

Supplementary Material

1. Supplementary Appendix S1. Ethical Approval Statement - Page 1.
2. Supplementary Table S1. Results of Univariate Logistic Regression Analysis- Page 2.
3. Supplementary Table S2. Multicollinearity VIF Analysis Results for Independent Variables-Pages 3.
4. Supplementary Table S3. P-values of pairwise DeLong test for AUROC of all models in the independent test set-Pages 4.
5. Supplementary Table S4. Comparison of model performance with and without SMOTE oversampling – Pages 5.
6. Supplementary Table S5. Comparison of model performance using MICE imputation versus complete case analysis- Pages 5.
7. Supplementary Table S6. Subgroup analysis results of the provincial cohort - Pages 6.
8. Supplementary Figure S1. ROC curves of all models in the training set - Pages 7.
9. Supplementary Figure S2. Precision-Recall curves of all models in the training set - Pages 8.
10. Supplementary Figure S3 Calibration curves of all models in the training set - Pages 9

.

Supplementary Appendix S1. Ethical Approval Statement.

This study was approved by the Ethics Committee of Guangdong Second Provincial General Hospital (approval number: 2025-KY-KZ-013-02). Anonymization and data confidentiality measures were strictly implemented throughout the study to protect participants' privacy, and all procedures were performed in accordance with the Declaration of Helsinki (1975, revised 1983).

Supplementary Table S1 Results of Univariate Logistic Regression Analysis

Variable	Beta	SE	z	OR (95% CI)	P.adj
gender_male	0.193	0.339	0.569	1.212 (0.624-2.362)	0.635
age	-0.379	0.173	-2.195	0.685 (0.483-0.954)	0.055
educationalLevel	-0.5	0.182	-2.744	0.607 (0.419-0.859)	0.014*
MMSEscore	0.327	0.184	1.778	1.387 (0.976-2.025)	0.115
PSQIScore	0.235	0.173	1.361	1.265 (0.905-1.788)	0.240
HAMAScore	0.372	0.172	2.167	1.451 (1.045-2.056)	0.055
HAMDScore	-0.223	0.178	-1.248	0.800 (0.559-1.131)	0.279
visuospatialExecutiveFunction	-1.296	0.251	-5.172	0.274 (0.161-0.432)	<0.001*
naming	0.125	0.173	0.722	1.133 (0.806-1.603)	0.593
attention	-1.127	0.245	-4.599	0.324 (0.192-0.506)	<0.001*
language	-0.85	0.202	-4.203	0.427 (0.282-0.626)	<0.001*
abstraction	-0.357	0.173	-2.067	0.699 (0.494-0.976)	0.066
delayedRecall	-1.063	0.209	-5.085	0.345 (0.224-0.510)	<0.001*
orientation	-1.747	0.408	-4.279	0.174 (0.068-0.348)	<0.001*
MoCATotalScore	-1.404	0.248	-5.664	0.246 (0.146-0.387)	<0.001*
overlapSaccadeAccuracy	-1.26	0.48	-2.627	0.284 (0.091-0.608)	0.018 *
overlapSaccadeLatency	0.417	0.228	1.826	1.517 (1.029-2.481)	0.109
meanOverlapSaccadeDuration	1.809	0.413	4.384	6.105 (2.876-14.540)	<0.001*
meanOverlapSaccadeVelocity	-1.764	0.326	-5.419	0.171 (0.085-0.308)	<0.001*
antiSaccadeAccuracy	-0.11	0.17	-0.648	0.896 (0.639-1.250)	0.600
meanAntiSaccadeDuration	0.991	0.265	3.744	2.693 (1.675-4.749)	0.001*
antiSaccadeCorrectionRate	-0.606	0.192	-3.164	0.545 (0.366-0.779)	0.004*
antiSaccadeLatency	0.113	0.172	0.661	1.120 (0.801-1.578)	0.600
meanAntiSaccadeCorrectionDuration	0.721	0.214	3.362	2.056 (1.384-3.221)	0.002*
meanAntiSaccadeVelocity	-1.688	0.319	-5.285	0.185 (0.093-0.327)	<0.001*
smoothPursuitOnsetDuration	1.289	0.816	1.58	3.630 (1.304-30.628)	0.165
smoothPursuitVelocity	-0.013	0.276	-0.048	0.987 (0.558-1.777)	0.961
smoothPursuitDeviationTimes	0.075	0.16	0.468	1.078 (0.790-1.492)	0.663
smoothPursuitTotalDeviation	0.084	0.16	0.524	1.088 (0.797-1.505)	0.644

Note: 1. Outcome variable: presence of PD (0 = Control, 1 = PD); 2. Beta: regression coefficient; 3. SE: standard error; 4. z: z-statistic; 5. OR (95% CI): odds ratio (95% confidence interval); an OR > 1 indicates that the factor is a risk factor for PD, while an OR < 1 indicates a protective factor; 6. P.adj: Benjamini-Hochberg (BH)-adjusted P value; P.adj < 0.05 indicates a statistically significant difference.

Supplementary Table S2 Multicollinearity VIF Analysis Results for Independent Variables

Variable	VIF	Tolerance	Collinearity_Level
MoCATotalScore	45.402	0.022	Severe multicollinearity
delayedRecall	8.022	0.125	Moderate multicollinearity
visuospatialExecutiveFunction	7.153	0.14	Moderate multicollinearity
attention	4.236	0.236	Low multicollinearity
language	4.135	0.242	Low multicollinearity
meanOverlapSaccadeVelocity	4.048	0.247	Low multicollinearity
meanAntiSaccadeVelocity	3.843	0.26	Low multicollinearity
meanOverlapSaccadeDuration	3.746	0.267	Low multicollinearity
meanAntiSaccadeDuration	3.295	0.303	Low multicollinearity
orientation	2.6	0.385	Low multicollinearity
meanAntiSaccadeCorrectionDuration	2.408	0.415	Low multicollinearity
antiSaccadeCorrectionRate	2.165	0.462	Low multicollinearity
educationalLevel	1.827	0.547	Low multicollinearity
overlapSaccadeAccuracy	1.671	0.598	Low multicollinearity

Note: 1) VIF (Variance Inflation Factor): Used to measure the degree of multicollinearity among independent variables, calculated as the extent to which one independent variable in the model can be explained by other independent variables; 2) Tolerance: Calculated as $1/VIF$, inversely related to VIF, lower tolerance indicates more severe collinearity; 3) Collinearity judgment criteria: $VIF < 5$ (Tolerance > 0.2) for low collinearity, $5 \leq VIF < 10$ ($0.1 \leq \text{Tolerance} \leq 0.2$) for moderate collinearity, $VIF \geq 10$ (Tolerance < 0.1) for severe collinearity; 4) Interpretation is based on commonly used clinical statistical standards. Variables with severe collinearity are recommended to be handled by removal, combination or variable replacement to avoid affecting model stability.

Supplementary Table S3 P-values of pairwise DeLong test for AUROC of all models in the independent test set

Model1	Model2	AUC1	AUC2	P_value
LR	RF	0.8438	0.9711	0.002*
LR	XGBoost	0.8438	0.9306	0.008*
LR	MLP	0.8438	0.8715	0.109
LR	SVM	0.8438	0.9144	0.028*
LR	DT	0.8438	0.8316	0.796
RF	XGBoost	0.9711	0.9306	0.094
RF	MLP	0.9711	0.8715	0.008*
RF	SVM	0.9711	0.9144	0.024*
RF	DT	0.9711	0.8316	0.003*
XGBoost	MLP	0.9306	0.8715	0.068
XGBoost	SVM	0.9306	0.9144	0.386
XGBoost	DT	0.9306	0.8316	0.005*
MLP	SVM	0.8715	0.9144	0.191
MLP	DT	0.8715	0.8316	0.412
SVM	DT	0.9144	0.8316	0.014*

Note: Pairwise comparisons were performed using the DeLong test. $p < 0.05$ indicates statistical significance.

Supplementary Table S4 Comparison of model performance with and without SMOTE oversampling.

Models	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Accuracy	Kappa	AUC	PR-AUC
No SMOTE	0.9541	0.9121	0.9286	0.9432	0.9286	0.9541	0.9412	0.9350	0.8686	0.9782	0.9869
With SMOTE	0.9450	0.9560	0.9626	0.9355	0.9626	0.9450	0.9537	0.9500	0.8994	0.9826	0.9896

Note: PPV = Positive Predictive Value; NPV = Negative Predictive Value; AUC = Area Under the Receiver Operating Characteristic Curve; PR-AUC = Area Under the Precision-Recall Curve.

Supplementary Table S5 Comparison of model performance using MICE imputation versus complete case analysis

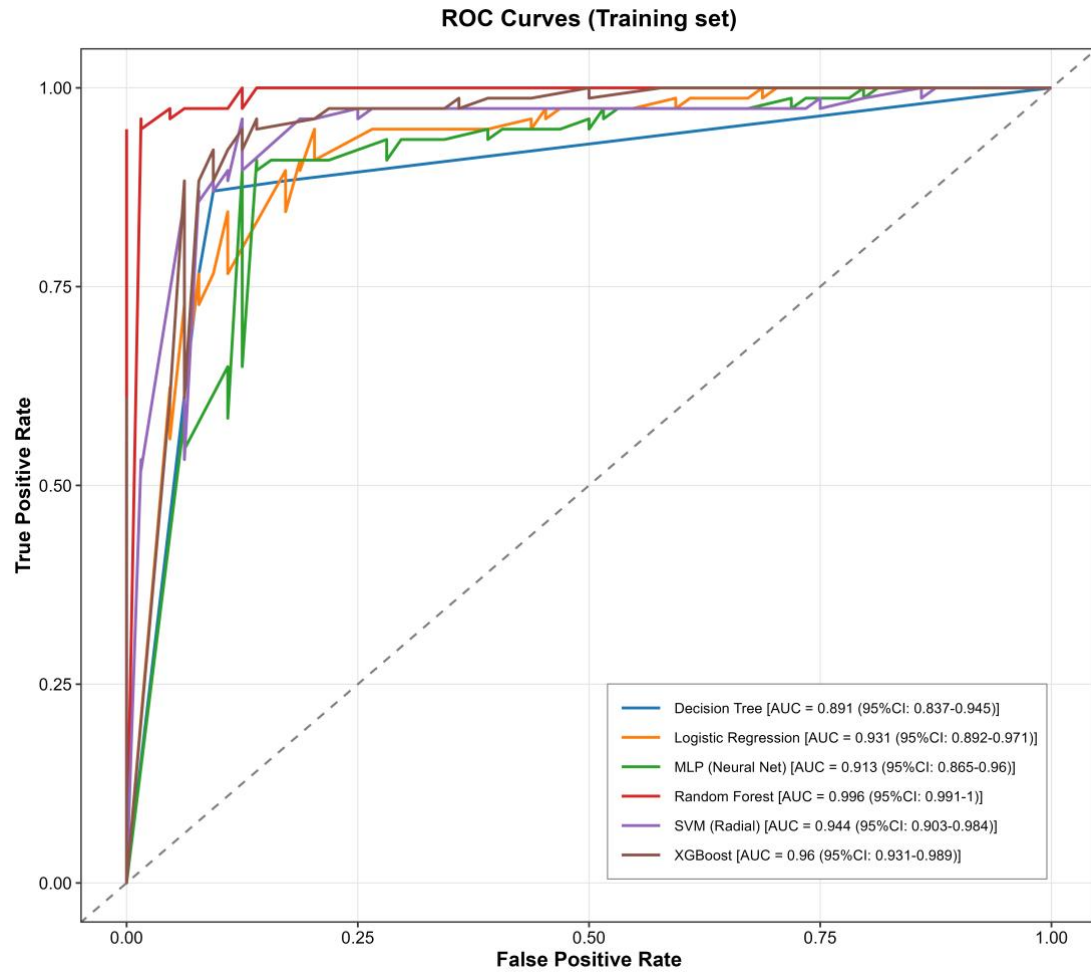
Models	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Accuracy	Kappa	AUC	PR-AUC
No Mice	0.961	0.984	0.987	0.955	0.987	0.961	0.974	0.972	0.943	0.996	0.971
With Mice	0.958	1.000	1.000	0.949	1.000	0.958	0.979	0.977	0.953	0.998	0.936

Note: PPV = Positive Predictive Value; NPV = Negative Predictive Value; AUC = Area Under the Receiver Operating Characteristic Curve; PR-AUC = Area Under the Precision-Recall Curve.

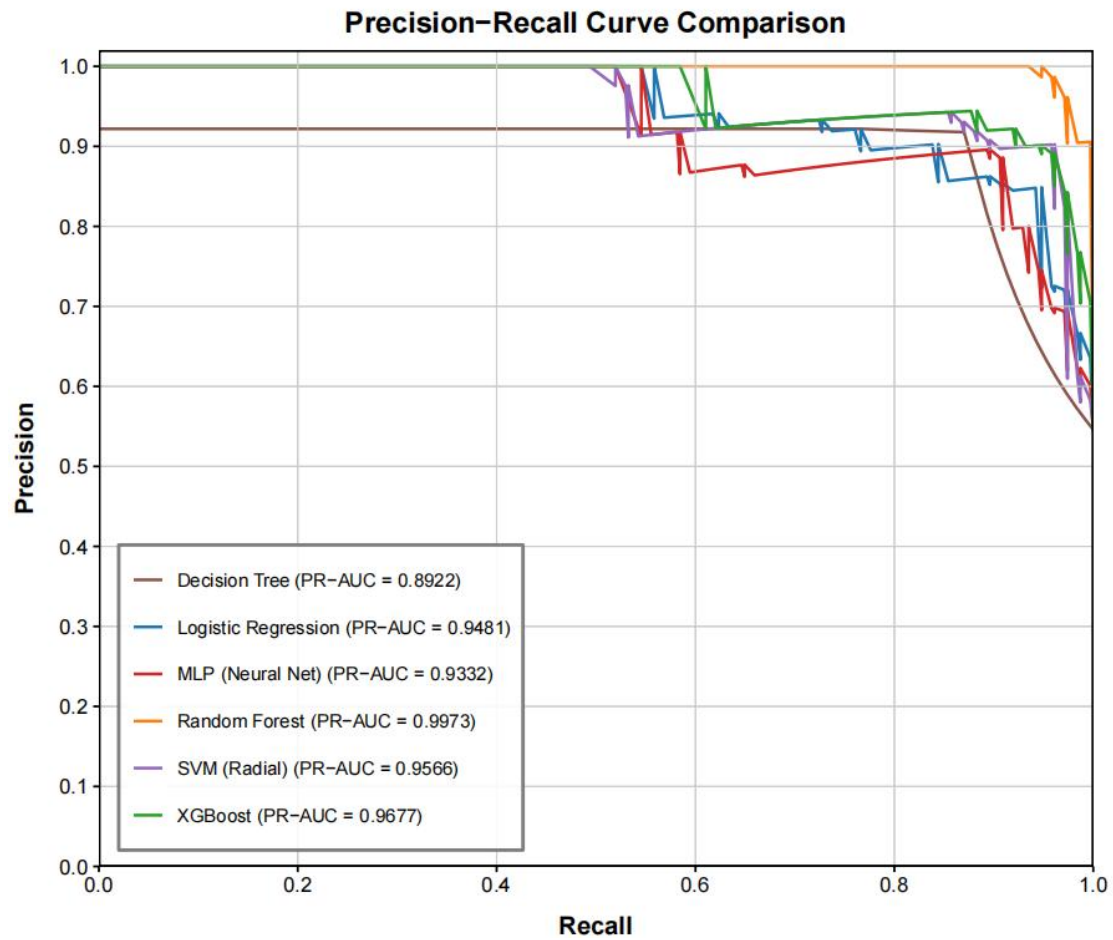
Supplementary Table S6. Subgroup analysis results of the provincial cohort

Subgroup	Sample_Size	N	Sensitivity	Specificity	PosPredValue	NegPredValue	Precision	Recall	F1	Accuracy	Kappa	AUC	PR_AUC
age	Low (<59)	70	0.978	0.920	0.957	0.958	0.957	0.978	0.967	0.957	0.906	0.998	0.999
age	Mid (59-69)	67	0.906	0.914	0.906	0.914	0.906	0.906	0.906	0.910	0.821	0.984	0.984
age	High (>=69)	63	0.875	1.000	1.000	0.886	1.000	0.875	0.933	0.937	0.873	0.963	0.978
gender	female	99	0.964	0.909	0.930	0.952	0.930	0.964	0.946	0.939	0.877	0.979	0.981
gender	male	101	0.926	0.894	0.909	0.913	0.909	0.926	0.917	0.911	0.821	0.957	0.973

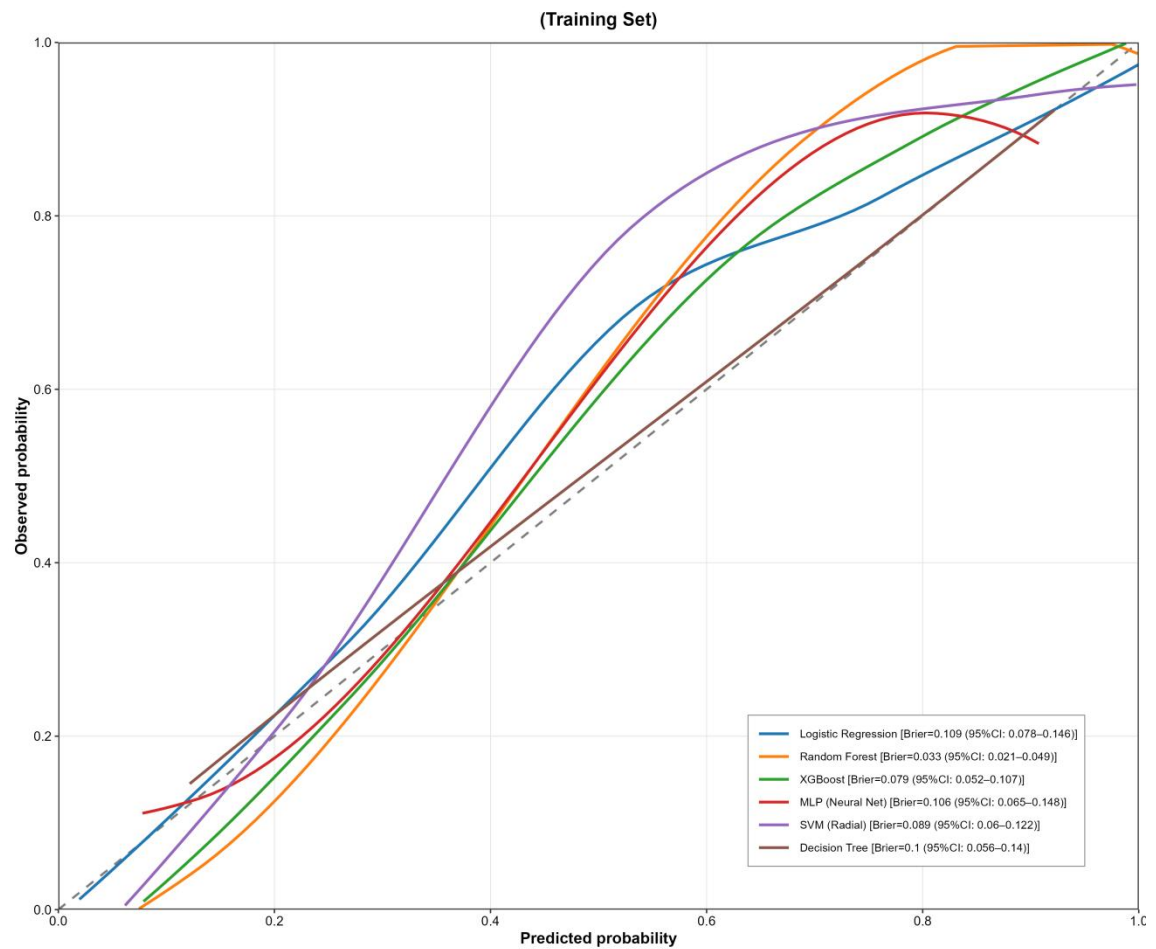
Note: PPV = Positive Predictive Value; NPV = Negative Predictive Value; AUC = Area Under the Receiver Operating Characteristic Curve; PR-AUC = Area Under the Precision-Recall Curve.



Supplementary Figure S1 ROC curves of all models in the training set. The AUROC of the RF model in the training set was 0.996 (95% CI 0.991-1.000).



Supplementary Figure S2 Precision-Recall (PR) curves of all models in the training set. The PR-AUC of the Random Forest model in the training set was 0.9973.



Supplementary Figure S3 Calibration curves of all models in the training set. The Brier score of the Random Forest model in the training set was 0.033 (95% CI 0.021-0.049).