NeuroImage: Clinical 42 (2024) 103571

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Longitudinal prognosis of Parkinson's outcomes using causal connectivity

Cooper J. Mellema [a,b,h], Kevin P. Nguyen [a,b,h], Alex Treacher [a,e,h], Aixa X. Andrade [a,b,h], Nader Pouratian [g,h], Vibhash D. Sharma [f,h], Padraig O'Suileabhain [f,h], Albert A. Montillo [a,b,c,d,h,*]

[a] Lyda Hill Department of Bioinformatics, United States
[b] Biomedical Engineering Department, United States
[c] Advanced Imaging Research Center, United States
[d] Radiology Department, United States
[e] Biophysics Department, United States
[f] Neurology Department, United States
[g] Neurosurgery Department, United States
[h] University of Texas Southwestern Medical Center, United States

ARTICLE INFO

Keywords:
fMRI
Connectivity
Functional connectivity
Effective connectivity
Parkinson's
Prognosis

ABSTRACT

Despite the prevalence of Parkinson's disease (PD), there are no clinically-accepted neuroimaging biomarkers to predict the trajectory of motor or cognitive decline or differentiate Parkinson's disease from atypical progressive parkinsonian diseases. Since abnormal connectivity in the motor circuit and basal ganglia have been previously shown as early markers of neurodegeneration, we hypothesize that patterns of interregional connectivity could be useful to form patient-specific predictive models of disease state and of PD progression. We use fMRI data from subjects with Multiple System Atrophy (MSA), Progressive Supranuclear Palsy (PSP), idiopathic PD, and healthy controls to construct predictive models for motor and cognitive decline and differentiate between the four subgroups. Further, we identify the specific connections most informative for progression and diagnosis. When predicting the one-year progression in the MDS-UPDRS-III[1*] and Montreal Cognitive assessment (MoCA), we achieve new state-of-the-art mean absolute error performance. Additionally, the balanced accuracy we achieve in the diagnosis of PD, MSA, PSP, versus healthy controls surpasses that attained in most clinics, underscoring the relevance of the brain connectivity features. Our models reveal the connectivity between deep nuclei, motor regions, and the thalamus as the most important for prediction. Collectively these results demonstrate the potential of fMRI connectivity as a prognostic biomarker for PD and increase our understanding of this disease.

1. Introduction

Neurodegenerative diseases remain difficult to diagnose until late in their course. Additionally, their future trajectories remain challenging to prognose, with unbiased assessments being time consuming and difficult to administer. One promising avenue for making informative prognostic and diagnostic decisions for patients is through neuroimaging techniques. Functional Magnetic Resonance Imaging (fMRI) is of particular interest because functional changes are an early and sensitive marker of neurodegeneration, that precede atrophy visible in structural MRI and diffusion MRI (Dadi et al., 2019), do not require expensive and ionizing radiation like PET and SPECT imaging (Leung et al., 2019; Mansu et al., 2017), and provide insights into the pathophysiology of these complex, multifaceted diseases (Noble et al., 2019; Smitha et al., 2017). Measures derived from fMRI, particularly interregional brain connectivity, have been strongly predictive in a wide variety of prognostic and diagnostic analyses for neurodegenerative diseases and reveal networks and signaling pathways associated with disease symptoms, prognoses, and subtypes (Smitha et al., 2017). Among neurodegenerative diseases, Parkinson's disease (PD) is the second most common (Jankovic and Tan, 2020; Pringsheim et al., 2014). It is a movement disorder with no clinically accepted neuroimaging measures of prognosis. The diagnosis of PD relies on a detailed clinical evaluation and differentiating it from other forms of Parkinsonism such as Multiple System Atrophy (MSA)

* Corresponding author.
E-mail address: Albert.Montillo@UTSouthwestern.edu (A.A. Montillo).

[1*] A summary of all acronyms used in this paper is presented in Appendix A. Supplementary Data.

https://doi.org/10.1016/j.nicl.2024.103571
Received 25 August 2023; Received in revised form 24 January 2024; Accepted 26 January 2024
Available online 6 February 2024

and Progressive Supranuclear Palsy (PSP) can be challenging. To fill this need, we develop predictive models using new measures of connectivity to predict an individual's long-term motor and cognitive trajectories – achieving state of the art performance. The same measures also differentiate between PD, MSA, PSP, and healthy controls. We use a new NIH dataset in which medication confounds have been minimized which not only allows to make more accurate predictions, but also enables the identification of connectivity features associated with true disease progression and not confounded by medication. Finally, we present an fMRI data augmentation method to construct more accurate machine learning predictive models from such finite datasets. The tools and analyses presented in this manuscript are a substantial step to address clinical needs and the implicated interregional brain connectivity measures provide new avenues for further research and potential therapeutic targets.

1.1. Parkinson's disease diagnosis

Parkinson's disease has two major symptomatic lookalikes, PSP and MSA. Despite the similar symptoms, these diseases have distinct pathophysiology and require distinct treatments (Hughes et al., 2002). In general clinical settings (i.e. outside of movement disorder specialty clinics), the diagnostic balanced accuracy is typically only 0.63 and even in movement disorder specialist clinics, the balanced accuracy is only 0.75 early in the disease presentation (i.e. de novo cases) (Hughes et al., 2002). The accuracy of diagnoses by specialists does increase as the disease progresses, but a tool which could diagnose the disease earlier would be preferable. Due to this low accuracy, there has been great interest in developing machine-learning based diagnostic models to supplement clinical findings and aid diagnosis. However, since the precise changes in the brain during the progression of Parkinson's disease are poorly understood (Jankovic and Tan, 2020; Pringsheim et al., 2014), model development has been impeded. Recently, evidence has emerged suggesting that *structural connectivity measures*, present in diffusion tensor imaging (DTI) MRI, may be useful for this diagnostic task with balanced accuracy up to 0.77 (Chougar et al., 2021). Our work's 0.68 balanced accuracy indicates that connectivity measures from functional MRI are also similarly discriminative, potentially helping to elucidate the differing pathophysiologies of PD, PSP, and MSA.

1.2. Parkinson's disease prognosis

Currently there is no clinically accepted tool to predict the trajectory of disease for individual patients with PD. Developing such a tool would enable physicians to inform patients of their prognosis and enrich clinical trials with fast progressors likely to show change even in their spans of just two years. Neuroprotective trials in PD have been hamstrung by the rate of participants disease progression being variable and often gradual, so that the control group deteriorates little over the few years that is feasible to conduct a trial, and the intervention group has limited opportunity to show efficacy. A prognostic tool would empower such clinical trials to identify effective candidate treatments. Currently, there are heuristics that guide PD prognosis including the unilaterality versus bilaterality of symptoms and age of onset, but these remain coarse and non-quantitative (Jankovic and Tan, 2020). Therefore, biomarkers for quantitative prognosis are needed (Marek et al., 2018). We expect the ability to predict prognosis will aid in the development of and iteration upon disease-modifying therapies, provide patients with more certain estimates of their progression, and provide a stepping-stone for future therapies by identifying connectivity biomarkers associated with progression. The minimal clinically important difference (CID) in PD patient's MDS-UPDRS-III score is 3.25, so models should target an error of less than 3.25 to be useful (Horváth et al., 2015). Previous studies have employed T1 imaging (Zeighami et al., 2019), SPECT imaging (Leung et al., 2019; Mansu et al., 2017), and DTI imaging (Taylor et al., 2018) to predict PD progression. However, these methods achieve mean absolute errors (MAEs) in MDS-UPDRS-III prediction only as low as 3.22, barely exceeding recommended bar for CID of 3.25 and do not offer additional connectivity insights into the pathophysiology of the disease. There has also been evidence of early and sensitive prognostic correlates in fMRI, but the majority of these studies have been confounded by medication (e.g., levodopa or one of its analogues) at the time of imaging and MDS-UPDRS-III evaluation. Previously identified functional correlates to diagnosis, specific symptom severity, medication response, and other non-prognostic neuroimaging biomarkers are numerous in the literature, suggesting a robust functional signal of PD pathophysiology (Amboni et al., 2015; Baggio et al., 2015; Hassan et al., 2017; Hou et al., 2017; Ng et al., 2017; Tuovinen et al., 2018). Specific functional correlates of long-term prognosis as measured with MDS-UPDRS-III or MoCA have been found by multiple researchers, confirming that there are not only correlates to current severity, but also future severity (Burciu et al., 2016; Hou et al., 2017; Manza et al., 2016; Olde Dubbelink et al., 2014; Simioni et al., 2016). However, there has been a dearth of methods employing effective connectivity (i.e., causal connectivity) as a prognostic biomarker, which our study addresses. Effective connectivity can have higher predictive power than functional connectivity, but more importantly it has direct interpretability in the discovered connections and allows greater insight into pathophysiology than purely functional measures (Bielczyk et al., 2019; Chockanathan et al., 2019; Friston et al., 2019; Mellema and Montillo, 2023; Yao et al., 2017). We hypothesize that using effective connectivity measures to make prognostic models will generate more accurate predictions of 1-year progression. Furthermore, we hypothesize that mining these predictive models' most important features will indicate brain regions and connections of potential pathological import in PD.

1.3. Contributions

The primary contributions of this work are five-fold.

1. We develop prognostic models with leading accuracy for predicting the 1-year change in motor symptomatology and for predicting the 1-year change in cognitive function.
2. We develop diagnostic models for PD, achieving balanced accuracy to differentiate PD, PSP, MSA and normal controls above that attained in most clinics.
3. We quantitatively compare four measures of functional and effective connectivity, including new measures which incorporate structural priors as well as traditional measures across the aforementioned diagnostic and prognostic predictive tasks.
4. We develop an anatomically realistic augmentation method for 4D fMRI–derived connectivity analysis and demonstrate that it can significantly improve prediction performance.
5. We identify the most strongly associated connections for each prognostic and diagnostic target, revealing what the models have learned and providing new insight into disease mechanism.

2. Materials and methods

2.1. Materials

This work uses fMRI data from the Parkinson's Disease Biomarkers Program (PDBP) (Ofori et al., 2016) to construct diagnostic and prognostic models. The PDBP dataset was chosen as the primary dataset due to its emphasis on fMRI data collection with patients in an OFF-medication state, since medication level has been a significant confound in PD outcome studies (Ng et al., 2017). Previous studies have found greater than 80 % concordance between task and rest fMRI signal and have successfully used task fMRI for resting analysis (Beheshtian et al., 2021; Kraus et al., 2021). Therefore, while only task-based fMRI was gathered by PDBP, we treated it as resting data in subsequent analysis, so that we can apply our connectivity analysis which has been

validated on resting state fMRI. The task used in this study followed a hand-grip paradigm (Ofori et al., 2016), suggesting motor regions (i.e. the primary regions involved in this task) would be the most affected by the task-as-rest assumption (Kraus et al., 2021). However, even with this task-as-rest assumption, prior work reliably identifies consistent network topologies and variants unique to the subject (Beheshtian et al., 2021; Kraus et al., 2021). Connectivity measures were derived from the baseline (year 0) fMRI scan of each subject and used as the inputs to form prognostic models that predict the diagnosis and the primary outcome tracked in PDBP, namely the 12 month change in UPDRS-III score.

The change in the UPDRS over a 1-year period was significant as tested with a paired *t*-test, p-value 2.53E-5. The net change in the MoCA score in the PD population was not significant as tested with the same paired *t*-test. As cognitive changes take longer in the progression of PD, it is expected that only some patients will have significant changes over a 1-year period. The MoCA results, therefore, should be interpreted as an ability to pick up on some patients who have significant changes amongst a larger population who do not meaningfully change over the time of the study. Previous results by Krishnan et al (Krishnan et al., 2017) suggest that changes less than 2 points on the MoCA scale are insignificant, meaning 25 of the 71 patients in this PDBP sample have a meaningful progression over this period. Our ability to predict changes, then, hinges on our ability to distinguish these "progressors" from non-progressors.

The subset of the PDBP dataset used includes all 146 subjects with fMRI, including 73 idiopathic PD, 45 Normal Control (NC), 21 PSP, and 7 MSA subjects. Subjects were enrolled in the study on average 3 years after initial diagnosis by a movement disorder expert, then followed prospectively for up to 3-years during which time the diagnosis was open to revision; the final diagnosis is used for classification purposes. All fMRI data was collected as described in **Supplemental Section 10.2.** Data was collected from 2 sites, with identical acquisition parameters and scanner make and model. The subjects were tested and scanned after 12–14 h of overnight withdrawal of antiparkinsonian medication. Further details on the cohort selection criteria can be found in Burciu et al. (Burciu et al., 2016). We performed 3 predictive tasks on PDBP. (1) For the diagnostic task, we built a 4-class classifier to distinguish between PD, PSP, MSA, and NC. For the prognostic measures, we (2) built a regression model to predict the 1-year change in motor skills as measured through a motor examination (MDS-UPDRS-III), and (3) an additional model to predict the change in cognitive score (MoCA).

The demographics of the subjects used in the subsequent diagnostic prediction experiments are shown in Table 1 (row a). For the motor progression task, we used the 63 PD subjects who had a recorded one-year change in the MDS-UPDRS-III. Demographics for the longitudinal motor progression prediction are shown in Table 1 (row b). For the longitudinal cognitive prediction task, we use all 71 PD subjects who had a recorded 1-year change in their Montreal Cognitive Assessment (MoCA). Demographics for the longitudinal cognitive progression prediction are shown in Table 1 (row c).

2.2. Methods

2.2.1. Derivation of the functional connectivity covariates from the functional MRI

All fMRI data was prepared with an in-house fMRI preprocessing pipeline with advanced motion correction, shown to be significantly better at removing motion artifact in PD than competing methods (Raval et al., 2022). Pipeline processing details are provided in **Supplemental Section 10.2**. Mean regional timeseries were extracted with the Schaefer atlas with 100 cortical regions and 35 additional subcortical regions included (Schaefer et al., 2018). The Schaefer atlas is a functional atlas whose regions are defined through the clustering of functional activity in fMRI. This functional atlas was chosen because: 1) a functional atlas tends to capture better functional variability than a purely anatomical atlas (Mellema et al., 2022), and 2) the Schaefer atlas groups regions into resting-state networks (RSNs), which facilitates inter- and intra-RSN partitioning and analysis. The cerebellum and striatum from the Automated Anatomical Labelling (AAL) atlas (Ofori et al., 2016; Rolls et al., 2020) were included because both structures contain signals of diagnostic importance and have been often overlooked in prior analyses (Stoodley et al., 2012).

From each mean regional timeseries, the timeseries was linearly detrended and z-score normalized, and a set of functional connectivity (FC) and effective connectivity (EC) measures were derived. The FC measures included Correlation (*Corr*), Partial Correlation (*PCorr*), and a machine learning measure of functional connectivity from an XGBoost predictive model ($ML.FC_{XGB}$). Correlation and partial correlation are commonly used measures of FC, while $ML.FC_{XGB}$ has been shown to potentially contain more predictive information for prognostic tasks than the other two FC measures (Mellema and Montillo, 2023). $ML.FC_{XGB}$ was shown to have the highest reproducibility across multiple scans of the same subject and higher predictive power across physiological and cognitive prediction targets. The EC measures included (1) Principal Components Granger Causality (*PC.GC*), which calculates a Granger-Causality score from a lower-dimensional Principal Components projection, and (2) Structurally Projected Granger Causality (*SP.GC*). *SP.GC* (Mellema and Montillo, 2023) calculates a Granger-Causality score from a lower-dimensional Structural Projection which uses a population-average diffusion-MRI derived tractography connectivity prior calculated on the Human Connectome Project (Yeh et al., 2018) in the initial projection. This prior is a diffusion-MRI atlas of connectivity developed from over 800 healthy subjects and represents a "standard structural connectivity" prior. *SP.GC* regularizes the brain connectivity by projecting brain activity into a lower dimensional space with a preference for communication along physically connected paths. *SP.GC* was shown to have higher reproducibility than other EC measures and having more predictive power for multiple targets. *PC.GC* was chosen because it also performed well in predictive tasks (Mellema and Montillo, 2023). Please note that there remains some controversy of the nature of the directed functional interactions inferred by Granger causal methods– particularly on if they truly can be called "causal". We have chosen here to adhere to common nomenclature of "Granger causality", "causal", and "effective connectivity", but note the active discussion of the nature of connections inferred by this technique (Barnett et al.,

Table 1
PDBP Demographics. Row **1a** lists the demographics of subjects with an available fMRI and diagnosis, Row **1b** shows the demographics of PD subjects with an available initial fMRI and recorded 1-year change in MDS-UPDRS-III score, and Row **1c** shows the demographics of PD subjects with an available initial fMRI and recorded 1-year change in MoCA.

	Target	# Subjects	Age	% Female	# PD/NC/PSP/MSA	
a	Diagnosis	146	65.1 ± 8.8	42.5 %	73/45/21/7	
					Initial score	**1 yr Δ score**
b	MDS-UPDRS-III	63	64.4 ± 9.0	31.6 %	40.0 ± 27.4	12.5 ± 6.5
c	MoCA	71	64.0 ± 8.9	31.0 %	25.89 ± 2.76	−0.25 ± 2.24

2018; Stokes and Purdon, 2017). Each connectivity measure produces a matrix containing the strength of connection between each region of interest (ROI); the value representing the connection strength between two ROIs is also called an 'edge' and the ROI is also called a 'node'. See **Supplemental Section 10.3** for further detail.

2.2.2. Augmentation of fMRI

Augmentation can improve model performance without introducing spurious findings by introducing realistic perturbations so a training dataset can be representative of more of the population variation than the original dataset (Chlap et al., 2021). We adapt the BLENDS method which generates realistic perturbations in fMRI by warping to new subject's anatomy who was not in the original training set (Nguyen et al., 2022). BLENDS assumes that brain anatomy (the specific shape of the gyri and sulci) is not the primary determinant in the disease process. Therefore, new fMRI can be constructed from existing by warping existing brain shape to that of another subject's with the same diagnosis. In PD, atrophy of deep nuclei but not cortical brain shape is associated with early diagnosis or progression (Zeighami et al., 2019), so the BLENDS methodology is appropriate as an augmentation approach. Augmentation was performed on the training data only and not for validation or testing. The additional anatomic augmentation samples were drawn from disease-matched PDBP subjects with T1 images and no corresponding fMRI.

BLENDS is suitable for deriving connectivity measures because it generates a new timeseries (4D) fMRI. However, the magnitude of the BLENDS-induced variation is small. Across the PDBP diagnostic subject set, the R^2 between each augmented and original timeseries is 0.79+/-0.02. Therefore, we adapted the BLENDS method by increasing the perturbation magnitude. This was achieved in two steps. First, the difference between the original timeseries and initial BLENDS-perturbed timeseries was calculated and the difference was scaled to a gain factor of 0.5, i.e., while the original fMRI timeseries was scaled to 0 mean and unit variance, the difference was scaled to 0 mean and 0.5 variance. Then, this scaled difference was added back to the initial fMRI timeseries to create the augmented timeseries. Finally, connectivity measures are calculated on the augmented timeseries as was done in the initial timeseries. The full augmentation pipeline is illustrated in Fig. 1**.** This augmentation drives the augmented vs. original connectivity R^2 down to as low as 0.26 as shown in Table 2. Augmentation without applying a gain factor resulted in an augmented vs. original connectivity R^2 which was greater than 0.98, that was deemed to be insufficient variation introduced by the augmentation method. The gain factor of 0.5 was empirically chosen to preserve a moderate R^2 for all connectivity measures.

2.2.3. Predictive model training using covariates from fMRI and clinicodemographics

We trained a separate XGBoost predictive model (Chen and Guestrin, 2016) for each of the PDBP targets as it has been shown to be highly efficacious for tabular data and apt to identify predictive signals in other neurodegenerative diseases (Marinescu et al., 2022; Torlay et al., 2017). This machine learning model paradigm has been shown to be a particularly efficacious method for making predictions on complex data where the relation (linear or nonlinear) between the input and target are unknown, particularly when the data is best represented as a wide matrix, as in our case (Marinescu et al., 2022; Torlay et al., 2017).

For an unbiased estimate of model performance, we employed a nested, 10x9 fold cross-validation partitioning, with separate training, validation, and test partitions to hyperparameter optimize (HPO) and evaluate our model. We performed a single stratified split into 10 folds, then held out 1 fold for testing, performed a 9 fold cross validation with the 9 remaining folds, and then permuted the testing fold. We got 10 test performances from 10 separately optimized models. The folds were split with stratification by the target (diagnosis, MDS-UPDRS-III, etc.) age, gender, and ethnicity. Within each inner loop, Bayesian hyperparameter

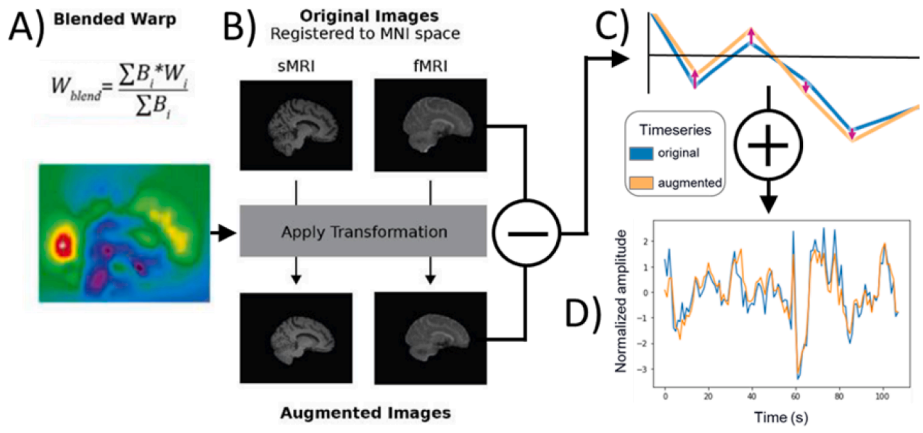

Fig. 1. Augmentation of fMRI timeseries data. **(A)** Initially a set of subjects' T1 sMRI (MPRAGE) is nonlinearly warped to MNI-152 space, and each subject's inverse transform, W_i, is stored in a library of warps. To create a perturbation for each subject fMRI from the original dataset, 5 of these warps are blended with a random spatial Gaussian mixing with blending matrices B_i to create a unique blended warp, W_{blend} **(B)** Next, the original subject had their fMRI perturbed by applying the blended warp to every frame to generate a scan with the original functional activity projected onto the new realistic brain shape (combination of 5 other subjects' brains). **(C)** This generates small but realistic distortions in the extracted mean regional timeseries. **(D)** The small differences in the timeseries were scaled up with a gain factor to generate larger timeseries differences, treating the first realistic distortion from **(C)** as analogous to the initial step in a gradient descent.

Table 2
Similarity post-augmentation shows the similarities of the measures obtained before augmentation and after augmentation. The R^2 for the entire timeseries is on the left, and the R^2 for the augmented connectivity measures are on the right.

R^2 between original and augmented timeseries	Connectivity measure	R^2 between original and augmented connectivity measure
0.79 ± 0.02	Correlation	0.74 ± 0.05
	Partial Correlation	0.78 ± 0.04
	PCA Granger Causality	0.49 ± 0.23
	Structurally Projected Granger Causality	0.26 ± 0.22

optimization (Liaw et al., 2018; Martinez-Cantin, 2014) was performed searching over 128 model configurations. This training setup has been shown to be more representative of real-world model performance after hyperparameter optimization than a single partition (single held-out test set) (Cawley and Talbot, 2010). Further model optimization details including the hyperparameters optimized (including L1 and L2 normalization) are in **Supplemental Section 10.4**.

For all models, each continuous feature was z-score normalized based on the training fold data, while each categorical feature was left as a category. Clinicodemographic covariates including site of data collection, age, gender, ethnicity, education level, and handedness were included as additional covariates to the FC or EC derived connectivity measures. The clinicodemographic data was preserved for each augmentation of a given subject. As the dimensionality of the flattened 1D-vector reformatting of an 2D EC or FC matrix contains up to 18,225 unique elements, we applied a dimensionality reduction per fold step using principal components analysis (PCA) fit on the validation data only where the number of components were chosen using the *kneedle* algorithm (Satopaa et al., 2011). We also compared with an alternative dimensionality reduction technique, univariate dimensionality reduction, in which the top 5 % of the most target correlated edges in the training data are retained for the model. This led to a training dimensionality of 50–150 samples by 20–30 or 50–60 features in the PCA and univariate approaches respectively. For all models, performance on the held out test data is reported. The test data was not seen in model training nor selection and additional checks were placed to ensure there was no leakage of training into validation and no leakage of train or validation into test.

2.2.4. Experiment specifications

2.2.4.1. Experiment 1: Diagnosis of PD vs PSP vs MSA vs normal control. Our first experiment predicted a diagnosis for each patient among these labels: PD, PSP, MSA, or NC. We used the available PDBP data which includes 45 NC patients, 73 PD patients, 21 PSP patients, and 7 MSA patients. We tested our ability to differentiate these diagnoses when using clinicodemographic features and EC and FC measures. These clinicodemographic features included both initial clinical scores (MoCA and UPDRS), as well as age, site, handedness, education level, gender, and ethnicity. The EC and FC measures used included the following: *Corr*, *PCorr*, $ML.FC_{XGB}$, *PC.GC*, and *SP.GC*. We also tested augmentation factors of 0 (no augmentation), augmentation to achieve a uniform distribution over diagnostic groups ('Match' augmentation), augmentation to a uniform distribution with 5x the least populous group, uniform with 10x, and uniform with 20x. We used the HPO training scheme as described above to optimize the XGBoost models, choosing the best model per inner validation fold with the highest balanced accuracy on those inner folds and evaluating its performance in the test partition from the outer fold not seen during model training nor model selection. We then report the test performances (balanced accuracy) on each of the 10 outer folds.

2.2.4.2. Experiment 2: Prognosis of motor symptom (MDS-UPDRS-III) progression. Our second experiment regressed the 1-year change in the MDS-UPDRS-III scores in 63 PD patients. MDS-UPDRS-III scores was acquired while patients were off PD medications for at least 12 h. We performed this regression with the clinicodemographic features and EC or FC measures and include baseline MDS-UPDRS-III as a covariate. We tested augmentation levels of: 0x (no augmentation), augmentation to achieve a uniform distribution over the range of MDS-UPDRS-III change ('Match'), augmentation to 5x the original sample size and sampled to a uniform distribution, then the same with 10x and 20x. We trained XGBoost models with Bayesian hyperparameter optimization (Liaw et al., 2018; Martinez-Cantin, 2014) and chose the hyperparameter configuration with lowest mean absolute error on the validation set. We report the test performance on the outer 10 folds of the best performing models on the inner validation folds.

2.2.4.3. Experiment 3: Prognosis of cognitive symptom (MoCA) progression. Our third experiment regressed the 1-year change in MoCA in the 71 PD patients with available MoCA scores. We used the initial MoCA score as an additional covariate and do not use the initial MDS-UPDRS-III score, but otherwise followed the same experimental setup as the 1-year regression of MDS-UPDRS-III.

We additionally tested if calculating the levodopa equivalent daily dose (LEDD) that each patient received aided in the prediction of UPDRS or MoCA progression. Of the 63 patients in the UPDRS prediction and the 71 in the MoCA prediction, only 19 had adequately collected medication data to effectively calculate the LEDD. LEDD was calculated with the formula outlined in **Supplemental Section 10.10**. The LEDD for these subjects had a mean of 137.8 mg, a median of 118.9 mg, and a range of 0–450.6 mg of Carbidopa-Levodopa IR equivalents. We reran our search process using only the highest performing augmentation strategy using LEDD as a covariate. Since XGBoost can handle missing data with high fidelity (Chen and Guestrin, 2016), this experiment tested whether we can leverage the sparse LEDD data for additional performance benefit.

3. Results

3.1. Experiment 1: Diagnosis of PD vs PSP vs MSA vs normal control (NC)

Fig. 2**a** shows the test performance as measured by balanced accuracy for assigning a diagnosis from PD, PSP, MSA, and NC. Chance *balanced* accuracy is 0.25. Using partial correlation connectivity at an augmentation factor of 10, when combined with clinicodemographic correlates, achieves the highest balanced accuracy of all combinations: 0.68 using univariate dimensionality reduction and 0.65 with the PCA dimensionality reduction. It also achieves an F1 score of 0.67. For this top performing model, the one-vs-all classification accuracy across the ten outer cross-validation folds of HC vs not HC was 0.85, PD vs not PD was 0.72, PSP vs not PD was 0.95, and MSA vs not MSA was 0.80. Note that the balanced accuracy better represents the overall performance, though individual one-vs all classification accuracies are provided for completeness. We performed a *t*-test to evaluate the null hypothesis that the distribution of balanced accuracy across all folds is different than the chance *balanced* accuracy. We then performed multiple comparisons correction using the Benjamini Hochberg procedure at a false-discovery rate (FDR) of 0.05. The resulting p-value was 7.6E-5 for the univariate case and 7.6E-5 for the PCA case. The differences between top-performing methods are not statistically significant. In general, univariate dimensionality reduction provided better performance than PCA dimensionality reduction. Additional metrics are presented in **Supplemental Section 10.5**.

3.2. Experiment 2: Prognosis of motor symptom (MDS-UPDRS-III) progression

The held-out test performance (as measured by MAE) predicting the 1-year change in MDS-UPDRS-III is shown in Fig. 2**b**. Chance MAE is 3.9 and was calculated by predicting the median change in MDS-UPDRS-III for all subjects. The best (lowest) MAE of 1.8 was achieved by the partial correlation connectivity metric using an augmentation factor of 20 and univariate dimensionality reduction, which also achieved an R^2 of 0.69. The corresponding FDR-corrected p-value (testing the H_0 that the MAE is truly different from the chance MAE of 3.9) was 6.0E-7. The best-predictor using an EC measure (SP.GC with PCA dimensionality reduction) was slightly worse, but still achieved a respectable MAE of 2.0, an R^2 of 0.59, and a p-value of 2.0E-6. On average across multiple

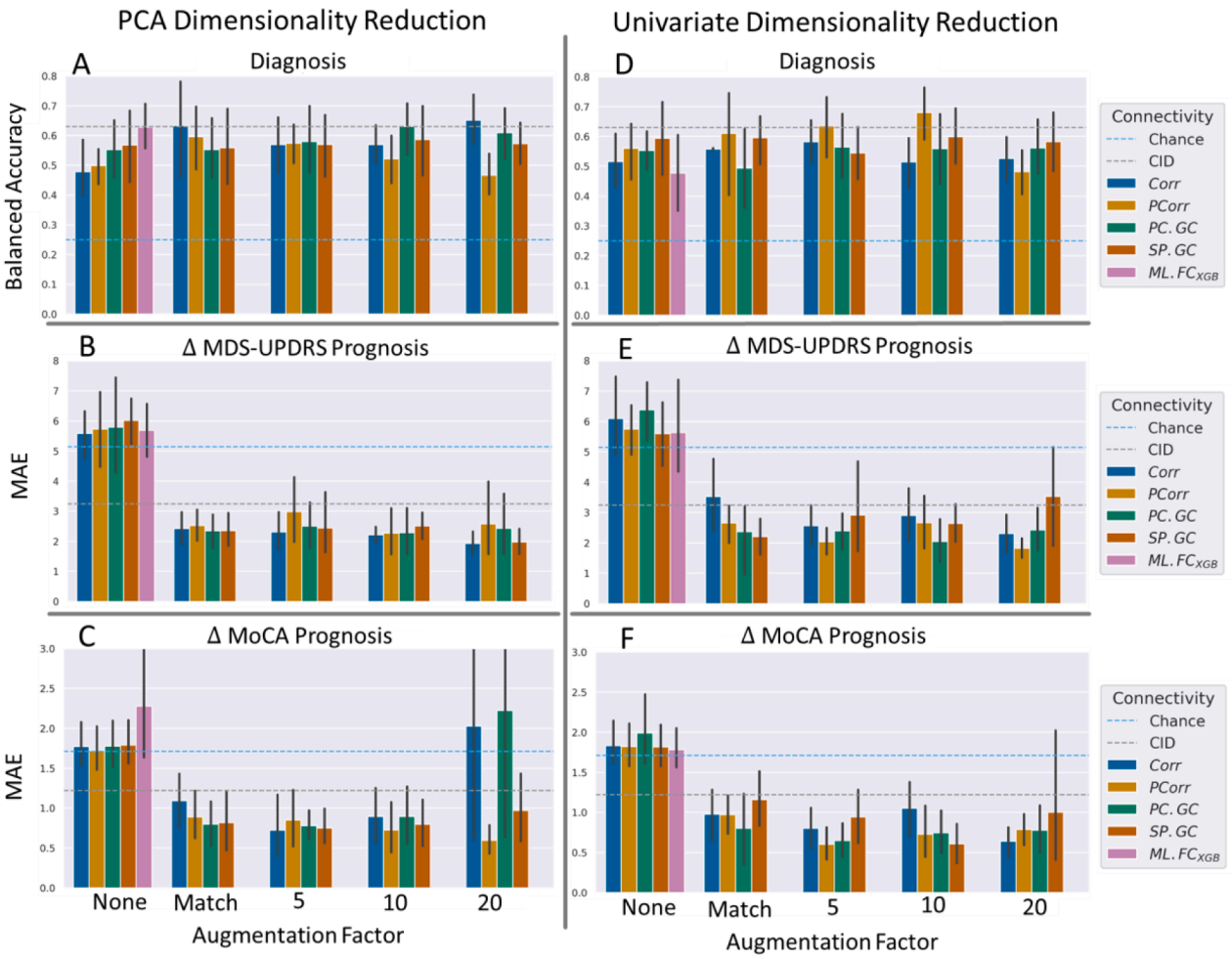

Fig. 2. Diagnosis and prognosis of PD. This figure shows the performance of models trained to predict various targets with different connectivity features and levels of augmentation. The left column (A-C) shows the predictive performance of models trained with a PCA dimensionality reduction step while the right column (D-F) shows the predictive performance of models trained with a univariate dimensionality reduction step. The first row (A and D) shows the balanced accuracy when making a NC/PD/PSP/MSA diagnostic classification, while the second and third rows show the mean absolute error (MAE) predicting the 1-year change in either MDS-UPDRS-III (B and E) or MoCA (C and F), respectively. The bar colors correspond to the connectivity method used. The dashed blue lines show chance accuracy. For the diagnosis task, standard clinical accuracy is also shown. For the prognostic tasks, the clinically important difference (CID) is also shown as dashed gray lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

connectivity measures and augmentation factors, the PCA dimensionality reduction technique outperformed the univariate dimensionality reduction technique. The top performing methods and measures were similar in outcome (i.e. no statistically significant differences between the top performers was observed). See **Supplemental Section 10.6** for additional metrics for every predictor.

3.3. *Experiment 3: Prognosis of cognitive symptom (MoCA) progression*

Fig. 2**c** shows held-out test performance of models predicting the 1-year change in MoCA. Chance performance was an MAE of 1.6. The lowest MAE of 0.60 across predictive feature sets and a best R^2 of 0.81 was achieved using a model with the partial correlation connectivity metric, PCA dimensionality reduction and an augmentation factor of 20x. The FDR corrected p-value was 4.0E-5. The best EC performance was achieved with the SP.GC features with univariate dimensionality reduction and an augmentation factor of 10x. This approach achieved an MAE of 0.61, R^2 of 0.79, and FDR corrected p-value of 8.6E-5. The models using the PCA dimensionality reduction performed equally as well as the univariate dimensionality reduction overall. Additional metrics are shown in **Supplemental Section 10.5**.

In our secondary experiment testing LEDD as a predictor, no statistical improvement was found when including LEDD as a covariate. This does not mean that there is no predictive information in LEDD as a covariate, but rather that this level of sparsity in the data doesn't permit us to appropriately learn the predictive signal, if any. These results are further outlined in **Supplemental Section 10.10**.

3.4. *Primary experiment summary*

Experiments 1–3 achieve results comparable to or significantly better than previous attempts at diagnosis or progression. Table 3 (Discussion) compares these results to other state of the art diagnostic and prognostic methods. Our MDS-UPDRS-III prognostic model improves on previous best MAE from 3.22 to 1.8, and our MoCA prognostic model improves on previous best MAE from 0.74 to 0.60.

3.5. *Augmentation sensitivity analysis*

The results of Experiments 1–3 demonstrate the significant advantages of fMRI augmentation when training predictive models. To probe further the augmentation benefits, we conducted a sensitivity study. This study showed the largest benefit from augmentation occurred between 1 and 3x, or 2-4x the original dataset size. This suggests that the

Table 3
Comparison to the literature. The performance of similar diagnostic and prognostic models in the literature is compared to the performances achieved in this paper. The performance of the best Partial Correlation and best SP.GC models was not significantly different from each other, but both are significantly better than competing models. The reference, data type, modeling approach, and final performance for each diagnostic and prognostic model are shown. This paper's results are highlighted in grey, and the best performances are highlighted as bold italic text (Chahine et al., 2019; Nguyen et al., 2021; Son et al., 2016).

Target	Author	Data				Model	Performance	
		Subjects	*Feature*	*Scan Type*	*On Meds*		*Balanced Acc*	
Diagnosis	(Chougar et al., 2021)	228	Regional values	DTI and T1	NA	Logistic regression	0.77	
	(Zhang et al., 2020)	363	Regional values	T1	Yes	Univariate	***0.83***	
	Our result	146	Clinico-demo	rest fMRI*	No	XGBoost	0.68	
							MAE	*R^2*
MDS-UPDRS-III prognosis	(Nguyen et al., 2021b)	82	ReHO & fALFF	rest fMRI	Yes	Logistic regression	14†¶	0.56
	(Chahine et al., 2019)	413	Volumetry + CSF	T1	Yes	LME	NA	0.44
	(K. H. Leung et al., 2019)	198	Regional values	SPECT	Yes	Deep learning	3.22	NA
	(Son et al., 2016)	100	Genetics + regional	SPECT	Yes	Linear regression	8.36†¶	NA
	Our result	63	Partial Corr	rest fMRI*	No	XGBoost	***1.8***	***0.69***
			SP.GC				***2.0***	***0.59***
							MAE	*R^2*	*Dx acc*
MoCA prognosis	(Lin et al., 2021)	131	Regional values	DTI	Yes	Random Forest	NA	NA	83.9
	(Zeighami et al., 2019)	222	Volumetry	T1	No	Univariate	0.74††	0.06§	NA
	(Silva-Batista et al., 2018)	39	Balance scale	Clinic	Yes	Univariate	NA	0.52¶	NA
	Our result	71	Partial Corr	rest fMRI*	No	XGBoost	***0.60***	***0.81***	NA
			SP.GC				***0.61***	***0.79***	NA

*= task-based treated as rest, † = RMSE, †† = most severe tertile versus least severe only, ¶ = current (not change over time), § = r provided, R2 approximated.

models begin to learn the variation in the disease presentation, when the number of subjects approaches 150–200 subjects, and that future diagnostic and prognostic PD studies might target upwards of 200 subjects. For further sensitivity study results see **Supplemental Section 10.6.**

Identification of the most predictive features (covariates).

To identify the multivariate features most associated with a given outcome, we analyzed the important principal components for each of the prognostic tasks, as the PCA reduced models performed better on average than the univariate-reduced models. This revealed the multivariate associations of both causal and correlative features to the prognostic outcomes. Unlike the prognostic models, the best diagnostic model used clinicodemographic features alone. The important diagnostic features are listed in the supplement.

To maximize the reproducibility of imaging biomarkers, we identify features with high importance across multiple models. This consensus approach of identifying important features has been shown to be more generalizable than features identified with a single model (Botvinik-Nezer, 2020; Mellema and Montillo, 2023). We take the 10 most predictive hyperparameter configurations on each inner loop of training and retrain the corresponding XGBoost models on the entire dataset each. This allows us to estimate distributions of importance across these features. We use the Gini feature importance weighted by the number of samples routed through any given decision node as our feature importance metric for these XGBoost models. We analyze in greater depth the best FC and best EC connectivity features, namely those from Partial Correlation and SP.GC, respectively. We perform this analysis across

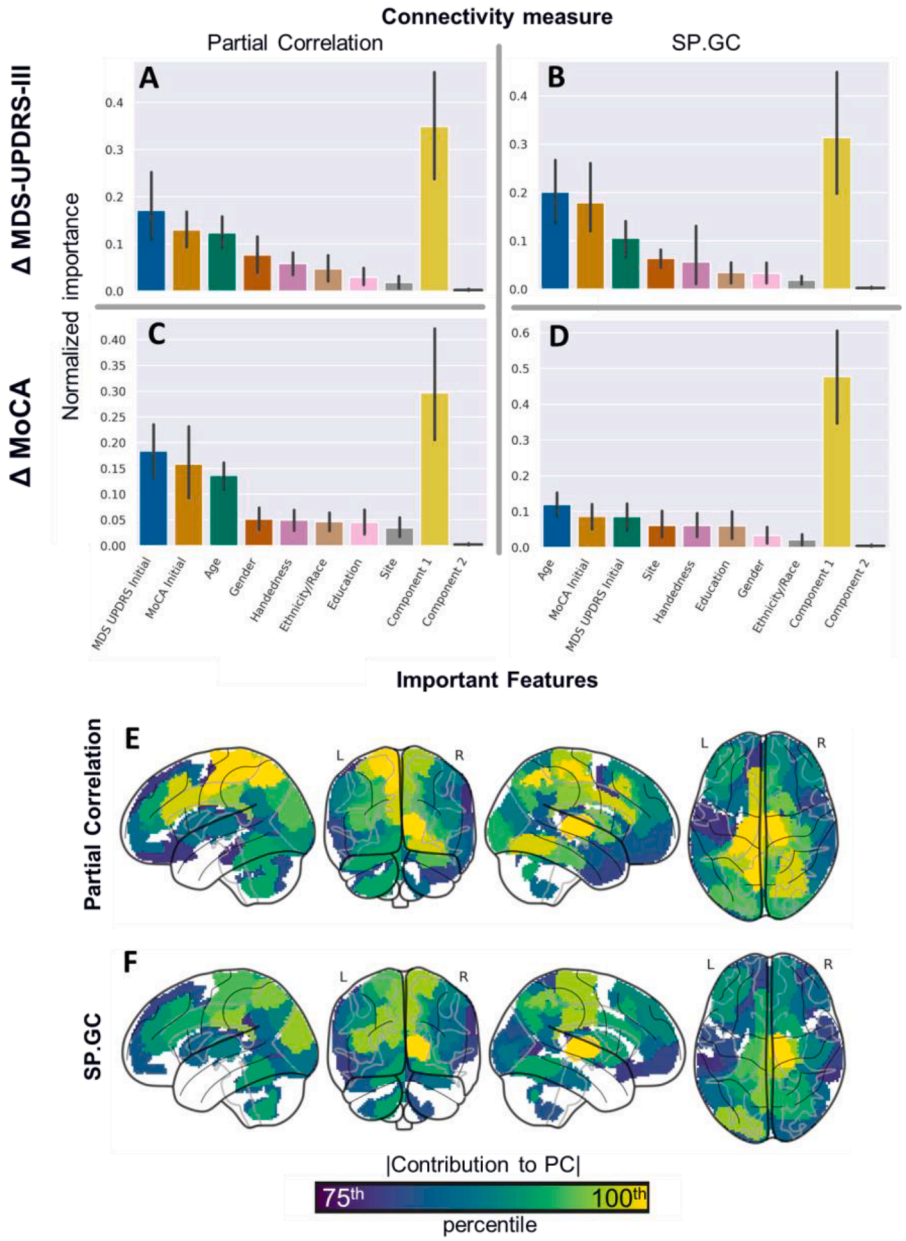

Fig. 3. Feature importances. This figure shows the importance of features and visualizes the connectivity features used by the prognostic models enumerated earlier. Panels **A-D** show the importance (and 95 % CI) of the features with the highest average importance across the top 10 PCA dimensionality reduced XGBoost models on the ΔMDS-UPDRS-III and ΔMoCA prediction tasks. The clinicodemographic features are presented left to right in descending order of importance followed by the connectivity features presented left to right in order of descending importance. The principal components were fit across all data, using the same components for both the ΔMDS-UPDRS-III and ΔMoCA tasks. **A** shows the feature importances from models using Partial Correlation features to predict ΔMDS-UPDRS-III. **B** shows the feature importances from models using SP.GC features to predict ΔMDS-UPDRS-III. **C** shows the feature importances from models using Partial Correlation features to predict ΔMoCA. **D** shows the feature importances from models using SP.GC features to predict ΔMoCA. **E** and **F** show the total weighted degree of each brain region in the connectivity graph of edge contribution to principal component 1. **E** shows the weighted degree of the Partial Correlation PC contributions. **F** shows the weighted degree of the SP.GC PC contributions.

both progression tasks (change in MDS-UPDRS-III and change in MoCA). Fig. 3 shows the calculated feature importances across these tasks. The Partial Correlation and SP.GC features for both the MoCA and MDS-UPDRS-III tasks use just the first component of the PCA decomposition, and that component is more important to the diagnostic model than any single feature (not pictured: the univariate R^2 between the first component and the target varies from 0.02 to 0.11). Fig. 3**A-D** shows that the total importance of the clinicodemographic features exceeds that of the connectivity features, but that the connectivity importance is highest in isolation.

As all models converged on using just the first principal component to achieve maximal validation performance, Fig. 3**E-F** shows a visualization of the first principal component. For context, each principal component weighs every edge in the set of connectivity matrices with a contribution to that component and the components are ordered by the variance across the data the component explains. The first component used by the model captures the most variance across the connectivity matrices. This visualization summarizes the weight per edge of the graph of connections between every ROI as a single, interpretable image. This is done by summing the weighted contribution of every edge feeding into and out of a ROI to get a scalar value per region. In this way the spatial contribution that the principal component represents can be

graphically depicted across the brain using the weighted degree of each brain region. For the component from the Partial Correlation connectivity measure, (Fig. 3**E**), edges are observed that involve the deep nuclei and thalamus, as well as primary and secondary motor areas which are of high importance (>90 percentile) (See **Supplemental Figure S4** for a secondary visualization showing the top *edges*). We also observe contributions from the occipital lobe and prefrontal cortex. The important regions in this component are similar to the default mode network (DMN), with the addition of several deep brain motor nuclei. The similarity with the DMN is also observed to have substantial involvement of the posterior cingulate cortex and precuneus. The component from the SP.GC connectivity measure (Fig. 3**F**) has greatest involvement of edges incident upon the deep nuclei and thalamus, with lesser contributions from similar regions as the Partial Correlation-based component. The primary difference between the Partial Correlation- and SP.GC-based components is the greater importance the SP.GC component places on the deep nuclei. This visualization indicates that connections from deep nuclei to prefrontal areas are significant in these important principal components.

When examining the components in ensemble, rather than individual nodes, the executive control network (ECN) is most strongly implicated in the prognostic task – specifically in the heavy involvement of intra-ECN connections in the component. Second most implicated on a network-level is the salience network, or ventral attention network, including connections to and from the salience network in addition to intra-salience network connections. We direct the reader to **Supplemental Section 10.9 and Supplemental Figure S5** for a more in-depth analysis of the intra and inter-network analysis and discussion.

4. Discussion

4.1. Comparisons to prior work

In the diagnostic task, we achieved a balanced accuracy of 0.68, which is well above chance *balanced* accuracy (0.25) as well as above the accuracy (0.63) attained in general clinics (Hughes et al., 2002). This diagnostic performance was achieved without significant contribution from data present in fMRI as described in the results section above (i.e. from clinicodemographic data alone). We note however that using modalities *other than fMRI*, higher performance has been reported. To appreciate this a comparison of the methods predicting the diagnosis of PD is shown in the top panel of Table 3. For the diagnosis task, the best balanced accuracy is 0.83, reported by Zhang and colleagues (Zhang et al., 2020), using T1-weighted (MPRAGE) MRI containing information primarily about the shape and volume of neuroanatomical structures. Meanwhile, Chougar et al. (Chougar et al., 2021) attained a balanced accuracy of 0.77 using diffusion-weighted MRI containing information primarily about the static structural connectivity between brain regions. Given that the clinicodemographic features used were similar, this suggests that the diagnostic signal present in the functional connectivity of 3 T fMRI is weaker than the signal present in other MRI contrasts. Further experiments are warranted to determine the degree of complementarity between the functional signals from fMRI with the structural ones in DTI and T1, and the benefits from their integration.

For the two longitudinal progression tasks of motor and cognitive decline, our results attain a new state-of-the-art level of performance. Previous attempts to perform these longitudinal measures of prognosis have found MAEs in the longitudinal prediction of MDS-UPDRS-III of 3.22 to 14.0 and R^2s of 0.44 to 0.56, using fMRI, SPECT, genetics, and T1 MRI. Using our recently proposed connectivity measures, we achieve an MAE of 1.8, an R^2 of 0.69 significantly greater than chance, with a p-value of 6.0E-7. Our results are reported on held our test data, using a nested-cross validation model construction approach with careful isolation of training, validation, and test. A comparison of the methods predicting MDS-UPDRS-III is shown in the middle panel of Table 3. The longitudinal prediction of MoCA is comparatively understudied. While studies have predicted the longitudinal progression of PD patients from "cognitively normal" to "minor cognitive impairment" (Lin et al., 2021; Silva-Batista et al., 2018), there is a dearth of attempts to regress more quantitative PD cognitive measures longitudinally. To our knowledge, such an attempt has only been published by Zeighami and colleagues to date (Zeighami et al., 2019). They used only the least and most severe terciles for their prediction, dropping the more difficult-to-predict inner tercile. We predict across all ranges of MoCA progression and achieve a superior performance including both a lower MAE and higher R^2 than previous attempts. A comparison of the methods predicting cognition is shown in the bottom panel of Table 3.

4.2. Significance

Overall, the proposed methods attain excellent (low) MAEs on the held-out test data for both predicting motor decline (change in MDS-UPDRS-III) and cognitive decline (change in MoCA). An ideal model would have an error in prediction less than the clinically important difference (CID) in scores; i.e. the difference where treatment decisions are significantly different. Horváth et al. suggests that the CID may be asymmetric, suggesting a CID for improvement of 3.25 and a CID for worsening of 4.63 (Horváth et al., 2015). We chose to use the more stringent suggested CID of 3.25 for our analysis here. Our model's MAE is below the smallest CID, therefore, we observe that our models ability to predict motor decline meets this bar for clinical utility. For the prediction of *cognitive decline*, Wu et al. identified a minimal CID in MoCA of 1.22 for stroke patients (Wu et al., 2019). While there is not yet a similar study identifying CID for cognitive decline for PD patients, there are measures of clinically meaningful cognitive decline in stroke patients. This threshold for a significant difference in cognitive scores in stroke serves as a proxy target until such time as additional work defines a cognitive CID in PD patients. Upon comparison then, we observe that our models predicting cognitive decline meet this bar for clinical utility, though further work needs to be done to better define thresholds for cognitive outcomes in PD specifically. Finally, we achieve a balanced accuracy for the *diagnosis* of PD versus other Parkinsonian syndromes which compares favorably to the accuracy achieved in non-specialty clinics (Hughes et al., 2002). These results meet our objectives of providing accurate diagnostic and prognostic tools using noninvasive biomarkers which can be further analyzed to potentially examine pathophysiologic differences. Our findings corroborate previous suggestions of a robust functional connectivity signal of PD pathophysiology and extend the previous work in univariate predictors of longitudinal prognosis (Burciu et al., 2016; Hou et al., 2017; Manza et al., 2016; Simioni et al., 2016) to a *multivariate* predictive model. While our work found a multivariate component using many regions and edges was predictive of progression, certain edges and nodes that are of particularly high importance to this component are noteworthy: Those implicated regions include: diffuse involvement within the executive control network, the salience/ventral attention network's connectivity to other large-scale brain networks, and an extensive involvement of all major brain networks to a lesser degree all are implicated by this work. This has been corroborated by previous work which notes broad degradation across diffuse cognitive networks (Mitchell et al., 2021; Tahmasian et al., 2015; Tuovinen et al., 2018). The involvement of these networks implies some level of dysfunction at a global, or near global level, is an early warning of faster disease progression. Singular nodes that were identified as being particularly important in the multivariate predictor include those in the thalamus, putamen, caudate, pallidum, primary and secondary motor areas, and prefrontal cortex – many of which have been similarly implicated in progression (Burciu et al., 2016; Simioni et al., 2016; Wu et al., 2012). The involvement of motor areas specifically should be interpreted with caution due to the "task-as-rest" assumption in our analysis, but the other implicated areas are important brain hubs of connectivity where subtle multivariate changes might be most anticipated to first be revealed. The network and nodal results

taken together suggest that early indication of severity and progression in both motor and cognitive domains may be indicated first with subtle, brain-wide changes in modulating networks and hubs, rather than concentrated to a singular motor-implicated region, as you might expect in a disease with many motor symptoms. The significance of these diffuse modulatory changes in many networks is harder to interpret, but indicates that we should consider broadening our searches for pathophysiologic understanding in future experiments and treatment development. Finally, while the confounding effect of medication can obfuscate the results of longitudinal studies, our results mitigate this effect by using fMRI and MDS-UPDRS-III scores acquired when the patients have weaned overnight from medication use.

Across predictive targets, we found that Partial Correlation and SP.GC measures of connectivity are robust to overfitting as they performed well, never dropping below chance performance, regardless of the level of augmentation. The differences in final performance between the best Partial Correlation and best SP.GC models were statistically indistinguishable and such performance similarity may indicate a convergence to a common underlying signal.

The impact of connectivity measure choice and augmentation amount applied varied depending on the prediction task. For the diagnostic task, augmentation provided minor performance benefit and the choice of connectivity measure had minimal effect. For the prediction of longitudinal motor decline using MDS-UPDRS-III augmentation significantly improved performance, with an augmentation factor of $\geq$ 5x producing the most accurate results. For the prediction of longitudinal cognitive decline with MoCA augmentation was beneficial and the choice of connectivity measure had a lesser effect.

4.3. Limitations and future directions

This work uses the PDBP dataset because the study acquired all measures including imaging and symptomatology in the OFF-medication state, which mitigates the strong levodopa medication confound known to affect both the brain scans and the MDS-UPDRS-III rating. The OFF-medication state is only a wean of 12 h. We acknowledge that while this has been shown sufficient for much of the effects of L-DOPA to wear off, it is not a true "OFF" state, as effects of longer-acting medications have not fully worn off. Also this approach does not mitigate long-term connectivity changes that could be caused by L-DOPA administration. However, a short wean is much safter and more comfortable for the patient, so extending the OFF-medication state could be problematic. Currently, PDBP is the largest database of PD with OFF-medication measures and has been relatively understudied. For increased clinical utility, it would be useful to extend the models developed herein to predict multiple years ahead, with data from a multi-year longitudinal study rather than the single year of tracking which is what is currently available in the PDBP fMRI data subset. Additionally, long fMRI acquisitions of at least 12 min (Birn et al., 2013) with true resting state, rather than task fMRI treated as rest, may also boost signal and further increase model prediction performance. It is likely that differences in performances across connectivity measures will become more apparent as the length of the fMRI acquisition increases and we move to rest, rather than task-based fMRI. Alternatively, the task-based fMRI in question here may have additional analysis opportunities with further extension of classical and our newly proposed connectivity measures, though the use of task-based paradigms limits broad applicability due to potential difficulties in consistently applying these tasks outside of research settings. We also acknowledge that PD diagnosis is defined here by extensive expert clinical evaluation, but a truly definitive diagnosis would use *postmortem* pathological diagnoses (Surmeier et al., 2017). The diagnostic accuracy of the PDBP study is likely similar or better than that found by Jankovic et al.; who found a PSP vs MSA vs PD vs other clinical diagnostic accuracy of 92.6 % confirmed by post-mortem studies when the diagnostician was a movement disorder specialist who followed the patient for several years (Jankovic et al., 2000). A consensus amongst multiple movement disorder experts tracking the patient over multiple years (such as used in this study) likely has even higher accuracy, though this has not yet been studied in the case of PD post-mortem diagnoses. However, this study does use a relatively short disease duration (3 years on average), and later evaluation of these patients when the differentiating phenotypes are more apparent would likely be more accurate and reflective of the true diagnosis of PD vs PSP vs MSA. Furthermore, the progression of 12.5 points on the MDS-UPDRS-III is high, though some of this is explained by measurements being made in the OFF state. It is also possible that this dataset, acquired at a movement-disorder specialty clinic, is biased towards faster progressing cases of PD. Further exploration with larger datasets from different contexts is warranted. Finally, external validation can be performed to further bolster confidence in the model performance and feature importances. Currently, other datasets do not control for ON versus OFF-medication confounds, but we hope more studies in the future will acquire measurements of subjects while off all medication. This could be more representative of the inherent parkinsonism heterogeneity and would likely permit training higher performing models that are even more likely to generalize to disparate patient populations.

5. Conclusions

We have applied both classical and new causal measures of brain connectivity derived from fMRI, presented a novel framework for fMRI timeseries data augmentation, and applied rigorous hyperparameter optimization to construct XGBoost models that achieve state of the art accuracy in predicting an individual's longitudinal decline in key metrics of Parkinson's disease. When predicting the 1-year change in motor function (MDS-UPDRS-III) and 1-year change in cognition (MoCA) for PD patients, we achieve an MAE of 1.8 and 0.60 respectively. These results surpass thresholds of clinical utility while using an imaging modality becoming more and more available in clinical settings. We identified specific connections that associate with improved prognosis. Our results confirm the presence of early, sensitive connectivity biomarkers of the progression of Parkinson's disease, capable of accurately predicting even 1-year decline in motor and cognitive performances. Prognostic tools such as this could be used for subject selection for disease-modifying trials that have sufficient power to detect change despite the typically short trial duration because of enrichment for rate of deterioration. These results therefore can help identify effective candidate therapies, create potential for furthering the pathophysiological understanding of PD, and ultimately pave a path to provide patients and clinicians alike with much-needed tools of clinically relevant prognostic indicators where none have previously existed.

Credit authorship contribution statement

Cooper J. Mellema: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kevin P. Nguyen:** Writing – review & editing, Validation, Software, Resources, Methodology. **Alex Treacher:** Writing – review & editing, Software, Project administration, Data curation. **Aixa X. Andrade:** Writing – review & editing, Software, Methodology, Data curation. **Nader Pouratian:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Vibhash D. Sharma:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Padraig O'Suileabhain:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Albert A. Montillo:** Visualization, Methodology, Investigation, Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used was open source. Links to data and code repository are provided in manuscript

Acknowledgements

The authors thank Drs. Debra Babcock and David Vaillancourt for their assistance in accessing and answering questions about the latest PDBP dataset. Cooper Mellema was supported by NIH NINDS F31 fellowship NS115348. Albert Montillo and Alex Treacher were supported by NIH NIA R01AG059288. Albert Montillo was additionally supported by NIH NIGMS R01GM144486, NIH NCI U01CA207091, the King Foundation, and the Lyda Hill Foundation. Figures for this manuscript were generated with the Matplotlib and Seaborn packages (Hunter, 2007; Michael L. Waskom, 2021). Analysis was performed with the XGBoost, sklearn, and Statsmodels packages (Chen and Guestrin, 2016; Pedregosa et al., 2011; Seabold and Perktold, 2010).

Materials and Ethics Statement

To facilitate reuse and extension, the authors are pleased to provide full source code for the connectivity measures and the analyses of this manuscript. The repository: https://github.com/DeepLearningForPrecisionHealthLab/CausalMeasures contains the implementation of the underlying causal measures. The code to apply these methods specifically to Parkinson's disease are in the repository: https://github.com/DeepLearningForPrecisionHealthLab/ParkinsonsPrognosis. The Parkinson's data came from the NIH PDBP dataset which is publicly available at (https://pdbp.ninds.nih.gov/our-data). The data used for this study were deidentified in accordance with NIH and HIPPA guidelines. All data was gathered with informed consent from all participants.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2024.103571.

References

Amboni, M., Tessitore, A., Esposito, F., Santangelo, G., Picillo, M., Vitale, C., Giordano, A., Erro, R., de Micco, R., Corbo, D., Tedeschi, G., Barone, P., 2015. Resting-state functional connectivity associated with mild cognitive impairment in Parkinson's disease. J. Neurol. 262, 425–434. https://doi.org/10.1007/s00415-014-7591-5.

Baggio, H.-C., Segura, B., Sala-Llonch, R., Marti, M.-J., Valldeoriola, F., Compta, Y., Tolosa, E., Junqué, C., 2015. Cognitive impairment and resting-state network connectivity in Parkinson's disease. Hum. Brain Mapp. 36, 199–212. https://doi.org/10.1002/hbm.22622.

Barnett, L., Barrett, A.B., Seth, A.K., 2018. Misunderstandings regarding the application of Granger causality in neuroscience. PNAS 115, E6676–E6677. https://doi.org/10.1073/pnas.1714497115.

Beheshtian, E., Jalilianhasanpour, R., Modir Shanechi, A., Sethi, V., Wang, G., Lindquist, M.A., Caffo, B.S., Agarwal, S., Pillai, J.J., Gujar, S.K., Sair, H.I., 2021. Identification of the Somatomotor Network from Language Task-based fMRI Compared with Resting-State fMRI in Patients with Brain Lesions. Radiology 301, 178–184. https://doi.org/10.1148/radiol.2021204594.

Bielczyk, N.Z., Uithol, S., van Mourik, T., Anderson, P., Glennon, J.C., Buitelaar, J.K., 2019. Disentangling causal webs in the brain using functional magnetic resonance imaging: A review of current approaches. Network Neuroscience (cambridge, Mass.) 3, 237–273. https://doi.org/10.1162/netn_a_00062.

Birn, R.M., Molloy, E.K., Patriat, R., Parker, T., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., 2013. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage 83, 550–558. https://doi.org/10.1016/j.neuroimage.2013.05.099.

Botvinik-Nezer, R.E.A., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582, 84–88. https://doi.org/10.1038/s41586-020-2314-9.

Burciu, R.G., Chung, J.W., Shukla, P., Ofori, E., Li, H., McFarland, N.R., Okun, M.S., Vaillancourt, D.E., 2016. Functional MRI of disease progression in Parkinson disease and atypical parkinsonian syndromes. Neurology 87, 709–717. https://doi.org/10.1212/WNL.0000000000002985.

Cawley, G.C., Talbot, N., 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Chahine, Lana M., Siderowf, Andrew, Barnes, Janel, Seedorff, Nicholas, Caspell-Garcia, Chelsea, Simuni, Tanya, et al., 2019. Predicting Progression in Parkinson's Disease Using Baseline and 1-Year Change Measures. J. Parkinson's Dis. 9 (4), 665–679. https://doi.org/10.3233/JPD-181518.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794.

Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A., 2021. A review of medical image data augmentation techniques for deep learning applications. J. Med. Imaging Radiat. Oncol. 65, 545–563.

Chockanathan, U., DSouza, A.M., Abidin, A.Z., Schifitto, G., Wismüller, A., 2019. Automated diagnosis of HIV-associated neurocognitive disorders using large-scale Granger causality analysis of resting-state functional MRI. Comput. Biol. Med. 106, 24–30. https://doi.org/10.1016/j.compbiomed.2019.01.006.

Chougar, L., Faouzi, J., Pyatigorskaya, N., Yahia-Cherif, L., Gaurav, R., Biondetti, E., Villotte, M., Valabrègue, R., Corvol, J.-C., Brice, A., Mariani, L.-L., Cormier, F., Vidailhet, M., Dupont, G., Piot, I., Grabli, D., Payan, C., Colliot, O., Degos, B., Lehéricy, S., 2021. Automated Categorization of Parkinsonian Syndromes Using Magnetic Resonance Imaging in a Clinical Setting. Mov. Disord. 36, 460–470. https://doi.org/10.1002/mds.28348.

Dadi, K., Rahim, M., Abraham, A., Chyzhyk, D., Milham, M., Thirion, B., Varoquaux, G., 2019. Benchmarking functional connectome-based predictive models for resting-state fMRI. Neuroimage 192, 115–134. https://doi.org/10.1016/j.neuroimage.2019.02.062.

Friston, K.J., Preller, K.H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., Zeidman, P., 2019. Dynamic causal modelling revisited. Neuroimage 199, 730–744. https://doi.org/10.1016/j.neuroimage.2017.02.045.

Hassan, M., Chaton, L., Benquet, P., Delval, A., Leroy, C., Plomhause, L., Moonen, A.J.H., Duits, A.A., Leentjens, A.F.G., van Kranen-Mastenbroek, V., Defebvre, L., Derambure, P., Wendling, F., Dujardin, K., 2017. Functional connectivity disruptions correlate with cognitive phenotypes in Parkinson's disease. NeuroImage: Clinical 14, 591–601. https://doi.org/10.1016/j.nicl.2017.03.002.

Horváth, K., Aschermann, Z., Ács, P., Deli, G., Janszky, J., Komoly, S., Balázs, É., Takács, K., Karádi, K., Kovács, N., 2015. Minimal clinically important difference on the Motor Examination part of MDS-UPDRS. Parkinsonism Relat. Disord. 21, 1421–1426. https://doi.org/10.1016/j.parkreldis.2015.10.006.

Hou, Y., Luo, C., Yang, J., Ou, R., Liu, W., Song, W., Gong, Q., Shang, H., 2017. Default-mode network connectivity in cognitively unimpaired drug-naïve patients with rigidity-dominant Parkinson's disease. J. Neurol. 264, 152–160. https://doi.org/10.1007/s00415-016-8331-9.

Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y., Lees, A.J., 2002. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain 125, 861–870.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.5281/zenodo.3264781.

Jankovic, J., Tan, E.K., 2020. Parkinson's disease: etiopathogenesis and treatment. J. Neurol. Neurosurg. Psychiatry 91, 795–808. https://doi.org/10.1136/jnnp-2019-322338.

Jankovic, J., Rajput, A.H., McDermott, M.P., Perl, D.P., for the Parkinson Study Group, 2000. The Evolution of Diagnosis in Early Parkinson Disease. Arch. Neurol. 57, 369–372. https://doi.org/10.1001/archneur.57.3.369.

Kraus, B.T., Perez, D., Ladwig, Z., Seitzman, B.A., Dworetsky, A., Petersen, S.E., Gratton, C., 2021. Network variants are similar between task and rest states. Neuroimage 229, 117743. https://doi.org/10.1016/j.neuroimage.2021.117743.

Krishnan, K., Rossetti, H., Hynan, L.S., Carter, K., Falkowski, J., Lacritz, L., Cullum, C.M., Weiner, M., 2017. Changes in Montreal Cognitive Assessment Scores Over Time. Assessment 24, 772–777. https://doi.org/10.1177/1073191116654217.

Leung, K.H., Salmanpour, M.R., Saberi, A., Klyuzhin, I.S., Sossi, V., Jha, A.K., Pomper, M. G., Du, Y., Rahmim, A., 2019. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC): Conference proceedings. IEEE, Piscataway, NJ.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I., 2018. Tune: A Research Platform for Distributed Model Selection and Training, 8 pp. http://arxiv.org/pdf/1807.05118v1.

Lin, H., Liu, Z., Yan, W., Zhang, D., Liu, J., Xu, B., Li, W., Zhang, Q., Cai, X., 2021. Brain connectivity markers in advanced Parkinson's disease for predicting mild cognitive impairment. Eur. Radiol. 31, 9324–9334. https://doi.org/10.1007/s00330-021-08086-3.

Mansu, K., Seong-Jin, S., Hyunjin, P., 2017. Imaging genetics approach to predict progression of Parkinson's diseases. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. In: IEEE Engineering in Medicine and Biology Society. Annual International Conference 2017, pp. 3922–3925, 10.1109/EMBC.2017.8037714.

Manza, P., Zhang, S., Li, C.-S.-R., Leung, H.-C., 2016. Resting-state functional connectivity of the striatum in early-stage Parkinson's disease: Cognitive decline and motor symptomatology. Hum. Brain Mapp. 37, 648–662. https://doi.org/10.1002/hbm.23056.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C.S., Caspell-Garcia, C., Simuni, T., Jennings, D., Tanner, C.M., Trojanowski, J.Q., Shaw, L.M., Seibyl, J.,

Schuff, N., Singleton, A., Kieburtz, K., Toga, A.W., Mollenhauer, B., Galasko, D., Chahine, L.M., Weintraub, D., Foroud, T., Tosun-Turgut, D., Poston, K., Arnedo, V., Frasier, M., Sherer, T., 2018. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. Ann. Clin. Transl. Neurol. 5, 1460–1477. https://doi.org/10.1002/acn3.644.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Eshaghi, A., Toni, T., Salaterski, M., Lunina, V., Ansart, M., Durrleman, S., Lu, P., Iddi, S., Li, D., Thompson, W.K., Donohue, M.C., Nahon, A., Levy, Y., Halbersberg, D., Cohen, M., Liao, H., Li, T., Yu, K., Zhu, H., Tamez-Pena, J. G., Ismail, A., Wood, T., Bravo, H.C., Nguyen, M., Sun, N., Feng, J., Yeo, B.T.T., Chen, G., Qi, K., Chen, S., Qiu, D., Buciuman, I., Kelner, A., Pop, R., Rimocea, D., Ghazi, M.M., Nielsen, M., Ourselin, S., Sorensen, L., Venkatraghavan, V., Liu, K., Rabe, C., Manser, P., Hill, S.M., Howlett, J., Huang, Z., Kiddle, S., Mukherjee, S., Rouanet, A., Taschler, B., Tom, B.D.M., White, S.R., Faux, N., Sedai, S., Oriol, J.d.V., Clemente, E.E.V., Estrada, K., Aksman, L., Altmann, A., Stonnington, C.M., Wang, Y., Wu, J., Devadas, V., Fourrier, C., Raket, L.L., Sotiras, A., Erus, G., Doshi, J., Davatzikos, C., Vogel, J., Doyle, A., Tam, A., Diaz-Papkovich, A., Jammeh, E., Koval, I., Moore, P., Lyons, T.J., Gallacher, J., Tohka, J., Ciszek, R., Jedynak, B., Pandya, K., Bilgel, M., Engels, W., Cole, J., Golland, P., Klein, S., Alexander, D.C., 2022. The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up. Machine Learning for Biomedical Imaging (MELBA).

Martinez-Cantin, R., 2014. BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. J. Mach. Learn. Res. 15, 3915–3919.

Mellema, C.J., Montillo, A., 2023. Novel Machine Learning Approaches for Improving the Reproducibility and Reliability of Functional and Effective Connectivity from Functional MRI:. Prepublished on arXiv. In press J. Neural Eng. https://doi.org/10.48550/ARXIV.2201.13378.

Mellema, C.J., Nguyen, K.P., Treacher, A., Montillo, A., 2022. Reproducible neuroimaging features for diagnosis of autism spectrum disorder with machine learning. Sci. Rep. 12, 3057. https://doi.org/10.1038/s41598-022-06459-2.

Mitchell, T., Lehéricy, S., Chiu, S.Y., Strafella, A.P., Stoessl, A.J., Vaillancourt, D.E., 2021. Emerging Neuroimaging Biomarkers Across Disease Stage in Parkinson Disease: A Review. JAMA Neurol. 78, 1262–1272. https://doi.org/10.1001/jamaneurol.2021.1312.

Ng, B., Varoquaux, G., Poline, J.B., Thirion, B., Greicius, M.D., Poston, K.L., 2017. Distinct alterations in Parkinson's medication-state and disease-state connectivity. NeuroImage: Clin. 16, 575–585. https://doi.org/10.1016/j.nicl.2017.09.004.

Nguyen, K.P., Raval, V., Treacher, A., Mellema, C., Yu, F.F., Pinho, M.C., Subramaniam, R.M., Dewey, R.B., Montillo, A.A., 2021. Predicting Parkinson's disease trajectory using clinical and neuroimaging baseline measures. Parkinsonism Relat. Disord. 85, 44–51. https://doi.org/10.1016/j.parkreldis.2021.02.026.

Nguyen, K.P., Raval, V., Minhajuddin, A., Carmody, T., Trivedi, M.H., Dewey, R.B., Montillo, A.A., 2022. The BLENDS method for data augmentation of 4-dimensional brain images. J. Brain Connect. https://doi.org/10.1089/brain.2021.0186.

Noble, S., Scheinost, D., Constable, R.T., 2019. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. Neuroimage 203, 116157. https://doi.org/10.1016/j.neuroimage.2019.116157.

Ofori, E., Du, G., Babcock, D., Huang, X., Vaillancourt, D.E., 2016. Parkinson's disease biomarkers program brain imaging repository. Neuroimage 124, 1120–1124. https://doi.org/10.1016/j.neuroimage.2015.05.005.

Olde Dubbelink, K.T., Schoonheim, M.M., Deijen, J.B., Twisk, J.W., Barkhof, F., Berendse, H.W., 2014. Functional connectivity and cognitive decline over 3 years in Parkinson disease. Neurology 83, 2046–2053.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pringsheim, T., Jette, N., Frolkis, A., Steeves, T.D.L., 2014. The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov. Disord. 29, 1583–1590. https://doi.org/10.1002/mds.25945.

Raval, V., Nguyen, K.P., Pinho, M., Dewey, R.B., Trivedi, M., Montillo, A.A., 2022. Pitfalls and recommended strategies and metrics for suppressing motion artifacts in functional MRI. Neuroinformatics. https://doi.org/10.1007/s12021-022-09565-8.

Rolls, E.T., Huang, C.-C., Lin, C.-P., Feng, J., Joliot, M., 2020. Automated anatomical labelling atlas 3. Neuroimage 206, 116189. https://doi.org/10.1016/j.neuroimage.2019.116189.

Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a "kneedle" in a Haystack: Detecting Knee Points in System Behavior. In: 2011 31st International Conference on Distributed Computing Systems Workshops. 2011 31st International Conference on Distributed Computing Systems Workshops (ICDCS Workshops), Minneapolis, MN, USA. 6/19/2011 - 6/23/2011. IEEE, pp. 166–171.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral cortex (New York, N.Y. : 1991) 28, 3095–3114. https://doi.org/10.1093/cercor/bhx179.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference.

Silva-Batista, C., Corcos, D.M., Kanegusuku, H., Piemonte, M.E.P., Gobbi, L.T.B., de Lima-Pardini, A.C., de Mello, M.T., Forjaz, C.L.M., Ugrinowitsch, C., 2018. Balance and fear of falling in subjects with Parkinson's disease is improved after exercises with motor complexity. Gait Posture 61, 90–97. https://doi.org/10.1016/j.gaitpost.2017.12.027.

Simioni, A.C., Dagher, A., Fellows, L.K., 2016. Compensatory striatal-cerebellar connectivity in mild-moderate Parkinson's disease. NeuroImage: Clinical 10, 54–62. https://doi.org/10.1016/j.nicl.2015.11.005.

Smitha, K.A., Akhil Raja, K., Arun, K.M., Rajesh, P.G., Thomas, B., Kapilamoorthy, T.R., Kesavadas, C., 2017. Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks. Neuroradiol. J. 30, 305–317. https://doi.org/10.1177/1971400917697342.

Son, S.J., Kim, M., Park, H., 2016. Imaging analysis of Parkinson's disease patients using SPECT and tractography. Sci. Rep. 6, 38070. https://doi.org/10.1038/srep38070.

Stokes, P.A., Purdon, P.L., 2017. A study of problems encountered in Granger causality analysis from a neuroscience perspective. PNAS 114, E7063–E7072. https://doi.org/10.1073/pnas.1704663114.

Stoodley, C.J., Valera, E.M., Schmahmann, J.D., 2012. Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. Neuroimage 59, 1560–1570. https://doi.org/10.1016/j.neuroimage.2011.08.065.

Surmeier, D.J., Obeso, J.A., Halliday, G.M., 2017. Selective neuronal vulnerability in Parkinson disease. Nat. Rev. Neurosci. 18, 101–113. https://doi.org/10.1038/nrn.2016.178.

Tahmasian, M., Bettray, L.M., van Eimeren, T., Drzezga, A., Timmermann, L., Eickhoff, C. R., Eickhoff, S.B., Eggers, C., 2015. A systematic review on the applications of resting-state fMRI in Parkinson's disease: Does dopamine replacement therapy play a role? Cortex 73, 80–105. https://doi.org/10.1016/j.cortex.2015.08.005.

Taylor, K.I., Sambataro, F., Boess, F., Bertolino, A., Dukart, J., 2018. Progressive Decline in Gray and White Matter Integrity in de novo Parkinson's Disease: An Analysis of Longitudinal Parkinson Progression Markers Initiative Diffusion Tensor Imaging Data. Front. Aging Neurosci. 10, 318. https://doi.org/10.3389/fnagi.2018.00318.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciu, M., 2017. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy. Brain Informatics 4, 159–169. https://doi.org/10.1007/s40708-017-0065-7.

Tuovinen, N., Seppi, K., de Pasquale, F., Müller, C., Nocker, M., Schocke, M., Gizewski, E. R., Kremser, C., Wenning, G.K., Poewe, W., Djamshidian, A., Scherfler, C., Seki, M., 2018. The reorganization of functional architecture in the early-stages of Parkinson's disease. Parkinsonism Relat. Disord. 50, 61–68. https://doi.org/10.1016/j.parkreldis.2018.02.013.

Waskom, M.L., 2021. seaborn: statistical data visualization. J. Open Source Software 6, 3021. https://doi.org/10.21105/joss.03021.

Wu, C.-Y., Hung, S.-J., Lin, K.-C., Chen, K.-H., Chen, P., Tsay, P.-K., 2019. Responsiveness, Minimal Clinically Important Difference, and Validity of the MoCA in Stroke Rehabilitation. Occup. Ther. Int. 2019, 2517658. https://doi.org/10.1155/2019/2517658.

Wu, T., Wang, J., Wang, C., Hallett, M., Zang, Y., Wu, X., Chan, P., 2012. Basal ganglia circuits changes in Parkinson's disease patients. Neurosci. Lett. 524, 55–59. https://doi.org/10.1016/j.neulet.2012.07.012.

Yao, Q., Zhu, D., Li, F., Xiao, C., Lin, X., Huang, Q., Shi, J., 2017. Altered Functional and Causal Connectivity of Cerebello-Cortical Circuits between Multiple System Atrophy (Parkinsonian Type) and Parkinson's Disease. Front. Aging Neurosci. 9, 266. https://doi.org/10.3389/fnagi.2017.00266.

Yeh, F.-C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J.C., Vettel, J.M., Verstynen, T., 2018. Population-averaged atlas of the macroscale human structural connectome and its network topology. Neuroimage 178, 57–68. https://doi.org/10.1016/j.neuroimage.2018.05.027.

Zeighami, Y., Fereshtehnejad, S.-M., Dadar, M., Collins, D.L., Postuma, R.B., Dagher, A., 2019. Assessment of a prognostic MRI biomarker in early de novo Parkinson's disease. NeuroImage: Clin. 24, 101986 https://doi.org/10.1016/j.nicl.2019.101986.

Zhang, Y., Wang, H., Xu, D., Hou, B., Lin, T., Shi, L., Luo, Y., You, H., Feng, F., 2020. Differential diagnosis of parkinsonian degenerative disorders in combination manual measurements with automated volumetry of the brain. Preprint, 29 pp.