

Supplementary Materials

Supplementary Table S1. Sensitivity analysis comparing balanced subsample (n = 200 per group) and full dataset results for HRV dispersion features in the comparison between paradoxical insomnia (PI) and normal sleep (N).

Sleep Stage	HRV features	q-value (Subsample)	q-value (Full dataset)	Rank-biserial (Subsample)	Rank-biserial (Full dataset)	Direction concordant
WASO	SampEn IQR	< 0.0001	< 0.0001	-0.323	-0.286	Yes
	ApEn IQR	0.0001	< 0.0001	-0.296	-0.259	Yes
	DFA α 2 IQR	0.0001	< 0.0001	-0.287	-0.267	Yes
	SDNN IQR	0.0002	< 0.0001	-0.276	-0.235	Yes
	pNN50 IQR	0.0002	< 0.0001	-0.275	-0.235	Yes
	SD2 IQR	0.0005	< 0.0001	-0.265	-0.222	Yes
	SD1 IQR	0.0007	< 0.0001	-0.256	-0.236	Yes
	RMSSD IQR	0.0007	< 0.0001	-0.256	-0.236	Yes
	AVNN IQR	0.0008	< 0.0001	-0.253	-0.220	Yes
	IHR IQR	0.0009	< 0.0001	-0.251	-0.224	Yes
	SD1/SD2 IQR	0.0034	< 0.0001	-0.229	-0.203	Yes
	HF IQR	0.0048	< 0.0001	-0.222	-0.202	Yes
	LF IQR	0.0061	< 0.0001	-0.216	-0.196	Yes
	DFA α 1 IQR	0.0064	< 0.0001	-0.213	-0.199	Yes
	DFA α 2 median	0.0079	0.0035	0.208	0.156	Yes
	Total Power IQR	0.0091	0.0005	-0.205	-0.169	Yes
	VLF median	0.0100	0.0015	0.203	0.173	Yes
	LF/HF IQR	0.0186	0.0009	-0.187	-0.166	Yes
	SD1/SD2 median	0.0310	0.0641	0.177	0.144	Yes

* HRV dispersion features represent the interquartile range (IQR) of 5-minute HRV windows within each sleep stage. q-values correspond to Benjamini–Hochberg false discovery rate–corrected p-values from Mann–Whitney U tests. Rank-biserial correlation indicates effect size for the group comparison. Direction concordance indicates whether the effect direction was consistent between the balanced subsample analysis and the full dataset analysis.

Supplementary Table S2. Sensitivity analysis comparing balanced subsample (n = 200 per group) and full dataset results for HRV dispersion features in the comparison between objective insomnia (OI) and paradoxical insomnia (PI).

Sleep Stage	HRV features	q-value (Subsample)	q-value (Full dataset)	Rank-biserial (Subsample)	Rank-biserial (Full dataset)	Direction concordant
WASO	DFA $\alpha 1$ IQR	0.0001	< 0.0001	-0.286	-0.338	Yes
	DFA $\alpha 2$ IQR	0.0001	< 0.0001	-0.281	-0.315	Yes
	SD1/SD2 IQR	0.0058	< 0.0001	-0.209	-0.234	Yes
	SampEn IQR	0.0064	< 0.0001	-0.205	-0.285	Yes
	SD2 IQR	0.0071	< 0.0001	-0.202	-0.261	Yes
	VLF median	0.0120	0.0009	0.190	0.256	Yes
	SDNN IQR	0.0142	< 0.0001	-0.186	-0.244	Yes
	ApEn IQR	0.0143	< 0.0001	-0.185	-0.290	Yes
	RMSSD IQR	0.0143	0.0005	-0.184	-0.223	Yes
	SD1 IQR	0.0143	0.0005	-0.184	-0.223	Yes

* HRV dispersion features represent the interquartile range (IQR) of 5-minute HRV windows within each sleep stage. q-values correspond to Benjamini–Hochberg false discovery rate–corrected p-values from Mann–Whitney U tests. Rank-biserial correlation indicates effect size for the group comparison. Direction concordance indicates whether the effect direction was consistent between the balanced subsample analysis and the full dataset analysis.

Supplementary Table S3. Baseline comparison of whole-night and stage-specific logistic regression and Random Forest models for PI vs. N classification.

Sleep Stage	Model	Precision	Recall	F1-score	Accuracy	10-Fold CV AP
Whole-night	Logistic Regression	0.489	0.383	0.430	0.492	0.551 ± 0.081
	Random Forest	0.542	0.433	0.481	0.533	0.580 ± 0.079
WASO	Logistic Regression	0.623	0.679	0.650	0.610	0.717 ± 0.088
	Random Forest	0.627	0.750	0.683	0.629	0.720 ± 0.074
NREM	Logistic Regression	0.574	0.517	0.544	0.567	0.569 ± 0.099
	Random Forest	0.509	0.467	0.487	0.508	0.592 ± 0.082
REM	Logistic Regression	0.540	0.567	0.553	0.542	0.599 ± 0.096
	Random Forest	0.500	0.467	0.483	0.500	0.555 ± 0.060

* Classification performance was compared across whole-night, WASO, NREM, and REM HRV representations using logistic regression and Random Forest classifiers for the PI vs. N comparison. Precision, recall, F1-score, and accuracy were evaluated on the held-out test set, and average precision (AP) was additionally estimated using 10-fold cross-validation and is presented as mean ± standard deviation. This supplementary comparison was included to determine whether the relative advantage of WASO-based classification reflected stage-specific information content beyond whole-night HRV summaries and whether a more flexible non-linear classifier provided additional value beyond a simpler linear baseline.

Supplementary Table S4. Baseline comparison of whole-night and stage-specific logistic regression and Random Forest models for PI vs. OI classification.

Sleep Stage	Model	Precision	Recall	F1-score	Accuracy	10-Fold CV AP
Whole-night	Logistic Regression	0.525	0.533	0.529	0.525	0.545 ± 0.062
	Random Forest	0.522	0.600	0.558	0.525	0.542 ± 0.059
WASO	Logistic Regression	0.596	0.500	0.544	0.591	0.666 ± 0.053
	Random Forest	0.619	0.464	0.531	0.600	0.634 ± 0.073
NREM	Logistic Regression	0.487	0.617	0.544	0.483	0.474 ± 0.052
	Random Forest	0.542	0.533	0.538	0.542	0.534 ± 0.083
REM	Logistic Regression	0.525	0.517	0.521	0.525	0.524 ± 0.046
	Random Forest	0.486	0.600	0.537	0.483	0.573 ± 0.086

* Classification performance was compared across whole-night, WASO, NREM, and REM HRV representations using logistic regression and Random Forest classifiers for the PI vs. OI comparison. Precision, recall, F1-score, and accuracy were evaluated on the held-out test set, and average precision (AP) was additionally estimated using 10-fold cross-validation and is presented as mean ± standard deviation. This supplementary comparison was included to determine whether the relative advantage of WASO-based classification reflected stage-specific information content beyond whole-night HRV summaries and whether a more flexible non-linear classifier provided additional value beyond a simpler linear baseline.