Table 7. Ablation study. In the original Mamba model, the sequence output from the Mamba backbone was aggregated using simple mean pooling over the temporal dimension before being passed to a classification head. Mamba-Attention, which augments Mamba with a trainable attention mechanism in Euclidean space. Instead of uniform averaging, this attention module learns to assign varying importance to each time step, allowing the model to focus on the most informative regions of the sequence. It is important to note that this attention mechanism operates entirely in Euclidean space and differs from our GaA, which is tailored for the Riemannian manifold structure. We also define Geo-Mamba as the counterpart that incorporates GeoMind's architectural improvements within Riemannian space, but without the GaA module. Our full proposed model, *GeoMind*, integrates both these enhancements along with the GaA module to fully leverage the underlying geometry of the data. Bold highlights the best performance.

Model	**HCP-WM**	**ADNI**	**OASIS**	**PPMI**	**ABIDE**
Mamba	96.76±0.86	74.40±5.43	87.09±0.75	67.93±10.69	66.34±0.27
Mamba-Attention	97.22±0.63	76.70±5.29	88.75±1.39	70.15±12.01	68.95±1.78
Geo-Mamba	97.25±0.65	79.60±2.80	89.26±2.29	70.97±8.02	69.75±2.70
GeoMind	**98.29±0.26**	**81.20±2.27**	**89.60±1.87**	**71.35±10.26**	**70.97±3.47**

Table 8. Multi-class classification results on the ADNI dataset.

(%)	GCN	GIN	GSN	MGNN	GNN-AK	SPDNet	MLP
Acc	50.00±6.51	51.60±5.20	52.80±5.31	48.80±5.31	52.40±6.56	52.40±5.20	46.40±7.42
Pre	36.08±14.22	39.75±13.50	53.23±10.33	40.58±10.28	46.07±9.48	37.01±8.89	46.52±12.03
F1	38.49±9.73	41.76±7.65	48.21±7.48	38.71±6.11	43.20±6.55	31.63±8.76	43.83±9.13
(%)	1D-CNN	RNN	LSTM	Mixer	TF	Mamba	*GeoMind*
Acc	46.00±5.44	45.60±6.25	46.00±7.43	48.40±4.18	52.00±6.93	47.20±6.14	**56.00±3.36**
Pre	36.40±9.72	40.95±11.29	25.89±11.28	48.06±12.84	47.63±19.50	38.55±13.23	**60.36±7.67**
F1	39.21±7.71	39.24±7.14	31.87±9.93	39.40±5.47	44.03±11.32	37.19±5.53	**50.83±5.73**

Table 9. Comparison between various Mamba configurations and the proposed *GeoMind* model on HCP-WM dataset (brain regions $N = 360$). The last column (*GeoMind*) highlights the best-performing setting. For a fair comparison, the hidden dimension and network depth of Mamba were adjusted to match the parameter scale of *GeoMind* (highlighted by underline).

	Mamba					***GeoMind***
Hidden dim	2048	1024	1024	1024	512	N
Network layer	5	5	4	2	8	2
Para (M)	132	33.71	27.05	14.07	13.93	14.60
Accuracy	97.22±0.63	97.06±0.62	96.76±0.86			**98.29±0.26**
Precision	97.27±0.62	97.09±0.60	96.80±0.84			**98.18±0.34**
F1	97.22±0.63	97.06±0.62	96.76±0.86			**98.16±0.35**

Table 10. Hyperparameter settings for different models. N denotes the number of brain regions. The code is available at anonymous GitHub for reproducibility. 'M-SGD' and 'M-Adam' represent Stochastic Gradient Descent (SGD) and Adam optimizers, respectively, equipped with manifold-aware updates that enforce geometric constraints (e.g., orthogonality).

Model	1D-CNN	RNN	LSTM	Mixer	TF	Mamba	GCN	GIN
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	5×10^{-5}	10^{-4}	10^{-4}
Weight decay	5×10^{-4}	5×10^{-4}	5×10^{-4}	5×10^{-4}	5×10^{-4}	0	5×10^{-4}	5×10^{-4}
Batch size	64	64	64	64	64	16	64	64
Epochs	300	300	300	300	300	300	300	300
Hidden dim	1024	1024	1024	1024	1024	1024	1024	1024
Layers	2	2	2	4	4	4	2	2
Model	GSN	MGNN	GNN-AK	SPDNet	MLP	STAGIN	NeuroGraph	*GeoMind*
Optimizer	Adam	Adam	Adam	M-SGD	Adam	Adam	Adam	M-Adam
Learning rate	10^{-2}	10^{-2}	10^{-3}	5×10^{-3}	10^{-4}	5×10^{-4}	10^{-5}	5×10^{-5}
Weight decay	0	5×10^{-4}	0	10^{-5}	5×10^{-4}	10^{-3}	5×10^{-4}	0
Batch size	16	16	128	32	64	3	16	16
Epochs	300	1000	100	100	300	100	100	300
Hidden dim	256	1024	128	[N, 64, 32]	1024	128	32	N
Layers	2	2	2	3	2	4	3	2