Supplementary Materials

Supplementary Tables

Table S1: Supplementary Table S1. Post-hoc pairwise group comparisons for demographic and clinical variables (Tukey HSD test, family-wise error rate controlled).

Variable	Group 1	Group 2	Mean Diff	95% CI	Adj. P	Significant
ENROLL_AGE	PDD	PDMCI	−1.45	[−4.31, 1.40]	0.456	No
ENROLL_AGE	PDD	PDNC	−5.49	[−8.14, −2.83]	< 0.001	Yes
ENROLL_AGE	PDMCI	PDNC	−4.03	[−5.47, −2.59]	< 0.001	Yes
Disease Duration	PDD	PDMCI	3.54	[−12.08, 19.15]	0.855	No
Disease Duration	PDD	PDNC	7.58	[−7.29, 22.45]	0.455	No
Disease Duration	PDMCI	PDNC	4.04	[−5.83, 13.91]	0.601	No
EDUCYRS	PDD	PDMCI	0.11	[−0.90, 1.11]	0.967	No
EDUCYRS	PDD	PDNC	0.31	[−0.62, 1.25]	0.714	No
EDUCYRS	PDMCI	PDNC	0.21	[−0.30, 0.71]	0.607	No
MoCA Total	PDD	PDMCI	5.15	[4.18, 6.11]	< 0.001	Yes
MoCA Total	PDD	PDNC	7.40	[6.51, 8.30]	< 0.001	Yes
MoCA Total	PDMCI	PDNC	2.26	[1.77, 2.74]	< 0.001	Yes
VFLT	PDD	PDMCI	0.98	[−1.76, 3.72]	0.677	No
VFLT	PDD	PDNC	1.47	[−1.10, 4.03]	0.372	No
VFLT	PDMCI	PDNC	0.49	[−0.97, 1.94]	0.715	No
JLO Total	PDD	PDMCI	−0.24	[−1.71, 1.23]	0.922	No
JLO Total	PDD	PDNC	−0.49	[−1.87, 0.88]	0.678	No
JLO Total	PDMCI	PDNC	−0.25	[−1.03, 0.53]	0.730	No
GDS Total Score	PDD	PDMCI	−0.60	[−1.08, −0.11]	0.011	Yes
GDS Total Score	PDD	PDNC	−1.30	[−1.76, −0.85]	< 0.001	Yes
GDS Total Score	PDMCI	PDNC	−0.71	[−0.95, −0.46]	< 0.001	Yes
UPDRS Total	PDD	PDMCI	−21.58	[−41.34, −1.81]	0.029	Yes
UPDRS Total	PDD	PDNC	−40.93	[−59.98, −21.89]	< 0.001	Yes
UPDRS Total	PDMCI	PDNC	−19.36	[−30.20, −8.51]	< 0.001	Yes
ESS Total Score	PDD	PDMCI	−1.36	[−2.62, −0.09]	0.032	Yes
ESS Total Score	PDD	PDNC	−3.56	[−4.74, −2.38]	< 0.001	Yes
ESS Total Score	PDMCI	PDNC	−2.21	[−2.84, −1.58]	< 0.001	Yes

Significant results ($P < 0.05$) are indicated as "Yes".

Table S2: Supplementary Table S2. Comparison with recent literature on cognitive classification/prediction in Parkinson's disease (PD).

Study (link to title)	**Data modality**	**Task (classes)**	**N**	**Main metric (as reported)**
Predicting dementia in people with Parkinson's disease (2025)[**?**]	Primary care EHR + lifestyle + polygenic score (no imaging)	Dementia risk (binary)	UKB & external	AUC ≈0.62 (UK Biobank), 0.58 (SIDIAP), 0.66 (PPMI)
An interpretable multiparametric radiomics model of basal ganglia to predict dementia conversion in Parkinson's disease (2023)[**?**]	MRI radiomics (T1/T2/FLAIR) + clinical	PDD conversion (binary)	262 (train/test split)	Test AUC 0.889 (combined radiomics+clinical) vs 0.722 (clinical-only)
Predicting cognitive decline in Parkinson's disease using FDG-PET–based supervised learning (2022)[**?**]	FDG-PET	PD-MCI converters vs stable MCI (binary)	Train 43; Val 19	Independent validation AUC 0.73; Sens 0.67, Spec 0.80
Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease (2022)[**?**]	Clinical + biofluid + genetic/epigenetic (no imaging)	(1) Cog. impairment vs preserved; (2) Dementia conversion (binary)	PPMI cohort	Best AUC 0.90–0.93 (impairment, combined vars); PDD vs PDMCI AUC 0.69–0.75
Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning (2017)[**?**]	Resting-state fMRI functional connectomics	PD-MCI vs PD-nonMCI (binary)	Train 70; Val 25	Balanced accuracy 82.6% (LOOCV); validation: PD-nonMCI 76.5%, PD-MCI 87.5% correct
Multimodal neuroimaging-based prediction of Parkinson's disease with mild cognitive impairment using machine learning technique (2024)[**?**]	Clinical + rs-fMRI + T1 MRI + plasma NfL	PDNC vs PDMCI (binary)	173	AUC 0.840; Acc 0.762; Sens 0.745; Spec 0.783
Our framework (this study)	**Routine clinical scales only (no imaging/biomarkers)**	**Three-class: PDNC/PDMCI/PDD**	**PPMI cohort**	**Accuracy 0.92; Macro-F1 0.79; Recall—PDNC 1.00, PDMCI 0.71, PDD 0.71; F1—PDNC 0.97, PDMCI 0.81, PDD 0.59 (test set $n = 206/68/14$)**

Notes: "Main metric" reproduces the authors' primary metric(s). Letters/words match the original article titles exactly; links point to the publishers' pages.

Supplementary Figures

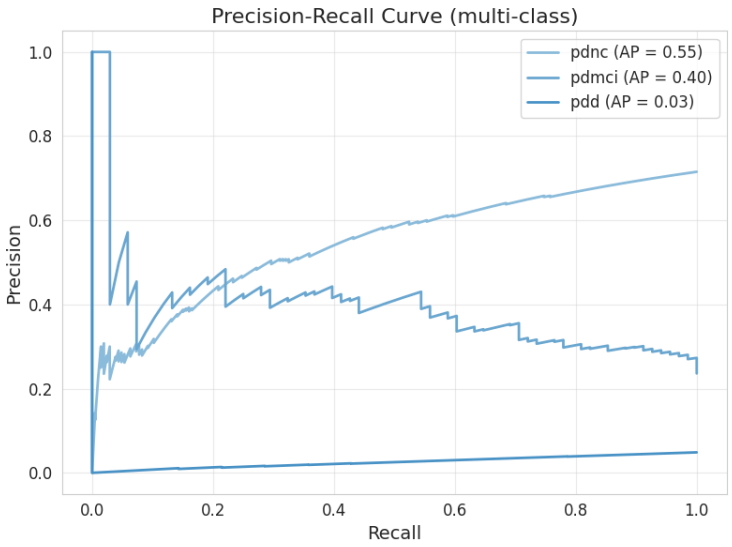


Figure S1: Supplementary Figure S1. Precision–Recall curves for SVM baseline by class.

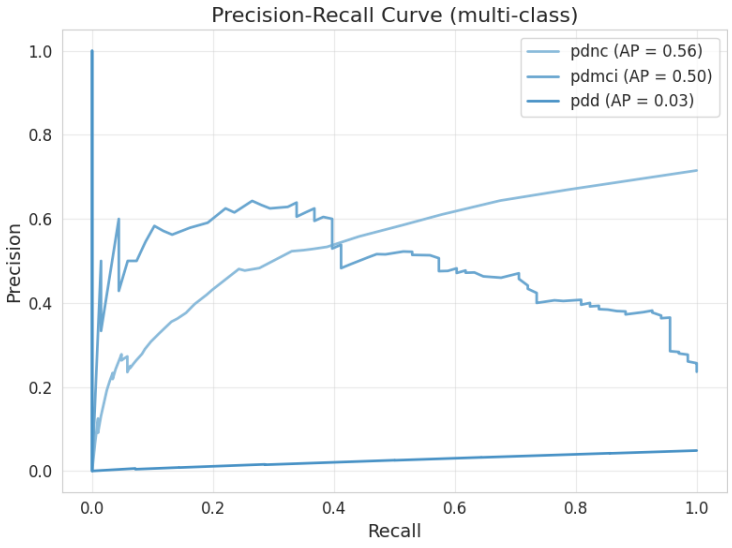


Figure S2: Supplementary Figure S2. Precision–Recall curves for Random Forest baseline by class.

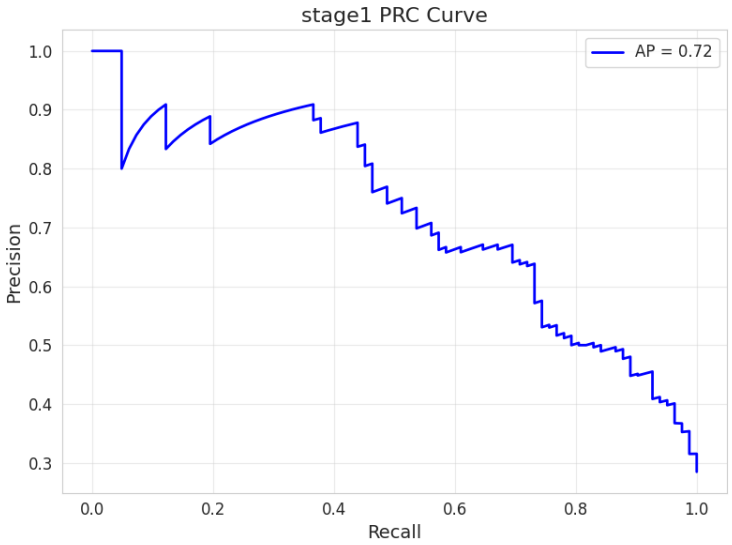


Figure S3: Supplementary Figure S3. Precision–Recall curve for Stage 1 classification. Average precision (AP) = 0.72.

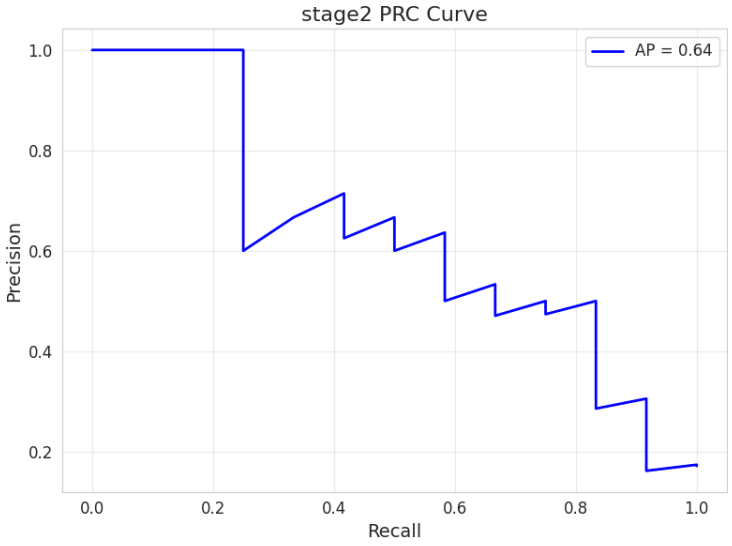


Figure S4: Supplementary Figure S4. Precision–Recall curve for Stage 2 classification. Average precision (AP) = 0.64.

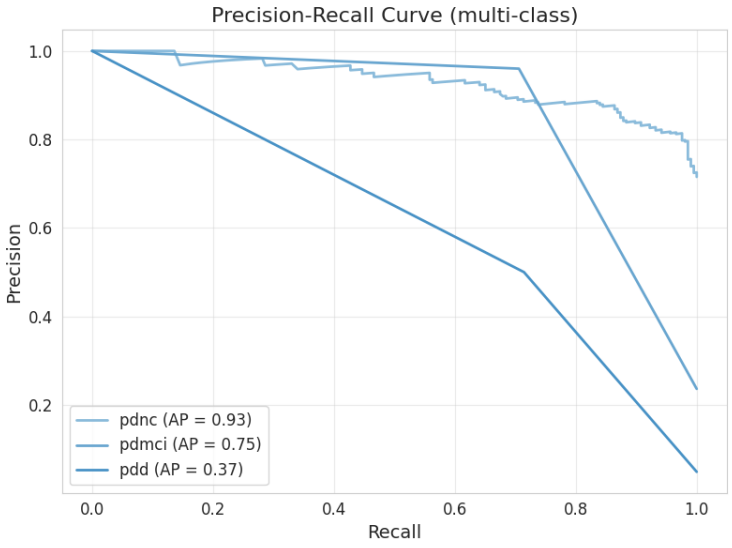
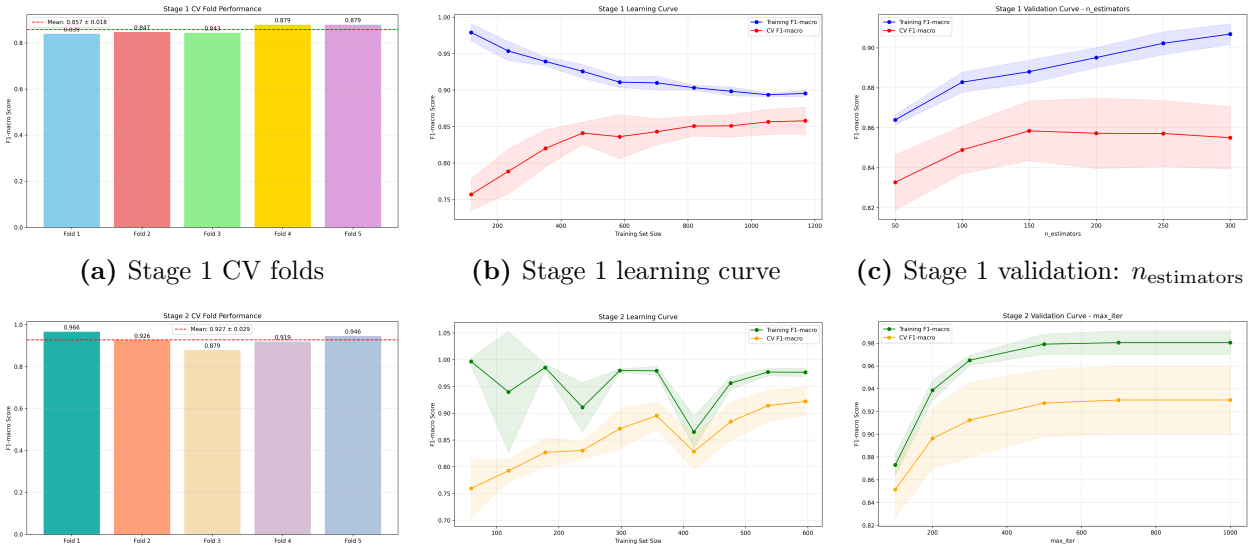


Figure S5: Supplementary Figure S5. Precision–Recall curves for the final three-class classification. Average precision (AP): PDNC = 0.93, PDMCI = 0.75, PDD = 0.37.



(a) Stage 1 CV folds **(b)** Stage 1 learning curve **(c)** Stage 1 validation: $n_{\text{estimators}}$

(d) Stage 2 CV folds **(e)** Stage 2 learning curve **(f)** Stage 2 validation: `max_iter`

Figure S6: Supplementary Figure S6. Cross-validation learning behavior of the two-stage pipeline. (a–c) Stage 1 (XGBoost): (a) fold-wise CV F1-macro, (b) learning curve (training vs. validation), (c) validation curve across $n_{\text{estimators}}$. (d–f) Stage 2 (MLP): (d) fold-wise CV F1-macro, (e) learning curve (training vs. validation), (f) validation curve across `max_iter`. These plots demonstrate that hyperparameter optimization was performed strictly within the training partition and that the independent 20% test set remained untouched during model development.

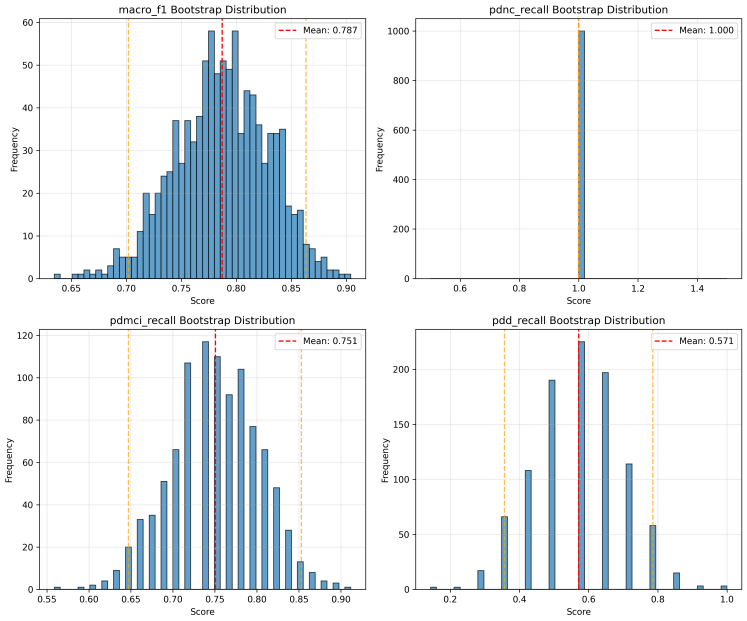


Figure S7: Supplementary Figure S7. Distributions of macro-F1 (top-left) and class-wise recall (top-right: PDNC; bottom-left: PDMCI; bottom-right: PDD) from 1,000 stratified bootstrap replicates of the held-out test set. Red dashed lines indicate mean values; orange dashed lines denote the 2.5th and 97.5th percentiles (95% confidence intervals). The analysis confirmed stable estimates for macro-F1 (mean 0.787, 95% CI [0.71–0.86]), PDNC recall (1.00, 95% CI [0.98–1.00]), and PDMCI recall (0.75, 95% CI [0.65–0.85]). PDD recall showed greater variability (mean 0.57, 95% CI [0.35–0.80]) due to the small sample size ($n = 79$), highlighting the variance and generalization challenge in this subgroup.

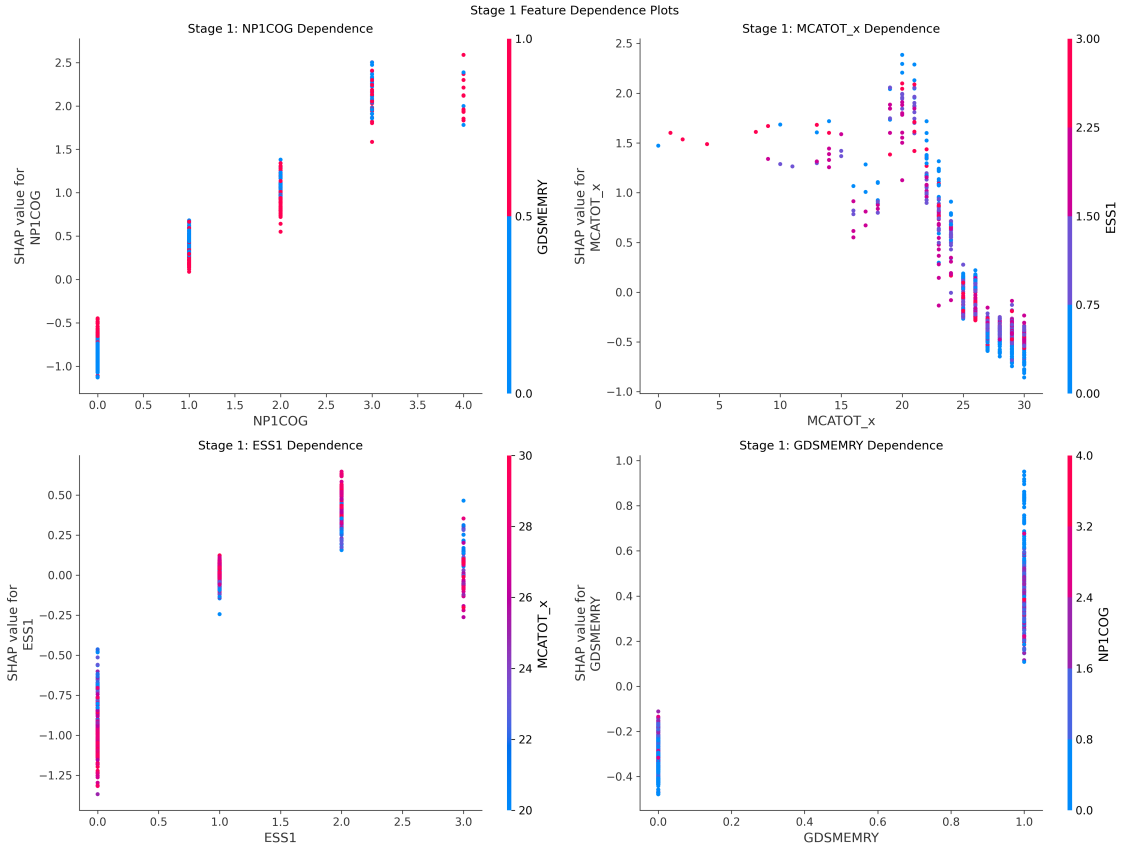


Figure S8: Supplementary Figure S8. Stage 1 directional SHAP dependence patterns (PDNC vs. non-PDNC). Higher MCATOT scores consistently contribute negatively to cognitive impairment probability, demonstrating protective effects of preserved cognitive performance. Conversely, NP1COG (cognitive complaints) shows positive contributions with threshold effects above score 2.0. ESS1 and GDSMEMRY demonstrate non-linear relationships with inflection points corresponding to clinical significance thresholds.

Stage 1 dependence patterns (Supplementary Figure S8) align with clinical expectations: objective cognitive performance (MCATOT) demonstrates sigmoid protective effects with critical inflection points around 24–26 corresponding to established MCI screening cutoffs. Subjective cognitive complaints (NP1COG) show positive contributions to impairment classification, particularly above threshold values of 2.0.

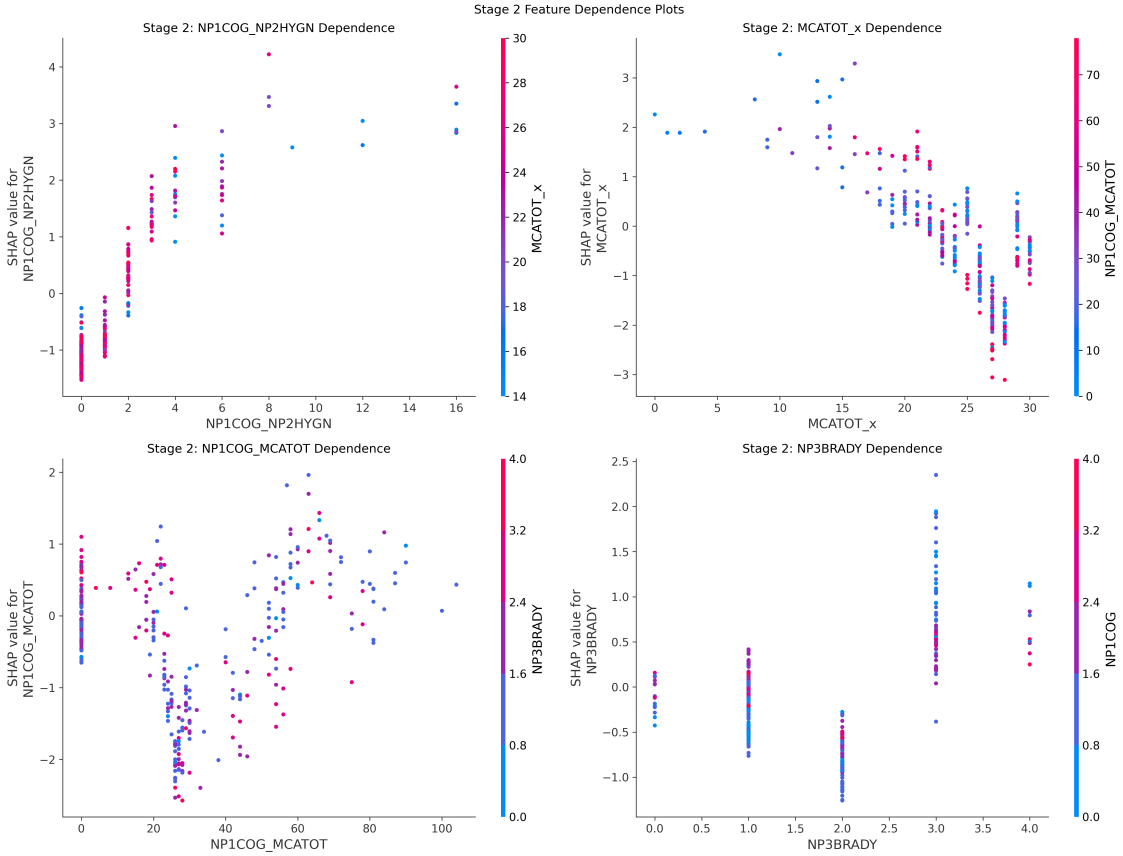


Figure S9: Supplementary Figure S9. Stage 2 directional SHAP dependence patterns (PDMCI vs. PDD). The interaction term `NP1COG_NP2HYGN` shows threshold effects where values > 2 sharply increase PDD probability, capturing synergy between cognitive complaints and functional dependence. MCATOT exhibits protective effects, while NP3BRADY demonstrates threshold-based contributions with larger impact at higher severity levels.

Table S3: Supplementary Table S3. Training configuration and hyperparameters for the two-stage classification pipeline.

Global settings	
Random seeds	42 (global, train-test split, SMOTE, XGB, MLP, CV)
Train:test split	0.8:0.2 (stratified)
Cross-validation	5-fold, stratified, scoring = macro F1
Feature scaling	z-score normalization (StandardScaler)
Missing-value handling	Features with >60% missing dropped; remaining imputed (median)
Feature selection	SHAP top-20 $\rightarrow$ RFE (final 10 features for Stage 1)
SMOTE strategy	Stage 1: 0.8; Stage 2: {0:400, 1:350}
Stage 1: XGBoost (PDNC vs. non-PDNC)	
Model type	`XGBClassifier`
Grid search space	$n_{\text{estimators}} \in \{100, 150, 200\}$; $max\_depth \in \{3, 4, 5\}$; $learning\_rate \in \{0.01, 0.05, 0.1, 0.2\}$; $scale\_pos\_weight \in \{1, 2, 3\}$; reg_alpha, $reg\_lambda \in \{0, 0.1, 1.0\}$
Best parameters	$learning\_rate = 0.05$; $max\_depth = 3$; $n_{\text{estimators}} = 200$; $reg\_alpha = 0.1$; $reg\_lambda = 0.1$; $scale\_pos\_weight = 1$
CV performance	$F1_{\text{macro}} = 0.8572 \pm 0.0176$
Decision threshold	0.203 (ensure non-PDNC recall ≥ 0.85)
Stage 2: MLP (PDMCI vs. PDD)	
Model type	`MLPClassifier`
Hidden layers	(100, 50)
Activation	`relu`
Solver	`adam`
Alpha	0.0001
Learning rate	`constant`, init = 0.001
Batch size	`auto`
Max iterations	500
Other params	`shuffle=True`, `momentum=0.9`, `nesterov=True`, `early_stopping=False`, `validation_fraction=0.1`, `tol=`1e$-$4, `n_iter_no_change=10`
Decision threshold	0.563
CV performance	$F1_{\text{macro}} = 0.9273 \pm 0.0293$

References