Title of submission to PLOS journals

Name1 Surname[1,2☯], Name2 Surname[2☯], Name3 Surname[2,3¤], Name4 Surname[2], Name5 Surname[2‡], Name6 Surname[2‡], Name7 Surname[1,2,3*], with the Lorem Ipsum Consortium[¶]

1 Affiliation Dept/Program/Center, Institution Name, City, State, Country
2 Affiliation Dept/Program/Center, Institution Name, City, State, Country
3 Affiliation Dept/Program/Center, Institution Name, City, State, Country

☯These authors contributed equally to this work.
‡These authors also contributed equally to this work.
¤Current Address: Dept/Program/Center, Institution Name, City, State, Country
†Deceased
¶Membership list can be found in the Acknowledgments section.
* correspondingauthor@institute.edu

Abstract

this section will contain the abstract

Author summary

Refer to comments in the Latex template as this may not be necessary.

Introduction

this section will contain the introduction text

Materials and methods

Cohort construction

Both the original study by Shu et al. [1] and the subsequent replication and reproducibility study by Arafe et al. [2] selected patients with Parkinson's disease (PD) from the PPMI dataset and matched their age, sex, and H&Y score from the first of two visits spanning approximately 36 months apart. For the first visit, each patient underwent an evaluation consisting of a clinical assessment and an MRI scan. They also had a follow-up clinical examination 3 years later. Patients were classified as 'progressive' if their H&Y score at the follow-up visit exceeded the score from 3 years prior; otherwise, they were classified as 'stable'. Regarding inclusion criteria, Shu et al. created a cohort by limiting their selection to patients with MRI data from a Siemens Verio 3T MRI machine, incorporating restrictions on repetition time, echo time, inversion time, field of view, matrix size, and slice thickness. Their final cohort comprised 144 patients, equally distributed between progressive and stable subjects. On the other hand, rather than a single cohort, Arafe et al. constructed 5 cohorts, each consisting of 72 stable and 72 progressive subjects. One cohort aimed to replicate the

Shu et al. cohort, while the other four were designed to assess the sensitivity of model predictions to the selection process. Table 1 shows the various filters Arafe et al. used to create their 5 cohorts.

Table 1. Summary of Arafe et al.'s PPMI filters used to construct their 5 cohorts. Source: [2]

	VRC	**SRC**	**MSRC**	**NFRC**	**FSC**
Research Group	PD	PD	PD	PD	PD
Acquisition Type	3D	3D	3D	3D	3D
Field Strength	3T	3T	3T	any	3T
Slice Thickness	1mm	1mm	1mm	1mm $\leq$ 1.2mm	1mm
Manufacturer	Siemens	Siemens	any	any	Siemens
Manufacturer model	Verio	any	any	any	any
Weighting	T1	T1	T1	T1	T1

Like Arafe et al. and Shu et al., as part of the selection process, we filtered the subjects in the PPMI database using the following inclusion criteria:

- **C1**: patient has a diagnosis of idiopathic PD;
- **C2**: PPMI database contains records of at least 2 visits spaced approximately 3 years apart;
- **C3**: PPMI database contains a T1-weighted MRI from the first visit determined by C2;
- **C4**: PPMI database contains H&Y scores for both visits.

We used a variation of the PPMI filters in Table 1 to construct 7 cohorts. Specifically, the selection process for Arafe et al.'s Multiple Scanner Replication Cohort (MSRC) and Functional State Cohort (FSC) form the basis for the remaining inclusion criteria. Like MSRC, our 7 cohorts have been constructed with the following additional criteria:

- **C5**: all MRI machine manufacturers and models are permissable;
- **C6**: scanner restrictions: slice thickness = 1 mm and field strength = 3T;

After filtering the PPMI dataset for Criteria 1 to 6, visit pairs were formed for the remaining subjects while ensuring C2 and C3 remained true for each visit pair. During this phase, an additional restriction was imposed for the formation of Cohorts 1 to 6 such that a patient's functional state (also called PD state), which may be "On" or "Off", must be the same for both visits. The functional state of a patient, "On" or "Off", is determined by their PD medication status during clinical examinations. As the classification into progressive or stable groups depends on the stability of H&Y scores over time, this restriction bolsters the comparability of these scores. Note that this restriction was also imposed on Arafe et al.'s FSC [2].

As the PPMI dataset is collected over an extended period of time and the protocol requires clinical evaluation in both "On" and "Off" functional states for every visit, many subjects have more than one visit pair and it is possible for a subject to be classified as both progressive and stable for different visit pairs. Therefore, after the creation of a cohort, validation checks were defined to ensure that any given subject was only included once, as either stable or progressive, in the resulting cohort.

One of our objectives was to create the largest possible cohorts that adhered to the stated restrictions. Thus, although Cohorts 1, 3, 5, and 7 were created by matching

patients from the progressive and stable classes on age, sex and H&Y scores from the first visit, Cohorts 2, 4, and 6 used no matching filter. This approach allowed for larger cohorts but necessitated the creation of demographics feature sets to evaluate the impact of this uneven distribution on the results. Cohorts 1 and 2 sampled visit pairs with PD state = "Off", Cohorts 5 and 6 sampled pairs with PD state = "On", and Cohorts 3 and 4 sampled visit pairs with either PD state = "Off" or PD state = "On" for both visits. Cohort 7 was defined without any restrictions on the PD state (see Fig 1).

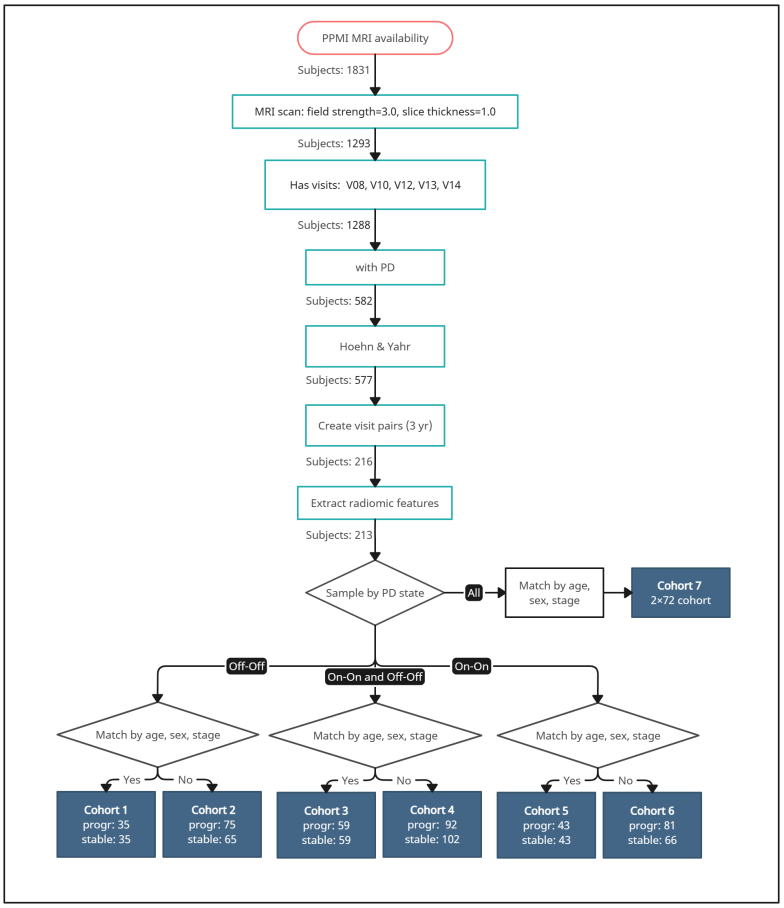

Fig 1. Cohort construction Process of filtering the PPMI dataset to construct 7 cohorts.

Results

Cohorts

The demographics for Cohorts 1 to 7 are summarized in Table 2. Cohorts 1 and 2 were sampled from the same set of visit pairs, comprising a maximum of 140 subjects with PD state = "Off" for both visits. Cohort 2 randomly sampled one visit pair for each of the available subjects and distributed them to each class as evenly as possible. This resulted in 65 stable subjects and 75 progressive subjects. The mean age and standard deviation for the stable subjects is 62.3±9.3 and 60.0±9.7 for the progressive subjects. There are 23 females and 42 males in the stable group, 28 females and 47 males in the

progressive group. Regarding H&Y scores, 45 stable subjects have a score of 1, 17 a score of 2, and 3 have a score of 3. For the progressive subjects, only 8 have an H&Y score = 1 and 67 have an H&Y score = 2. Cohort 1 is the smallest of the constructed cohorts with only 35 stable and 35 progressive subjects for a total of 70 subjects. The stable and progressive groups are composed of 13 females and 22 males, and 15 females and 20 males respectively. The mean age and standard deviation for the stable subjects is 60.5±7.6 and 61.9±8.2 for the progressive subjects. There are 18 subjects with H&Y score = 1 and 17 with score = 2 for both the stable and progressive groups.

Cohorts 3 and 4 sampled from a set of visit pairs for 194 subjects with either PD state = "Off" or PD state ="On" for both visits. Cohort 3 has 118 subjects with 59 subjects in each class. The stable group has 21 females and 38 males, whereas the progressive group has 25 females and 34 males. The mean age and standard deviation for the stable subjects and progressive subjects is 60.6±8.7 and 62.0±9.4 respectively. Both the stable and progressive groups have 26 subjects with H&Y score = 1 and 33 subjects with H&Y score = 2. Cohort 4 is the largest cohort and is composed of 102 stable subjects and 92 progressive subjects. The mean age and standard deviation for the stable and progressive groups are very similar to each other at 62.3±10.0 and 62.3±9.6 respectively. Cohort 4 has 38 females and 64 males in the stable group, whereas the progressive group has 35 females and 57 males. The number of subjects in the stable group that have H&Y score = 1 is 64, H&Y score = 2 is 34, and H&Y score = 3 is 4. The number of subjects in the progressive group that have an H&Y score of 1, 2, or 3 are 9, 83, and 0 respectively.

Cohorts 5 and 6 sampled from the set of visit pairs defined for the 147 subjects with PD state = "On" for both visits. Cohort 5 has 86 subjects evenly distributed into stable and progressive groups of 43 each. The stable group is composed of 18 females and 25 males, and the progressive group has 16 females and 27 males. The mean age and standard deviation for the stable group is 63.0±8.9 and 62.7±10.5 for the progressive group. There are 17 subjects in each group with H&Y score = 1 and 26 subjects in each group with H&Y score = 2. Cohort 6 has 66 stable subjects and 81 progressive subjects for a total of 147 subjects. The stable group is composed of 24 females and 42 males with a mean age and standard deviation of 62.7±10.4, and the progressive group has 33 females and 48 males with a mean age of 63.5±9.5. The stable group has 36 subjects with H&Y score = 1, 28 subjects with H&Y score = 2, and 2 subjects with H&Y score = 3. In contrast, the progressive group has 7 subjects with H&Y score = 1, 74 subjects with H&Y score = 2, and 0 subjects with H&Y score = 3.

Cohort 7 is the only cohort that samples from a set of visit pairs for 213 subjects without any restrictions for the PD state of a patient. This cohort was developed as a reference cohort for comparison with those created by [1] and [2]. Cohort 7 has 72 stable and 72 progressive subjects for a total of 144 subjects. There are 28 females and 44 males in the stable group and 30 females and 42 males in the progressive group. The mean age and standard deviation is 61.6±8.5 and 62.6±9.2 for the stable and progressive groups respectively. The stable and progressive groups each have 23 subjects with H&Y score = 1 and 49 subjects with H&Y score = 2.

Table 2. Summary of the seven constructed cohorts.

	Cohort 1		**Cohort 2**		**Cohort 3**		**Cohort 4**		**Cohort 5**		**Cohort 6**		**Cohort 7**	
	Stable	Progr	Stable	Progr	Stable	Progr	Stable	Progr	Stable	Progr	Stable	Progr	Stable	Progr
Subjects, No.	35	35	65	75	59	59	102	92	43	43	66	81	72	72
F/M No.	13/22	15/20	23/42	28/47	21/38	25/34	38/64	35/57	18/25	16/27	24/42	33/48	28/44	30/42
Age, mean SD	60.5 ± 7.6	61.9 ± 8.2	62.3 ± 9.3	60.0 ± 9.7	60.6 ± 8.7	62.0 ± 9.4	62.3 ± 10.0	62.3 ± 9.6	63.0 ± 8.9	62.7 ± 10.5	62.7 ± 10.4	63.5 ± 9.5	61.6 ± 8.5	62.6 ± 9.2
Hoehn & Yahr Stage 1 (n)	18	18	45	8	26	26	64	9	17	17	36	7	23	23
Hoehn & Yahr Stage 2 (n)	17	17	17	67	33	33	34	83	26	26	28	74	49	49
Hoehn & Yahr Stage 3 (n)	0	0	3	0	0	0	4	0	0	0	2	0	0	0

Table 3. ROC AUC scores for Cohorts 1 to 7 for the feature sets F1 to F9

ROC AUC scores	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Cohort 6	Cohort 7
F1 score ± std							
SVM	0.437 ± 0.102	0.757 ± 0.057	0.361 ± 0.068	0.767 ± 0.075	0.282 ± 0.015	0.785 ± 0.080	0.422 ± 0.078
DT	0.414 ± 0.086	0.815 ± 0.081	0.436 ± 0.112	0.777 ± 0.044	0.382 ± 0.110	0.759 ± 0.041	0.513 ± 0.035
kNN	0.404 ± 0.069	0.818 ± 0.095	0.372 ± 0.109	0.741 ± 0.064	0.285 ± 0.099	0.763 ± 0.037	0.450 ± 0.114
GNB	0.433 ± 0.135	0.790 ± 0.087	0.369 ± 0.116	0.749 ± 0.074	0.214 ± 0.018	0.688 ± 0.087	0.436 ± 0.058
F2 score ± std							
SVM	0.551 ± 0.088	0.472 ± 0.082	0.588 ± 0.171	0.587 ± 0.095	0.479 ± 0.122	0.586 ± 0.103	0.479 ± 0.088
DT	0.502 ± 0.176	0.482 ± 0.066	0.585 ± 0.106	0.502 ± 0.095	0.573 ± 0.127	0.517 ± 0.071	0.559 ± 0.072
kNN	0.508 ± 0.094	0.505 ± 0.071	0.512 ± 0.095	0.571 ± 0.085	0.491 ± 0.086	0.579 ± 0.027	0.569 ± 0.049
GNB	0.571 ± 0.135	0.488 ± 0.076	0.463 ± 0.056	0.555 ± 0.063	0.469 ± 0.090	0.548 ± 0.025	0.467 ± 0.082
F3 score ± std							
SVM	0.522 ± 0.136	0.800 ± 0.047	0.425 ± 0.113	0.786 ± 0.064	0.515 ± 0.161	0.746 ± 0.087	0.503 ± 0.071
DT	0.467 ± 0.155	0.709 ± 0.044	0.607 ± 0.104	0.792 ± 0.065	0.622 ± 0.074	0.750 ± 0.071	0.565 ± 0.054
kNN	0.480 ± 0.048	0.839 ± 0.096	0.452 ± 0.076	0.803 ± 0.071	0.447 ± 0.112	0.787 ± 0.055	0.560 ± 0.046
GNB	0.531 ± 0.129	0.745 ± 0.094	0.421 ± 0.083	0.775 ± 0.066	0.413 ± 0.048	0.647 ± 0.064	0.453 ± 0.080
F4 score ± std							
SVM	0.490 ± 0.102	0.549 ± 0.072	0.354 ± 0.194	0.514 ± 0.061	0.498 ± 0.122	0.470 ± 0.133	0.455 ± 0.088
DT	0.447 ± 0.086	0.559 ± 0.125	0.501 ± 0.090	0.468 ± 0.082	0.404 ± 0.101	0.522 ± 0.083	0.517 ± 0.078
kNN	0.524 ± 0.129	0.534 ± 0.097	0.386 ± 0.095	0.469 ± 0.059	0.465 ± 0.108	0.493 ± 0.107	0.561 ± 0.089
GNB	0.453 ± 0.105	0.539 ± 0.085	0.317 ± 0.108	0.499 ± 0.062	0.483 ± 0.099	0.457 ± 0.048	0.463 ± 0.079
F5 score ± std							
SVM	0.457 ± 0.163	0.494 ± 0.121	0.541 ± 0.082	0.45 ± 0.082	0.482 ± 0.125	0.459 ± 0.139	0.475 ± 0.116
DT	0.511 ± 0.159	0.508 ± 0.182	0.527 ± 0.094	0.46 ± 0.071	0.465 ± 0.126	0.453 ± 0.09	0.406 ± 0.099
kNN	0.531 ± 0.191	0.456 ± 0.104	0.489 ± 0.108	0.574 ± 0.076	0.478 ± 0.142	0.471 ± 0.119	0.452 ± 0.064
GNB	0.48 ± 0.214	0.521 ± 0.114	0.374 ± 0.062	0.403 ± 0.104	0.468 ± 0.128	0.449 ± 0.091	0.519 ± 0.102
F6 score ± std							
SVM	0.457 ± 0.186	0.458 ± 0.097	0.49 ± 0.122	0.481 ± 0.074	0.461 ± 0.103	0.463 ± 0.114	0.424 ± 0.123
DT	0.451 ± 0.212	0.479 ± 0.106	0.52 ± 0.122	0.474 ± 0.082	0.457 ± 0.145	0.457 ± 0.123	0.539 ± 0.107
kNN	0.509 ± 0.271	0.465 ± 0.123	0.512 ± 0.193	0.536 ± 0.115	0.431 ± 0.148	0.527 ± 0.041	0.467 ± 0.102
GNB	0.571 ± 0.251	0.428 ± 0.069	0.384 ± 0.147	0.469 ± 0.127	0.484 ± 0.076	0.483 ± 0.108	0.439 ± 0.093
F7 score ± std							
SVM	0.549 ± 0.182	0.547 ± 0.08	0.355 ± 0.191	0.511 ± 0.07	0.382 ± 0.143	0.601 ± 0.079	0.512 ± 0.17
DT	0.554 ± 0.219	0.524 ± 0.13	0.423 ± 0.066	0.494 ± 0.064	0.337 ± 0.14	0.52 ± 0.116	0.474 ± 0.069
kNN	0.494 ± 0.187	0.488 ± 0.079	0.475 ± 0.132	0.509 ± 0.05	0.386 ± 0.12	0.466 ± 0.092	0.578 ± 0.121
GNB	0.446 ± 0.251	0.486 ± 0.073	0.279 ± 0.109	0.505 ± 0.113	0.427 ± 0.158	0.474 ± 0.126	0.452 ± 0.112
F8 score ± std							
SVM	0.577 ± 0.124	0.548 ± 0.138	0.585 ± 0.199	0.491 ± 0.095	0.458 ± 0.169	0.522 ± 0.131	0.432 ± 0.15
DT	0.571 ± 0.164	0.528 ± 0.104	0.462 ± 0.169	0.546 ± 0.116	0.364 ± 0.179	0.459 ± 0.098	0.391 ± 0.089
kNN	0.48 ± 0.217	0.543 ± 0.174	0.5 ± 0.145	0.471 ± 0.071	0.533 ± 0.159	0.47 ± 0.085	0.451 ± 0.109
GNB	0.457 ± 0.219	0.514 ± 0.091	0.364 ± 0.107	0.53 ± 0.145	0.411 ± 0.198	0.534 ± 0.081	0.483 ± 0.116
F9 score ± std							
SVM	0.486 ± 0.238	0.52 ± 0.103	0.413 ± 0.104	0.493 ± 0.167	0.378 ± 0.181	0.399 ± 0.105	0.497 ± 0.059
DT	0.557 ± 0.162	0.5 ± 0.08	0.446 ± 0.098	0.529 ± 0.087	0.395 ± 0.14	0.421 ± 0.064	0.518 ± 0.131
kNN	0.531 ± 0.196	0.441 ± 0.046	0.427 ± 0.081	0.563 ± 0.072	0.395 ± 0.156	0.545 ± 0.083	0.445 ± 0.101
GNB	0.469 ± 0.204	0.442 ± 0.087	0.313 ± 0.141	0.484 ± 0.089	0.479 ± 0.189	0.425 ± 0.088	0.5 ± 0.115

Feature Extraction

Experiments

Discussion

This section discusses the findings.

Conclusion

The conclusion is here.

Supporting information

S1 Fig. Bold the title sentence. Add descriptive text after the title of the item (optional).

Acknowledgments

References

1. Shu ZY, Cui SJ, Wu X, Xu Y, Huang P, Pang PP, et al. Predicting the progression of Parkinson's disease using conventional MRI and machine learning: An application of radiomic biomarkers in whole-brain white matter. Magnetic Resonance in Medicine. 2021;85(3):1611–1624.
2. Arafe M, Bhagwat N, Chatelain Y, Dugré M, Sokołowski A, Wang M, et al. Predicting Parkinson's disease progression using MRI-based white matter radiomic biomarker and machine learning: a reproducibility and replicability study. bioRxiv. 2023;doi:10.1101/2023.05.05.539590.