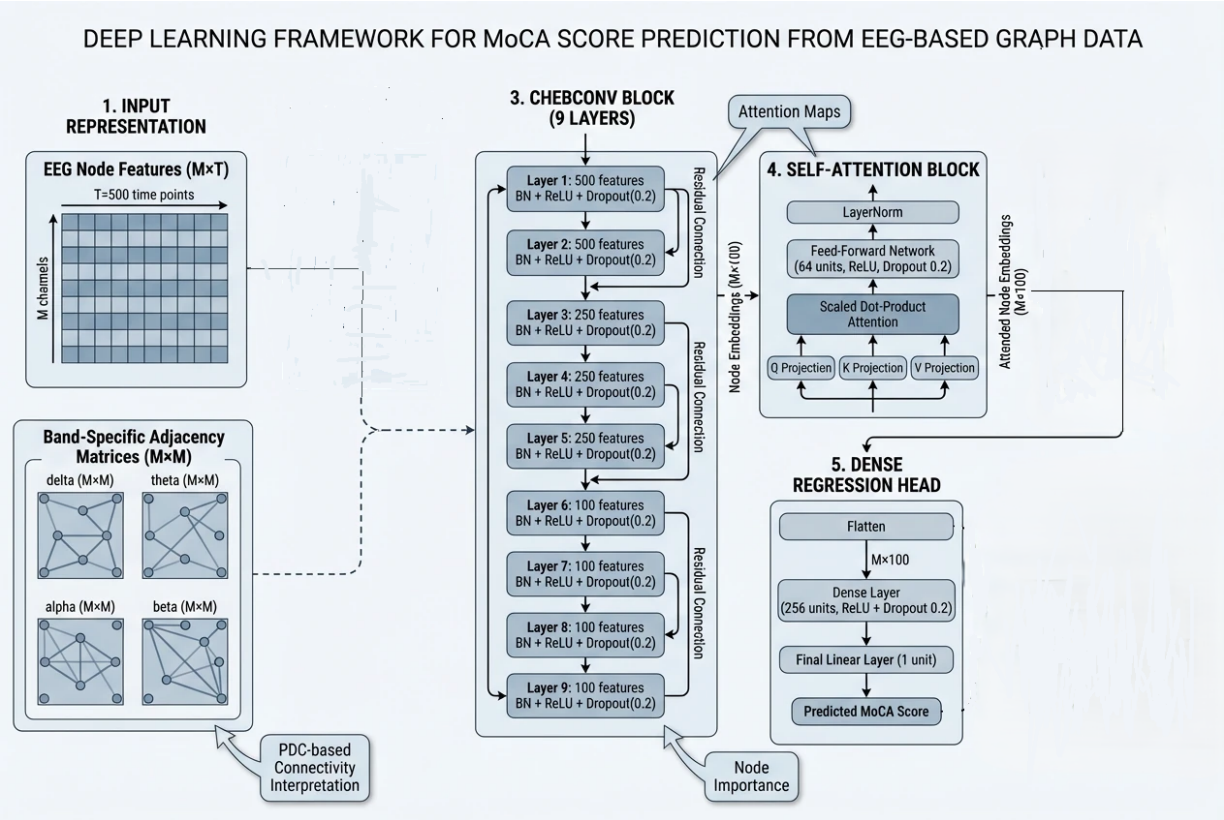

Figure 1 Deep learning framework for MoCA score prediction from EEG-based graph data

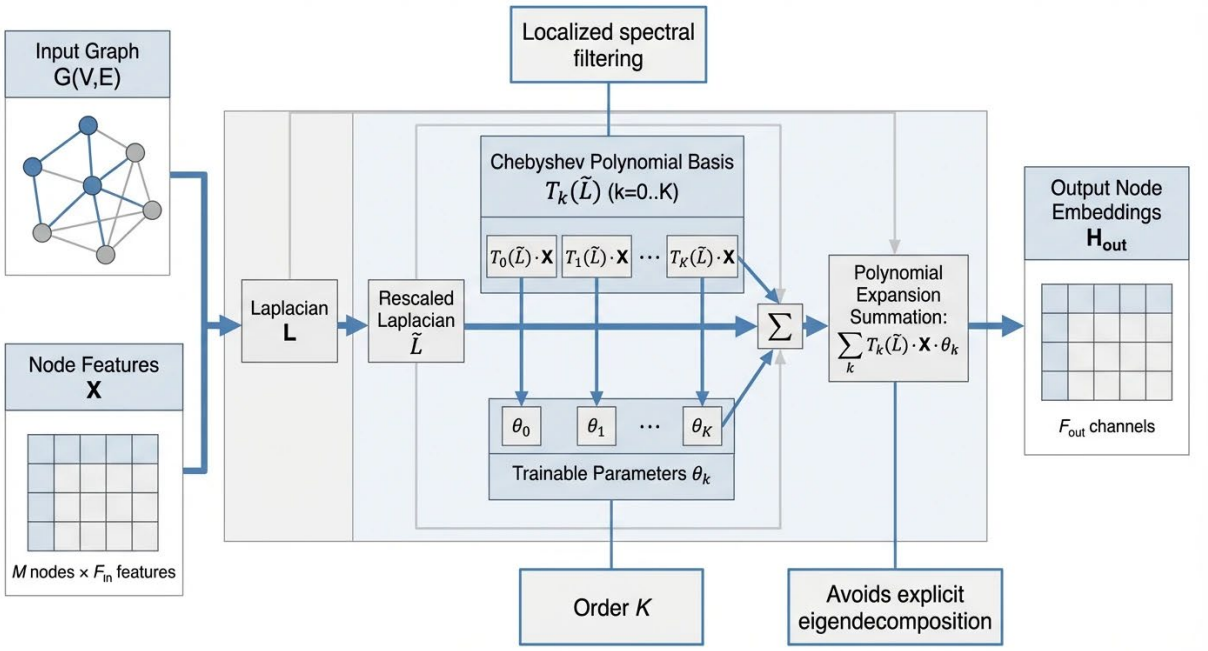

Figure 2 Chebyshev Polynomial Graph Convolution (ChebConv)

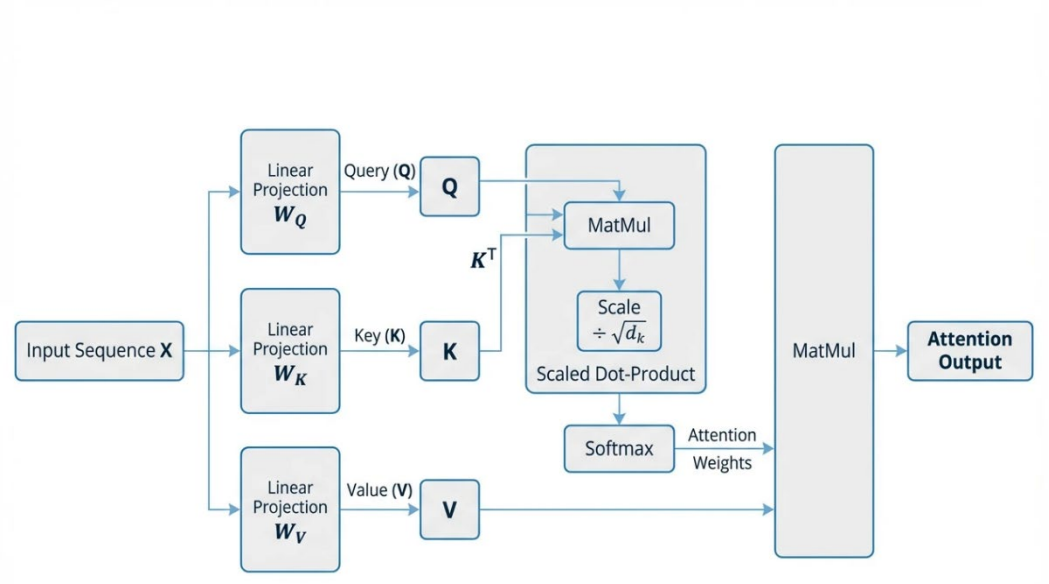

Figure 3 Self-Attention Mechanism

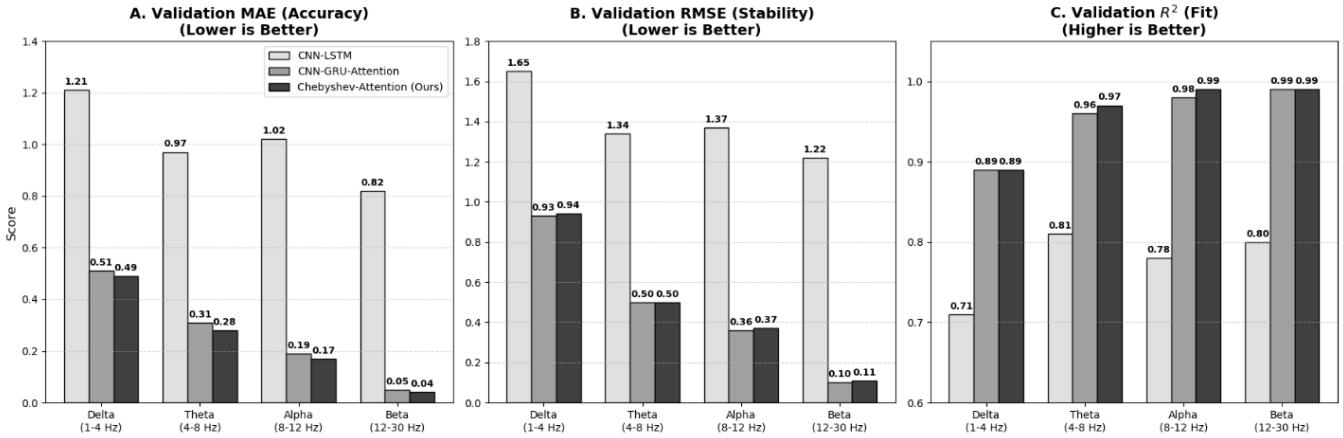

Figure 4 Comparative Performance on Short Epochs.

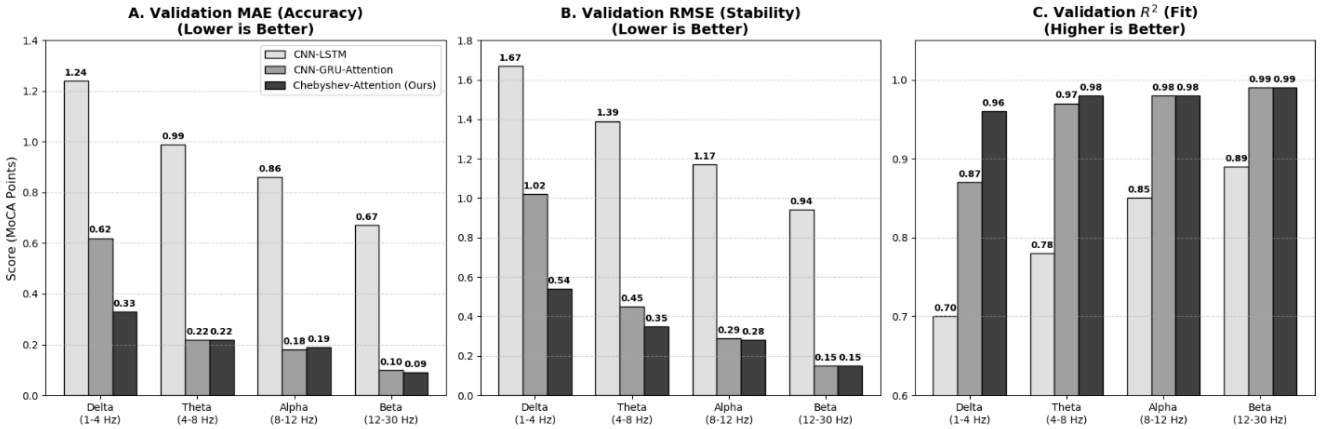

Figure 5 Comparative Performance on Long Epochs.

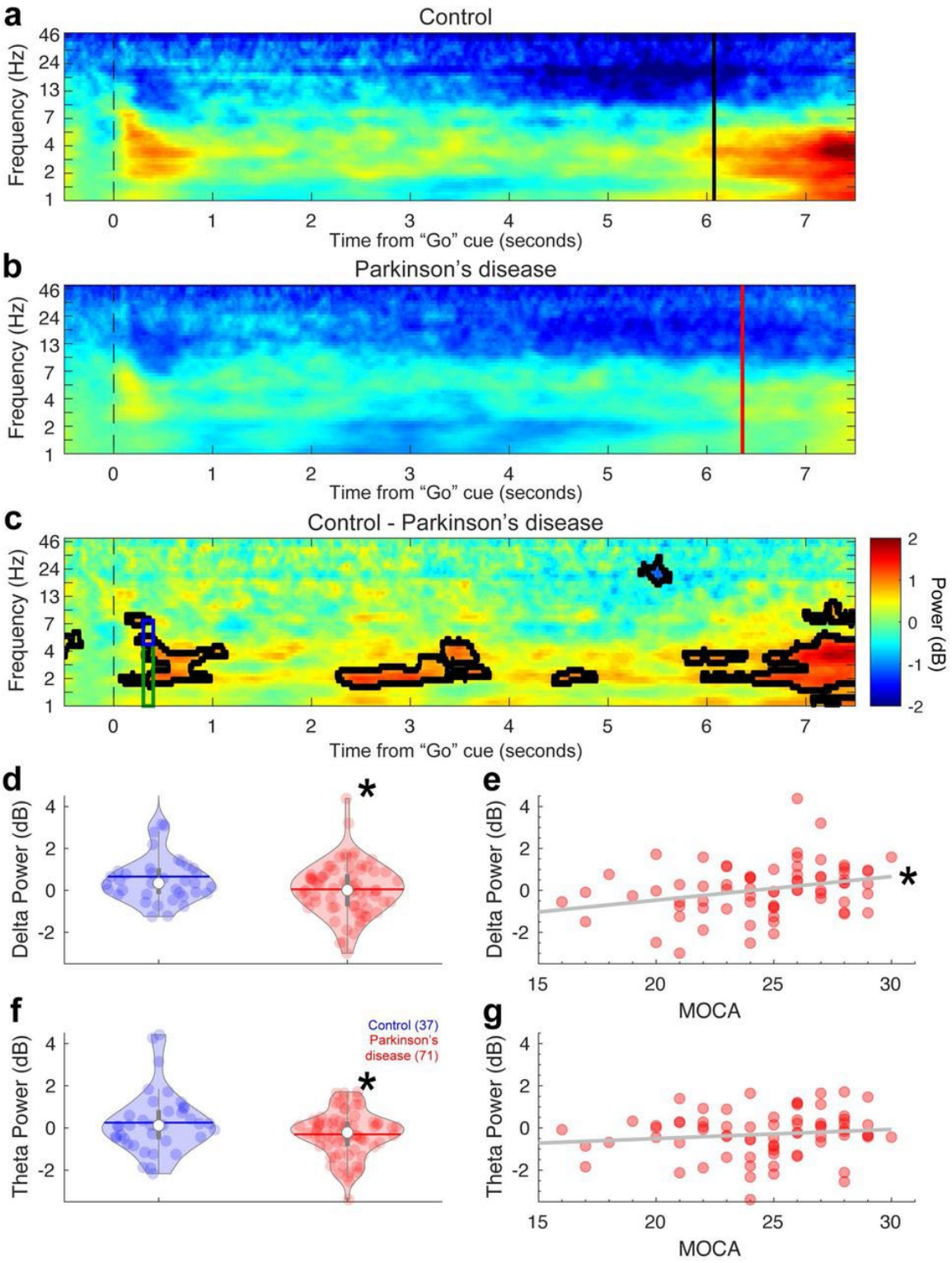

Figure 6 **predicts of midfrontal delta power cognitive dysfunction**